**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

# INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|---|---|---|
| **(51) International Patent Classification $^6$ :**<br>**C12N 15/31, C07K 14/315, 16/12, C12Q 1/68** | **A2** | **(11) International Publication Number:** **WO 98/50555**<br>**(43) International Publication Date:** 12 November 1998 (12.11.98) |

**(21) International Application Number:** PCT/US98/08985

**(22) International Filing Date:** 4 May 1998 (04.05.98)

**(30) Priority Data:**

| | | |
|---|---|---|
| 60/044,031 | 6 May 1997 (06.05.97) | US |
| 60/046,655 | 16 May 1997 (16.05.97) | US |
| 60/066,009 | 14 November 1997 (14.11.97) | US |

**(71) Applicant** *(for all designated States except US)*: HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

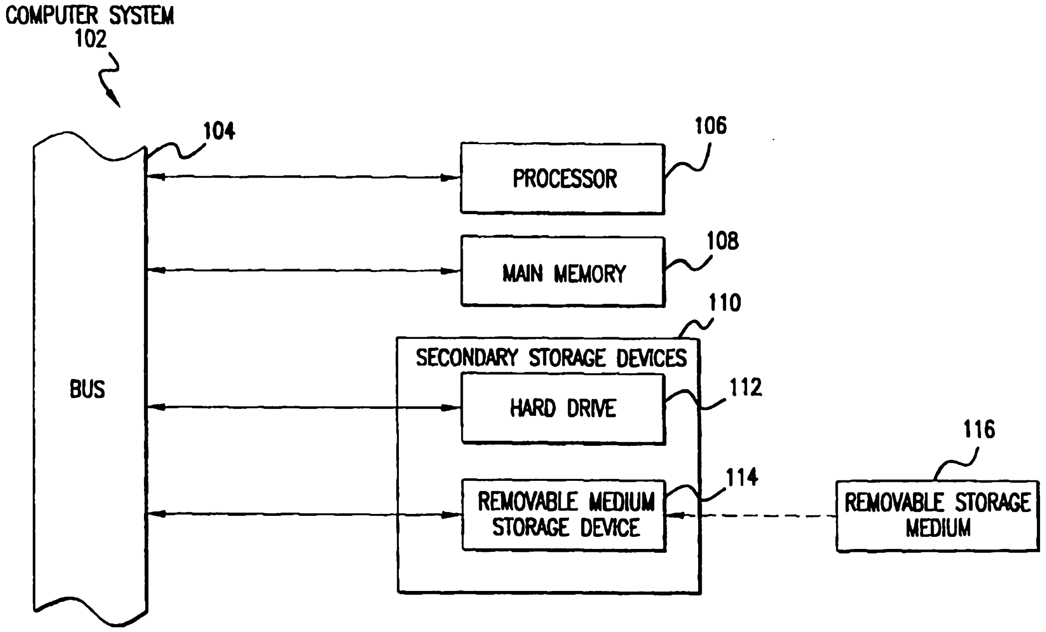
**(72) Inventors; and**
**(75) Inventors/Applicants** *(for US only)*: KUNSCH, Charles, A. [US/US]; 4083 Spalding Hollow, Norcross, GA 30092 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). BARASH, Steven, C. [US/US]; 582 College Parkway #303, Rockville, MD 20850 (US).

**(74) Agents:** BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

**(81) Designated States:** AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

**Published**
*Without international search report and to be republished upon receipt of that report.*

**(54) Title:** *ENTEROCOCCUS FAECALIS* POLYNUCLEOTIDES AND POLYPEPTIDES



## (57) Abstract

The present invention provides polynucleotide sequences of the genome of *Enterococcus faecalis*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynucleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

## *FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | |
|---|---|
| **AL** | Albania |
| **AM** | Armenia |
| **AT** | Austria |
| **AU** | Australia |
| **AZ** | Azerbaijan |
| **BA** | Bosnia and Herzegovina |
| **BB** | Barbados |
| **BE** | Belgium |
| **BF** | Burkina Faso |
| **BG** | Bulgaria |
| **BJ** | Benin |
| **BR** | Brazil |
| **BY** | Belarus |
| **CA** | Canada |
| **CF** | Central African Republic |
| **CG** | Congo |
| **CH** | Switzerland |
| **CI** | Côte d'Ivoire |
| **CM** | Cameroon |
| **CN** | China |
| **CU** | Cuba |
| **CZ** | Czech Republic |
| **DE** | Germany |
| **DK** | Denmark |
| **EE** | Estonia |
| **ES** | Spain |
| **FI** | Finland |
| **FR** | France |
| **GA** | Gabon |
| **GB** | United Kingdom |
| **GE** | Georgia |
| **GH** | Ghana |
| **GN** | Guinea |
| **GR** | Greece |
| **HU** | Hungary |
| **IE** | Ireland |
| **IL** | Israel |
| **IS** | Iceland |
| **IT** | Italy |
| **JP** | Japan |
| **KE** | Kenya |
| **KG** | Kyrgyzstan |
| **KP** | Democratic People's Republic of Korea |
| **KR** | Republic of Korea |
| **KZ** | Kazakstan |
| **LC** | Saint Lucia |
| **LI** | Liechtenstein |
| **LK** | Sri Lanka |
| **LR** | Liberia |
| **LS** | Lesotho |
| **LT** | Lithuania |
| **LU** | Luxembourg |
| **LV** | Latvia |
| **MC** | Monaco |
| **MD** | Republic of Moldova |
| **MG** | Madagascar |
| **MK** | The former Yugoslav Republic of Macedonia |
| **ML** | Mali |
| **MN** | Mongolia |
| **MR** | Mauritania |
| **MW** | Malawi |
| **MX** | Mexico |
| **NE** | Niger |
| **NL** | Netherlands |
| **NO** | Norway |
| **NZ** | New Zealand |
| **PL** | Poland |
| **PT** | Portugal |
| **RO** | Romania |
| **RU** | Russian Federation |
| **SD** | Sudan |
| **SE** | Sweden |
| **SG** | Singapore |
| **SI** | Slovenia |
| **SK** | Slovakia |
| **SN** | Senegal |
| **SZ** | Swaziland |
| **TD** | Chad |
| **TG** | Togo |
| **TJ** | Tajikistan |
| **TM** | Turkmenistan |
| **TR** | Turkey |
| **TT** | Trinidad and Tobago |
| **UA** | Ukraine |
| **UG** | Uganda |
| **US** | United States of America |
| **UZ** | Uzbekistan |
| **VN** | Viet Nam |
| **YU** | Yugoslavia |
| **ZW** | Zimbabwe |

# *Enterococcus faecalis* Polynucleotides and Polypeptides

## FIELD OF THE INVENTION

The present invention relates to the field of molecular biology. In particular, it relates to, among other things, nucleotide sequences of *Enterococcus faecalis*, contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequences, and uses of the polynucleotides and sequences thereof, such as in fermentation, polypeptide production, assays and pharmaceutical development, among others.

## BACKGROUND OF THE INVENTION

Enterococci have been recognized as being pathogenic for humans since the turn of the century when they were first described by Thiercelin in 1988 as microscopic organisms. The genus Enterococcus includes the species Enterococcus faecalis or *E. faecalis* which is the most common pathogen in the group, accounting for 80 - 90 percent of all enterococcal infections. *See* Lewis et al. (1990) Eur J. Clin Microbiol Infect Dis. 9:111-117.

The incidence of enterococcal infections has increased in recent years and enterococci are now the second most frequently reported nosocomial pathogens. Enterococcal infection is of particular concern because of its resistance to antibiotics. Recent attention has focused on enterococci not only because of their increasing role in nosocomial infections, but also because of their remarkable and increasing resistance to antimicrobial agents. These factors are mutually reinforcing since resistance allows enterococci to survive in an environment in which antimicrobial agents are heavily used; the hospital setting provides the antibiotics which eliminate or suppress susceptible bacteria, thereby providing a selective advantage for resistant organisms, and the hospital also provides the potential for dissemination of resistant enterococci via the usual routes of hand and environmental contamination.

Antimicrobial resistance can be divided into two general types, inherent or intrinsic property and that which is acquired. The genes for intrinsic resistance, like other species characteristics, appear to reside on the chromosome. Acquired resistance results from either a mutation in the existing DNA or acquisition of new DNA. The various inherent traits expressed by enterococci include resistance to semisynthetic penicillinase-resistant penicillins, cephalosporins, low levels of aminoglycosides, and low levels of clindamycin. Examples of acquired resistance include resistance to chloramphenicol, erythromycin, high levels of clindamycin, tetracycline, high levels of aminoglycosides, penicillin by means of penicillinase, fluoroquinolones, and vancomycin. Resistance to high levels of penicillin without penicillinase and resistance to fluoroquinolones are not known to be plasmid or transposon mediated and presumably are due to mutation(s).

Although the main reservoir for enterococci in humans is the gastrointestinal tract, the bacteria can also reside in the gallbladder, urethra and vagina.

*E. faecalis* has emerged as an important pathogen in endocarditis, bacteremia, urinary tract infections (UTIs), intraabdominal infections, soft tissue infections, and neonatal sepsis (Lewis 1990, *supra*). In the 1970s and 1980s enterococci became firmly established as major nosocomial pathogens. They are now the fourth leading cause of hospital-acquired infection and the third leading cause of bacteremia in the United States. Fatality ratios for enterococcal bactermia range from 12% to 68%, with death due to enterococcal sepsis in 4 to 50% of these cases. *See* Emori, T.G. (1993) Clin. Microbiol. Rev. 6:428-442.

The ability of enterococci to colonize the gastrointestinal tract, plus the many intrinsic and acquired resistance traits, means that these organisms, which usually seem to have relatively low intrinsic virulence, are given an excellent opportunity to become secondary invaders. Since nosocomial isolates of enterococci have displayed resistance to essentially every useful antimicrobial agent, it will likely become increasingly difficult to successfully treat and control enterococcal infections. Particularly when the various resistance genes come together in a single strain, an event almost certain to occur at some time in the future.

The etiology of diseases mediated or exacerbated by Enterococcus faecalis, involves the programmed expression of *E. faecalis* genes, and that characterizing these genes and their patterns of expression would dramatically add to our understanding of the organism and its host interactions. Knowledge of the *E. faecalis* gene and genomic organization would improve our understanding of disease etiology and lead to improved and new ways of preventing, treating and diagnosing diseases. Thus, there is a need to characterize the genome of *E. faecalis* and for polynucleotides of this organism.

## SUMMARY OF THE INVENTION

The present invention is based on the sequencing of fragments of the *Enterococcus faecalis* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS:1-982.

The present invention provides the nucleotide sequence of hundreds of contigs of the *Enterococcus faecalis* genome, which are listed in tables below and set out in the

Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS:1-982.

The present invention further provides nucleotide sequences which are at least 95%, 96%, 97%, 98%, and 99%, identical to the nucleotide sequences of SEQ ID NOS:1-982.

The nucleotide sequence of SEQ ID NOS:1-982, a representative fragment thereof, or a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NOS:1-982 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

The present invention further provides systems, particularly computer-based systems which contain the sequence information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Enterococcus faecalis* genome.

Another embodiment of the present invention is directed to fragments of the *Enterococcus faecalis* genome having particular structural or functional attributes. Such fragments of the *Enterococcus faecalis* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs, fragments which modulate the expression of an operably linked ORF, hereinafter referred to as expression modulating fragments or EMFs, and fragments which can be used to diagnose the presence of *Enterococcus faecalis* in a sample, hereinafter referred to as diagnostic fragments or DFs.

Each of the ORFs in fragments of the *Enterococcus faecalis* genome disclosed in Tables 1-3, and the EMFs found 5' prime of the initiation codon, can be used in numerous ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

The present invention further includes recombinant constructs comprising one or more fragments of the *Enterococcus faecalis* genome of the present invention. The recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Enterococcus faecalis* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Enterococcus faecalis* genome of the present invention. The host cells can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a procaryotic cell such as a bacterial cell.

The present invention is further directed to isolated polypeptides and proteins encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present invention from cells which have been altered to express them.

The invention further provides methods of obtaining homologs of the fragments of the *Enterococcus faecalis* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both monoclonal and polyclonal antibodies.

The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the antibodies of the present invention, or one or more of the DFs of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the antibodies, or one of the DFs of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of bound antibodies or hybridized DFs.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of: (a)contacting an agent with an isolated protein encoded by one of the ORFs of the present invention; and (b)determining whether the agent binds to said protein.

The present genomic sequences of *Enterococcus faecalis* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Enterococcus faecalis* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to *Enterococcus faecalis* researchers and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes has and will greatly enhance the ability to analyze and understand chromosomal organization. In particular, sequenced contigs and genomes will provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the ability to do comparative genomic and molecular phylogeny.

## DESCRIPTION OF THE FIGURES

**FIGURE 1** is a block diagram of a computer system (102) that can be used to implement computer-based systems of the present invention.

**FIGURE 2** is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Enterococcus faecalis* genome of the present invention. Both Macintosh and Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of sequence files. The program Sequis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Enterococcus faecalis* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using Extrseq, a Unix utility for retrieving sequences from an SQL database. The resulting sequence file is processed by seq_filter to trim portions of the sequences with more than 1% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research ( TIGR ) for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly step is loaded into the database with the lassie program. Identification of open reading frames (ORFs) is accomplished by processing contigs with GeneMark, described in Borodovsky, M. and Mclninch, J.D. (1993) *Comput. Chem.*, 17:123133. The ORFs are searched against *E. faecalis* sequences from GenBank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al.*, *J. Mol. Biol. 215:* 403-410 (1990)). Results of the ORF determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3.

## DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

The present invention is based on the sequencing of fragments of the *Enterococcus faecalis* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS: 1-982. (As used herein, the "primary sequence" refers to the nucleotide sequence represented by the IUPAC nomenclature system.)

In addition to the aforementioned *Enterococcus faecalis* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS: 1-982 , or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS:1-982" refers to any portion of the SEQ ID NOS: 1-982 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Enterococcus faecalis* open reading frames (ORFs ), expression modulating fragment ( EMFs ) and fragments which can be used to diagnose the presence of *Enterococcus faecalis* in a sample (DFs ). A non-limiting identification of preferred representative fragments is provided in Tables 1-3. As discussed in detail below, the information provided in SEQ ID NOS:1-982 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled in the art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Enterococcus faecalis* proteins.

The present invention is further directed to nucleic acid molecules encoding portions or fragments of the nucleotide sequences described herein. Fragments include portions of the nucleotide sequences of Table 1-3 and SEQ ID NOS:1-982, at least 10 contiguous nucleotides in length selected from any two integers, one of which representing a 5' nucleotide position and a second of which representing a 3' nucleotide position, where the first nucleotide for each nucleotide sequence in SEQ ID NOS:1-982 is position 1. That is, every combination of a 5' and 3' nucleotide position that a fragment at least 10 contiguous nucleotides in length could occupy is included in the invention. At least means a fragment may be 10 contiguous nucleotide bases in length or any integer between 10 and the length of an entire nucleotide sequence of SEQ ID NOS:1-982 minus 1. Therefore, included in the invention are contiguous fragments specified by any 5' and 3' nucleotide base positions of a nucleotide sequences of SEQ ID NOS:1-982 wherein the contiguous fragment is any integer between 10 and the length of an entire nucleotide sequence minus 1.

Further, the invention includes polynucleotides comprising fragments specified by size, in nucleotides, rather than by nucleotide positions. The invention includes any fragment size, in contiguous nucleotides, selected from integers between 10 and the length of an entire nucleotide sequence minus 1. Preferred sizes of contiguous nucleotide fragments include 20 nucleotides, 30 nucleotides, 40 nucleotides, 50 nucleotides. Other preferred sizes of contiguous nucleotide fragments, which may be useful as diagnostic probes and primers, include fragments 50-300 nucleotides in length which include, as discussed above, fragment sizes representing each integer between 50-300. Larger fragments are also useful according to the present invention corresponding to most, if not all, of the nucleotide sequences shown in SEQ ID NOS:1-982. The preferred sizes are, of course, meant to exemplify not limit the present invention as all size fragments, representing any integer between 10 and the length of an entire nucleotide sequence minus 1, of each SEQ ID NO:, are included in the invention.

The present invention also provides for the exclusion of any fragment, specified by 5' and 3' base positions or by size in nucleotide bases as described above for any nucleotide sequence of SEQ ID NOS:1-982. Any number of fragments of nucleotide sequences in SEQ ID NOS:1-982, specified by 5' and 3' base positions or by size in nucleotides, as described above, may be excluded from the present invention.

While the presently disclosed sequences of SEQ ID NOS:1-982 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances, further investigation of a fragment or sequence of the invention may reveal a nucleotide sequence error present in a nucleotide sequence disclosed in SEQ ID NOS:1-982. However, once the present invention is made available (*i.e.*, once the information in SEQ ID NOS:1-982 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS: 1-982 will be well within the skill of the art. The present disclosure makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotides may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

Even if all of the very rare sequencing errors in SEQ ID NOS: 1-982 were corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-982.

As discussed elsewhere herein, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning and sequencing DNA. A wide variety of *Enterococcus faecalis* strains that can be used to prepare *E. faecalis* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC). While the present invention is enabled by the sequences and other information herein disclosed, the *E. faecalis* strain that provided the DNA of the present Sequence Listing, Strain V586, kindly provided by Dr. Michael Gilmore, University of Oklahoma, has been deposited in the ATCC, as a convenience to those of skill in the art. The *E. faecalis* strain V586 was deposited 2 May 1997 at the ATCC, 10801 University Blvd. Manassas, VA 20110-2209, and given accession number 55969. The provision of the deposits is not a waiver of any rights of the inventors or their assignees in the present subject matter.

The nucleotide sequences of the genomes from different strains of *Enterococcus faecalis* differ somewhat. However, the nucleotide sequences of the genomes of all *Enterococcus faecalis* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS: 1-982. Nearly all will be at least 99% identical and the great majority will be 99.9% identical.

The present application is further directed to nucleic acid molecules at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleic acid sequence shown in SEQ ID NOS:

1-982. The above nucleic acid sequences are included irrespective of whether they encode a polypeptide having *E. faecalis* activity. This is because even where a particular nucleic acid molecule does not encode a polypeptide having *E. faecalis* activity, one of skill in the art would still know how to use the nucleic acid molecule, for instance, as a hybridization probe. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having *E. faecalis* activity include, *inter alia*, isolating an *E. faecalis* gene or allelic variants thereof from a DNA library, and detecting *E. faecalis* mRNA expression samples, environmental samples, suspected of containing *E. faecalis* by Northern Blot analysis.

Preferred, are nucleic acid molecules having sequences at least 90%, 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequence shown in SEQ ID NOS: 1-982, which do, in fact, encode a polypeptide having *E. faecalis* protein activity By "a polypeptide having *E. faecalis* activity" is intended polypeptides exhibiting activity similar, but not necessarily identical, to an activity of the *E. faecalis* protein of the invention, as measured in a particular biological assay suitable for measuring activity of the specified protein.

Due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the nucleic acid molecules having a sequence at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the nucleic acid sequences shown in SEQ ID NOS: 1-982 will encode a polypeptide having *E. faecalis* protein activity. In fact, since degenerate variants of these nucleotide sequences all encode the same polypeptide, this will be clear to the skilled artisan even without performing the above described comparison assay. It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode a polypeptide having *E. faecalis* protein activity. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect protein function (e.g., replacing one aliphatic amino acid with a second aliphatic amino acid), as further described below.

The biological activity or function of the polypeptides of the present invention are expected to be similar or identical to polypeptides from other bacteria that share a high degree of structural identity/similarity. Tables 1 and 2 lists accession numbers and descriptions for the closest matching sequences of polypeptides available through Genbank. It is therefore expected that the biological activity or function of the polypeptides of the present invention will be similar or identical to those polypeptides from other bacterial genuses, species, or strains listed in Tables 1 and 2.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the *E. faecalis* polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted, inserted, or substituted with another nucleotide. The query sequence may be an entire sequence shown in SEQ ID NOS: 1-982, the ORF (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the presence invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. *See* Brutlag et al. (1990) Comp. App. Biosci. 6:237-245. In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by first converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the lenght of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only nucleotides outside the 5' and 3' nucleotides of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 nucleotide subject sequence is aligned to a 100 nucleotide query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 nucleotides at 5' end. The 10 unpaired nucleotides represent 10% of the sequence (number of nucleotides at the 5' and 3' ends not matched/total number of nucleotides in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 nucleotides were perfectly matched the final percent identity would be 90%. In another example, a 90 nucleotide subject sequence is compared with a 100 nucleotide query sequence. This time the deletions are internal deletions so that there are no nucleotides on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only nucleotides 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

## COMPUTER RELATED EMBODIMENTS

The nucleotide sequences provided in SEQ ID NOS:1-982, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-982 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, provided refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; *i.e.*, a nucleotide sequence provided in SEQ ID NOS:1-982 , a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide of SEQ ID NOS:1-982. Such a manufacture provides a large portion of the *Enterococcus faecalis* genome and parts thereof (*e.g.*, a *Enterococcus faecalis* open reading frame (ORF)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the *Enterococcus faecalis* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD- ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently know methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention. A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially- available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-982, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS: 1-982 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol. 215*:403-410 (1990)) and BLAZE (Brutlag *et al.*, *Comp. Chem. 17*:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Enterococcus faecalis* genome which contain homology to ORFs or proteins from both *Enterococcus faecalis* and from other organisms. Among the ORFs discussed herein are protein encoding fragments of the *Enterococcus faecalis* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful metabolites, proteins to be used as vaccines or in the generation of immuno-therapeutic reagents, or as drug screening targets.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, commercially important fragments of the *Enterococcus faecalis* genome.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer- based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Enterococcus faecalis* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the *Enterococcus faecalis* genome. In the present examples, implementing software which implement the BLAST algorithm, described in Altschul *et al.* (1990) *J. Mol. Biol. 215:* 403-410, is used to identify open reading frames within the *Enterococcus faecalis* genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, *etc.* A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, *etc.*) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for accessing and processing the genomic sequence (such as search tools, comparing tools, *etc.*) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

## BIOCHEMICAL EMBODIMENTS

Other embodiments of the present invention are directed to isolated fragments of the *Enterococcus faecalis* genome. The fragments of the *Enterococcus faecalis* genome of the present invention include, but are not limited to fragments which encode peptides, hereinafter open reading frames (ORFs), fragments which modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Enterococcus faecalis* in a sample, hereinafter diagnostic fragments (DFs).

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Enterococcus faecalis* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS:1-982, to representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

A variety of purification means can be used to generate the isolated fragments of the present invention. These include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Enterococcus faecalis* DNA can be enzymatically sheared to produce fragments of 15-20 kb in length. These fragments can then be used to generate a *Enterococcus faecalis* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS:1-982. Well known and routine techniques of PCR cloning then can be used to isolate the ORF from the lambda DNA library or *Enterococcus faecalis* genomic DNA. Thus, given the availability of SEQ ID NOS:1-982, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS:1-982 using methods set out above, those of skill will be enabled by the present disclosure to isolate any ORF-containing or other nucleic acid fragment of the present invention.

The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA. As used herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence translatable into protein. Each sequence of SEQ ID NOS:1-982, however, begins and ends with a termination codon. For purposes of numbering and reference to polynucleotide and polypeptide sequences the entire sequence of each sequence of SEQ ID NOS:1-982 is included with the first nucleotide being positon 1. Therefore, for reference purposes the numbering used in the present invention is that provided in the sequence listing for SEQ ID NOS:1-982.

Tables 1, 2, and 3 list ORFs in the *Enterococcus faecalis* genomic contigs of the present invention that were identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It will be appreciated that other criteria can be used, in accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive, or more selective lists.

Table 1 sets out ORFs in the *Enterococcus faecalis* contigs of the present invention that over a continuous region of at least 50 bases are 95% or more identical (by BLAST analysis) to a nucleotide sequence available through GenBank in March, 1997.

Table 2 sets out ORFs in the *Enterococcus faecalis* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through GenBank in March, 1997.

Table 3 sets out ORFs in the *Enterococcus faecalis* contigs of the present invention that do not match significantly, by BLASTP analysis, a polypeptide sequence available through GenBank in March, 1997.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column indicates the coordinate of the first nucleotide of the ORF, counting from the 5' end of the contig strand; the fourth column indicates the coordinate of the final nucleotide of the ORF, counting from the 5' end of the contig strand.

In Tables 1 and 2, column five lists the Reference for the closest matching sequence available through GenBank. These reference numbers are the database entry numbers commonly used by those of skill in the art, who will be familiar with their denominators. Descriptions of the nomenclature are available from the National Center for Biotechnology Information. Column six in Tables 1 and 2 provides the gene name of the matching sequence.

In Table 1, column seven provides the nucleotide BLAST percent identity score from the comparison of the ORF and the GenBank sequence, column eight indicates the length in nucleotides of the highest scoring segment pair identified by the BLAST identity analysis, and column nine provides the total length of the ORF in nucleotides.

In Table 2, column seven provides the protein BLAST percent similarity of the highest scoring segment pair identified, column eight provides the percent identity of the highest scoring segment pair, and column nine provides the total length of the ORF in nucleotides.

The concepts of percent identity and percent similarity of two polypeptide sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (*e.g.*, at positions 1, 3 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (*i.e.*, possessed similar biochemical characteristics). Many programs for analysis of nucleotide or amino acid sequence similarity, such as fasta and BLAST specifically list percent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the percent identity of the highest scoring segment pair in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in the pertinent literature highlighted by the citations provided below.

It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Enterococcus faecalis* genome other than those listed in Tables 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are fragments which induce the expression or an operably linked ORF in response to a specific regulatory factor or physiological event.

EMF sequences can be identified within the contigs of the *Enterococcus faecalis* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the *Enterococcus faecalis* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention. Further, the two methods can be combined and used together.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below.

A sequence which is suspected as being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host in examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Enterococcus faecalis* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Enterococcus faecalis* genome, such as by using well-known computer analysis software, and by generating and testing probes or amplification primers consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS:1-982, a representative fragment thereof, or a nucleotide sequence at least 99% and preferably 99.9% identical to SEQ ID NOS:1-982, with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated.

Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (*i.e.*, sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Enterococcus faecalis* origin isolated by using part or all of the fragments in question as a probe or primer.

Each of the ORFs of the *Enterococcus faecalis* genome disclosed in Tables 1, 2 and 3, and the EMFs found 5 to the ORFs, can be used as polynucleotide reagents in numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particularly *Enterococcus faecalis*. Especially preferred in this regard are ORFs such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Enterococcus faecalis*. Also particularly preferred are ORFs that can be used to distinguish between strains of *Enterococcus faecalis*, particularly those that distinguish medically important strain, such as drug-resistant strains.

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation, or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for example, Lee *et al.*, *Nucl. Acids Res. 6*:3073 (1979); Cooney *et al.*, *Science 241*:456 (1988); and Dervan *et al.*, *Science 251*:1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, *J. Neurochem. 56*:560 (1991) and *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)).

The present invention further provides recombinant constructs comprising one or more fragments of the *Enterococcus faecalis* genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a fragment of the *Enterococcus faecalis* genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBS SK (+ or -), pBS KS (+ or -), pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (available from Stratagene) pSVK3, pBPV, pMSG, pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein- I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

The present invention further provides host cells containing any one of the isolated fragments of the *Enterococcus faecalis* genomic fragments and contigs of the present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or a procaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE, dextran mediated transfection and electroporation, which are described in, for instance, Davis, L. *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986).

A host cell containing one of the fragments of the *Enterococcus faecalis* genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF. The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs depicted in Tables 2 and 3 which encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Such short fragments as may be obtained most readily by synthesis are useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polypeptides and proteins to isolate and purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography.

The polypeptides and proteins of the present invention also can be purified from cells which have been altered to express the desired polypeptide or protein. Preferred polypeptides and proteins of the present invention are polypeptides and proteins coded for by the polynucleotides of SEQ ID NOS:1-982, wherein the polypeptides and proteins are coded in the same frame as the termination codon at the end of each sequence of SEQ ID NOS:1-982. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of the *E. faecalis* polypeptide can be substantially purified by the one-step method described by Smith et al. (1988) Gene 67:31-40. Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies directed against the polypeptides of the invention in methods which are well known in the art of protein purification.

The invention further provides for isolated *E. faecalis* polypeptides comprising an amino acid sequence selected from the group including: (a) the amino acid sequence of a full-length *E. faecalis* polypeptide having the complete amino acid sequence from the first methionine codon to the termination codon of each sequence listed in SEQ ID NOS:1-982, wherein said termination codon is at the end of each SEQ ID NO: and said first methionine is the first methionine in frame with said termination codon; and (b) the amino acid sequence of a full-length *E. faecalis* polypeptide having the complete amino acid sequence in (a) excepting the N-terminal methionine.

The polypeptides of the present invention also include polypeptides having an amino acid sequence at least 80% identical, more preferably at least 90% identical, and still more preferably 95%, 96%, 97%, 98% or 99% identical to those described in (a) and (b) above.

The present invention is further directed to polynucleotide encoding portions or fragments of the amino acid sequences described herein as well as to portions or fragments of the isolated amino acid sequences described herein. Fragments include portions of the amino acid sequences described herein, are at least 5 contiguous amino acid in length, are selected from any two integers, one of which representing a N-terminal position. The initiation codon of the polypeptides of the present inventions position 1. The initiation codon (positon 1) for purposes of the present invention is the first methionine codon of each sequence of SEQ ID NOS:1-982 which is in frame with the termination codon at the end of each said sequence. Every combination of a N-terminal and C-terminal position that a fragment at least 5 contiguous amino acid residues in length could occupy, on any given amino acid sequence encoded by a sequence of SEQ ID NOS:1-982 is included in the invention, i.e., from initiation codon up to the termination codon. At least means a fragment may be 5 contiguous amino acid residues in length or any integer between 5 and the number of residues in a full length amino acid sequence minus 1. Therefore, included in the invention are contiguous fragments specified by any N-terminal and C-terminal positions of amino acid sequence set forth in SEQ ID NOS:1-982 wherein the contiguous fragment is any integer between 5 and the number of residues in a full length sequence minus 1.

Further, the invention includes polypeptides comprising fragments specified by size, in amino acid residues, rather than by N-terminal and C-terminal positions. The invention includes any fragment size, in contiguous amino acid residues, selected from integers between 5 and the number of residues in a full length sequence minus 1. Preferred sizes of contiguous polypeptide fragments include about 5 amino acid residues, about 10 amino acid residues, about 20 amino acid residues, about 30 amino acid residues, about 40 amino acid residues, about 50 amino acid residues, about 100 amino acid residues, about 200 amino acid residues, about 300 amino acid residues, and about 400 amino acid residues. The preferred sizes are, of course, meant to exemplify, not limit, the present invention as all size fragments representing any integer between 5 and the number of residues in a full length sequence minus 1 are included in the invention. The present invention also provides for the exclusion of any fragments specified by N-terminal and C-terminal positions or by size in amino acid residues as described above. Any number of fragments specified by N-terminal and C-terminal positions or by size in amino acid residues as described above may be excluded.

The above fragments need not be active since they would be useful, for example, in immunoassays, in epitope mapping, epitope tagging, to generate antibodies to a particular portion of the protein, as vaccines, and as molecular weight markers.

Further polypeptides of the present invention include polypeptides which have at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98% or 99% similarity to those described above.

A further embodiment of the invention relates to a polypeptide which comprises the amino acid sequence of a *E. faecalis* polypeptide having an amino acid sequence which contains at least one conservative amino acid substitution, but not more than 50 conservative amino acid substitutions, not more than 40 conservative amino acid substitutions, not more than 30 conservative amino acid substitutions, and not more than 20 conservative amino acid substitutions. Also provided are polypeptides which comprise the amino acid sequence of a *E. faecalis* polypeptide, having at least one, but not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 conservative amino acid substitutions.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to the amino acid sequences encoded by the sequences of SEQ ID NOS:1-982, as described hererin, can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al., (1990) Comp. App. Biosci. 6:237-245. In a sequence alignment the query and subject sequences are both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, the results, in percent identity, must be manually corrected. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query amino acid residues outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not match/align with the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected. No other manual corrections are to made for the purposes of the present invention.

The above polypeptide sequences are included irrespective of whether they have their normal biological activity. This is because even where a particular polypeptide molecule does not have biological activity, one of skill in the art would still know how to use the polypeptide, for instance, as a vaccine or to generate antibodies. Other uses of the polypeptides of the present invention that do not have *E. faecalis* activity include, *inter alia*, as epitope tags, in epitope mapping, and as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods known to those of skill in the art.

As described below, the polypeptides of the present invention can also be used to raise polyclonal and monoclonal antibodies, which are useful in assays for detecting *E. faecalis* protein expression or as agonists and antagonists capable of enhancing or inhibiting *E. faecalis* protein function. Further, such polypeptides can be used in the yeast two-hybrid system to "capture" *E. faecalis* protein binding proteins which are also candidate agonists and antagonists according to the present invention. *See, e.g.,* Fields et al. (1989) Nature 340:245-246.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (*e.g.*, yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Enterococcus faecalis* genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

Recombinant expression vehicle or "vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancer and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual,* 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), alpha-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *E. coli, B. subtilis, Salmonella typhimurium* and various species within the genera *Pseudomonas* and *Streptomyces*. Others may, also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available form Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, *Cell 23:*175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5 flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are substantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity, and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

The invention further provides methods of obtaining homologs from other strains of *Enterococcus faecalis*, of the fragments of the *Enterococcus faecalis* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Enterococcus faecalis* is defined as a homolog of a fragment of the *Enterococcus faecalis* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Enterococcus faecalis* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among those are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS:1-982 or from a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ ID NOS:1-982 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis *et al.*, *PCR Protocols*, Academic Press, San Diego, CA (1990)).

When using primers derived from SEQ ID NOS:1-982 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS: 1-982, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, annealing at 50-60°C in 6X SSPC and 50% formamide, and washing at 50- 65°C in 0.5X SSPC) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency conditions (*e.g.*, hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences which are greater than 40-50% homologous to the primer will also be amplified.

When using DNA probes derived from SEQ ID NOS:1-982, or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-982, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, hybridizing at 50- 65°C in 5X SSPC and 50% formamide, and washing at 50- 65°C in 0.5X SSPC), sequences having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (*e.g.*, hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacteria which are closely related to *Enterococcus faecalis*.

## ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and industrial purposes consistent with the type of putative identification of the polypeptide. Such identifications permit one skilled in the art to use the *Enterococcus faecalis* ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., MacMillan Publications, Ltd. NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESES, Tramper *et al.*, Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

### 1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

The various metabolic pathways present in *Enterococcus faecalis* can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS:1-982.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts *et al.*, *Symbiosis 21*:79 (1986) and Voragen *et al.* in *Biocatalysts In Agricultural Biotechnology*, Whitaker *et al.*, Eds., *American Chemical Society Symposium Series 389*:93 (1989) .

The metabolism of sugars is an important aspect of the primary metabolism of *Enterococcus faecalis*. Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerases such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulonic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger *et al.*, *Biotechnology* 6(A), Rhine *et al.*, Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir *et al.*, *Biotechnology Letters 1*:21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis *et al.*, beginning on page 357 in GENE MANIPULATIONS AND FUNGI; Benett *et al.,* Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for quantitative determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrosylates. This application is described in Owusu *et al.*, *Biochem. et Biophysica. Acta. 872:*83 (1986), for instance.

The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.*, *Biotechnology, The Textbook of Industrial Microbiology*, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose- produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, Starch *40*:307 (1988).

Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See Faultman *et al.*, Acid Proteases Structure Function and Biology, Tang, J., ed., Plenum Press, New York (1977) and Godfrey *et al.*, Industrial Enzymes, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.*, Report Industrial Enzymes by 1990, Hel Hepner & Associates, London (1986)).

Another class of commercially usable proteins of the present invention are the microbial lipases, described by, for instance, Macrae *et al.*, *Philosophical Transactions of the Chiral Society of London 310*:227 (1985) and Poserke, *Journal of the American Oil Chemist Society 61*:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application of lipases include the use as a detergent additive to facilitate the removal of fats from fabrics in the course of the washing procedures.

The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates. Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.*, *Recent Advances in the Generation of Chiral Intermediates Using Enzymes*, CRC Press, Boca Raton, Florida (1990)). The following reactions catalyzed by enzymes are of interest to organic chemists: hydrolysis of carboxylic acid esters, phosphate esters, amides and nitriles, esterification reactions, trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other hand, has been described in detail by Bud *et al.*, Chemistry in Britain (1987), p. 127.

Amino transferases, enzymes involved in the biosynthesis and metabolism of amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo- selective synthesis of only L-amino acids and generally possess uniformly high catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology 136*:479 (1987).

Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination.

### 2. Generation of Antibodies

As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety of procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be used to generate an antibody which selectively binds the protein.

*E. faecalis* protein-specific antibodies for use in the present invention can be raised against the intact *E. faecalis* protein or an antigenic polypeptide fragment thereof, which may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier.

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules, single chain whole antibodies, and antibody fragments. Antibody fragments of the present invention include Fab and F(ab')2 and other fragments including single-chain Fvs (scFv) and disulfide-linked Fvs (sdFv). Also included in the present invention are chimeric and humanized monoclonal antibodies and polyclonal antibodies specific for the polypeptides of the present invention. The antibodies of the present invention may be prepared by any of a variety of methods. For example, cells expressing a polypeptide of the present invention or an antigenic fragment thereof can be administered to an animal in order to induce the production of sera containing polyclonal antibodies. For example, a preparation of *E. faecalis* polypeptide or fragment thereof is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In a preferred method, the antibodies of the present invention are monoclonal antibodies or binding fragments thereof. Such monoclonal antibodies can be prepared using hybridoma technology. *See, e.g.,* Harlow et al., ANTIBODIES: A LABORATORY MANUAL, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988); Hammerling, et al., in: MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS 563-681 (Elsevier, N.Y., 1981). Fab and F(ab')2 fragments may be produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments). Alternatively, *E. faecalis* polypeptide-binding fragments, chimeric, and humanized antibodies can be produced through the application of recombinant DNA technology or through synthetic chemistry using methods known in the art.

Alternatively, additional antibodies capable of binding to the polypeptide antigen of the present invention may be produced in a two-step procedure through the use of anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and that, therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, *E. faecalis* polypeptide-specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the *E. faecalis* polypeptide-specific antibody can be blocked by the *E. faecalis* polypeptide antigen. Such antibodies comprise anti-idiotypic antibodies to the *E. faecalis* polypeptide-specific antibody and can be used to immunize an animal to induce formation of further *E. faecalis* polypeptide-specific antibodies.

Antibodies and fragements thereof of the present invention may be described by the portion of a polypeptide of the present invention recognized or specifically bound by the antibody. Antibody binding fragements of a polypeptide of the present invention may be described or specified in the same manner as for polypeptide fragements discussed above., i.e, by N-terminal and C-terminal positions or by size in contiguous amino acid residues. Any number of antibody binding fragments, of a polypeptide of the present invention, specified by N-terminal and C-terminal positions or by size in amino acid residues, as described above, may also be excluded from the present invention. Therefore, the present invention includes antibodies the specifically bind a particuarlly discribed fragement of a polypeptide of the present invention and allows for the exclusion of the same.

Antibodies and fragements thereof of the present invention may also be described or specified in terms of their cross-reactivity. Antibodies and fragements that do not bind polypeptides of any other species of *Enterococcus* other than *E. faecalis* are included in the present invention. Likewise, antibodies and fragements that bind only species of *Enterococcus*, i.e. antibodies and fragements that do not bind bacteria from any genus other than *Enterococcus*, are included in the present invention.

### 3. Diagnostic and Detection Assays and Kits

The present invention further relates to methods for assaying enterococcal infection in an animal by detecting the expression of genes encoding enterococcal polypeptides of the present invention. The methods comprise analyzing tissue or body fluid from the animal for *Enterococcus*-specific antibodies, nucleic acids, or proteins. Analysis of nucleic acid specific to *Enterococcus* is assayed by PCR or hybridization techniques using nucleic acid sequences of the present invention as either hybridization probes or primers. *See, e.g.,* Sambrook et al. Molecular cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, 2nd ed., 1989, page 54 reference); Eremeeva et al. (1994) J. Clin. Microbiol. 32:803-810 (describing differentiation among spotted fever group *Rickettsiae* species by analysis of restriction fragment length polymorphism of PCR-amplified DNA) and Chen et al. 1994 J. Clin. Microbiol. 32:589-595 (detecting *B. burgdorferi* nucleic acids *via* PCR).

Where diagnosis of a disease state related to infection with *Enterococcus* has already been made, the present invention is useful for monitoring progression or regression of the disease state whereby patients exhibiting enhanced *Enterococcus* gene expression will experience a worse clinical outcome relative to patients expressing these gene(s) at a lower level.

By "biological sample" is intended any biological sample obtained from an animal, cell line, tissue culture, or other source which contains *Enterococcus* polypeptide, mRNA, or DNA. Biological samples include body fluids (such as saliva, blood, plasma, urine, mucus, synovial fluid, etc.) tissues (such as muscle, skin, and cartilage) and any other biological source suspected of containing *Enterococcus* polypeptides or nucleic acids. Methods for obtaining biological samples such as tissue are well known in the art.

The present invention is useful for detecting diseases related to *Enterococcus* infections in animals. Preferred animals include monkeys, apes, cats, dogs, birds, cows, pigs, mice, horses, rabbits and humans. Particularly preferred are humans.

Total RNA can be isolated from a biological sample using any suitable technique such as the single-step guanidinium-thiocyanate-phenol-chloroform method described in Chomczynski et al. (1987) Anal. Biochem. 162:156-159. mRNA encoding *Enterococcus* polypeptides having sufficient homology to the nucleic acid sequences identified in SEQ ID NOS:1-982 to allow for hybridization between complementary sequences are then assayed using any appropriate method. These include Northern blot analysis, S1 nuclease mapping, the polymerase chain reaction (PCR), reverse transcription in combination with the polymerase chain reaction (RT-PCR), and reverse transcription in combination with the ligase chain reaction (RT-LCR).

Northern blot analysis can be performed as described in Harada et al. (1990) Cell 63:303-312. Briefly, total RNA is prepared from a biological sample as described above. For the Northern blot, the RNA is denatured in an appropriate buffer (such as glyoxal/dimethyl sulfoxide/sodium phosphate buffer), subjected to agarose gel electrophoresis, and transferred onto a nitrocellulose filter. After the RNAs have been linked to the filter by a UV linker, the filter is prehybridized in a solution containing formamide, SSC, Denhardt's solution, denatured salmon sperm, SDS, and sodium phosphate buffer. A *E. faecalis* polynucleotide sequence shown in SEQ ID NOS:1-982 labeled according to any appropriate method (such as the $^{32}$P-multiprimed DNA labeling system (Amersham)) is used as probe. After hybridization overnight, the filter is washed and exposed to x-ray film. DNA for use as probe according to the present invention is described in the sections above and will preferably at least 15 nucleotides in length.

S1 mapping can be performed as described in Fujita et al. (1987) Cell 49:357-367. To prepare probe DNA for use in S1 mapping, the sense strand of an above-described *E. faecalis* DNA sequence of the present invention is used as a template to synthesize labeled antisense DNA. The antisense DNA can then be digested using an appropriate restriction endonuclease to generate further DNA probes of a desired length. Such antisense probes are useful for visualizing protected bands corresponding to the target mRNA (*i.e.*, mRNA encoding *Enterococcus* polypeptides).

Levels of mRNA encoding *Enterococcus* polypeptides are assayed, for *e.g.*, using the RT-PCR method described in Makino et al. (1990) Technique 2:295-301. By this method, the radioactivities of the "amplicons" in the polyacrylamide gel bands are linearly related to the initial concentration of the target mRNA. Briefly, this method involves adding total RNA isolated from a biological sample in a reaction mixture containing a RT primer and appropriate buffer. After incubating for primer annealing, the mixture can be supplemented with a RT buffer, dNTPs, DTT, RNase inhibitor and reverse transcriptase. After incubation to achieve reverse transcription of the RNA, the RT products are then subject to PCR using labeled primers. Alternatively, rather than labeling the primers, a labeled dNTP can be included in the PCR reaction mixture. PCR amplification can be performed in a DNA thermal cycler according to conventional techniques. After a suitable number of rounds to achieve amplification, the PCR reaction mixture is electrophoresed on a polyacrylamide gel. After drying the gel, the radioactivity of the appropriate bands (corresponding to the mRNA encoding the *Enterococcus* polypeptides of the present invention) are quantified using an imaging analyzer. RT and PCR reaction ingredients and conditions, reagent and gel concentrations, and labeling methods are well known in the art. Variations on the RT-PCR method will be apparent to the skilled artisan. Other PCR methods that can detect the nucleic acid of the present invention can be found in PCR PRIMER: A LABORATORY MANUAL (C.W. Dieffenbach et al. eds., Cold Spring Harbor Lab Press, 1995).

The polynucleotides of the present invention, including both DNA and RNA, may be used to detect polynucleotides of the present invention or Enterococcal species including *E. faecalis* using bio chip technology. The present invention includes both high density chip arrays (>1000 oligonucleotides per cm$^2$) and low density chip arrays (<1000 oligonucleotides per $cm^2$). Bio chips comprising arrays of polynucleotides of the present invention may be used to detect Enterococcal species, including *E. faecalis*, in biological and environmental samples and to diagnose an animal, including humans, with an *E. faecalis* or other Enterococcal infection. The bio chips of the present invention may comprise polynucleotide sequences of other pathogens including bacteria, viral, parasitic, and fungal polynucleotide sequences, in addition to the polynucleotide sequences of the present invention, for use in rapid diffenertial pathogenic detection and diagnosis. The bio chips can also be used to monitor an *E. faecalis* or other Enterococcal infections and to monitor the genetic changes (deletions, insertions, mismatches, etc.) in response to drug therapy in the clinic and drug development in the laboratory. The bio chip technology comprising arrays of polynucleotides of the present invention may also be used to simultaneously monitor the expression of a multiplicity of genes, including those of the present invention. The polynucleotides used to comprise a selected array may be specified in the same manner as for the fragements, i.e, by their 5' and 3' positions or length in contigious base pairs and include from. Methods and particular uses of the polynucleotides of the present invention to detect Enterococcal species, including *E. faecalis*, using bio chip technology include those known in the art and those of: U.S. Patent Nos. 5510270, 5545531, 5445934, 5677195, 5532128, 5556752, 5527681, 5451683, 5424186, 5607646, 5658732 and World Patent Nos. WO/9710365, WO/9511995, WO/9743447, WO/9535505, each incorporated herein in their entireties.

Biosensors using the polynucleotides of the present invention may also be used to detect, diagnose, and monitor *E. faecalis* or other Enterococcal species and infections thereof. Biosensors using the polynucleotides of the present invention may also be used to detect particular polynucleotides of the present invention. Biosensors using the polynucleotides of the present invention may also be used to monitor the genetic changes (deletions, insertions, mismatches, etc.) in response to drug therapy in the clinic and drug development in the laboratory. Methods and particular uses of the polynucleotides of the present invention to detect Enterococcal species, including *E. faecalis*, using biosenors include those known in the art and those of: U.S. Patent Nos 5721102, 5658732, 5631170, and World Patent Nos. WO97/35011, WO/9720203, each incorporated herein in their entireties.

Thus, the present invention includes both bio chips and biosensors comprising polynucleotides of the present invention and methods of their use.

Assaying *Enterococcus* polypeptide levels in a biological sample can occur using any art-known method, such as antibody-based techniques. For example, *Enterococcus* polypeptide expression in tissues can be studied with classical immunohistological methods. In these, the specific recognition is provided by the primary antibody (polyclonal or monoclonal) but the secondary detection system can utilize fluorescent, enzyme, or other conjugated secondary antibodies. As a result, an immunohistological staining of tissue section for pathological examination is obtained. Tissues can also be extracted, *e.g.*, with urea and neutral detergent, for the liberation of *Enterococcus* polypeptides for Western-blot or dot/slot assay. *See, e.g.,* Jalkanen, M. et al. (1985) J. Cell. Biol. 101:976-985; Jalkanen, M. et al. (1987) J. Cell . Biol. 105:3087-3096. In this technique, which is based on the use of cationic solid phases, quantitation of a *Enterococcus* polypeptide can be accomplished using an isolated *Enterococcus* polypeptide as a standard. This technique can also be applied to body fluids.

Other antibody-based methods useful for detecting *Enterococcus* polypeptide gene expression include immunoassays, such as the ELISA and the radioimmunoassay (RIA). For example, a *Enterococcus* polypeptide-specific monoclonal antibodies can be used both as an immunoabsorbent and as an enzyme-labeled probe to detect and quantify a *Enterococcus* polypeptide. The amount of a *Enterococcus* polypeptide present in the sample can be calculated by reference to the amount present in a standard preparation using a linear regression computer algorithm. Such an ELISA is described in Iacobelli et al. (1988) Breast Cancer Research and Treatment 11:19-30. In another ELISA assay, two distinct specific monoclonal antibodies can be used to detect *Enterococcus* polypeptides in a body fluid. In this assay, one of the antibodies is used as the immunoabsorbent and the other as the enzyme-labeled probe.

The above techniques may be conducted essentially as a "one-step" or "two-step" assay. The "one-step" assay involves contacting the *Enterococcus* polypeptide with immobilized antibody and, without washing, contacting the mixture with the labeled antibody. The "two-step" assay involves washing before contacting the mixture with the labeled antibody. Other conventional methods may also be employed as suitable. It is usually desirable to immobilize one component of the assay system on a support, thereby allowing other components of the system to be brought into contact with the component and readily removed from the sample. Variations of the above and other immunological methods included in the present invention can also be found in Harlow et al., ANTIBODIES: A LABORATORY MANUAL, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988).

Suitable enzyme labels include, for example, those from the oxidase group, which catalyze the production of hydrogen peroxide by reacting with substrate. Glucose oxidase is particularly preferred as it has good stability and its substrate (glucose) is readily available. Activity of an oxidase label may be assayed by measuring the concentration of hydrogen peroxide formed by the enzyme-labeled antibody/substrate reaction. Besides enzymes, other suitable labels include radioisotopes, such as iodine ($^{125}$I, $^{121}$I), carbon

($^{14}C$), sulphur ($^{35}S$), tritium ($^{3}H$), indium ($^{112}In$), and technetium ($^{99m}Tc$), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

Further suitable labels for the *Enterococcus* polypeptide-specific antibodies of the present invention are provided below. Examples of suitable enzyme labels include malate dehydrogenase, Enterococcal nuclease, delta-5-steroid isomerase, yeast-alcohol dehydrogenase, alpha-glycerol phosphate dehydrogenase, triose phosphate isomerase, peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase, and acetylcholine esterase.

Examples of suitable radioisotopic labels include $^{3}H$, $^{111}In$, $^{125}I$, $^{131}I$, $^{32}P$, $^{35}S$, $^{14}C$, $^{51}Cr$, $^{57}To$, $^{58}Co$, $^{59}Fe$, $^{75}Se$, $^{152}Eu$, $^{90}Y$, $^{67}Cu$, $^{217}Ci$, $^{211}At$, $^{212}Pb$, $^{47}Sc$, $^{109}Pd$, etc. $^{111}In$ is a preferred isotope where *in vivo* imaging is used since its avoids the problem of dehalogenation of the $^{125}I$ or $^{131}I$-labeled monoclonal antibody by the liver. In addition, this radionucleotide has a more favorable gamma emission energy for imaging. *See, e.g.*, Perkins et al. (1985) Eur. J. Nucl. Med. 10:296-301; Carasquillo et al. (1987) J. Nucl. Med. 28:281-287. For example, $^{111}In$ coupled to monoclonal antibodies with 1-(P-isothiocyanatobenzyl)-DPTA has shown little uptake in non-tumors tissues, particularly the liver, and therefore enhances specificity of tumor localization. See, Esteban et al. (1987) J. Nucl. Med. 28:861-870.

Examples of suitable non-radioactive isotopic labels include $^{157}Gd$, $^{55}Mn$, $^{162}Dy$, $^{52}Tr$, and $^{56}Fe$.

Examples of suitable fluorescent labels include an $^{152}Eu$ label, a fluorescein label, an isothiocyanate label, a rhodamine label, a phycoerythrin label, a phycocyanin label, an allophycocyanin label, an o-phthaldehyde label, and a fluorescamine label.

Examples of suitable toxin labels include, *Pseudomonas* toxin, diphtheria toxin, ricin, and cholera toxin.

Examples of chemiluminescent labels include a luminal label, an isoluminal label, an aromatic acridinium ester label, an imidazole label, an acridinium salt label, an oxalate ester label, a luciferin label, a luciferase label, and an aequorin label.

Examples of nuclear magnetic resonance contrasting agents include heavy metal nuclei such as Gd, Mn, and iron.

Typical techniques for binding the above-described labels to antibodies are provided by Kennedy et al. (1976) Clin. Chim. Acta 70:1-31, and Schurs et al. (1977) Clin. Chim. Acta 81:1-40. Coupling techniques mentioned in the latter are the glutaraldehyde method, the periodate method, the dimaleimide method, the m-maleimidobenzyl-N-hydroxy-succinimide ester method, all of which methods are incorporated by reference herein.

In a related aspect, the invention includes a diagnostic kit for use in screening serum containing antibodies specific against *E. faecalis* infection. Such a kit may include an isolated *E. faecalis* antigen comprising an epitope which is specifically immunoreactive with at least one anti-*E. faecalis* antibody. Such a kit also includes means for detecting the binding of said antibody to the antigen. In specific embodiments, the kit may include a recombinantly produced or chemically synthesized peptide or polypeptide antigen. The peptide or polypeptide antigen may be attached to a solid support.

In a more specific embodiment, the detecting means of the above-described kit includes a solid support to which said peptide or polypeptide antigen is attached. Such a kit may also include a non-attached reporter-labeled anti-human antibody. In this embodiment, binding of the antibody to the *E. faecalis* antigen can be detected by binding of the reporter labeled antibody to the anti-*E. faecalis* polypeptide antibody.

In a related aspect, the invention includes a method of detecting *E. faecalis* infection in a subject. This detection method includes reacting a body fluid, preferably serum, from the subject with an isolated *E. faecalis* antigen, and examining the antigen for the presence of bound antibody. In a specific embodiment, the method includes a polypeptide antigen attached to a solid support, and serum is reacted with the support. Subsequently, the support is reacted with a reporter-labeled anti-human antibody. The support is then examined for the presence of reporter-labeled antibody.

The solid surface reagent employed in the above assays and kits is prepared by known techniques for attaching protein material to solid support material, such as polymeric beads, dip sticks, 96-well plates or filter material. These attachment methods generally include non-specific adsorption of the protein to the support or covalent attachment of the protein , typically through a free amine group, to a chemically reactive group on the solid support, such as an activated carboxyl, hydroxyl, or aldehyde group. Alternatively, streptavidin coated plates can be used in conjunction with biotinylated antigen(s).

The polypeptides and antibodies of the present invention, including fragments thereof, may be used to detect Enterococcal species including *E. faecalis* using bio chip and biosensor technology. Bio chip and biosensors of the present invention may comprise the polypeptides of the present invention to detect antibodies, which specifically recognize Enterococcal species, including *E. faecalis*. Bio chip and biosensors of the present invention may also comprise antibodies which specifically recognize the polypeptides of the present invention to detect Enterococcal species, including *E. faecalis* or specific polypeptides of the present invention. Bio chips or biosensors comprising polypeptides or antibodies of the present invention may be used to detect

Enterococcal species, including *E. faecalis*, in biological and environmental samples and to diagnose an animal, including humans, with an *E. faecalis* or other Enterococcal infection. Thus, the present invention includes both bio chips and biosensors comprising polypeptides or antibodies of the present invention and methods of their use.

The bio chips of the present invention may further comprise polypeptide sequences of other pathogens including bacteria, viral, parasitic, and fungal polypeptide sequences, in addition to the polypeptide sequences of the present invention, for use in rapid diffenertial pathogenic detection and diagnosis. The bio chips of the present invention may further comprise antibodies or fragements thereof specific for other pathogens including bacteria, viral, parasitic, and fungal polypeptide sequences, in addition to the antibodies or fragements thereof of the present invention, for use in rapid diffenertial pathogenic detection and diagnosis. The bio chips and biosensors of the present invention may also be used to monitor an *E. faecalis* or other Enterococcal infection and to monitor the genetic changes (amio acid deletions, insertions, substitutions, etc.) in response to drug therapy in the clinic and drug development in the laboratory. The bio chip and biosensors comprising polypeptides or antibodies of the present invention may also be used to simultaneously monitor the expression of a multiplicity of polypeptides, including those of the present invention. The polypeptides used to comprise a bio chip or biosensor of the present invention may be specified in the same manner as for the fragements, i.e, by their N-terminal and C-terminal positions or length in contigious amino acid residue. Methods and particular uses of the polypeptides and antibodies of the present invention to detect Enterococcal species, including *E. faecalis*, or specific polypeptides using bio chip and biosensor technology include those known in the art, those of the U.S. Patent Nos. and World Patent Nos. listed above for bio chips and biosensors using polynucleotides of the present invention, and those of: U.S. Patent Nos. 5658732, 5135852, 5567301, 5677196, 5690894 and World Patent Nos. WO9729366, WO9612957, each incorporated herein in their entireties.

### 4. Screening Assay for Binding Agents

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Enterococcus faecalis* fragment and contigs herein described.

In general, such methods comprise steps of:

(a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the *Enterococcus faecalis* genome; and

(b) determining whether the agent binds to said protein or said fragment.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, "Application of Synthetic Peptides: Antisense Peptides," in *Synthetic Peptides, A User's Guide*, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, *Biochemistry 28:*9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, *Nucl. Acids Res. 6*:3073 (1979); Cooney *et al.*, *Science 241*:456 (1988); and Dervan *et al.*, *Science 251:*1360 (1991)) or to the mRNA itself (antisense - Okano, *J. Neurochem. 56:*560 (1991); *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression,* CRC Press, Boca Raton, FL (1988)). Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

### 5. Pharmaceutical Compositions and Vaccines

The present invention further provides pharmaceutical agents which can be used to modulate the growth or pathogenicity of *Enterococcus faecalis*, or another related organism, *in vivo* or *in vitro*. As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical compositions. As used herein, the "pharmaceutical agents of the present invention" refers the pharmaceutical agents which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

As used herein, a pharmaceutical agent is said to "modulate the growth and/or pathogenicity of *Enterococcus faecalis* or a related organism, *in vivo* or *in vitro*," when the agent reduces the rate of growth, rate of division, or viability of the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth by binding to an important protein thus blocking the biological activity of the protein, while other agents may bind to a component of the outer surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs of the present invention and serve as a vaccine. The development and use of a vaccine based on outer membrane components are well known in the art.

As used herein, a "related organism" is a broad term which refers to any organism whose growth can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a vaccine. As such, related organisms do not need to be bacterial but may be fungal or viral pathogens.

The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body weight daily, taking into account the routes of administration, symptoms, *etc.*

The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, *etc.* The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, *etc.* Moieties capable of mediating such effects are disclosed in, among other sources, REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal stability, biological half-life, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers also may be effected in this way and can be assayed by methods well known to the skilled artisan.

The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (*e.g.*, inhalation, intravenously, intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple injections.

In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, *etc.* In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from about 1 pg/kg to 10 mg/kg (body weight of patient), although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each compound, or (2) the serum concentrations of each compound can be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

The administration of the agent(s) of the invention may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organisms growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic administration of the compound(s) serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

The agents of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in a mixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, *e.g.*, human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16th Ed., Osol, A., Ed., Mack Publishing, Easton PA (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention. The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to control the duration of action by controlled release preparations is to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-microcapsules and poly(methylmethacylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES (1980).

The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

The present invention also provides vaccines comprising one or more polypeptides of the present invention. Heterogeneity in the composition of a vaccine may be provided by combining *E. faecalis* polypeptides of the present invention. Multi-component vaccines of this type are desirable because they are likely to be more effective in eliciting protective immune responses against multiple species and strains of the *Enterococcus* genus than single polypeptide vaccines.

Multi-component vaccines are known in the art to elicit antibody production to numerous immunogenic components. *See, e.g.*, Decker et al. (1996) J. Infect. Dis. 174:S270-275. In addition, a hepatitis B, diphtheria, tetanus, pertussis tetravalent vaccine has recently been demonstrated to elicit protective levels of antibodies in human infants against all four pathogenic agents. *See, e.g.*, Aristegui, J. et al. (1997) Vaccine 15:7-9.

The present invention in addition to single-component vaccines includes multi-component vaccines. These vaccines comprise more than one polypeptide, immunogen or antigen. Thus, a multi-component vaccine would be a vaccine comprising more than one of the *E. faecalis* polypeptides of the present invention.

Further within the scope of the invention are whole cell and whole viral vaccines. Such vaccines may be produced recombinantly and involve the expression of one or more of the *E. faecalis* polypeptides described in SEQ ID NOS:1-982. For example, the *E.*

*faecalis* polypeptides of the present invention may be either secreted or localized intracellular, on the cell surface, or in the periplasmic space. Further, when a recombinant virus is used, the *E. faecalis* polypeptides of the present invention may, for example, be localized in the viral envelope, on the surface of the capsid, or internally within the capsid. Whole cells vaccines which employ cells expressing heterologous proteins are known in the art. *See, e.g.*, Robinson, K. et al. (1997) Nature Biotech. 15:653-657; Sirard, J. et al. (1997) Infect. Immun. 65:2029-2033; Chabalgoity, J. et al. (1997) Infect. Immun. 65:2402-2412 . These cells may be administered live or may be killed prior to administration. Chabalgoity, J. et al., *supra*, for example, report the successful use in mice of a live attenuated *Salmonella* vaccine strain which expresses a portion of a platyhelminth fatty acid-binding protein as a fusion protein on its cells surface.

A multi-component vaccine can also be prepared using techniques known in the art by combining one or more *E. faecalis* polypeptides of the present invention, or fragments thereof, with additional non-Enterococcal components (*e.g.*, diphtheria toxin or tetanus toxin, and/or other compounds known to elicit an immune response). Such vaccines are useful for eliciting protective immune responses to both members of the *Enterococcus* genus and non-Enterococcal pathogenic agents.

The vaccines of the present invention also include DNA vaccines. DNA vaccines are currently being developed for a number of infectious diseases. *See, et al.*, Boyer, et al. (1997) Nat. Med. 3:526-532; reviewed in Spier, R. (1996) Vaccine 14:1285-1288. Such DNA vaccines contain a nucleotide sequence encoding one or more *E. faecalis* polypeptides of the present invention oriented in a manner that allows for expression of the subject polypeptide. For example, the direct administration of plasmid DNA encoding *B. burgdorgeri* OspA has been shown to elicit protective immunity in mice against borrelial challenge. *See,* Luke et al. (1997) J. Infect. Dis. 175:91-97.

The present invention also relates to the administration of a vaccine which is co-administered with a molecule capable of modulating immune responses. Kim et al. (1997) Nature Biotech. 15:641-646, for example, report the enhancement of immune responses produced by DNA immunizations when DNA sequences encoding molecules which stimulate the immune response are co-administered. In a similar fashion, the vaccines of the present invention may be co-administered with either nucleic acids encoding immune modulators or the immune modulators themselves. These immune modulators include granulocyte macrophage colony stimulating factor (GM-CSF) and CD86.

The vaccines of the present invention may be used to confer resistance to Enterococcal infection by either passive or active immunization. When the vaccines of the present invention are used to confer resistance to Enterococcal infection through active immunization, a vaccine of the present invention is administered to an animal to elicit a protective immune response which either prevents or attenuates a Enterococcal infection. When the vaccines of the present invention are used to confer resistance to Enterococcal infection through passive immunization, the vaccine is provided to a host animal (*e.g.*, human, dog, or mouse), and the antisera elicited by this antisera is recovered and directly provided to a recipient suspected of having an infection caused by a member of the *Enterococcus* genus.

The ability to label antibodies, or fragments of antibodies, with toxin molecules provides an additional method for treating Enterococcal infections when passive immunization is conducted. In this embodiment, antibodies, or fragments of antibodies, capable of recognizing the *E. faecalis* polypeptides disclosed herein, or fragments thereof, as well as other *Enterococcus* proteins, are labeled with toxin molecules prior to their administration to the patient. When such toxin derivatized antibodies bind to *Enterococcus* cells, toxin moieties will be localized to these cells and will cause their death.

The present invention thus concerns and provides a means for preventing or attenuating a Enterococcal infection resulting from organisms which have antigens that are recognized and bound by antisera produced in response to the polypeptides of the present invention. As used herein, a vaccine is said to prevent or attenuate a disease if its administration to an animal results either in the total or partial attenuation (*i.e.*, suppression) of a symptom or condition of the disease, or in the total or partial immunity of the animal to the disease.

The administration of the vaccine (or the antisera which it elicits) may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the compound(s) are provided in advance of any symptoms of Enterococcal infection. The prophylactic administration of the compound(s) serves to prevent or attenuate any subsequent infection. When provided therapeutically, the compound(s) is provided upon or after the detection of symptoms which indicate that an animal may be infected with a member of the *Enterococcus* genus. The therapeutic administration of the compound(s) serves to attenuate any actual infection. Thus, the *E. faecalis* polypeptides, and fragments thereof, of the present invention may be provided either prior to the onset of infection (so as to prevent or attenuate an anticipated infection) or after the initiation of an actual infection.

The polypeptides of the invention, whether encoding a portion of a native protein or a functional derivative thereof, may be administered in pure form or may be coupled to a macromolecular carrier. Example of such carriers are proteins and carbohydrates. Suitable proteins which may act as macromolecular carrier for enhancing the immunogenicity of the polypeptides of the present invention include keyhole limpet hemacyanin (KLH) tetanus toxoid, pertussis toxin, bovine serum albumin, and ovalbumin. Methods for coupling the polypeptides of the present invention to such macromolecular carriers are disclosed in Harlow et al., ANTIBODIES: A LABORATORY MANUAL, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988).

A composition is said to be "pharmacologically or physiologically acceptable" if its administration can be tolerated by a recipient animal and is otherwise suitable for administration to that animal. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

While in all instances the vaccine of the present invention is administered as a pharmacologically acceptable compound, one skilled in the art would recognize that the composition of a pharmacologically acceptable compound varies with the animal to which it is administered. For example, a vaccine intended for human use will generally not be co-administered with Freund's adjuvant. Further, the level of purity of the *E. faecalis* polypeptides of the present invention will normally be higher when administered to a human than when administered to a non-human animal.

As would be understood by one of ordinary skill in the art, when the vaccine of the present invention is provided to an animal, it may be in a composition which may contain salts, buffers, adjuvants, or other substances which are desirable for improving the efficacy of the composition. Adjuvants are substances that can be used to specifically augment a specific immune response. These substances generally perform two functions: (1) they protect the antigen(s) from being rapidly catabolized after administration and (2) they nonspecifically stimulate immune responses.

Normally, the adjuvant and the composition are mixed prior to presentation to the immune system, or presented separately, but into the same site of the animal being immunized. Adjuvants can be loosely divided into several groups based upon their composition. These groups include oil adjuvants (for example, Freund's complete and incomplete), mineral salts (for example, $AlK(SO_4)_2$, $AlNa(SO_4)_2$, $AlNH_4(SO_4)$, silica, kaolin, and carbon), polynucleotides (for example, poly IC and poly AU acids), and certain natural substances (for example, wax D from *Mycobacterium tuberculosis*, as well as substances found in *Corynebacterium parvum*, or *Bordetella pertussis*, and members of the genus *Brucella*. Other substances useful as adjuvants are the saponins such as, for example, Quil A. (Superfos A/S, Denmark). Preferred adjuvants for use in the present invention include aluminum salts, such as $AlK(SO_4)_2$, $AlNa(SO_4)_2$, and $AlNH_4(SO_4)$.

Examples of materials suitable for use in vaccine compositions are provided in REMINGTON'S PHARMACEUTICAL SCIENCES 1324-1341 (A. Osol, ed, Mack Publishing Co, Easton, PA, (1980) (incorporated herein by reference).

The therapeutic compositions of the present invention can be administered parenterally by injection, rapid infusion, nasopharyngeal absorption (intranasopharangeally), dermoabsorption, or orally. The compositions may alternatively be administered intramuscularly, or intravenously. Compositions for parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Carriers or occlusive dressings can be used to increase skin permeability and enhance antigen absorption. Liquid dosage forms for oral administration may generally comprise a liposome solution containing the liquid dosage form. Suitable forms for suspending liposomes include emulsions, suspensions, solutions, syrups, and elixirs containing inert diluents commonly used in the art, such as purified water. Besides the inert diluents, such compositions can also include adjuvants, wetting agents, emulsifying and suspending agents, or sweetening, flavoring, or perfuming agents.

Therapeutic compositions of the present invention can also be administered in encapsulated form. For example, intranasal immunization using vaccines encapsulated in biodegradable microsphere composed of poly(DL-lactide-co-glycolide). *See*, Shahin, R. et al. (1995) Infect. Immun. 63:1195-1200. Similarly, orally administered encapsulated *Salmonella typhimurium* antigens can also be used. Allaoui-Attarki, K. et al. (1997) Infect. Immun. 65:853-857. Encapsulated vaccines of the present invention can be administered by a variety of routes including those involving contacting the vaccine with mucous membranes (*e.g.*, intranasally, intracolonicly, intraduodenally).

Many different techniques exist for the timing of the immunizations when a multiple administration regimen is utilized. It is possible to use the compositions of the invention more than once to increase the levels and diversities of expression of the immunoglobulin repertoire expressed by the immunized animal. Typically, if multiple immunizations are given, they will be given one to two months apart.

According to the present invention, an "effective amount" of a therapeutic composition is one which is sufficient to achieve a desired biological effect. Generally, the dosage needed to provide an effective amount of the composition will vary depending upon such factors as the animal's or human's age, condition, sex, and extent of disease, if any, and other variables which can be adjusted by one of ordinary skill in the art.

The antigenic preparations of the invention can be administered by either single or multiple dosages of an effective amount. Effective amounts of the compositions of the invention can vary from 0.01-1,000 μg/ml per dose, more preferably 0.1-500 μg/ml per dose, and most preferably 10-300 μg/ml per dose.

### 6. Shot-Gun Approach to Megabase DNA Sequencing

The present invention further demonstrates that a large genome can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing protocols.

Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present disclosure.

## ILLUSTRATIVE EXAMPLES

## LIBRARIES AND SEQUENCING

### 1. Shotgun Sequencing Probability Analysis

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Landerman and Waterman, *Genomics 2:*231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability, P0, that any given base in a sequence of size L, in nucleotides, is not sequenced after a certain amount, n, in nucleotides, of random sequence has been determined can be calculated by the equation P0 = e-m, where m is L/n, the fold coverage. For instance, for a genome of 2.8 Mb, m=1 when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point, P0 = e-1 = 0.37. The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence L has not been determined and, therefore, is equivalent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage, approximately 37% of a polynucleotide of size L, in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a 2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

Similarly, the total gap length, G, is determined by the equation G = Le-m, and the average gap size, g, follows the equation, g = L/n. Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a polynucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, *Genomics* 2: 231 (1988).

### 2. Random Library Construction

In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

*Enterococcus faecalis* DNA is prepared by phenol extraction. A mixture containing 200 μg DNA in 1.0 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 50% glycerol is processed through a nebulizer (IPI Medical Products) with a stream of nitrogen adjusted to 35 Kpa for 2 minutes. The sonicated DNA is ethanol precipitated and redissolved in 500 μl TE buffer.

To create blunt-ends, a 100 μl aliquot of the resuspended DNA is digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200 μl BAL31 buffer. The digested DNA is phenol-extracted, ethanol-precipitated, redissolved in 100 μl TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size is excised from the gel, and the LGT agarose is melted and the resulting solution is extracted with phenol to separate the agarose from the DNA. DNA is ethanol precipitated and redissolved in 20 μl of TE buffer for ligation to vector.

A two-step ligation procedure is used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50 ul) contains 2 μg of DNA fragments, 2 μg pUC18 DNA (Pharmacia) cut with SmaI and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and is incubated at 14°C for 4 hr. The ligation mixture then is phenol extracted and ethanol precipitated, and the precipitated DNA is dissolved in 20 μl TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder are visualized by ethidium bromide-staining and UV illumination and identified by size as insert (I), vector (v), v+I, v+2i, v+3i, *etc.* The portion of the gel containing v+I DNA is excised and the v+I DNA is recovered and resuspended into 20 μl TE. The v+I DNA then is blunt-ended by T4 polymerase treatment for 5 min. at 37°C in a reaction mixture (50 ul) containing the v+I linears, 500 μM each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+I linears are dissolved in 20 μl TE. The final ligation to produce circles is carried out in a 50 μl reaction containing 5 μl of v+I linears and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture is stored at -20°C.

This two-stage procedure results in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

Since deviation from randomness can arise from propagation the DNA in the host, *E. coli* host cells deficient in all recombination and restriction functions (A. Greener, *Strategies 3 (1)*:5 (1990)) are used to prevent rearrangements, deletions, and loss of clones by restriction. Furthermore, transformed cells are plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating is carried out as follows. A 100 µl aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) is thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7 µl aliquot of 1.42 M beta-mercaptoethanol is added to the aliquot of cells to a final concentration of 25 mM. Cells are incubated on ice for 10 min. A 1 µl aliquot of the final ligation is added to the cells and incubated on ice for 30 min. The cells are heat pulsed for 30 sec. at 42°C and placed back on ice for 2 min. The outgrowth period in liquid culture is eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation mixture is plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml MgC12 (1 M), and 1 ml MgSO4/100 ml SOB agar. The 15 ml top layer is poured just prior to plating. Our titer is approximately 100 colonies/10 µl aliquot of transformation.

All colonies are picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products are deleted from the library, resulting in a slight increase in gap number over that expected.

### 3. Random DNA Sequencing

High quality double stranded DNA plasmid templates are prepared using a "boiling bead" method developed in collaboration with Advanced Genetic Technology Corp. (Gaithersburg, MD) (Adams *et al.*, *Science 252*:1651 (1991); Adams *et al.*, *Nature 355*:632 (1992)). Plasmid preparation is performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Template concentration is determined using Hoechst Dye and a Millipore Cytofluor. DNA concentrations are not adjusted, but low-yielding templates are identified where possible and not sequenced.

Templates are also prepared from an *Enterococcus faecalis* lambda genomic library in the vector DASH II (Stratagene). In particular, *Enterococcus faecalis* DNA (> 100 kb) is partially digested in a reaction mixture (200 ul) containing 50 µg DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23°C. The digested DNA was phenol-extracted and fractionated by sucrose density gradient centrifugation. Fractions of the sucrose gradient containing 15 to 25 kb are recovered in a final volume of 6 ul. One µl of fragments is used with 1 µl of lambda DASHII vector (Stratagene) in the recommended ligation reaction. One µl of the ligation mixture is used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract (Stratagene, #227711). Phage are plated directly without amplification from the packaging mixture (after dilution with 500 µl of recommended SM buffer and chloroform treatment). Yield is about 2.5x103 pfu/ul. An amplified library is prepared by infecting restructure NM539 host E. coli cells eitn approximately 1x104 phage particles and recovering the progeny phages particles. The recovered phage is stored frozen in 7% dimethylsulfoxide. The phage titer is approximately 1x109 pfu/ml.

For high throughput sequencing of individual lambda phage clones, liquid lysates (100 µl) are prepared from randomly selected plaques (from the unamplified library) and template is prepared by long-range PCR using T7 and T3 vector-specific primers.

Sequencing reactions are carried out on plasmid and/or PCR templates using the AB Catalyst LabStation with Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers (Adams *et al.*, *Nature 368*:474 (1994)). Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. T7 and T3 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are performed by eight individuals using an average of fourteen AB 373 DNA Sequencers per day. All sequencing reactions are analyzed using the Stretch modification of the AB 373, primarily using a 34 cm well-to-read distance. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

Richards *et al.*, Chapter 28 in AUTOMATED DNA SEQUENCING AND ANALYSIS, M. D. Adams, C. Fields, J. C. Venter, Eds., Academic Press, London, (1994) described the value of using sequence from both ends of sequencing templates to facilitate ordering of contigs in shotgun assembly projects of lambda and cosmid clones. We balance the desirability of both-end sequencing (including the reduced cost of lower total number of templates) against shorter read-lengths for sequencing reactions performed with the M13RP1 (reverse) primer compared to the M13-21 (forward) primer. Approximately one-half of the templates are sequenced from both ends. Random reverse sequencing reactions are done based on successful forward sequencing reactions. Some M13RP1 sequences are obtained in a semi-directed fashion: M13-21: sequences pointing outward at the ends of contigs are chosen for M13RP1 sequencing in an effort to specifically order contigs.

### 4. Protocol for Automated Cycle Sequencing

The sequencing was carried out using ABI Catalyst robots and AB 373 Automated DNA Sequencers. The Catalyst robot is a publicly available sophisticated pipetting and temperature control robot which has been developed specifically for DNA sequencing reactions. The Catalyst combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-labelled sequencing primers, and reaction buffer. Reaction mixes and templates are combined in the wells of an aluminum 96-well thermocycling plate. Thirty consecutive cycles of linear amplification (i.*e.*., one primer synthesis) steps are performed including denaturation, annealing of primer and template, and extension; *i.e.*, DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without the need for an oil overlay.

Two sequencing protocols are used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer is labelled with a different fluorescent dye, permitting the four individual reactions to be combined into one lane of the 373 DNA Sequencer for electrophoresis, detection, and base-calling. ABI currently supplies pre-mixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR- generated templates with both dye-primers and dye-terminators with approximately equal fidelity, although plasmid templates generally give longer usable sequences.

Thirty-two reactions are loaded per AB373 Sequencer each day, for a total of 960 samples. Electrophoresis is run overnight following the manufacturer's protocols, and the data is collected for twelve hours. Following electrophoresis and fluorescence detection, the ABI 373 performs automatic lane tracking and base-calling. The lane-tracking is confirmed visually. Each sequence electropherogram (or fluorescence lane trace) is inspected visually and assessed for quality. Trailing sequences of low quality are removed and the sequence itself is loaded via software to a Sybase database (archived daily to 8mm tape). Leading vector polylinker sequence is removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 are around 400 bp and depend mostly on the quality of the template used for the sequencing reaction. ABI 373 Sequencers converted to Stretch Liners provide a longer electrophoresis path prior to fluorescence detection and increase the average number of usable bases to 500-600 bp.

## INFORMATICS

### 1. Data Management

A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database management system and was designed to automate data flow wherever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen is based on a Unix platform, it was necessary to design and implement a variety of multi- user, client-server applications which allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

### 2. Assembly

An assembly engine (TIGR Assembler) developed for the rapid and accurate assembly of thousands of sequence fragments is employed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than 104 fragments, the algorithm builds a hash table of 10 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology 164:*765 (1988)). The contig is extended by the fragment only if strict criteria for the quality of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched end, and the minimum percentage match. These criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repetitive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of alignments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain range of base pairs (definable for each clone based on the known clone size range for a given library).

The process resulted in 982 contigs as represented by SEQ ID NOs:1-982.

### 3. Identifying Genes

The predicted coding regions of the *Enterococcus faecalis* genome were initially defined with the program GeneMark, which finds ORFs using a probabilistic classification technique. The predicted coding region sequences were used in searches against a database of all *Enterococcus faecali* nucleotide sequences from GenBank (March, 1997), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and compared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases. ORFs that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3.

## ILLUSTRATIVE APPLICATIONS

### 1. Production of an Antibody to a *Enterococcus faecalis* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can be chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows.

### 2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., *Nature 256*:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., *Meth. Enzymol. 70:*419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al., Basic Methods in Molecular Biology,* Elsevier, New York. Section 21-2 (1989).

### 3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al., J. Clin.* Endocrinol. Metab. 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in: *Handbook of Experimental Immunology,* Wier, D., ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0. 1 to 0. 2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980)

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi- quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, antibodies are useful in various animal models of enterococcal disease as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a potential immunotherapeutic or immunoprophylactic reagent.

### 4. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Enterococcus faecalis* genome, such as those of Tables 1-3 and SEQ ID NOS:1-982 can be used, in accordance with the present invention, to prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

### 5. Isolation of a Selected DNA Clone From the Deposited Sample of E. faecalis

Three approaches can be used to isolate a *E. faecalis* clone comprising a polynucleotide of the present invention from any *E. faecalis* genomic DNA library. The *E. faecalis* strain V586 has been deposited as a convienent source for obtaining a *E. faecalis* strain although a wide varity of strains *E. faecalis* strains can be used which are known in the art.

*E. faecalis* genomic DNA is prepared using the following method. A 20ml overnight bacterial culture grown in a rich medium (e.g., Trypticase Soy Broth, Brain Heart Infusion broth or Super broth), pelleted, ished two times with TES (3OmM Tris-pH 8.0, 25mM EDTA, 50mM NaCl), and resuspended in 5ml high salt TES (2.5M NaCl). Lysostaphin is added to final concentration of approx 50ug/ml and the mixture is rotated slowly 1 hour at 37C to make protoplast cells. The solution is then placed in incubator (or place in a shaking water bath) and warmed to 55C. Five hundred micro liter of 20% sarcosyl in TES (final concentration 2%) is then added to lyse the cells. Next, guanidine HCl is added to a final concentration of 7M (3.69g in 5.5 ml). The mixture is swirled slowly at 55C for 60-90 min (solution should clear). A CsCl gradient is then set up in

SW41 ultra clear tubes using 2.0ml 5.7M CsCl and overlaying with 2.85M CsCl. The gradient is carefully overlayed with the DNA-containing GuHCl solution. The gradient is spun at 30,000 rpm, 20C for 24 hr and the lower DNA band is collected. The volume is increased to 5 ml with TE buffer. The DNA is then treated with protease K (10 ug/ml) overnight at 37 C, and precipitated with ethanol. The precipitated DNA is resuspended in a desired buffer.

In the first method, a plasmid is directly isolated by screening a plasmid *E. faecalis* genomic DNA library using a polynucleotide probe corresponding to a polynucleotide of the present invention. Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with $^{32}$P-$\gamma$-ATP using T4 polynucleotide kinase and purified according to routine methods. (*See, e.g.*, Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The library is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art. *See, e.g.*, Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL (Cold Spring Harbor, N.Y. 2nd ed. 1989); Ausubel et al., CURRENT PROTOCALS IN MOLECULAR BIOLOGY (John Wiley and Sons, N.Y. 1989). The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening. *See, e.g.*, Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL (Cold Spring Harbor, N.Y. 2nd ed. 1989); Ausubel et al., CURRENT PROTOCALS IN MOLECULAR BIOLOGY (John Wiley and Sons, N.Y. 1989) or other techniques known to those of skill in the art.

Alternatively, two primers of 15-25 nucleotides derived from the 5' and 3' ends of a polynucleotide of SEQ ID NOS:1-982 are synthesized and used to amplify the desired DNA by PCR using a *E. faecalis* genomic DNA prep as a template. PCR is carried out under routine conditions, for instance, in 25 $\mu$l of reaction mixture with 0.5 ug of the above DNA template. A convenient reaction mixture is 1.5-5 mM $MgCl_2$, 0.01% (w/v) gelatin, 20 $\mu$M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the

DNA product.

Finally, overlapping oligos of the DNA sequences of SEQ ID NOS:1-982 can be chemically synthesized and used to generate a nucleotide sequence of desired length using PCR methods known in the art.

### 6(a). Expression and Purification Enterococcal polypeptides in E. coli

The bacterial expression vector pQE60 was used for bacterial expression of some of the polypeptide fragements of the present invention which were used in the soft tissue and systemic infection models discussed below. (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311). pQE60 encodes ampicillin antibiotic resistance ("Ampr") and contains a bacterial origin of replication ("ori"), an IPTG inducible promoter, a ribosome binding site ("RBS"), six codons encoding histidine residues that allow affinity purification using nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin (QIAGEN, Inc., *supra*) and suitable single restriction enzyme cleavage sites. These elements are arranged such that an inserted DNA fragment encoding a polypeptide expresses that polypeptide with the six His residues (i.e., a "6 X His tag") covalently linked to the carboxyl terminus of that polypeptide.

The DNA sequence encoding the desired portion of a *E. faecalis* protein of the present invention was amplified from *E. faecalis* genomic DNA using PCR oligonucleotide primers which anneal to the 5' and 3' sequences coding for the portions of the *E. faecalis* polynucleotide shown in SEQ ID NOS:1-982. Additional nucleotides containing restriction sites to facilitate cloning in the pQE60 vector are added to the 5' and 3' sequences, respectively.

For cloning the mature protein, the 5' primer has a sequence containing an appropriate restriction site followed by nucleotides of the amino terminal coding sequence of the desired *E. faecalis* polynucleotide sequence in SEQ ID NOS:1-982. One of ordinary skill in the art would appreciate that the point in the protein coding sequence where the 5' and 3' primers begin may be varied to amplify a DNA segment encoding any desired portion of the complete protein shorter or longer than the mature form. The 3' primer has a sequence containing an appropriate restriction site followed by nucleotides complementary to the 3' end of the polypeptide coding sequence of SEQ ID NOS:1-982, excluding a stop codon, with the coding sequence aligned with the restriction site so as to maintain its reading frame with that of the six His codons in the pQE60 vector.

The amplified *E. faecalis* DNA fragment and the vector pQE60 were digested with restriction enzymes which recognize the sites in the primers and the digested DNAs were then ligated together. The *E. faecalis* DNA was inserted into the restricted pQE60 vector in a manner which places the *E. faecalis* protein coding region downstream from the IPTG-inducible promoter and in-frame with an initiating AUG and the six histidine codons.

The ligation mixture was transformed into competent *E. coli* cells using standard procedures such as those described by Sambrook et al., *supra*.. *E. coli* strain M15/rep4, containing multiple copies of the plasmid pREP4, which expresses the lac repressor and confers kanamycin resistance ("Kanr"), was used in carrying out the illustrative example described herein. This strain, which was only one of many that are suitable for expressing a *E. faecalis* polypeptide, is available commercially (QIAGEN, Inc., *supra*). Transformants were identified by their ability to grow on LB agar plates in the presence of ampicillin and kanamycin. Plasmid DNA was isolated from resistant colonies and the identity of the cloned DNA confirmed by restriction analysis, PCR and DNA sequencing.

Clones containing the desired constructs were grown overnight ("O/N") in liquid culture in LB media supplemented with both ampicillin (100 µg/ml) and kanamycin (25 µg/ml). The O/N culture was used to inoculate a large culture, at a dilution of approximately 1:25 to 1:250. The cells were grown to an optical density at 600 nm ("OD600") of between 0.4 and 0.6. Isopropyl-β-D-thiogalactopyranoside ("IPTG") was then added to a final concentration of 1 mM to induce transcription from the lac repressor sensitive promoter, by inactivating the lacI repressor. Cells subsequently were incubated further for 3 to 4 hours. Cells then were harvested by centrifugation.

The cells were then stirred for 3-4 hours at 4°C in 6M guanidine-HCl, pH 8. The cell debris was removed by centrifugation, and the supernatant containing the *E. faecalis* polypeptide was loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity were purified in a simple one-step procedure (for details see: The QIAexpressionist, 1995, QIAGEN, Inc., *supra*). Briefly the supernatant was loaded onto the column in 6 M guanidine-HCl, pH 8, the column was first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the *E. faecalis* polypeptide was eluted with 6 M guanidine-HCl, pH 5.

The purified protein was then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein could be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins can be eluted by the addition of 250 mM immidazole. Immidazole was removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein was stored at 4° C or frozen at -80° C.

Some of the polypeptide of the present invention were prepared using a non-denaturing protein purification method. For these polypeptides, the cell pellet from each liter of culture was resuspended in 25 mls of Lysis Buffer A at 4°C (Lysis Buffer A = 50 mM Na-phosphate, 300 mM NaCl, 10 mM 2-mercaptoethanol, 10% Glycerol, pH 7.5 with 1 tablet of Complete EDTA-free protease inhibitor cocktail (Boehringer Mannheim #1873580) per 50 ml of buffer). Absorbance at 550 nm was approximately 10-20 O.D./ml. The suspension was then put through three freeze/thaw cycles from -70°C (using a ethanol-dry ice bath) up to room temperature. The cells were lysed via sonication in short 10 sec bursts over 3 minutes at approximately 80W while kept on ice. The sonicated sample was then centrifuged at 15,000 RPM for 30 minutes at 4°C. The supernatant was passed through a column containing 1.0 ml of CL-4B resin to pre-clear the sample of any proteins that may bind to agarose non-specifically, and the flow-through fraction was collected.

The pre-cleared flow-through was applied to a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (Quiagen, Inc., *supra*). Proteins with a 6 X His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure. Briefly, the supernatant was loaded onto the column in Lysis Buffer A at 4°C, the column was first washed with 10 volumes of Lysis Buffer A until the A280 of the eluate returns to the baseline. Then, the column was washed with 5 volumes of 40 mM Imidazole (92% Lysis Buffer A / 8% Buffer B) (Buffer B = 50 mM Na-Phosphate, 300 mM NaCl, 10% Glycerol, 10 mM 2-mercaptoethanol, 500 mM Imidazole, pH of the final buffer should be 7.5). The protein was eluted off of the column with a series of increasing Imidazole solutions made by adjusting the ratios of Lysis Buffer A to Buffer B. Three different concentrations were used: 3 volumes of 75 mM Imidazole, 3 volumes of 150 mM Imidazole, 5 volumes of 500 mM Imidazole. The fractions containing the purified protein were analyzed using 8 %, 10 % or 14% SDS-PAGE depending on the protein size. The purified protein was then dialyzed 2X against phosphate-buffered saline (PBS) in order to place it into an easily workable buffer. The purified protein was stored at 4° C or frozen at -80°.

The following alternative method may be used to purify *E. faecalis* expressed in *E coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells are harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfuidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 x *g* for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 x *g* centrifugation for 15 min., the pellet is discarded and the *E. faecalis* polypeptide-containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 x *g*) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded *E. faecalis* polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 mm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the *E. faecalis* polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant $A_{280}$ monitoring of the effluent. Fractions containing the *E. faecalis* polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant *E. faecalis* polypeptide exhibits greater than 95% purity after the above refolding and purification steps. No major contaminant bands are observed from Commassie blue stained 16% SDS-PAGE gel when 5 μg of purified protein is loaded. The purified protein is also tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

### 6(b). Alternative Expression and Purification Enterococcal polypeptides in E. coli

Tthe vector pQE10 was alternatively used to clone and express some of the polypeptides of the present invention for use in the soft tissue and systemic infection models discussed below. The difference being such that an inserted DNA fragment encoding a polypeptide expresses that polypeptide with the six His residues (i.e., a "6 X His tag") covalently linked to the amino terminus of that polypeptide. The bacterial expression vector pQE10 (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311) was used in this example . The components of the pQE10 plasmid are arranged such that the inserted DNA sequence encoding a polypeptide of the present invention expresses the polypeptide with the six His residues (*i.e.*, a "6 X His tag")) covalently linked to the amino terminus.

The DNA sequences encoding the desired portions of a polypeptide of SEQ ID NOS:1-982 were amplified using PCR oligonucleotide primers from genomic *E. faecalis* DNA. The PCR primers anneal to the nucleotide sequences encoding the desired amino acid sequence of a polypeptide of the present invention. Additional nucleotides containing restriction sites to facilitate cloning in the pQE10 vector were added to the 5' and 3' primer sequences, respectively.

For cloning a polypeptide of the present invention, the 5' and 3' primers were selected to amplify their respective nucleotide coding sequences. One of ordinary skill in the art would appreciate that the point in the protein coding sequence where the 5' and 3' primers begins may be varied to amplify a DNA segment encoding any desired portion of a polypeptide of the present invention. The 5' primer was designed so the coding sequence of the 6 X His tag is aligned with the restriction site so as to maintain its reading frame with that of *E. faecalis* polypeptide. The 3' was designed to include an stop codon. The amplified DNA fragment was then cloned, and the protein expressed, as described above for the pQE60 plasmid.

The DNA sequences encoding the amino acid sequences of SEQ ID NOS:1-982 may also be cloned and expressed as fusion proteins by a protocol similar to that described directly above, wherein the pET-32b(+) vector (Novagen, 601 Science Drive, Madison, WI 53711) is preferentially used in place of pQE10.

The above methods are not limited to the polypeptide fragements actually produced. The above method, like the methods below, can be used to produce either full length polypeptides or desired fragements therof.

### 6(c). Alternative Expression and Purification of Enterococcal polypeptides in E. coli

The bacterial expression vector pQE60 is used for bacterial expression in this example (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311). However, in this example, the polypeptide coding sequence is inserted such that translation of the six His codons is prevented and, therefore, the polypeptide is produced with no 6 X His tag.

The DNA sequence encoding the desired portion of the *E. faecalis* amino acid sequence is amplified from an *E. faecalis* genomic DNA prep the deposited DNA clones using PCR oligonucleotide primers which anneal to the 5' and 3' nucleotide sequences corresponding to the desired portion of the *E. faecalis* polypeptides. Additional nucleotides containing restriction sites to facilitate cloning in the pQE60 vector are added to the 5' and 3' primer sequences.

For cloning a *E. faecalis* polypeptides of the present invention, 5' and 3' primers are selected to amplify their respective nucleotide coding sequences. One of ordinary skill in the art would appreciate that the point in the protein coding sequence where the 5' and 3' primers begin may be varied to amplify a DNA segment encoding any desired portion of a polypeptide of the present invention. The 3' and 5' primers contain appropriate restriction sites followed by nucleotides complementary to the 5' and 3' ends of the coding sequence respectively. The 3' primer is additionally designed to include an in-frame stop codon.

The amplified *E. faecalis* DNA fragments and the vector pQE60 are digested with restriction enzymes recognizing the sites in the primers and the digested DNAs are then ligated together. Insertion of the *E. faecalis* DNA into the restricted pQE60 vector places the *E. faecalis* protein coding region including its associated stop codon downstream from the IPTG-inducible promoter and in-frame with an initiating AUG. The associated stop codon prevents translation of the six histidine codons downstream of the insertion point.

The ligation mixture is transformed into competent *E. coli* cells using standard procedures such as those described by Sambrook et al. *E. coli* strain M15/rep4, containing multiple copies of the plasmid pREP4, which expresses the lac repressor and confers kanamycin resistance ("Kanr"), is used in carrying out the illustrative example described herein. This strain, which is only one of many that are suitable for expressing *E. faecalis* polypeptide, is available commercially (QIAGEN, Inc., *supra*). Transformants are identified by their ability to grow on LB plates in the presence of ampicillin and kanamycin. Plasmid DNA is isolated from resistant colonies and the identity of the cloned DNA confirmed by restriction analysis, PCR and DNA sequencing.

Clones containing the desired constructs are grown overnight ("O/N") in liquid culture in LB media supplemented with both ampicillin (100 µg/ml) and kanamycin (25 µg/ml). The O/N culture is used to inoculate a large culture, at a dilution of approximately 1:25 to 1:250. The cells are grown to an optical density at 600 nm ("OD600") of between 0.4 and 0.6. isopropyl-b-D-thiogalactopyranoside ("IPTG") is then added to a final concentration of 1 mM to induce transcription from the *lac* repressor sensitive promoter, by inactivating the lacI repressor. Cells subsequently are incubated further for 3 to 4 hours. Cells then are harvested by centrifugation.

To purify the *E. faecalis* polypeptide, the cells are then stirred for 3-4 hours at 4°C in 6M guanidine-HCl, pH 8. The cell debris is removed by centrifugation, and the supernatant containing the *E. faecalis* polypeptide is dialyzed against 50 mM Na-acetate buffer pH 6, supplemented with 200 mM NaCl. Alternatively, the protein can be successfully refolded by dialyzing it against 500 mM NaCl, 20% glycerol, 25 mM Tris/HCl pH 7.4, containing protease inhibitors. After renaturation the protein can be purified by ion exchange, hydrophobic interaction and size exclusion chromatography. Alternatively, an affinity chromatography step such as an antibody column can be used to obtain pure *E. faecalis* polypeptide. The purified protein is stored at 4° C or frozen at -80° C.

The following alternative method may be used to purify *E. faecalis* polypeptides expressed in *E coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells are harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells ware then lysed by passing the solution through a microfluidizer (Microfuidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 x *g* for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 x *g* centrifugation for 15 min., the pellet is discarded and the *E. faecalis* polypeptide-containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 x *g*) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded *E. faecalis* polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 μm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 mm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the *E. faecalis* polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant $A_{280}$ monitoring of the effluent. Fractions containing the *E. faecalis* polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant *E. faecalis* polypeptide exhibits greater than 95% purity after the above refolding and purification steps. No major contaminant bands are observed from Commassie blue stained 16% SDS-PAGE gel when 5 μg of purified protein is loaded. The purified protein is also tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

### 6(d). Cloning and Expression of E. faecalis in Other Bacteria

*E. faecalis* polypeptides can also be produced in: *E. faecalis* using the methods of S. Skinner et al., (1988) Mol. Microbiol. 2:289-297 or J. I. Moreno (1996) Protein Expr.

Purif. 8(3):332-340; *Lactobacillus* using the methods of C. Rush et al., 1997 Appl. Microbiol. Biotechnol. 47(5):537-542; or in *Bacillus subtilis* using the methods Chang et al., U.S. Patent No. 4,952,508.

### 7. Cloning and Expression in COS Cells

A *E. faecalis* expression plasmid is made by cloning a portion of the DNA encoding a *E. faecalis* polypeptide into the expression vector pDNAI/Amp or pDNAIII (which can be obtained from Invitrogen, Inc.). The expression vector pDNAI/amp contains: (1) an *E. coli* origin of replication effective for propagation in *E. coli* and other prokaryotic cells; (2) an ampicillin resistance gene for selection of plasmid-containing prokaryotic cells; (3) an SV40 origin of replication for propagation in eukaryotic cells; (4) a CMV promoter, a polylinker, an SV40 intron; (5) several codons encoding a hemagglutinin fragment (i.e., an "HA" tag to facilitate purification) followed by a termination codon and polyadenylation signal arranged so that a DNA can be conveniently placed under expression control of the CMV promoter and operably linked to the SV40 intron and the polyadenylation signal by means of restriction sites in the polylinker. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein described by Wilson et al. 1984 Cell 37:767. The fusion of the HA tag to the target protein allows easy detection and recovery of the recombinant protein with an antibody that recognizes the HA epitope. pDNAIII contains, in addition, the selectable neomycin marker.

A DNA fragment encoding a *E. faecalis* polypeptide is cloned into the polylinker region of the vector so that recombinant protein expression is directed by the CMV promoter. The plasmid construction strategy is as follows. The DNA from a *E. faecalis* genomic DNA prep is amplified using primers that contain convenient restriction sites, much as described above for construction of vectors for expression of *E. faecalis* in *E. coli*. The 5' primer contains a Kozak sequence, an AUG start codon, and nucleotides of the 5' coding region of the *E. faecalis* polypeptide. The 3' primer, contains nucleotides complementary to the 3' coding sequence of the *E. faecalis* DNA, a stop codon, and a convenient restriction site.

The PCR amplified DNA fragment and the vector, pDNAI/Amp, are digested with appropriate restriction enzymes and then ligated. The ligation mixture is transformed into an appropriate *E. coli* strain such as SURE™ (Stratagene Cloning Systems, La Jolla, CA 92037), and the transformed culture is plated on ampicillin media plates which then are incubated to allow growth of ampicillin resistant colonies. Plasmid DNA is isolated from resistant colonies and examined by restriction analysis or other means for the presence of the fragment encoding the *E. faecalis* polypeptide

For expression of a recombinant *E. faecalis* polypeptide, COS cells are transfected with an expression vector, as described above, using DEAE-dextran, as described, for instance, by Sambrook et al. (*supra*). Cells are incubated under conditions for expression of *E. faecalis* by the vector.

Expression of the *E. faecalis*-HA fusion protein is detected by radiolabeling and immunoprecipitation, using methods described in, for example Harlow et al., *supra*.. To this end, two days after transfection, the cells are labeled by incubation in media containing $^{35}$S-cysteine for 8 hours. The cells and the media are collected, and the cells are washed and the lysed with detergent-containing RIPA buffer: 150 mM NaCl, 1% NP-40, 0.1% SDS, 1% NP-40, 0.5% DOC, 50 mM TRIS, pH 7.5, as described by Wilson et al. (*supra* ). Proteins are precipitated from the cell lysate and from the culture media using an HA-specific monoclonal antibody. The precipitated proteins then are analyzed by SDS-PAGE and autoradiography. An expression product of the expected size is seen in the cell lysate, which is not seen in negative controls.

### 8. Cloning and Expression in CHO Cells

The vector pC4 is used for the expression of *E. faecalis* polypeptide in this example. Plasmid pC4 is a derivative of the plasmid pSV2-dhfr (ATCC Accession No. 37146). The plasmid contains the mouse DHFR gene under control of the SV40 early promoter. Chinese hamster ovary cells or other cells lacking dihydrofolate activity that are transfected with these plasmids can be selected by growing the cells in a selective medium (alpha minus MEM, Life Technologies) supplemented with the chemotherapeutic agent methotrexate. The amplification of the DHFR genes in cells resistant to methotrexate (MTX) has been well documented. *See, e.g.*, Alt et al., 1978, J. Biol. Chem. 253:1357-1370; Hamlin et al., 1990, Biochem. et Biophys. Acta, 1097:107-143; Page et al., 1991, Biotechnology 9:64-68. Cells grown in increasing concentrations of MTX develop resistance to the drug by overproducing the target enzyme, DHFR, as a result of amplification of the DHFR gene. If a second gene is linked to the DHFR gene, it is usually co-amplified and over-expressed. It is known in the art that this approach may be used to develop cell lines carrying more than 1,000 copies of the amplified gene(s). Subsequently, when the methotrexate is withdrawn, cell lines are obtained which contain the amplified gene integrated into one or more chromosome(s) of the host cell.

Plasmid pC4 contains the strong promoter of the long terminal repeat (LTR) of the Rouse Sarcoma Virus, for expressing a polypeptide of interest, Cullen, et al. (1985) Mol. Cell. Biol. 5:438-447; plus a fragment isolated from the enhancer of the immediate early gene of human cytomegalovirus (CMV), Boshart, et al., 1985, Cell 41:521-530. Downstream of the promoter are the following single restriction enzyme cleavage sites that allow the integration of the genes: *Bam* HI, *Xba* I, and *Asp* 718. Behind these cloning sites the plasmid contains the 3' intron and polyadenylation site of the rat preproinsulin gene. Other high efficiency promoters can also be used for the expression, e.g., the human ß-actin promoter, the SV40 early or late promoters or the long terminal repeats from other retroviruses, e.g., HIV and HTLVI. Clontech's Tet-Off and Tet-On gene expression systems and similar systems can be used to express the *E. faecalis* polypeptide in a regulated way in mammalian cells (Gossen et al., 1992, Proc. Natl. Acad. Sci. USA 89:5547-5551. For the polyadenylation of the mRNA other signals, e.g., from the human growth hormone or globin genes can be used as well. Stable cell lines carrying a gene of interest integrated into the chromosomes can also be selected upon co-transfection with a selectable marker such as gpt, G418 or hygromycin. It is advantageous to use more than one selectable marker in the beginning, e.g., G418 plus methotrexate.

The plasmid pC4 is digested with the restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel. The DNA sequence encoding the *E. faecalis* polypeptide is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of the desired portion of the gene. A 5' primer containing a restriction site, a Kozak sequence, an AUG start codon, and nucleotides of the 5' coding region of the *E. faecalis* polypeptide is synthesized and used. A 3' primer, containing a restriction site, stop codon, and nucleotides complementary to the 3' coding sequence of the *E. faecalis* polypeptides is synthesized and used. The amplified fragment is digested with the restriction endonucleases and then purified again on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC4 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene are used for transfection. Five μg of the expression plasmid pC4 is cotransfected with 0.5 μg of the plasmid pSVneo using a lipid-mediated transfection agent such as Lipofectin™ or LipofectAMINE.™ (LifeTechnologies Gaithersburg, MD). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 µM, 2 µM, 5 µM, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100-200 µM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

### 9. Quantitative Murine Soft Tissue Infection Model for E. faecalis

Compositions of the present invention, including polypeptides and peptides, are assayed for their ability to function as vaccines or to enhance/stimulate an immune response to a bacterial species (e.g., *E. faecalis*) using the following quantitative murine soft tissue infection model. Mice (e.g., NIH Swiss female mice, approximately 7 weeks old) are first treated with a biologically protective effective amount, or immune enhancing/stimulating effective amount of a composition of the present invention using methods known in the art, such as those discussed above. *See,e.g.,* Harlow et al., ANTIBODIES: A LABORATORY MANUAL, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988). An example of an appropriate starting dose is 20ug per animal.

The desired bacterial species used to challenge the mice, such as *E. faecalis*, is grown as an overnight culture. The culture is diluted to a concentration of 5 X $10^8$ cfu/ml, in an appropriate media, mixed well, serially diluted, and titered. The desired doses are further diliuted 1:2 with sterilized Cytodex 3 microcarrier beads preswollen in sterile PBS (3g/100ml). Mice are anesthetize briefly until docile, but still mobile and injected with 0.2 ml of the Cytodex 3 bead/bacterial mixture into each animal subcutaneously in the inguinal region. After four days, counting the day of injection as day one, mice are sacrificed and the contents of the abscess is excised and placed in a 15 ml conical tube containing 1.0ml of sterile PBS. The contents of the abscess is then enzymatically treated and plated as follows.

The abscess is first disrupted by vortexing with sterilized glass beads placed in the tubes. 3.0mls of prepared enzyme mixture (1.0ml Collagenase D (4.0 mg/ml), 1.0ml Trypsin (6.0 mg/ml) and 8.0 mls PBS) is then added to each tube followed by a 20 min. incubation at 37C. The solution is then centrifuged and the supernatant drawn off. 0.5 ml dH20 is then added and the tubes are vortexed and then incubated for 10 min. at room temperature. 0.5 ml media is then added and samples are serially diluted and plated onto agar plates, and grown overnight at 37C. Plates with distinct and separate colonies are then counted, compared to positive and negative control samples, and quantified. The method can be used to ıdentify composition and determine appropriate and effective doses for humans and other animals by comparing the effective doses of compositions of the present invention with compositions known in the art to be effective in both mice and humans. Doses for the effective treatment of humans and other animals, using compositions of the present invention, are extrapolated using the data from the above experiments of mice. It is appreciated that further studies in humans and other animals may be needed to determine the most effective doses using methods of clinical practice known in the art.

### 10. Murine Systemic Neutropenic Model for E. faecalis Infection

Compositions of the present invention, including polypeptides and peptides, are assayed for their ability to function as vaccines or to enhance/stimulate an immune response to a bacterial species (e.g., *E. faecalis*) using the following qualitative murine systemic neutropenic model. Mice (e.g., NIH Swiss female mice, approximately 7 weeks old) are first treated with a biologically protective effective amount, or immune enhancing/stimulating effective amount of a composition of the present invention using methods known in the art, such as those discussed above. *See,e.g.,* Harlow et al., ANTIBODIES: A LABORATORY MANUAL, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988). An example of an appropriate starting dose is 20ug per animal.
Mice are then injected with 250 - 300 mg/kg cyclophosphamide intraperitonially. Counting the day of C.P. injection as day one, the mice are left untreated for 5 days to begin recovery of PMNL'S.

The desired bacterial species used to challenge the mice, such as *E. faecalis*, is grown as an overnight culture. The culture is diluted to a concentration of 5 X $10^8$ cfu/ml, in an appropriate media, mixed well, serially diluted, and titered. The desired doses are further diliuted 1:2 in 4% Brewer's yeast in media.
Mice are injected with the bacteria/brewer's yeast challenge intraperitonially. The Brewer's yeast solution alone is used as a control. The mice are then monitered twice daily for the first week following challenge, and once a day for the next week to ascertain morbidity and mortality. Mice remaining at the end of the experiment are sacrificed. The method can be used to identify compositions and determine appropriate and effective doses for humans and other animals by comparing the effective doses of compositions of the present invention with compositions known in the art to be effective in both mice and humans. Doses for the effective treatment of humans and other animals, using compositions of the present invention, are extrapolated using the data from the above experiments of mice. It is appreciated that further studies in humans

and other animals may be needed to determine the most effective doses using methods of clinical practice known in the art.

The disclosure of all publications (including patents, patent applications, journal articles, laboratory manuals, books, or other documents) cited herein are hereby incorporated by reference in their entireties.

The present invention is not to be limited in scope by the specific embodiments described herein, which are intended as single illustrations of individual aspects of the invention. Functionally equivalent methods and components are within the scope of the invention, in addition to those shown and described herein and will become apparant to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims.

### Table 1: *E. faecalis* - Coding regions containing known sequences

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| 3 | 2 | 423 | 1226 | gb\|U24692\| | "Enterococcus faecalis pyrimidine biosynthesis D (pyrD) gene, complete cds" | 99 | 229 |
| 47 | 14 | 17085 | 16216 | gb\|M81466\| | "Enterococcus faecalis RecA protein (recA) gene, partial cds" | 98 | 308 |
| 52 | 1 | 50 | 1441 | emb\|X62755\|SFNPRG | S.faecalis npr gene for NADH peroxidase | 98 | 1374 |
| 52 | 2 | 2456 | 1494 | emb\|X62755\|SFNPRG | S.faecalis npr gene for NADH peroxidase | 100 | 209 |
| 61 | 1 | 2 | 358 | gb\|U35369\| | "Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>" | 99 | 318 |
| 61 | 2 | 467 | 1975 | gb\|U35369\| | "Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>" | 98 | 1297 |
| 61 | 3 | 1749 | 1967 | gb\|U35369\| | "Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>" | 100 | 136 |
| 61 | 4 | 1990 | 2949 | gb\|U35369\| | "Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D- | 100 | 960 |

## Table 1: *E. faecalis* - Coding regions containing known sequences

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| | | | | | Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>" | | |
| 61 | 5 | 2112 | 2399 | gb\|U35369\| | "Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>" | 100 | 288 |
| 61 | 6 | 2922 | 3794 | gb\|U35369\| | "Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>" | 100 | 873 |
| 61 | 7 | 3671 | 4762 | gb\|U35369\| | "Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>" | 99 | 1092 |
| 61 | 8 | 4312 | 3860 | gb\|U35369\| | "Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>" | 100 | 453 |
| 61 | 9 | 4653 | 5783 | gb\|U35369\| | "Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), | 100 | 1131 |

Table 1: *E. faecalis* - Coding regions containing known sequences

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| | | | | | D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>" | | |
| 61 | 10 | 5750 | 6397 | gb\|U35369\| | "Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>" | 99 | 648 |
| 61 | 11 | 7158 | 6784 | gb\|U35369\| | "Enterococcus faecalis vancomycin resistance genes, response regulator (vanRB), protein histidine kinase (vanSB), D,D-carboxypeptidase (vanYB), putative D-2-hydroxyacid dehydrogenase (vanHB), D-Ala:D-Lac ligase (vanB), and putative D,D-dipeptidase (vanX>" | 100 | 161 |
| 67 | 1 | 3 | 809 | gb\|U24692\| | "Enterococcus faecalis pyrimidine biosynthesis D (pyrD) gene, complete cds" | 98 | 807 |
| 67 | 2 | 781 | 1512 | gb\|U24692\| | "Enterococcus faecalis pyrimidine biosynthesis D (pyrD) gene, complete cds" | 93 | 92 |
| 69 | 1 | 1 | 228 | gb\|U60038\| | "Enterococcus faecalis major cold-shock protein (cspA) gene, partial cds" | 100 | 136 |
| 72 | 15 | 15814 | 19737 | emb\|X62656\|EFASP1 | "E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes" | 92 | 2504 |
| 72 | 16 | 19739 | 20155 | emb\|X62657\|EFORF3 | E.faecalis plasmid pAD1 DNA for orf3 | 96 | 341 |
| 75 | 1 | 3 | 365 | emb\|Z19137\|EFPTSHGN | E.faecalis of ptsH gene encoding HPr | 100 | 267 |
| 83 | 12 | 8766 | 7432 | emb\|X78425\|EFPBP5 | E.faecalis pbp5 gene | 98 | 416 |
| 83 | 13 | 8869 | 9699 | emb\|X78425\|EFPBP5 | E.faecalis pbp5 gene | 99 | 819 |
| 83 | 14 | 9612 | 10913 | emb\|X78425\|EFPBP5 | E.faecalis pbp5 gene | 99 | 1203 |
| 83 | 15 | 10943 | 11746 | emb\|X78425\|EFPBP5 | E.faecalis pbp5 gene | 97 | 286 |

**Table 1: *E. faecalis* - Coding regions containing known sequences**

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| 84 | 2 | 1657 | 3558 | emb\|X86176\|EFRPODDNE | E.faecalis dnaE and rpoD gene | 99 | 797 |
| 84 | 3 | 3649 | 4773 | emb\|X86176\|EFRPODDNE | E.faecalis dnaE and rpoD gene | 99 | 1125 |
| 84 | 4 | 4913 | 7000 | emb\|X86176\|EFRPODDNE | E.faecalis dnaE and rpoD gene | 99 | 301 |
| 104 | 2 | 4018 | 2900 | gb\|U36195\| | "Enterococcus faecalis pyrAa gene, partial cds" | 93 | 310 |
| 108 | 7 | 5875 | 5183 | gb\|M58002\| | "Streptococcus faecalis bacterial cell wall hydrolase gene, complete cds" | 98 | 252 |
| 145 | 8 | 8193 | 7234 | gb\|U03756\| | "Enterococcus faecalis endocarditis specific antigen gene, complete cds" | 99 | 960 |
| 145 | 9 | 8836 | 8147 | gb\|U03756\| | "Enterococcus faecalis endocarditis specific antigen gene, complete cds" | 100 | 132 |
| 147 | 3 | 2096 | 3418 | emb\|X68847\|SFNOXAA | S.faecalis nox gene for NADH oxidase | 99 | 1301 |
| 154 | 4 | 2160 | 2492 | emb\|X17092\|PPRRA | Plasmid pAM-beta-1 (from S.faecalis) replication region DNA | 93 | 294 |
| 154 | 10 | 5935 | 6294 | gb\|U17153\| | "Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds" | 99 | 355 |
| 154 | 11 | 6279 | 6584 | gb\|U17153\| | "Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds" | 98 | 89 |
| 154 | 12 | 7882 | 7097 | gb\|U86375\| | "Enterococcus faecalis ermB regulator and adenine methylase (ermB) genes, complete cds" | 99 | 736 |
| 154 | 13 | 8750 | 8043 | gb\|U17153\| | "Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds" | 99 | 498 |
| 159 | 1 | 158 | 1483 | gb\|M58002\| | "Streptococcus faecalis bacterial cell wall hydrolase gene, complete cds" | 98 | 1323 |
| 159 | 2 | 807 | 157 | gb\|M58002\| | "Streptococcus faecalis bacterial cell wall hydrolase gene, complete cds" | 99 | 651 |
| 159 | 3 | 1395 | 2192 | gb\|M58002\| | "Streptococcus faecalis bacterial cell wall hydrolase gene, complete cds" | 93 | 350 |

Table 1: *E. faecalis* - Coding regions containing known sequences

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| 216 | 2 | 282 | 1841 | gb\|M90060\| | "Streptococcus faecalis H+ ATPase a (atpB),b (atpF),c (atpE),alpha (atpA), beta (atpD),gamma (atpG),delta (atpH),and epsilon (atpC) subunits, complete cds" | 81 | 1558 |
| 216 | 4 | 2809 | 2967 | gb\|M90060\| | "Streptococcus faecalis H+ ATPase a (atpB),b (atpF),c (atpE),alpha (atpA), beta (atpD),gamma (atpG),delta (atpH),and epsilon (atpC) subunits, complete cds" | 86 | 132 |
| 216 | 5 | 2940 | 4244 | gb\|M90060\| | "Streptococcus faecalis H+ ATPase a (atpB),b (atpF),c (atpE),alpha (atpA), beta (atpD),gamma (atpG),delta (atpH),and epsilon (atpC) subunits, complete cds" | 83 | 1293 |
| 238 | 3 | 1814 | 2218 | gb\|M38386\| | "Streptococcus faecalis mtlF enzymeIII, mannitol-mtlD-phosphate- dehydrogenase" | 96 | 302 |
| 238 | 4 | 2182 | 2670 | gb\|M38386\| | "Streptococcus faecalis mtlF enzymeIII, mannitol-mtlD-phosphate- dehydrogenase" | 98 | 480 |
| 238 | 5 | 2634 | 3839 | gb\|M38386\| | "Streptococcus faecalis mtlF enzymeIII, mannitol-mtlD-phosphate- dehydrogenase" | 96 | 459 |
| 261 | 2 | 1397 | 510 | emb\|Z12296\|EFSPREG | E.faecalis sprE gene for serine proteinase homologue | 98 | 888 |
| 261 | 3 | 2474 | 1413 | dbj\|D85393\|ENEGE1E | "Enterococcus faecalis DNA for gelatinase, complete cds" | 98 | 1051 |
| 261 | 4 | 2974 | 2417 | dbj\|D85393\|ENEGE1E | "Enterococcus faecalis DNA for gelatinase, complete cds" | 97 | 516 |
| 275 | 3 | 1472 | 1044 | gb\|L23802\| | "Enterococcus faecalis pore forming, cell wall enzyme, regulatory, and dehydroquinase homologue proteins (ebsA,ebsB,ebsC,and ebsD) genes, complete cds with repeat region" | 98 | 422 |
| 275 | 4 | 1581 | 2018 | gb\|L23802\| | "Enterococcus faecalis pore forming, cell wall enzyme, regulatory, and dehydroquinase homologue proteins | 97 | 438 |

Table 1: *E. faecalis* - Coding regions containing known sequences

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| | | | | | (ebsA,ebsB,ebsC,and ebsD) genes, complete cds with repeat region" | | |
| 275 | 5 | 2789 | 2148 | gb\|L23802\| | "Enterococcus faecalis pore forming, cell wall enzyme, regulatory, and dehydroquinase homologue proteins (ebsA,ebsB,ebsC,and ebsD) genes, complete cds with repeat region" | 98 | 642 |
| 275 | 6 | 3475 | 2660 | gb\|L23802\| | "Enterococcus faecalis pore forming, cell wall enzyme, regulatory, and dehydroquinase homologue proteins (ebsA,ebsB,ebsC,and ebsD) genes, complete cds with repeat region" | 98 | 790 |
| 287 | 2 | 1565 | 558 | emb\|X17092\|PPRRA | Plasmid pAM-beta-1 (from S.faecalis) replication region DNA | 97 | 991 |
| 287 | 3 | 2049 | 1582 | emb\|X17092\|PPRRA | Plasmid pAM-beta-1 (from S.faecalis) replication region DNA | 97 | 461 |
| 287 | 6 | 2639 | 3346 | gb\|U17153\| | "Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds" | 99 | 498 |
| 294 | 11 | 4519 | 4211 | gb\|U17153\| | "Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds" | 100 | 50 |
| 302 | 1 | 1 | 1755 | emb\|X62658\|EFSEA1 | E.faecalis plasmid pAD1 sea1 gene and orfy | 83 | 1755 |
| 302 | 2 | 2310 | 2687 | emb\|X17214\|SFPASA1 | S. faecalis plasmid pAD1 asa1 gene for aggregation substance and ORF 1 | 100 | 378 |
| 302 | 3 | 2865 | 3329 | emb\|X17214\|SFPASA1 | S. faecalis plasmid pAD1 asa1 gene for aggregation substance and ORF 1 | 99 | 463 |
| 316 | 4 | 2724 | 2110 | gb\|M13771\| | "Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds" | 100 | 248 |
| 346 | 5 | 2224 | 2880 | emb\|X62755\|SFNPRG | S.faecalis npr gene for NADH peroxidase | 98 | 351 |

Table 1: *E. faecalis* - Coding regions containing known sequences

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| 349 | 2 | 686 | 907 | dbj\|D78257\|D78257 | "Enterococcus faecalis plasmid pYI17 genes for BacA, BacB, ORF3, ORF4, ORF5, ORF6, ORF7, ORF8, ORF9, ORF10, ORF11,partial cds" | 83 | 200 |
| 355 | 1 | 3 | 1166 | emb\|X17214\|SFPASA1 | S. faecalis plasmid pAD1 asa1 gene for aggregation substance and ORF 1 | 97 | 1100 |
| 355 | 2 | 1102 | 1548 | emb\|X17214\|SFPASA1 | S. faecalis plasmid pAD1 asa1 gene for aggregation substance and ORF 1 | 94 | 432 |
| 355 | 3 | 1663 | 2037 | emb\|X62657\|EFORF3 | E.faecalis plasmid pAD1 DNA for orf3 | 99 | 337 |
| 355 | 4 | 2035 | 2445 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 99 | 411 |
| 355 | 5 | 2558 | 2851 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 96 | 280 |
| 355 | 6 | 2838 | 3299 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 97 | 430 |
| 355 | 7 | 3236 | 3739 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 97 | 279 |
| 355 | 8 | 3696 | 4529 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 97 | 537 |
| 355 | 9 | 4587 | 5870 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 98 | 718 |
| 355 | 10 | 5843 | 6490 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 99 | 224 |
| 355 | 11 | 6471 | 6890 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 96 | 361 |
| 355 | 12 | 6881 | 7204 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 98 | 324 |
| 355 | 13 | 7191 | 8231 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 98 | 984 |
| 355 | 14 | 8218 | 8496 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 99 | 279 |
| 355 | 15 | 8412 | 8885 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading | 100 | 474 |

**Table 1:** ***E. faecalis* - Coding regions containing known sequences**

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| | | | | | frames" | | |
| 355 | 17 | 9479 | 9952 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 98 | 417 |
| 365 | 1 | 3 | 380 | gb\|M13771\| | "Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds" | 100 | 248 |
| 370 | 1 | 1 | 1299 | dbj\|D78016\|ENEPPD1A | "Enterococcus faecalis Plasmid pPD1 genes for REPB, REPA, TRAC, TRAB, TRAA, iPD1, TRAE, TRAF, complete cds and partial cds" | 73 | 1267 |
| 407 | 3 | 963 | 2162 | gb\|U38590\| | "Enterococcus faecalis plasmid pCF10 PrgN, PrgO, and PrgP genes, complete cds" | 98 | 257 |
| 407 | 5 | 3811 | 4131 | gb\|U38590\| | "Enterococcus faecalis plasmid pCF10 PrgN, PrgO, and PrgP genes, complete cds" | 86 | 317 |
| 417 | 1 | 42 | 419 | gb\|U00681\| | "Enterococcus faecalis plasmid pAD1 TraB (traB) gene, complete cds (traC) and (repA) genes, partial cds" | 98 | 304 |
| 417 | 2 | 313 | 41 | gb\|U00681\| | "Enterococcus faecalis plasmid pAD1 TraB (traB) gene, complete cds (traC) and (repA) genes, partial cds" | 97 | 198 |
| 417 | 3 | 440 | 754 | gb\|U00681\| | "Enterococcus faecalis plasmid pAD1 TraB (traB) gene, complete cds (traC) and (repA) genes, partial cds" | 100 | 219 |
| 426 | 1 | 112 | 462 | emb\|Z49243\|EF4110SOD | E.faecalis partial sod gene for superoxide dismutase (strain=BM4110) | 98 | 291 |
| 426 | 2 | 628 | 419 | emb\|Z49243\|EF4110SOD | E.faecalis partial sod gene for superoxide dismutase (strain=BM4110) | 100 | 148 |
| 426 | 3 | 456 | 725 | emb\|Z49243\|EF4110SOD | E.faecalis partial sod gene for superoxide dismutase (strain=BM4110) | 100 | 148 |
| 429 | 1 | 840 | 79 | emb\|X62658\|EFSEA1 | E.faecalis plasmid pAD1 seal gene and orfy | 98 | 737 |
| 429 | 2 | 1087 | 767 | emb\|X62658\|EFSEA1 | E.faecalis plasmid pAD1 seal gene and orfy | 99 | 321 |
| 429 | 4 | 2765 | 2460 | gb\|U17153\| | "Enterococcus faecalis plasmid pjh1 | 98 | 89 |

Table 1: *E. faecalis* - Coding regions containing known sequences

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| | | | | | tetracycline resistant (tetL) gene, complete cds" | | |
| 429 | 5 | 3166 | 2750 | gb\|U17153\| | "Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds" | 99 | 413 |
| 435 | 5 | 2731 | 2324 | gb\|M38052\| | "Enterococcus faecalis cytolysin B transport protein gene, complete cds" | 97 | 97 |
| 459 | 2 | 1330 | 1067 | gb\|M13771\| | "Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds" | 99 | 248 |
| 506 | 1 | 1242 | 4 | emb\|X17214\|SFPASA1 | S. faecalis plasmid pAD1 asa1 gene for aggregation substance and ORF 1 | 99 | 1144 |
| 514 | 3 | 1496 | 1113 | gb\|M13771\| | "Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds" | 100 | 248 |
| 527 | 2 | 1733 | 1371 | gb\|U17153\| | "Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds" | 98 | 153 |
| 544 | 1 | 309 | 4 | gb\|U38590\| | "Enterococcus faecalis plasmid pCF10 PrgN, PrgO, and PrgP genes, complete cds" | 95 | 306 |
| 561 | 1 | 3 | 761 | dbj\|D78016\|ENEPPD1A | "Enterococcus faecalis Plasmid pPD1 genes for REPB, REPA, TRAC, TRAB, TRAA, iPD1, TRAE, TRAF, complete cds and partial cds" | 77 | 528 |
| 561 | 2 | 772 | 1566 | gb\|U00681\| | "Enterococcus faecalis plasmid pAD1 TraB (traB) gene, complete cds (traC) and (repA) genes, partial cds" | 99 | 795 |
| 566 | 3 | 874 | 2037 | dbj\|D78016\|ENEPPD1A | "Enterococcus faecalis Plasmid pPD1 genes for REPB, REPA, TRAC, TRAB, TRAA, iPD1, TRAE, TRAF, complete cds and partial cds" | 90 | 1160 |
| 581 | 1 | 398 | 3 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading | 100 | 393 |

Table 1: *E. faecalis* - Coding regions containing known sequences

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| | | | | | frames" | | |
| 581 | 2 | 908 | 540 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 100 | 369 |
| 597 | 1 | 573 | 7 | gb\|M38052\| | "Enterococcus faecalis cytolysin B transport protein gene, complete cds" | 99 | 566 |
| 597 | 2 | 1247 | 516 | gb\|M38052\| | "Enterococcus faecalis cytolysin B transport protein gene, complete cds" | 97 | 701 |
| 604 | 7 | 3265 | 2903 | gb\|U17153\| | "Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds" | 100 | 143 |
| 618 | 1 | 1 | 534 | gb\|M13771\| | "Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds" | 99 | 470 |
| 622 | 1 | 864 | 16 | gb\|M13771\| | "Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds" | 99 | 849 |
| 622 | 2 | 1317 | 862 | gb\|M13771\| | "Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds" | 99 | 256 |
| 622 | 3 | 1586 | 1311 | gb\|M13771\| | "Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds" | 99 | 248 |
| 624 | 6 | 5641 | 8001 | gb\|U66286\| | "Enterococcus faecalis gyrase A (gyrA) gene, partial cds" | 98 | 219 |
| 635 | 1 | 516 | 953 | dbj\|D78257\|D78257 | "Enterococcus faecalis plasmid pYI17 genes for BacA, BacB, ORF3, ORF4, ORF5, ORF6, ORF7, ORF8, ORF9, ORF10, ORF11,partial cds" | 94 | 404 |

Table 1: *E. faecalis* - Coding regions containing known sequences

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| 635 | 2 | 920 | 1222 | dbj\|D78257\|D78257 | "Enterococcus faecalis plasmid pYI17 genes for BacA, BacB, ORF3, ORF4, ORF5, ORF6, ORF7, ORF8, ORF9, ORF10, ORF11,partial cds" | 83 | 299 |
| 637 | 1 | 3 | 545 | emb\|X62656\|EFASP1 | "E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes" | 92 | 506 |
| 658 | 2 | 1198 | 365 | gb\|M38052\| | "Enterococcus faecalis cytolysin B transport protein gene, complete cds" | 100 | 819 |
| 658 | 3 | 1446 | 1189 | gb\|M38052\| | "Enterococcus faecalis cytolysin B transport protein gene, complete cds" | 98 | 258 |
| 664 | 1 | 490 | 65 | emb\|X62658\|EFSEA1 | E.faecalis plasmid pAD1 sea1 gene and orfy | 88 | 423 |
| 664 | 2 | 737 | 417 | emb\|X62658\|EFSEA1 | E.faecalis plasmid pAD1 sea1 gene and orfy | 94 | 321 |
| 743 | 1 | 561 | 4 | dbj\|D78016\|ENEPPD1A | "Enterococcus faecalis Plasmid pPD1 genes for REPB, REPA, TRAC, TRAB, TRAA, iPD1, TRAE, TRAF, complete cds and partial cds" | 87 | 305 |
| 747 | 2 | 1139 | 324 | gb\|M38052\| | "Enterococcus faecalis cytolysin B transport protein gene, complete cds" | 99 | 691 |
| 747 | 3 | 577 | 783 | gb\|M38052\| | "Enterococcus faecalis cytolysin B transport protein gene, complete cds" | 100 | 207 |
| 747 | 4 | 1474 | 1133 | gb\|M13771\| | "Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds" | 99 | 248 |
| 777 | 1 | 401 | 3 | gb\|M38052\| | "Enterococcus faecalis cytolysin B transport protein gene, complete cds" | 100 | 335 |
| 816 | 1 | 793 | 512 | gb\|M13771\| | "Streptococcus faecalis 6'-aminoglycoside acetyltransferase phosphotransferase (AAC(6')-APH(2')) bifunctional resistance protein, complete cds" | 100 | 243 |
| 842 | 1 | 418 | 89 | emb\|X17214\|SFPASA1 | S. faecalis plasmid pAD1 asa1 gene for aggregation substance and ORF 1 | 91 | 303 |
| 842 | 2 | 856 | 605 | emb\|X62658\|EFSEA1 | E.faecalis plasmid pAD1 sea1 gene and orfy | 92 | 246 |

Table 1: *E. faecalis* - Coding regions containing known sequences

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| 847 | 1 | 1481 | 3 | emb\|X62658\|EFSEA1 | E.faecalis plasmid pAD1 seal gene and orfy | 92 | 1479 |
| 864 | 1 | 36 | 1106 | emb\|X62658\|EFSEA1 | E.faecalis plasmid pAD1 seal gene and orfy | 93 | 945 |
| 864 | 2 | 1571 | 3550 | emb\|X62656\|EFASP1 | "E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes" | 96 | 1979 |
| 872 | 1 | 263 | 3 | gb\|U17153\| | "Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds" | 98 | 261 |
| 874 | 1 | 833 | 693 | dbj\|D31675\|ENE16RNA8 | "Enterococcus faecalis 16S ribosomal RNA, partial sequence" | 100 | 98 |
| 878 | 1 | 302 | 30 | gb\|U17153\| | "Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds" | 94 | 94 |
| 878 | 2 | 263 | 445 | gb\|U17153\| | "Enterococcus faecalis plasmid pjh1 tetracycline resistant (tetL) gene, complete cds" | 99 | 181 |
| 921 | 1 | 748 | 26 | emb\|X62658\|EFSEA1 | E.faecalis plasmid pAD1 seal gene and orfy | 95 | 612 |
| 929 | 1 | 484 | 2 | emb\|X62658\|EFSEA1 | E.faecalis plasmid pAD1 seal gene and orfy | 99 | 409 |
| 946 | 1 | 3 | 422 | emb\|X62657\|EFORF3 | E.faecalis plasmid pAD1 DNA for orf3 | 99 | 341 |
| 946 | 2 | 420 | 830 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 98 | 411 |
| 946 | 3 | 866 | 1123 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 96 | 230 |
| 947 | 1 | 112 | 498 | emb\|X62656\|EFASP1 | "E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes" | 96 | 378 |
| 951 | 1 | 484 | 26 | emb\|X62658\|EFSEA1 | E.faecalis plasmid pAD1 seal gene and orfy | 95 | 353 |
| 956 | 1 | 3 | 545 | emb\|X62656\|EFASP1 | "E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes" | 96 | 543 |
| 956 | 2 | 524 | 721 | emb\|X62656\|EFASP1 | "E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes" | 94 | 161 |
| 957 | 1 | 616 | 2 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading frames" | 99 | 615 |
| 957 | 2 | 42 | 686 | emb\|X96977\|EFPAD1ORF | "E.faecalis plasmid pAD1, open reading | 99 | 595 |

Table 1: *E. faecalis* - Coding regions containing known sequences

| Contig ID | Orf ID | Start (nt) | Stop (nt) | Match Accession | Match Gene Name | Percent Indent | HSP nt length |
|---|---|---|---|---|---|---|---|
| | | | | | frames" | | |
| 968 | 1 | 1 | 456 | emb\|X62656\|EFASP1 | "E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes" | 96 | 366 |
| 968 | 2 | 339 | 641 | emb\|X62656\|EFASP1 | "E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes" | 95 | 158 |
| 968 | 3 | 395 | 658 | emb\|X62656\|EFASP1 | "E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes" | 94 | 126 |
| 977 | 1 | 5 | 943 | emb\|X17214\|SFPASA1 | S. faecalis plasmid pAD1 asa1 gene for aggregation substance and ORF 1 | 99 | 847 |
| 982 | 1 | 376 | 2 | emb\|X62658\|EFSEA1 | E.faecalis plasmid pAD1 sea1 gene and orfy | 95 | 365 |
| 985 | 1 | 85 | 471 | emb\|X62656\|EFASP1 | "E.faecalis plasmid pPD1 asp1 and URFs pd57, pd125 and pd113 genes" | 91 | 362 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 137 | 3 | 3208 | 2003 | gi\|152947 | transposase [Staphylococcus aureus] | 100 | 100 |
| 154 | 14 | 9166 | 9750 | gi\|141861 | traA gene product [Plasmid pAD1] | 100 | 100 |
| 276 | 16 | 11268 | 11047 | gnl\|PID\|e284733 | C34B7.2 [Caenorhabditis elegans] | 100 | 71 |
| 287 | 1 | 485 | 234 | gi\|152947 | transposase [Staphylococcus aureus] | 100 | 100 |
| 287 | 7 | 3454 | 3765 | gi\|152947 | transposase [Staphylococcus aureus] | 100 | 100 |
| 292 | 6 | 3001 | 4185 | gi\|488330 | alpha-amylase [unidentified cloning vector] | 100 | 100 |
| 429 | 3 | 2013 | 1654 | gi\|141863 | regulatory protein [Plasmid pAD1] | 100 | 100 |
| 604 | 3 | 1243 | 1043 | gi\|559860 | clyLs [Plasmid pAD1] | 100 | 98 |
| 604 | 4 | 1492 | 1268 | gi\|559859 | clyLl [Plasmid pAD1] | 100 | 100 |
| 656 | 7 | 7592 | 6834 | gi\|488339 | alpha-amylase [unidentified cloning vector] | 100 | 100 |
| 658 | 1 | 312 | 4 | gi\|152947 | transposase [Staphylococcus aureus] | 100 | 100 |
| 674 | 3 | 1236 | 1589 | gi\|1196996 | unknown protein [Transposon Tn10] | 100 | 98 |
| 700 | 1 | 375 | 4 | gi\|152947 | transposase [Staphylococcus aureus] | 100 | 100 |
| 961 | 1 | 1 | 450 | gi\|152947 | transposase [Staphylococcus aureus] | 100 | 100 |
| 72 | 17 | 20153 | 21040 | gi\|150556 | surface protein [Plasmid pCF10] | 99 | 99 |
| 99 | 5 | 3117 | 1933 | gi\|1006839 | malic enzyme [Streptococcus bovis] | 99 | 99 |
| 154 | 3 | 1955 | 1491 | gi\|149482 | transposase [Lactococcus lactis] | 99 | 99 |
| 326 | 3 | 3030 | 1774 | pir\|S16989\|S16989 | dihydrolipoamide S-acetyltransferase (EC 2.3.1.12) - Enterococcus faecalis | 99 | 98 |
| 407 | 6 | 4636 | 4235 | gi\|141859 | replication-associated protein [Plasmid pAD1] | 99 | 99 |
| 692 | 1 | 3 | 485 | gi\|559861 | clyM [Plasmid pAD1] | 99 | 99 |
| 99 | 6 | 3904 | 3134 | gi\|1146122 | L-malate permease [Streptococcus bovis] | 98 | 98 |
| 326 | 4 | 3358 | 3002 | pir\|S16989\|S16989 | dihydrolipoamide S-acetyltransferase (EC 2.3.1.12) - Enterococcus faecalis | 98 | 97 |
| 346 | 1 | 606 | 4 | gi\|1146122 | L-malate permease [Streptococcus bovis] | 98 | 98 |
| 367 | 31 | 14415 | 13999 | gi\|1644226 | ribosomal protein S10 [Bacillus subtilis] | 98 | 88 |
| 367 | 6 | 2797 | 2495 | gi\|142459 | initiation factor 1 [Bacillus subtilis] | 97 | 88 |
| 407 | 9 | 5454 | 4894 | gi\|141858 | replication-associated protein [Plasmid | 97 | 97 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | pAD1] | | |
| 497 | 6 | 3514 | 3762 | gi\|532552 | ORF19 [Enterococcus faecalis] | 97 | 87 |
| 558 | 1 | 1 | 399 | gi\|46638 | ORF 2 (AA 1 - 236) [Staphylococcus aureus] | 97 | 97 |
| 829 | 1 | 169 | 2 | gnl\|PID\|e283110 | femD [Staphylococcus aureus] | 97 | 86 |
| 407 | 8 | 4970 | 4599 | gi\|141858 | replication-associated protein [Plasmid pAD1] | 96 | 96 |
| 777 | 2 | 1102 | 380 | gi\|559861 | clyM [Plasmid pAD1] | 96 | 96 |
| 23 | 33 | 20797 | 21126 | gnl\|PID\|e223402 | DNA topoisomerase IV C subunit [Streptococcus pneumoniae] | 95 | 80 |
| 32 | 5 | 3454 | 3071 | gi\|147194 | phnA protein [Escherichia coli] | 95 | 87 |
| 95 | 8 | 5493 | 6875 | gi\|391682 | Na+ -ATPase beta subunit [Enterococcus hirae] | 95 | 89 |
| 138 | 25 | 16587 | 16745 | gi\|143136 | L-lactate dehydrogenase [Bacillus megaterium] | 95 | 70 |
| 367 | 20 | 9198 | 8797 | gi\|40150 | L14 protein (AA 1-122) [Bacillus subtilis] | 95 | 90 |
| 367 | 21 | 9519 | 9223 | gi\|1044973 | ribosomal protein L17 [Bacillus subtilis] | 95 | 89 |
| 439 | 2 | 846 | 1241 | gi\|488334 | alpha-amylase [unidentified cloning vector] | 95 | 94 |
| 604 | 1 | 792 | 4 | gi\|559861 | clyM [Plasmid pAD1] | 95 | 93 |
| 722 | 1 | 1 | 504 | gi\|47453 | ribosomal protein S12 [Streptococcus pneumoniae] | 95 | 94 |
| 17 | 8 | 7317 | 7676 | gi\|532554 | ORF21 [Enterococcus faecalis] | 94 | 86 |
| 95 | 2 | 1288 | 1791 | gi\|416405 | Na+-ATPase K subunit [Enterococcus hirae] | 94 | 88 |
| 97 | 3 | 2481 | 1432 | gi\|1750264 | heat shock protein 70 [Streptococcus pneumoniae] | 94 | 90 |
| 117 | 5 | 2700 | 3842 | gi\|467376 | unknown [Bacillus subtilis] | 94 | 89 |
| 327 | 3 | 3283 | 3762 | gi\|153566 | ORF (19K protein) [Enterococcus faecalis] | 94 | 87 |
| 327 | 5 | 4782 | 5054 | gi\|153568 | H+ ATPase [Enterococcus faecalis] | 94 | 82 |
| 387 | 4 | 3608 | 1728 | gi\|153661 | translational initiation factor IF2 [Enterococcus faecium] sp\|P18311\|IF2_ENTFC INITIATION FACTOR IF-2. | 94 | 88 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 455 | 1 | 2 | 259 | gi\|532549 | ORF16 [Enterococcus faecalis] | 94 | 82 |
| 97 | 2 | 1444 | 677 | gi\|450684 | dnaK gene product [Lactococcus lactis] | 93 | 83 |
| 188 | 2 | 1690 | 1911 | gi\|43865 | nifJ gene product [Klebsiella pneumoniae] | 93 | 78 |
| 216 | 6 | 4234 | 4680 | gi\|153574 | H+ ATPase [Enterococcus faecalis] | 93 | 86 |
| 298 | 2 | 2798 | 1221 | gi\|143012 | GMP synthetase [Bacillus subtilis] | 93 | 86 |
| 329 | 2 | 1538 | 771 | gi\|153826 | adhesin B [Streptococcus sanguis] | 93 | 83 |
| 367 | 15 | 7675 | 7247 | gi\|1044978 | ribosomal protein S8 [Bacillus subtilis] | 93 | 82 |
| 722 | 2 | 527 | 1030 | gi\|1644222 | ribosomal protein S7 [Bacillus subtilis] | 93 | 83 |
| 803 | 1 | 657 | 151 | gi\|1196998 | unknown protein [Transposon Tn10] | 93 | 93 |
| 962 | 1 | 130 | 636 | gi\|152947 | transposase [Staphylococcus aureus] | 93 | 92 |
| 237 | 12 | 6056 | 6385 | gi\|963038 | ArpU [Enterococcus hirae] | 92 | 76 |
| 309 | 4 | 8218 | 4541 | gi\|402363 | RNA polymerase beta-subunit [Bacillus subtilis] sp\|P37870\|RPOB_BACSU DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC .7.7.6) (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT). | 92 | 82 |
| 329 | 4 | 2529 | 1717 | gi\|310632 | hydrophobic membrane protein [Streptococcus gordonii] sp\|P42361\|P29K_STRGC 29 KD MEMBRANE PROTEIN IN PSAA 5'REGION ORF1). | 92 | 78 |
| 367 | 4 | 1942 | 1544 | gi\|142462 | ribosomal protein S11 [Bacillus subtilis] | 92 | 82 |
| 367 | 8 | 3648 | 3457 | pir\|C44859\|C44859 | adenylate kinase - Bacillus sp. (fragment) | 92 | 88 |
| 367 | 12 | 6183 | 5641 | gi\|1044981 | ribosomal protein S5 [Bacillus subtilis] | 92 | 81 |
| 367 | 17 | 8427 | 7885 | pir\|A29102\|R5BS5F | ribosomal protein L5 - Bacillus stearothermophilus | 92 | 83 |
| 527 | 1 | 1404 | 373 | gi\|153092 | replication protein [Staphylococcus aureus] | 92 | 81 |
| 701 | 1 | 2 | 352 | gi\|143793 | tyrosyl-tRNA synthetase [Bacillus caldotenax] | 92 | 74 |
| 23 | 28 | 17420 | 17566 | sp\|P45692\|EUTX_SALTY | ETHANOLAMINE UTILIZATION PROTEIN EUTX (FRAGMENT). | 91 | 73 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 57 | 5 | 4129 | 4701 | gi\|1595810 | type-I signal peptidase SpsB [Staphylococcus aureus] | 91 | 67 |
| 57 | 12 | 13281 | 13970 | gnl\|PID\|e254999 | phenylalany-tRNA synthetase beta subunit [Bacillus subtilis] | 91 | 75 |
| 156 | 5 | 4609 | 6474 | gi\|1303804 | YqeQ [Bacillus subtilis] | 91 | 79 |
| 216 | 3 | 1848 | 2765 | gi\|153572 | H+ ATPase [Enterococcus faecalis] | 91 | 81 |
| 367 | 24 | 10802 | 10128 | gi\|1165309 | S3 [Bacillus subtilis] | 91 | 78 |
| 415 | 1 | 452 | 883 | pir\|B56272\|B56272 | probable pheromone-responsive regulatory protein R - Enterococcus faecalis plasmid pCF10 | 91 | 90 |
| 466 | 2 | 1313 | 2065 | gi\|142443 | adenylosuccinate synthetase [Bacillus subtilis] sp\|P29726\|PURA_BACSU ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4) IMP--ASPARTATE LIGASE). | 91 | 79 |
| 545 | 1 | 1 | 345 | gi\|532549 | ORF16 [Enterococcus faecalis] | 91 | 80 |
| 572 | 1 | 8 | 652 | gi\|347998 | uracil phosphoribosyltransferase [Streptococcus salivarius] sp\|P36399\|UPP_STRSL PROBABLE URACIL PHOSPHORIBOSYLTRANSFERASE (EC .4.2.9) (UMP PYROPHOSPHORYLASE) (UPRTASE). | 91 | 78 |
| 599 | 1 | 8 | 343 | gi\|42029 | ORF1 gene product [Escherichia coli] | 91 | 75 |
| 600 | 2 | 585 | 779 | pir\|B48396\|B48396 | ribosomal protein L33 - Bacillus stearothermophilus | 91 | 81 |
| 652 | 1 | 394 | 2 | gi\|535662 | transposase [Insertion sequence IS1251] | 91 | 81 |
| 1 | 4 | 3465 | 2557 | gi\|1644224 | elongation factor Tu [Bacillus subtilis] | 90 | 83 |
| 17 | 19 | 14844 | 17297 | gi\|532549 | ORF16 [Enterococcus faecalis] | 90 | 77 |
| 52 | 3 | 2650 | 2811 | gi\|473902 | alpha-acetolactate synthase [Lactococcus lactis] | 90 | 68 |
| 74 | 9 | 5870 | 5469 | gi\|1653508 | hypothetical protein [Synechocystis sp.] | 90 | 52 |
| 75 | 3 | 1177 | 2091 | gi\|153615 | phosphoenolpyruvate:sugar phosphotransferase system enzyme I Streptococcus salivarius] | 90 | 83 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 117 | 10 | 6591 | 8126 | gi\|924848 | inosine monophosphate dehydrogenase [Streptococcus pyogenes] pir\|JC4372\|JC4372 IMP dehydrogenase (EC 1.1.1.205) - Streptococcus yogenes | 90 | 80 |
| 276 | 1 | 577 | 95 | gi\|530798 | LysB [Bacteriophage phi-LC3] | 90 | 72 |
| 287 | 5 | 2611 | 2441 | gi\|1333835 | copS gene product [Streptococcus pyogenes] | 90 | 78 |
| 290 | 1 | 1 | 708 | gi\|897795 | 30S ribosomal protein [Pediococcus acidilactici] sp\|P49668\|RS2_PEDAC 30S RIBOSOMAL PROTEIN S2. | 90 | 75 |
| 309 | 3 | 4401 | 1093 | gnl\|PID\|e187579 | DNA-directed RNA polymerase [Listeria innocua] | 90 | 81 |
| 367 | 22 | 9731 | 9513 | pir\|A02825\|R5BS29 | ribosomal protein L29 - Bacillus stearothermophilus | 90 | 76 |
| 452 | 4 | 2224 | 2508 | gi\|434759 | ORF [Homo sapiens] | 90 | 54 |
| 455 | 2 | 2776 | 323 | gi\|532549 | ORF16 [Enterococcus faecalis] | 90 | 77 |
| 623 | 1 | 3 | 221 | gi\|460259 | enolase [Bacillus subtilis] | 90 | 80 |
| 624 | 5 | 3612 | 5615 | gnl\|PID\|e208213 | DNA gyrase [Streptococcus pneumoniae] | 90 | 81 |
| 853 | 2 | 752 | 282 | gnl\|PID\|e13389 | translation initiation factor IF3 (AA 1-172) [Bacillus stearothermophilus] | 90 | 82 |
| 966 | 1 | 1 | 462 | gi\|532549 | ORF16 [Enterococcus faecalis] | 90 | 83 |
| 1 | 3 | 2596 | 2219 | gi\|1661195 | elongation factor-Tu [Streptococcus mutans] | 89 | 78 |
| 1 | 5 | 4314 | 3556 | gi\|1644223 | elongation factor G [Bacillus subtilis] | 89 | 79 |
| 23 | 21 | 13990 | 14295 | gi\|466518 | pduA [Salmonella typhimurium] | 89 | 75 |
| 23 | 32 | 19927 | 20799 | gnl\|PID\|e208211 | DNA topoisomerase IV [Streptococcus pneumoniae] | 89 | 83 |
| 42 | 2 | 349 | 1989 | gi\|287871 | groEL gene product [Lactococcus lactis] | 89 | 79 |
| 45 | 15 | 11835 | 12167 | gi\|150554 | surface exclusion protein [Plasmid pCF10] | 89 | 68 |
| 53 | 2 | 685 | 1797 | gnl\|PID\|e221213 | ClpX protein [Bacillus subtilis] | 89 | 81 |
| 86 | 4 | 3374 | 4024 | gi\|537286 | triosephosphate isomerase [Lactococcus lactis] | 89 | 78 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 95 | 7 | 3677 | 5506 | gi\|912449 | Na+ -ATPase alpha subunit [Enterococcus hirae] | 89 | 80 |
| 128 | 18 | 11348 | 11013 | gi\|466473 | cellobiose phosphotransferase enzyme II' [Bacillus tearothermophilus] | 89 | 60 |
| 132 | 1 | 180 | 2180 | gi\|153854 | uvs402 protein [Streptococcus pneumoniae] | 89 | 78 |
| 342 | 1 | 783 | 4 | gi\|1041115 | TRAC [Plasmid pPD1] | 89 | 79 |
| 367 | 23 | 10146 | 9691 | sp\|P14577\|RL16_BACSU | 50S RIBOSOMAL PROTEIN L16. | 89 | 80 |
| 367 | 27 | 12377 | 11541 | gi\|1165306 | L2 [Bacillus subtilis] | 89 | 79 |
| 435 | 4 | 2424 | 2215 | gi\|559863 | clyA [Plasmid pAD1] | 89 | 89 |
| 466 | 3 | 1972 | 2736 | gi\|467328 | adenylosuccinate synthetase [Bacillus subtilis] | 89 | 75 |
| 512 | 3 | 999 | 1607 | gi\|1477776 | ClpP [Bacillus subtilis] | 89 | 73 |
| 518 | 1 | 1 | 174 | gi\|786163 | Ribosomal Protein L10 [Bacillus subtilis] | 89 | 76 |
| 604 | 2 | 1000 | 713 | gi\|559861 | clyM [Plasmid pAD1] | 89 | 89 |
| 615 | 2 | 888 | 691 | gi\|467469 | unknown [Bacillus subtilis] | 89 | 75 |
| 677 | 2 | 992 | 429 | gi\|1389732 | S-adenosylmethionine synthetase [Bacillus subtilis] | 89 | 76 |
| 677 | 3 | 1315 | 950 | gi\|1020317 | S-adenosylmethionine synthetase [Staphylococcus aureus] | 89 | 73 |
| 722 | 3 | 1102 | 1278 | pir\|PW0010\|PW0010 | translation elongation factor G - Bacillus stearothermophilus (fragment) | 89 | 72 |
| 850 | 1 | 464 | 3 | gi\|142521 | deoxyribodipyrimidine photolyase [Bacillus subtilis] gnl\|PID\|e255102 deoxyribodipyrimidine photolyase [Bacillus ubtilis] | 89 | 72 |
| 17 | 5 | 3711 | 4751 | gi\|532554 | ORF21 [Enterococcus faecalis] | 88 | 72 |
| 37 | 5 | 3322 | 3717 | gi\|1216488 | uncharacterized open reading frame; hypothetical protein displaying similarity to a Bacillus subtilis hypothetical protein (Ylm [Streptococcus mutans] | 88 | 75 |
| 39 | 6 | 2454 | 2630 | sp\|P49865\|NTPR_ENT | NTPR PROTEIN (FRAGMENT). | 88 | 77 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | HR | | | |
| 48 | 3 | 1740 | 2666 | gi\|557492 | dihydroxynapthoic acid (DHNA) synthetase [Bacillus subtilis] gi\|143186 dihydroxynapthoic acid (DHNA) synthetase [Bacillus ubtilis] | 88 | 75 |
| 63 | 5 | 2753 | 3607 | gi\|1064814 | homologous to sp:PHOP_BACSUB [Bacillus subtilis] | 88 | 77 |
| 86 | 2 | 1004 | 2047 | gi\|153763 | plasmin receptor [Streptococcus pyogenes] | 88 | 79 |
| 104 | 6 | 6431 | 6213 | gi\|431231 | uracil permease [Bacillus caldolyticus] | 88 | 60 |
| 110 | 19 | 18174 | 16891 | gi\|217040 | acid glycoprotein [Streptococcus pyogenes] | 88 | 72 |
| 145 | 10 | 9040 | 8834 | gi\|393268 | 29-kiloDalton protein [Streptococcus pneumoniae] sp\|P42362\|P29K_STRPN 29 KD MEMBRANE PROTEIN IN PSAA 5'REGION ORF1). | 88 | 71 |
| 151 | 1 | 1620 | 316 | gi\|143366 | adenylosuccinate lyase (PUR-B) [Bacillus subtilis] pir\|C29326\|WZBSDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus ubtilis | 88 | 78 |
| 171 | 10 | 9676 | 10119 | gi\|1591672 | phosphate transport system ATP-binding protein [Methanococcus jannaschii] | 88 | 63 |
| 190 | 3 | 1997 | 975 | gi\|532554 | ORF21 [Enterococcus faecalis] | 88 | 76 |
| 229 | 6 | 5712 | 5954 | gi\|143648 | ribosomal protein L28 [Bacillus subtilis] | 88 | 70 |
| 270 | 2 | 895 | 1869 | gi\|1303828 | YqfJ [Bacillus subtilis] | 88 | 75 |
| 275 | 7 | 3761 | 3552 | gi\|425474 | SMDR1 [Schistosoma mansoni] | 88 | 72 |
| 293 | 1 | 614 | 3 | gi\|1783246 | highly homologous to many ATP-binding transport proteins; hypothetical [Bacillus subtilis] | 88 | 80 |
| 367 | 1 | 485 | 72 | gi\|142464 | ribosomal protein L17 [Bacillus subtilis] | 88 | 76 |
| 367 | 5 | 2335 | 1961 | gi\|1044989 | ribosomal protein S13 [Bacillus subtilis] | 88 | 80 |
| 367 | 16 | 7887 | 7681 | pir\|S48688\|S48688 | ribosomal protein S14 - Bacillus stearothermophilus | 88 | 83 |
| 598 | 1 | 1006 | 23 | gi\|565287 | transposase-like protein of PS3IS [thermophilic bacterium PS3] | 88 | 66 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | pir\|JC4292\|JC4292 insertion sequence element 1341 - *thermophilic acterium PS-3* | | |
| 600 | 3 | 1640 | 882 | gi\|763052 | integrase [Bacteriophage T270] | 88 | 68 |
| 669 | 1 | 2 | 514 | gi\|153801 | enzyme scr-II [Streptococcus mutans] | 88 | 75 |
| 808 | 2 | 624 | 394 | gi\|1574781 | exodeoxyribonuclease V (recB) [Haemophilus influenzae] | 88 | 77 |
| 871 | 1 | 714 | 229 | gi\|1574120 | branched-chain-amino-acid transaminase [Haemophilus influenzae] | 88 | 79 |
| 979 | 1 | 1 | 384 | gnl\|PID\|e187579 | DNA-directed RNA polymerase [Listeria innocua] | 88 | 78 |
| 983 | 1 | 34 | 282 | gi\|40026 | homologous to E.coli gidA [Bacillus subtilis] | 88 | 78 |
| 47 | 5 | 6799 | 5810 | gi\|532204 | prs [Listeria monocytogenes] | 87 | 79 |
| 69 | 3 | 2033 | 750 | gi\|1377831 | unknown [Bacillus subtilis] | 87 | 74 |
| 73 | 2 | 1432 | 167 | gi\|143434 | Rho Factor [Bacillus subtilis] | 87 | 76 |
| 76 | 5 | 2412 | 3740 | gi\|496283 | lysin [Bacteriophage Tuc2009] | 87 | 75 |
| 88 | 3 | 1600 | 2016 | gnl\|PID\|e137596 | heat shock induced protein HtpO [Lactobacillus leichmannii] | 87 | 75 |
| 89 | 7 | 6003 | 5608 | gi\|1695686 | pyruvate carboxylase [Bacillus stearothermophilus] | 87 | 77 |
| 93 | 1 | 283 | 119 | gi\|1124825 | unknown protein [Chlamydia trachomatis] | 87 | 56 |
| 104 | 1 | 2945 | 3 | gnl\|PID\|e199387 | carbamoyl-phosphate synthase [Lactobacillus plantarum] | 87 | 75 |
| 124 | 4 | 3191 | 2274 | gi\|995767 | UDP-glucose pyrophosphorylase [Streptococcus pyogenes] | 87 | 76 |
| 273 | 2 | 608 | 1108 | gi\|1184680 | polynucleotide phosphorylase [Bacillus subtilis] | 87 | 76 |
| 293 | 2 | 1020 | 532 | gi\|153741 | ATP-binding protein [Streptococcus mutans] | 87 | 74 |
| 326 | 5 | 4534 | 3533 | gi\|143378 | pyruvate decarboxylase (E-1) beta subunit [Bacillus subtilis] gi\|1377836 pyruvate decarboxylase E-1 beta subunit [Bacillus ubtilis] | 87 | 74 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 334 | 3 | 3182 | 3340 | pir\|A36324\|A36324 | growth arrest-specific protein - mouse | 87 | 50 |
| 337 | 1 | 1382 | 186 | gi\|308861 | GTG start codon [Lactococcus lactis] | 87 | 75 |
| 338 | 8 | 6925 | 5723 | gi\|149575 | L(+)-lactate dehydrogenase [Lactobacillus casei] sp\|P00343\|LDH_LACCA L-LACTATE DEHYDROGENASE (EC 1.1.1.27). {SUB -326} | 87 | 73 |
| 367 | 18 | 8782 | 8450 | pir\|A02819\|R5BS24 | ribosomal protein L24 - Bacillus stearothermophilus | 87 | 70 |
| 388 | 2 | 410 | 183 | gnl\|PID\|e225674 | unknown [Schizosaccharomyces pombe] | 87 | 75 |
| 440 | 1 | 466 | 1797 | gi\|520754 | putative [Bacillus subtilis] | 87 | 75 |
| 508 | 1 | 694 | 137 | gi\|496558 | orfX [Bacillus subtilis] | 87 | 73 |
| 654 | 3 | 530 | 802 | pir\|A47079\|A47079 | heat shock protein DnaJ - Lactococcus lactis | 87 | 70 |
| 18 | 1 | 3 | 413 | gi\|46912 | ribosomal protein L13 [Staphylococcus carnosus] | 86 | 70 |
| 18 | 2 | 406 | 819 | pir\|S08564\|R3BS9 | ribosomal protein S9 - Bacillus stearothermophilus | 86 | 73 |
| 50 | 1 | 84 | 1148 | gi\|452398 | threonine synthase [Bacillus sp.] | 86 | 74 |
| 74 | 14 | 10547 | 10080 | gi\|1314299 | ORF6; putative glutamyl-tRNA-transferase; similar to glutamyl-tRNA-transferase from Bacillus subtilis [Listeria monocytogenes] | 86 | 74 |
| 95 | 5 | 3176 | 3406 | gi\|487276 | Na+ -ATPase subunit C [Enterococcus hirae] | 86 | 62 |
| 114 | 8 | 9216 | 10313 | gi\|853776 | peptide chain release factor 1 [Bacillus subtilis] pir\|S55437\|S55437 peptide chain release factor 1 - Bacillus ubtilis | 86 | 69 |
| 115 | 2 | 501 | 899 | gi\|551879 | ORF 1 [Lactococcus lactis] | 86 | 70 |
| 164 | 26 | 25639 | 25842 | pir\|S34762\|S34762 | L-serine dehydratase beta chain - Clostridium sp. | 86 | 81 |
| 243 | 2 | 2143 | 1082 | gi\|143607 | sporulation protein [Bacillus subtilis] | 86 | 70 |
| 255 | 1 | 2 | 196 | gi\|755604 | unknown [Bacillus subtilis] | 86 | 64 |
| 257 | 3 | 3565 | 983 | gi\|928832 | ORF259; putative [Lactococcus lactis phage BK5-T] | 86 | 66 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| **Contig ID** | **ORF ID** | **Start (nt)** | **Stop (nt)** | **Match accession** | **Match gene name** | **% Sim** | **% Ident** |
|---|---|---|---|---|---|---|---|
| 273 | 3 | 943 | 1314 | gi\|1184680 | polynucleotide phosphorylase [Bacillus subtilis] | 86 | 65 |
| 288 | 2 | 554 | 1087 | gi\|153033 | tagatose 6-phosphate isomerase [Staphylococcus aureus] pir\|B38158\|B38158 galactose-6-phosphate isomerase 19K chain - taphylococcus aureus | 86 | 74 |
| 327 | 7 | 5183 | 5722 | gi\|153569 | H+ ATPase [Enterococcus faecalis] | 86 | 71 |
| 345 | 7 | 5111 | 5620 | gi\|1314294 | ORF1; putative 17 kDa protein [Listeria monocytogenes] | 86 | 63 |
| 350 | 3 | 1900 | 2781 | gi\|511015 | dihydroorotate dehydrogenase A [Lactococcus lactis] sp\|P54321\|PYDA_LACLC DIHYDROOROTATE DEHYDROGENASE A (EC 1.3.3.1) DIHYDROOROTATE OXIDASE A) (DHODEHASE A). | 86 | 73 |
| 363 | 3 | 3328 | 4233 | gi\|1657517 | hypothetical protein [Escherichia coli] | 86 | 59 |
| 367 | 25 | 11216 | 10851 | gi\|1165308 | L22 [Bacillus subtilis] | 86 | 68 |
| 367 | 26 | 11534 | 11220 | gi\|1165307 | S19 [Bacillus subtilis] | 86 | 77 |
| 367 | 30 | 13995 | 13453 | gi\|1165303 | L3 [Bacillus subtilis] | 86 | 75 |
| 393 | 1 | 1 | 660 | sp\|P33898\|G3P3_ECOLI | GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE C (EC 1.2.1.12) (GAPDH-C). | 86 | 77 |
| 396 | 1 | 1 | 192 | gi\|944942 | RipX [Bacillus subtilis] | 86 | 77 |
| 438 | 3 | 1279 | 1560 | gi\|1001878 | CspL protein [Listeria monocytogenes] | 86 | 75 |
| 510 | 1 | 1008 | 199 | gi\|473795 | 'ORF' [Escherichia coli] | 86 | 71 |
| 510 | 2 | 1912 | 962 | gi\|473794 | 'ORF' [Escherichia coli] | 86 | 76 |
| 539 | 1 | 705 | 4 | gi\|467477 | unknown [Bacillus subtilis] | 86 | 79 |
| 570 | 2 | 2069 | 1023 | gi\|881511 | Ccpa protein [Lactobacillus casei] | 86 | 72 |
| 654 | 2 | 240 | 575 | pir\|A47079\|A47079 | heat shock protein DnaJ - Lactococcus lactis | 86 | 77 |
| 677 | 1 | 431 | 102 | gi\|1389732 | S-adenosylmethionine synthetase [Bacillus subtilis] | 86 | 80 |
| 984 | 1 | 1 | 147 | pir\|A56922\|A56922 | transcription factor shn - fruit fly (Drosophila melanogaster) | 86 | 73 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 5 | 11 | 7720 | 8487 | gi\|41015 | aspartate-tRNA ligase [Escherichia coli] | 85 | 71 |
| 34 | 2 | 2133 | 1711 | gi\|47828 | pyruvate kinase [Bacillus stearothermophilus] | 85 | 75 |
| 97 | 4 | 2666 | 2517 | pir\|S39341\|S39341 | grpE protein - Lactococcus lactis | 85 | 66 |
| 103 | 2 | 1263 | 946 | gi\|143364 | phosphoribosyl aminoimidazole carboxylase I (PUR-E) [Bacillus ubtilis] | 85 | 68 |
| 103 | 3 | 1465 | 1169 | gi\|143364 | phosphoribosyl aminoimidazole carboxylase I (PUR-E) [Bacillus ubtilis] | 85 | 67 |
| 129 | 3 | 2395 | 3258 | gi\|143766 | (thrSv) (EC 6.1.1.3) [Bacillus subtilis] | 85 | 67 |
| 129 | 4 | 3240 | 4445 | gi\|143766 | (thrSv) (EC 6.1.1.3) [Bacillus subtilis] | 85 | 78 |
| 188 | 1 | 86 | 1447 | gnl\|PID\|e214721 | glutamine synthetase [Staphylococcus aureus] | 85 | 71 |
| 217 | 3 | 673 | 1086 | gi\|520540 | unknown [Bacillus subtilis] | 85 | 72 |
| 241 | 2 | 1715 | 1086 | gi\|495089 | recombinase [Staphylococcus aureus] | 85 | 68 |
| 285 | 2 | 712 | 993 | gi\|40014 | pot. ORF 446 (aa 1-446) [Bacillus subtilis] | 85 | 77 |
| 293 | 3 | 1149 | 1595 | gi\|755604 | unknown [Bacillus subtilis] | 85 | 66 |
| 300 | 2 | 2738 | 2220 | gi\|289261 | comE ORF2 [Bacillus subtilis] | 85 | 72 |
| 305 | 2 | 1853 | 2695 | pir\|S09411\|S09411 | spoIIIE protein - Bacillus subtilis | 85 | 70 |
| 322 | 1 | 1 | 171 | gi\|153562 | aspartate beta-semialdehyde dehydrogenase (EC 1.2.1.11) Streptococcus mutans] | 85 | 67 |
| 327 | 4 | 4056 | 4784 | gi\|153567 | H+ ATPase [Enterococcus faecalis] | 85 | 66 |
| 367 | 10 | 5417 | 4959 | pir\|A02795\|R5BS15 | ribosomal protein L15 - Bacillus stearothermophilus | 85 | 76 |
| 383 | 3 | 3168 | 2953 | gnl\|PID\|e274577 | csp [Lactobacillus plantarum] | 85 | 79 |
| 404 | 3 | 3069 | 2101 | gi\|143402 | recombination protein (ttg start codon) [Bacillus subtilis] gi\|1303923 RecN [Bacillus subtilis] | 85 | 72 |
| 469 | 1 | 2 | 724 | gi\|508979 | GTP-binding protein [Bacillus subtilis] | 85 | 78 |
| 488 | 1 | 1 | 996 | gi\|532548 | ORF15 [Enterococcus faecalis] | 85 | 67 |
| 535 | 5 | 6468 | 4849 | gi\|634107 | kdpB [Escherichia coli] | 85 | 68 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 584 | 3 | 732 | 562 | gi\|467374 | single strand DNA binding protein [Bacillus subtilis] sp\|P37455\|SSB_BACSU SINGLE-STRAND BINDING PROTEIN (SSB) HELIX-DESTABILIZING PROTEIN). | 85 | 75 |
| 695 | 1 | 78 | 500 | gi\|499384 | orf189 [Bacillus subtilis] | 85 | 75 |
| 836 | 1 | 1 | 357 | gi\|153801 | enzyme scr-II [Streptococcus mutans] | 85 | 69 |
| 17 | 20 | 17212 | 18813 | gi\|532548 | ORF15 [Enterococcus faecalis] | 84 | 68 |
| 23 | 31 | 18728 | 19987 | gnl\|PID\|e208211 | DNA topoisomerase IV [Streptococcus pneumoniae] | 84 | 68 |
| 34 | 3 | 3112 | 2144 | gi\|143312 | 6-phospho-1-fructokinase (gtg start codon; EC 2.7.1.11) [Bacillus tearothermophilus] | 84 | 69 |
| 36 | 1 | 1 | 1152 | gi\|1644223 | elongation factor G [Bacillus subtilis] | 84 | 73 |
| 49 | 12 | 6730 | 8190 | gi\|456319 | 74kDa protein [Bacteriophage FC1] | 84 | 65 |
| 51 | 2 | 1379 | 1663 | gi\|468207 | Submitter comments: A Mg2+ transporting P-type ATPase highly omologous with mgtB ATPase at 80 min on Salmonella chromosome. ediates the influx of Mg2+ only. Transcription regulated by xtracellular Mg2+ [Salmonella typhimurium] | 84 | 71 |
| 95 | 6 | 3330 | 3707 | gi\|487277 | Na+ -ATPase subunit G [Enterococcus hirae] | 84 | 64 |
| 104 | 5 | 6250 | 5459 | gnl\|PID\|e199440 | aspartate carbamoyltransferase, aspartate transcarbamylase, carbamylaspartotranskinase [Lactobacillus plantarum] | 84 | 65 |
| 105 | 6 | 4605 | 5273 | gi\|467411 | recombination protein [Bacillus subtilis] | 84 | 65 |
| 114 | 11 | 12278 | 12997 | gi\|556886 | serine hydroxymethyltransferase [Bacillus subtilis] pir\|S49363\|S49363 serine hydroxymethyltransferase - Bacillus ubtilis | 84 | 74 |
| 117 | 2 | 705 | 1484 | gi\|580906 | B.subtilis genes rpmH, rnpA, 50kd, gidA and gidB [Bacillus subtilis] gi\|467381 regulation of Spo0J and Orf283 (probable) | 84 | 70 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
|  |  |  |  |  | [Bacillus ubtilis] |  |  |
| 121 | 2 | 1274 | 2119 | gi\|290643 | ATPase [Enterococcus hirae] | 84 | 67 |
| 121 | 6 | 5016 | 5219 | gi\|153765 | DNA polymerase I [Streptococcus pneumoniae] | 84 | 66 |
| 128 | 27 | 22456 | 20453 | gi\|437916 | isoleucyl-tRNA synthetase [Staphylococcus aureus] | 84 | 71 |
| 130 | 1 | 2 | 133 | gi\|1237013 | ORF2 [Bacillus subtilis] | 84 | 74 |
| 138 | 35 | 26712 | 25777 | gi\|143795 | transfer RNA-Tyr synthetase [Bacillus subtilis] | 84 | 69 |
| 164 | 28 | 26378 | 27277 | gnl\|PID\|e247026 | orf6 [Lactobacillus sake] | 84 | 72 |
| 171 | 1 | 158 | 2719 | gi\|499335 | secA protein [Staphylococcus carnosus] | 84 | 68 |
| 210 | 5 | 4870 | 3884 | gi\|950062 | hypothetical yeast protein 1 [Mycoplasma capricolum] pir\|S48578\|S48578 hypothetical protein - Mycoplasma capricolum SGC3) (fragment) | 84 | 75 |
| 217 | 7 | 5222 | 3546 | gi\|143597 | CTP synthetase [Bacillus subtilis] | 84 | 68 |
| 243 | 1 | 1088 | 126 | gi\|143608 | sporulation protein [Bacillus subtilis] | 84 | 70 |
| 275 | 1 | 578 | 48 | gi\|1103865 | formyl-tetrahydrofolate synthetase [Streptococcus mutans] | 84 | 72 |
| 281 | 1 | 333 | 698 | gi\|1303962 | YqjK [Bacillus subtilis] | 84 | 68 |
| 292 | 23 | 18340 | 18038 | gi\|142988 | membrane transport protein [Bacillus stearothermophilus] pir\|A42478\|A42478 glutamine transport protein glnQ - Bacillus tearothermophilus | 84 | 61 |
| 309 | 2 | 1114 | 722 | gi\|1644219 | RNA polymerase beta' subunit [Bacillus subtilis] | 84 | 72 |
| 315 | 1 | 668 | 3 | gi\|149601 | thymidylate synthase (EC 2.1.1.45) [Lactobacillus casei] | 84 | 72 |
| 334 | 6 | 5375 | 6862 | gi\|1354211 | PET112-like protein [Bacillus subtilis] | 84 | 71 |
| 338 | 10 | 7585 | 10479 | gi\|467444 | transcription-repair coupling factor [Bacillus subtilis] sp\|P37474\|MFD_BACSU TRANSCRIPTION-REPAIR COUPLING FACTOR | 84 | 68 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | (TRCF). | | |
| 338 | 14 | 12713 | 13018 | gi\|467448 | unknown [Bacillus subtilis] | 84 | 64 |
| 340 | 3 | 1068 | 2273 | gi\|40046 | phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus] ir\|S15936\|NUBSSA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus | 84 | 69 |
| 375 | 2 | 1430 | 1780 | gi\|1402531 | ORF10 [Enterococcus faecalis] | 84 | 64 |
| 381 | 1 | 2 | 1279 | gnl\|PID\|e208212 | DNA topoisomerase IV [Streptococcus pneumoniae] | 84 | 67 |
| 421 | 1 | 5 | 151 | gi\|710632 | beta-glucosidase [Bacillus subtilis] | 84 | 73 |
| 421 | 3 | 1229 | 1465 | gi\|710632 | beta-glucosidase [Bacillus subtilis] | 84 | 65 |
| 445 | 1 | 1080 | 190 | gi\|46985 | glucose-1-phosphate thymidylyltransferase [Salmonella enterica] ir\|S23342\|S23342 hypothetical protein 6.1 - Salmonella choleraesuis p\|P55254\|RFBA_SALAN GLUCOSE-1-PHOSPHATE THYMIDYLYLTRANSFERASE (EC 7.7.24) (DTDP-GLUCOSE SYNTHASE) (DTDP-GLUCOSE PYROPHOSPHO | 84 | 71 |
| 466 | 9 | 10467 | 11006 | gi\|147403 | mannose permease subunit II-P-Man [Escherichia coli] | 84 | 61 |
| 497 | 2 | 469 | 1680 | gi\|1220529 | methyl transferase [Streptococcus pneumoniae] | 84 | 72 |
| 545 | 2 | 309 | 2171 | gi\|532548 | ORF15 [Enterococcus faecalis] | 84 | 68 |
| 550 | 5 | 2744 | 2265 | gi\|455528 | ORF2 [Streptococcus thermophilus bacteriophage] | 84 | 54 |
| 637 | 5 | 2679 | 3545 | gnl\|PID\|e236571 | cell wall anchoring signal [Enterococcus faecalis] | 84 | 72 |
| 653 | 3 | 1023 | 736 | gi\|1408584 | LtrC [Lactococcus lactis lactis] | 84 | 72 |
| 674 | 1 | 763 | 254 | gi\|467452 | unknown [Bacillus subtilis] | 84 | 66 |
| 788 | 1 | 165 | 500 | gi\|1196907 | daunorubicin resistance protein [Streptomyces peucetius] | 84 | 66 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 796 | 1 | 1 | 366 | gi\|496504 | orf beta [Streptococcus pyogenes] | 84 | 67 |
| 812 | 1 | 2 | 415 | gi\|511075 | ORF2 [Streptococcus agalactiae] | 84 | 73 |
| 935 | 2 | 1317 | 949 | gnl\|PID\|e247026 | orf6 [Lactobacillus sake] | 84 | 73 |
| 954 | 1 | 3 | 470 | gi\|40019 | ORF 821 (aa 1-821) [Bacillus subtilis] | 84 | 67 |
| 17 | 3 | 2922 | 3311 | gi\|532555 | ORF22 [Enterococcus faecalis] | 83 | 69 |
| 17 | 12 | 8919 | 10130 | gi\|532553 | ORF20 [Enterococcus faecalis] | 83 | 64 |
| 17 | 30 | 30339 | 29137 | gi\|467416 | unknown [Bacillus subtilis] | 83 | 69 |
| 22 | 4 | 3208 | 3453 | gi\|467469 | unknown [Bacillus subtilis] | 83 | 64 |
| 28 | 3 | 6158 | 3471 | pir\|A26738\|SYBSVS | valine--tRNA ligase (EC 6.1.1.9) - Bacillus stearothermophilus | 83 | 70 |
| 75 | 2 | 359 | 1405 | gi\|310628 | phosphoenolpyruvate:sugar phosphotransferase system enzyme I Streptococcus mutans] | 83 | 72 |
| 78 | 4 | 6971 | 5841 | gi\|155571 | alcohol dehydrogenase I (adhA) (EC 1.1.1.1) [Zymomonas mobilis] pir\|A35260\|A35260 alcohol dehydrogenase (EC 1.1.1.1) I - Zymomonas obilis | 83 | 72 |
| 95 | 9 | 6859 | 7521 | gi\|487280 | Na+ -ATPase subunit D [Enterococcus hirae] | 83 | 66 |
| 98 | 3 | 2785 | 4008 | gi\|984803 | ATPase [Bacillus subtilis] | 83 | 71 |
| 107 | 3 | 1467 | 988 | sp\|P37214\|ERA_STRMU | GTP-BINDING PROTEIN ERA HOMOLOG. | 83 | 73 |
| 122 | 4 | 2781 | 3047 | gi\|467436 | unknown [Bacillus subtilis] | 83 | 60 |
| 128 | 3 | 1572 | 2633 | gi\|559471 | pyruvate,orthophosphate dikinase [Mesembryanthemum crystallinum] pir\|S49497\|S49497 pyruvate,orthophosphate dikinase (EC 2.7.9.1) - ommon ice plant | 83 | 64 |
| 128 | 34 | 28154 | 26844 | gi\|142941 | ftsZ [Bacillus subtilis] | 83 | 69 |
| 141 | 2 | 555 | 809 | pir\|S03556\|R3BS18 | ribosomal protein S18 - Bacillus stearothermophilus | 83 | 71 |
| 173 | 8 | 6237 | 7241 | gi\|451216 | Mannosephosphate Isomerase [Streptococcus mutans] | 83 | 70 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 190 | 6 | 3124 | 2738 | gi\|532555 | ORF22 [Enterococcus faecalis] | 83 | 69 |
| 273 | 1 | 29 | 436 | gnl\|PID\|e269878 | ribosomal protein S15 [Bacillus subtilis] | 83 | 71 |
| 334 | 1 | 3 | 920 | gnl\|PID\|e248484 | X-1 [Homo sapiens] | 83 | 71 |
| 350 | 4 | 2723 | 2941 | gi\|511015 | dihydroorotate dehydrogenase A [Lactococcus lactis] sp\|P54321\|PYDA_LACLC DIHYDROOROTATE DEHYDROGENASE A (EC 1.3.3.1) DIHYDROOROTATE OXIDASE A) (DHODEHASE A). | 83 | 66 |
| 367 | 2 | 1218 | 529 | gi\|142463 | RNA polymerase alpha-core-subunit [Bacillus subtilis] | 83 | 70 |
| 399 | 23 | 21538 | 22989 | gi\|40025 | homologous to E.coli 50K [Bacillus subtilis] | 83 | 67 |
| 399 | 25 | 25046 | 25879 | gi\|43939 | D-glucitol-6-P-Dehydrogenase [Klebsiella pneumoniae] ir\|S50186\|S50186 sorbitol-6-phosphate 2-dehydrogenase (EC 1.1.140) - Klebsiella pneumoniae | 83 | 61 |
| 401 | 7 | 5097 | 5864 | gi\|755153 | ATP-binding protein [Bacillus subtilis] | 83 | 64 |
| 438 | 2 | 217 | 681 | gi\|530798 | LysB [Bacteriophage phi-LC3] | 83 | 67 |
| 497 | 4 | 2191 | 3402 | gi\|532553 | ORF20 [Enterococcus faecalis] | 83 | 63 |
| 539 | 2 | 2182 | 782 | gi\|467475 | unknown [Bacillus subtilis] | 83 | 67 |
| 563 | 1 | 2 | 1084 | gi\|142521 | deoxyribodipyrimidine photolyase [Bacillus subtilis] gnl\|PID\|e255102 deoxyribodipyrimidine photolyase [Bacillus ubtilis] | 83 | 68 |
| 565 | 4 | 1018 | 1206 | gi\|1123066 | weak similarity to bovine cAMP-dependant protein kinase II-B-binding protein (PIR:A39782) [Caenorhabditis elegans] | 83 | 58 |
| 577 | 1 | 1 | 561 | gi\|1303854 | YqgG [Bacillus subtilis] | 83 | 63 |
| 635 | 3 | 1210 | 1527 | gi\|1402526 | ORF5 [Enterococcus faecalis] | 83 | 65 |
| 644 | 1 | 2 | 442 | gi\|153801 | enzyme scr-II [Streptococcus mutans] | 83 | 69 |
| 655 | 3 | 848 | 1246 | gi\|147404 | mannose permease subunit II-M-Man [Escherichia coli] | 83 | 66 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 675 | 1 | 1 | 621 | gi\|467470 | lysyl-tRNA thynthetase [Bacillus subtilis] | 83 | 71 |
| 763 | 2 | 374 | 640 | gi\|145851 | envM [Escherichia coli] | 83 | 61 |
| 774 | 1 | 658 | 2 | gi\|1256145 | YbbP [Bacillus subtilis] | 83 | 60 |
| 3 | 1 | 58 | 327 | gi\|312443 | carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus] | 82 | 70 |
| 5 | 10 | 6389 | 7708 | sp\|P30053\|SYH_STREQ | HISTIDYL-TRNA SYNTHETASE (EC 6.1.1.21) (HISTIDINE--TRNA LIGASE) (HISRS). | 82 | 71 |
| 27 | 4 | 1906 | 1145 | gi\|1303960 | YqjI [Bacillus subtilis] | 82 | 71 |
| 32 | 2 | 1333 | 965 | gi\|1303839 | YqfR [Bacillus subtilis] | 82 | 60 |
| 34 | 1 | 1643 | 324 | gnl\|PID\|e218042 | pyruvate kinase [Lactobacillus delbrueckii] | 82 | 68 |
| 55 | 9 | 4182 | 5054 | gi\|1685110 | tetrahydrofolate dehydrogenase/cyclohydrolase [Streptococcus thermophilus] | 82 | 70 |
| 62 | 7 | 4644 | 4210 | gi\|143723 | putative [Bacillus subtilis] | 82 | 66 |
| 88 | 2 | 995 | 1624 | gi\|535349 | CodW [Bacillus subtilis] | 82 | 66 |
| 94 | 7 | 4790 | 3432 | gi\|1146247 | asparaginyl-tRNA synthetase [Bacillus subtilis] | 82 | 67 |
| 110 | 23 | 21590 | 20742 | gi\|467403 | seryl-tRNA synthetase [Bacillus subtilis] | 82 | 69 |
| 114 | 7 | 8623 | 9228 | gi\|703442 | thymidine kinase [Streptococcus gordonii] | 82 | 68 |
| 123 | 6 | 4499 | 4996 | gi\|467356 | unknown [Bacillus subtilis] | 82 | 68 |
| 130 | 3 | 1413 | 2381 | gi\|308851 | ATP binding protein [Lactococcus lactis] | 82 | 64 |
| 144 | 3 | 3292 | 2339 | gnl\|PID\|e183449 | putative ATP-binding protein of ABC-type [Bacillus subtilis] | 82 | 62 |
| 144 | 7 | 5331 | 5110 | gi\|335495 | A23R; putative [Vaccinia virus] | 82 | 47 |
| 159 | 4 | 2533 | 5010 | gi\|143148 | transfer RNA-Leu synthetase [Bacillus subtilis] | 82 | 71 |
| 159 | 6 | 5845 | 5387 | gi\|467354 | unknown [Bacillus subtilis] | 82 | 55 |
| 171 | 8 | 8510 | 9349 | gi\|1591672 | phosphate transport system ATP-binding protein [Methanococcus jannaschii] | 82 | 61 |
| 222 | 5 | 2158 | 3402 | gi\|143444 | RNase PH [Bacillus subtilis] | 82 | 66 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 254 | 6 | 1621 | 1112 | gi\|49316 | ORF2 gene product [Bacillus subtilis] | 82 | 61 |
| 279 | 12 | 9839 | 8442 | gi\|1237019 | Srb [Bacillus subtilis] | 82 | 67 |
| 288 | 1 | 22 | 546 | gi\|149393 | lacA [Lactococcus lactis] | 82 | 73 |
| 345 | 8 | 5608 | 8118 | gi\|442360 | ClpC adenosine triphosphatase [Bacillus subtilis] | 82 | 63 |
| 367 | 3 | 1472 | 1110 | gi\|142463 | RNA polymerase alpha-core-subunit [Bacillus subtilis] | 82 | 75 |
| 367 | 9 | 4961 | 3660 | gi\|44073 | SecY protein [Lactococcus lactis] | 82 | 65 |
| 367 | 28 | 12719 | 12411 | pir\|A02815\|R5BS23 | ribosomal protein L23 - Bacillus stearothermophilus | 82 | 66 |
| 367 | 29 | 13330 | 12701 | gi\|1165304 | L4 [Bacillus subtilis] | 82 | 67 |
| 379 | 5 | 4396 | 3107 | gi\|887820 | UUG start; possible frameshift at end? [Escherichia coli] | 82 | 71 |
| 393 | 2 | 1145 | 711 | gi\|1303993 | YqkL [Bacillus subtilis] | 82 | 67 |
| 416 | 1 | 3 | 650 | gi\|475113 | sucrase [Pediococcus pentosaceus] | 82 | 69 |
| 477 | 1 | 1 | 1209 | gi\|309663 | signaling protein [Plasmid pCF10] | 82 | 62 |
| 497 | 7 | 3760 | 4275 | gi\|532551 | ORF18 [Enterococcus faecalis] | 82 | 67 |
| 535 | 3 | 4275 | 1666 | gi\|1747434 | KdpD [Clostridium acetobutylicum] | 82 | 62 |
| 587 | 1 | 488 | 108 | gi\|1303840 | YqfS [Bacillus subtilis] | 82 | 71 |
| 623 | 2 | 122 | 1348 | gi\|460259 | enolase [Bacillus subtilis] | 82 | 67 |
| 656 | 1 | 1 | 1908 | gi\|1184680 | polynucleotide phosphorylase [Bacillus subtilis] | 82 | 69 |
| 687 | 1 | 227 | 1252 | gi\|40218 | PRPP synthetase (AA 1-317) [Bacillus subtilis] | 82 | 64 |
| 728 | 1 | 3 | 527 | gi\|1146183 | putative [Bacillus subtilis] | 82 | 65 |
| 741 | 1 | 3 | 704 | gi\|153804 | sucrose-6-phosphate hydrolase [Streptococcus mutans] | 82 | 66 |
| 846 | 1 | 458 | 3 | gnl\|PID\|e221400 | tex gene product [Bordetella pertussis] | 82 | 76 |
| 865 | 1 | 18 | 308 | gi\|416006 | orf CJO1.2 [Campylobacter jejuni] | 82 | 57 |
| 876 | 1 | 207 | 689 | gi\|1064795 | function unknown [Bacillus subtilis] | 82 | 62 |
| 925 | 1 | 436 | 128 | gi\|1773195 | hypothetical [Escherichia coli] | 82 | 74 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 983 | 2 | 280 | 474 | gi\|40026 | homologous to E.coli gidA [Bacillus subtilis] | 82 | 78 |
| 12 | 3 | 4778 | 5788 | gi\|1100074 | tryptophanyl-tRNA synthetase [Clostridium longisporum] | 81 | 68 |
| 31 | 4 | 2984 | 4456 | gi\|849026 | hypothetical 54.6-kDa protein [Bacillus subtilis] | 81 | 68 |
| 34 | 6 | 6707 | 6910 | gi\|606067 | ORF_f444 [Escherichia coli] | 81 | 54 |
| 37 | 1 | 1 | 144 | gi\|1303854 | YqgG [Bacillus subtilis] | 81 | 59 |
| 37 | 3 | 2671 | 1958 | gi\|40056 | phoP gene product [Bacillus subtilis] | 81 | 61 |
| 57 | 3 | 1733 | 3220 | gi\|1657506 | hypothetical protein [Escherichia coli] | 81 | 66 |
| 60 | 5 | 5564 | 4440 | gi\|143370 | phosphoribosylpyrophosphate amidotransferase (PUR-F; EC 2.4.2.14) Bacillus subtilis] | 81 | 63 |
| 73 | 3 | 2706 | 1450 | gi\|853767 | UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus ubtilis] | 81 | 61 |
| 88 | 4 | 1977 | 2732 | gnl\|PID\|e137596 | heat shock induced protein HtpO [Lactobacillus leichmannii] | 81 | 67 |
| 88 | 5 | 2723 | 3040 | gi\|535350 | CodX [Bacillus subtilis] | 81 | 65 |
| 101 | 4 | 3091 | 2435 | gi\|1109687 | ProZ [Bacillus subtilis] | 81 | 60 |
| 101 | 7 | 5884 | 4661 | gi\|1109684 | ProV [Bacillus subtilis] | 81 | 64 |
| 101 | 9 | 7501 | 7965 | gi\|1001768 | queuosine biosynthesis protein QueA [Synechocystis sp.] | 81 | 47 |
| 116 | 5 | 2766 | 3395 | gi\|1146234 | dihydrodipicolinate reductase [Bacillus subtilis] | 81 | 66 |
| 121 | 5 | 4811 | 5074 | gi\|153765 | DNA polymerase I [Streptococcus pneumoniae] | 81 | 64 |
| 121 | 7 | 5203 | 7488 | gi\|153765 | DNA polymerase I [Streptococcus pneumoniae] | 81 | 70 |
| 127 | 5 | 5103 | 3826 | gi\|290561 | o188 [Escherichia coli] | 81 | 48 |
| 147 | 1 | 299 | 1279 | gi\|467462 | cysteine synthetase A [Bacillus subtilis] | 81 | 65 |
| 147 | 2 | 1370 | 1861 | gnl\|PID\|e281583 | hypothetical 16.4 kd protein [Bacillus | 81 | 63 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | subtilis] | | |
| 154 | 1 | 168 | 638 | gi\|149533 | conjugated bile acid hydrolase [Lactobacillus plantarum] | 81 | 66 |
| 154 | 2 | 1074 | 1277 | gnl\|PID\|e242898 | aBIR [Lactococcus lactis] | 81 | 59 |
| 158 | 14 | 13790 | 12324 | gi\|558559 | pyrimidine nucleoside phosphorylase [Bacillus subtilis] | 81 | 71 |
| 164 | 5 | 2469 | 3035 | gi\|727436 | putative 20-kDa protein [Lactococcus lactis] | 81 | 61 |
| 223 | 8 | 5293 | 6153 | gnl\|PID\|e254976 | hypothetical protein [Bacillus subtilis] | 81 | 66 |
| 238 | 1 | 185 | 937 | gi\|622991 | mannitol transport protein [Bacillus stearothermophilus] sp\|P50852\|PTMB_BACST PTS SYSTEM, MANNITOL-SPECIFIC IIBC COMPONENT EIIBC-MTL) (MANNITOL- PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE NZYME II, BC COMPONENT) (EC 2.7.1.69) (EII-MTL). | 81 | 68 |
| 276 | 7 | 3109 | 2819 | pir\|A41207\|A41207 | collagen 13, nonfibrillar - freshwater sponge (Ephydatia muelleri) (fragment) | 81 | 77 |
| 307 | 2 | 1983 | 3617 | gi\|153742 | dextran glucosidase [Streptococcus mutans] | 81 | 69 |
| 322 | 2 | 122 | 286 | gi\|296147 | Asd protein [Bacillus subtilis] | 81 | 63 |
| 326 | 6 | 5352 | 4513 | gi\|40041 | pyruvate dehydrogenase (lipoamide) [Bacillus stearothermophilus] ir\|S10798\|DEBSPF pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) pha chain - Bacillus stearothermophilus | 81 | 69 |
| 329 | 3 | 1774 | 1448 | gi\|1117994 | surface antigen A variant precursor [Streptococcus pneumoniae] | 81 | 72 |
| 346 | 3 | 1056 | 1199 | gi\|536970 | ORF_f543 [Escherichia coli] | 81 | 43 |
| 362 | 4 | 1131 | 2213 | gi\|1001826 | cadmium-transporting ATPase [Synechocystis sp.] | 81 | 64 |
| 391 | 3 | 1345 | 575 | gi\|1184967 | ScrR [Streptococcus mutans] | 81 | 66 |
| 441 | 3 | 1873 | 3447 | gi\|1742675 | Phosphotransferase system enzyme II (EC 2.7.1.69) MalX [Escherichia coli] | 81 | 64 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 556 | 2 | 1062 | 493 | gi\|1553037 | RecN [Bacillus subtilis] | 81 | 66 |
| 710 | 2 | 361 | 816 | gi\|1303840 | YqfS [Bacillus subtilis] | 81 | 68 |
| 804 | 1 | 403 | 2 | gi\|149533 | conjugated bile acid hydrolase [Lactobacillus plantarum] | 81 | 68 |
| 5 | 7 | 3311 | 4255 | gi\|407881 | stringent response-like protein [Streptococcus equisimilis] pir\|S39975\|S39975 stringent response-like protein - Streptococcus quisimilis | 80 | 62 |
| 17 | 10 | 8283 | 8438 | gi\|1326394 | B0218.7 gene product [Caenorhabditis elegans] | 80 | 53 |
| 17 | 15 | 12258 | 12776 | gi\|532551 | ORF18 [Enterococcus faecalis] | 80 | 63 |
| 22 | 1 | 3 | 2180 | gi\|44027 | Tma protein [Lactococcus lactis] | 80 | 70 |
| 37 | 6 | 3707 | 5140 | pir\|B47154\|B47154 | signal recognition particle 54K chain homolog Ffh - Bacillus subtilis | 80 | 64 |
| 42 | 1 | 2 | 259 | gi\|1066157 | chaperonin-10 [Thermus aquaticus thermophilus] | 80 | 66 |
| 49 | 16 | 11106 | 11309 | gi\|1136430 | similar to hypothetical protein YM9959.11C of S.cerevisiae. [Homo sapiens] | 80 | 53 |
| 60 | 4 | 4465 | 3407 | gi\|143371 | phosphoribosyl aminoimidazole synthetase (PUR-M) [Bacillus subtilis] pir\|H29326\|AJBSCL phosphoribosylformylglycinamidine cyclo-ligase EC 6.3.3.1) - Bacillus subtilis | 80 | 62 |
| 60 | 9 | 9023 | 8745 | pir\|E29326\|E29326 | hypothetical protein (pur operon) - Bacillus subtilis | 80 | 50 |
| 66 | 1 | 1 | 783 | gi\|520753 | DNA topoisomerase I [Bacillus subtilis] | 80 | 66 |
| 80 | 3 | 2519 | 1821 | gnl\|PID\|e236074 | beta-phosphoglucomutase [Lactococcus lactis] | 80 | 62 |
| 83 | 9 | 6268 | 5378 | gi\|1070079 | R08B4.1 [Caenorhabditis elegans] | 80 | 72 |
| 89 | 18 | 19093 | 18845 | gi\|39451 | type III restriction endonuclease [Bacillus cereus] ir\|S15518\|JC1116 type III site-specific deoxyribonuclease (EC | 80 | 72 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | 1.21.5) - Bacillus cereus (fragment) | | |
| 97 | 1 | 366 | 4 | gi\|148506 | dnaJ [Erysipelothrix rhusiopathiae] | 80 | 70 |
| 107 | 2 | 1094 | 591 | sp\|P37214\|ERA_STRMU | GTP-BINDING PROTEIN ERA HOMOLOG. | 80 | 64 |
| 114 | 3 | 1474 | 5076 | gi\|43863 | pyruvate-flavodoxin oxidoreductase [Klebsiella pneumoniae] ir\|S01997\|QQKBFP pyruvate (flavodoxin) dehydrogenase (EC 1.2.99.-) Klebsiella pneumoniae | 80 | 62 |
| 117 | 3 | 1456 | 2367 | gi\|40031 | spoOJ93 gene product [Bacillus subtilis] | 80 | 56 |
| 126 | 3 | 1857 | 709 | gi\|551854 | ORF2 [Erwinia herbicola] | 80 | 68 |
| 128 | 28 | 23265 | 22447 | gi\|437916 | isoleucyl-tRNA synthetase [Staphylococcus aureus] | 80 | 63 |
| 133 | 10 | 9128 | 9856 | gi\|520844 | orf4 [Bacillus subtilis] | 80 | 63 |
| 158 | 4 | 3926 | 2703 | gi\|944943 | phosphopentomutase [Bacillus subtilis] | 80 | 64 |
| 172 | 5 | 3732 | 3920 | sp\|P20182\|YT14_STRFR | HYPOTHETICAL 29.1 KD PROTEIN IN TRANSPOSON TN4556. | 80 | 63 |
| 180 | 16 | 15548 | 16393 | gi\|1773200 | hypothetical protein [Escherichia coli] | 80 | 66 |
| 181 | 10 | 8597 | 7407 | gi\|143806 | AroF [Bacillus subtilis] | 80 | 64 |
| 194 | 4 | 1580 | 1957 | gi\|47394 | 5-oxoprolyl-peptidase [Streptococcus pyogenes] | 80 | 66 |
| 213 | 5 | 3515 | 4078 | gnl\|PID\|e199384 | pyrR gene product [Lactobacillus plantarum] | 80 | 65 |
| 217 | 11 | 7724 | 8395 | gi\|1561567 | Unknown [Bacillus subtilis] | 80 | 65 |
| 218 | 6 | 4843 | 5331 | gi\|1574120 | branched-chain-amino-acid transaminase [Haemophilus influenzae] | 80 | 64 |
| 225 | 8 | 6092 | 5829 | gi\|530459 | similar to phosphotransferase EII [Mycoplasma capricolum] | 80 | 52 |
| 229 | 2 | 1170 | 178 | gi\|1502419 | PlsX [Bacillus subtilis] | 80 | 59 |
| 243 | 3 | 2545 | 2150 | gi\|1732315 | transport system permease homolog [Listeria monocytogenes] | 80 | 64 |
| 275 | 2 | 694 | 939 | gi\|1256629 | cold-shock protein [Bacillus subtilis] | 80 | 65 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 307 | 3 | 3607 | 3888 | gi\|1321625 | exo-alpha-1, 4-glucosidase [Bacillus stearothermophilus] | 80 | 73 |
| 322 | 3 | 284 | 1090 | gi\|142828 | aspartate semialdehyde dehydrogenase [Bacillus subtilis] sp\|Q04797\|DHAS_BACSU ASPARTATE-SEMIALDEHYDE DEHYDROGENASE (EC .2.1.11) (ASA DEHYDROGENASE). | 80 | 62 |
| 349 | 1 | 2 | 616 | gi\|495089 | recombinase [Staphylococcus aureus] | 80 | 65 |
| 367 | 7 | 3511 | 2924 | gi\|44074 | adenylate kinase [Lactococcus lactis] | 80 | 64 |
| 386 | 7 | 4305 | 5306 | gi\|149396 | lacD [Lactococcus lactis] | 80 | 64 |
| 394 | 3 | 2642 | 3757 | pir\|B39096\|B39096 | alkaline phosphatase (EC 3.1.3.1) III precursor - Bacillus subtilis | 80 | 64 |
| 399 | 17 | 12070 | 13488 | gi\|1591862 | oxaloacetate decarboxylase, alpha subunit [Methanococcus jannaschii] | 80 | 61 |
| 399 | 24 | 22979 | 24907 | gi\|40026 | homologous to E.coli gidA [Bacillus subtilis] | 80 | 67 |
| 435 | 3 | 2217 | 2032 | gi\|559863 | clyA [Plasmid pAD1] | 80 | 78 |
| 466 | 1 | 3 | 1208 | gi\|467330 | replicative DNA helicase [Bacillus subtilis] | 80 | 61 |
| 475 | 4 | 3402 | 2947 | gi\|532547 | ORF14 [Enterococcus faecalis] | 80 | 68 |
| 491 | 4 | 3844 | 4392 | gi\|473892 | large-conductance mechanosensitive channel [Escherichia coli] gi\|473420 yhdC [Escherichia coli] | 80 | 56 |
| 605 | 2 | 1252 | 338 | gi\|580875 | ipa-57d gene product [Bacillus subtilis] | 80 | 69 |
| 615 | 1 | 760 | 14 | gi\|467469 | unknown [Bacillus subtilis] | 80 | 66 |
| 668 | 1 | 117 | 587 | pir\|S16974\|R5BS7F | ribosomal protein L9 - Bacillus stearothermophilus | 80 | 71 |
| 684 | 2 | 694 | 464 | gi\|786314 | Highly similar to Glycogen debranching enzyme 4-alpha-glucanotransferase, Swiss Prot. accession number P35573) Saccharomyces cerevisiae] | 80 | 33 |
| 767 | 1 | 1 | 480 | gi\|41828 | istB gene product [Escherichia coli] | 80 | 52 |
| 818 | 1 | 1 | 357 | gi\|1743856 | intrageneric coaggregation-relevant | 80 | 66 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | adhesin [Streptococcus gordonii] | | |
| 833 | 1 | 325 | 95 | gi\|1561567 | Unknown [Bacillus subtilis] | 80 | 68 |
| 934 | 1 | 394 | 56 | gi\|1001706 | ABC transporter subunit [Synechocystis sp.] | 80 | 63 |
| 948 | 1 | 465 | 4 | gi\|1773196 | similar to B. stearothermophilus N-carbamyl-L-amino acid amidohydrolase [Escherichia coli] | 80 | 59 |
| 949 | 1 | 61 | 411 | gi\|1330380 | Similar to cystathionine gamma-lyase [Caenorhabditis elegans] | 80 | 61 |
| 20 | 2 | 468 | 1262 | gi\|1256698 | chitinase [Serratia marcescens] | 79 | 67 |
| 22 | 3 | 2420 | 3238 | gi\|467460 | unknown [Bacillus subtilis] | 79 | 59 |
| 24 | 1 | 39 | 1109 | gi\|1303821 | YqfE [Bacillus subtilis] | 79 | 61 |
| 26 | 1 | 214 | 873 | gi\|403984 | deoxyguanosine kinase/deoxyadenosine kinase(I) subunit Lactobacillus acidophilus] | 79 | 68 |
| 47 | 8 | 10268 | 8106 | gi\|153657 | mismatch repair protein [Streptococcus pneumoniae] pir\|A33589\|A33589 mismatch repair protein hexB - Streptococcus neumoniae | 79 | 63 |
| 48 | 9 | 9905 | 9198 | gi\|290566 | f213 [Escherichia coli] | 79 | 53 |
| 58 | 4 | 4677 | 3694 | gi\|1653179 | hydrogenase subunit [Synechocystis sp.] | 79 | 52 |
| 63 | 6 | 3605 | 5443 | gi\|1064813 | homologous to sp:PHOR_BACSU [Bacillus subtilis] | 79 | 55 |
| 88 | 8 | 5493 | 4771 | gnl\|PID\|e208252 | unidentified [Streptococcus pneumoniae] | 79 | 57 |
| 146 | 8 | 6649 | 5609 | gi\|153676 | tagatose 1,6-aldolase [Streptococcus mutans] | 79 | 63 |
| 149 | 4 | 2554 | 1976 | gi\|1216490 | DNA/pantothenate metabolism flavoprotein [Streptococcus mutans] | 79 | 64 |
| 158 | 2 | 1859 | 1143 | gi\|1276873 | DeoD [Streptococcus thermophilus] | 79 | 67 |
| 179 | 19 | 19022 | 18417 | gi\|467372 | 3'-exo-deoxyribonuclease [Bacillus subtilis] | 79 | 61 |
| 222 | 2 | 982 | 230 | gi\|142988 | membrane transport protein [Bacillus | 79 | 59 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | stearothermophilus] pir\|A42478\|A42478 glutamine transport protein glnQ - Bacillus tearothermophilus | | |
| 228 | 6 | 4060 | 3401 | gi\|413950 | ipa-26d gene product [Bacillus subtilis] | 79 | 55 |
| 229 | 3 | 3270 | 1219 | gnl\|PID\|e186699 | MmsA [Streptococcus pneumoniae] | 79 | 62 |
| 238 | 7 | 5750 | 5100 | gi\|596046 | L8003.16 gene product [Saccharomyces cerevisiae] | 79 | 55 |
| 269 | 10 | 6664 | 5489 | gi\|1303788 | YqeH [Bacillus subtilis] | 79 | 63 |
| 274 | 1 | 1 | 1143 | gi\|153062 | helicase [Staphylococcus aureus] | 79 | 65 |
| 290 | 9 | 7364 | 8779 | gi\|466882 | pps1; B1496_C2_189 [Mycobacterium leprae] | 79 | 64 |
| 292 | 22 | 18122 | 17595 | gi\|1303951 | YqiZ [Bacillus subtilis] | 79 | 61 |
| 316 | 3 | 864 | 2003 | gi\|1146207 | putative [Bacillus subtilis] | 79 | 58 |
| 326 | 2 | 1772 | 360 | gi\|40044 | dihydrolipoamide dehydrogenase [Bacillus stearothermophilus] ir\|S13839\|S13839 dihydrolipoamide dehydrogenase (EC 1.8.1.4) - cillus stearothermophilus | 79 | 65 |
| 363 | 5 | 5738 | 7180 | gi\|1657519 | hypothetical protein [Escherichia coli] | 79 | 63 |
| 367 | 11 | 5668 | 5447 | gi\|216337 | ORF for L30 ribosomal protein [Bacillus subtilis] | 79 | 63 |
| 375 | 5 | 4346 | 3393 | gi\|1644203 | unknown [Bacillus subtilis] | 79 | 62 |
| 406 | 2 | 666 | 1481 | gi\|49316 | ORF2 gene product [Bacillus subtilis] | 79 | 58 |
| 460 | 7 | 4973 | 5860 | gi\|1276664 | acetyl-CoA carboxylase carboxytransferase beta subunit [Porphyra purpurea] | 79 | 62 |
| 486 | 1 | 380 | 3 | gi\|1256618 | transport protein [Bacillus subtilis] | 79 | 63 |
| 488 | 3 | 987 | 1997 | gi\|532547 | ORF14 [Enterococcus faecalis] | 79 | 69 |
| 500 | 2 | 1358 | 681 | gi\|535662 | transposase [Insertion sequence IS1251] | 79 | 75 |
| 523 | 3 | 1803 | 820 | gi\|142981 | ORF5; This ORF includes a region (aa23-103) containing a potential ron-sulphur centre homologous to a region of Rhodospirillum rubrum nd Chromatium vinosum; putative [Bacillus stearothermophilus] pir\|PQ0299\|PQ0299 | 79 | 62 |

**Table 2:** ***E. faecalis*** **- Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | hypothetical protein 5 (gldA 3' region) - | | |
| 552 | 2 | 2401 | 902 | gi\|887851 | ORF_o479 [Escherichia coli] | 79 | 63 |
| 587 | 2 | 622 | 434 | gi\|1303840 | YqfS [Bacillus subtilis] | 79 | 66 |
| 612 | 1 | 1 | 378 | gi\|1064791 | function umknown [Bacillus subtilis] | 79 | 56 |
| 654 | 1 | 2 | 286 | pir\|A47079\|A47079 | heat shock protein DnaJ - Lactococcus lactis | 79 | 75 |
| 701 | 2 | 325 | 534 | gi\|143793 | tyrosyl-tRNA synthetase [Bacillus caldotenax] | 79 | 63 |
| 708 | 2 | 369 | 566 | gi\|488430 | alcohol dehydrogenase 2 [Entamoeba histolytica] | 79 | 66 |
| 840 | 1 | 140 | 1078 | gi\|1573250 | aspartate aminotransferase (aspC) [Haemophilus influenzae] | 79 | 65 |
| 5 | 9 | 5555 | 6049 | gi\|407880 | ORF1 [Streptococcus equisimilis] | 78 | 58 |
| 33 | 4 | 3755 | 4597 | gi\|1742846 | NH(3)-dependent NAD(+) synthetase (EC 6.3.5.1) (Nitrogen-regulatory protein). [Escherichia coli] | 78 | 64 |
| 60 | 7 | 8100 | 5854 | gi\|143369 | phosphoribosylformyl glycinamidine synthetase II (PUR-Q) [Bacillus ubtilis] | 78 | 62 |
| 65 | 4 | 3407 | 2625 | gi\|1661179 | high affinity branched chain amino acid transport protein [Streptococcus mutans] | 78 | 67 |
| 76 | 7 | 5760 | 4747 | gi\|1161061 | dioxygenase [Methylobacterium extorquens] | 78 | 62 |
| 81 | 11 | 7141 | 6824 | gi\|1072380 | ORF3 [Lactococcus lactis] | 78 | 67 |
| 83 | 5 | 2559 | 2843 | gi\|1256896 | L9606.1 gene product [Saccharomyces cerevisiae] | 78 | 52 |
| 85 | 4 | 4298 | 3288 | gi\|142612 | branched chain alpha-keto acid dehydrogenase E1-beta [Bacillus ubtilis] | 78 | 61 |
| 85 | 8 | 6723 | 6307 | gi\|1303941 | YqiV [Bacillus subtilis] | 78 | 62 |
| 88 | 10 | 6477 | 6689 | gi\|222585 | nucleocapsid protein [Sialodacryoadenitis virus] | 78 | 57 |
| 93 | 5 | 1838 | 2641 | gi\|405133 | putative [Bacillus subtilis] | 78 | 51 |
| 117 | 1 | 3 | 707 | gi\|40027 | homologous to E.coli gidB [Bacillus | 78 | 64 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | subtilis] | | |
| 117 | 11 | 9624 | 8338 | gi\|467403 | seryl-tRNA synthetase [Bacillus subtilis] | 78 | 63 |
| 132 | 2 | 2323 | 2024 | gi\|683484 | fusion protein [Mumps virus] | 78 | 63 |
| 133 | 3 | 2241 | 3413 | gi\|405622 | unknown [Bacillus subtilis] | 78 | 63 |
| 150 | 2 | 568 | 1425 | gnl\|PID\|e185373 | ceuD gene product [Campylobacter coli] | 78 | 52 |
| 155 | 2 | 604 | 1182 | gi\|285628 | transcription antitermination factor NusG [Bacillus subtilis] pir\|S39859\|S39859 transcription antitermination factor NusG - acillus subtilis | 78 | 61 |
| 156 | 2 | 308 | 2629 | gi\|1573874 | ATP-dependent protease binding subunit (clpB) [Haemophilus influenzae] | 78 | 59 |
| 158 | 3 | 2719 | 1868 | gi\|1638804 | purine nucleoside phosphorylase [Bacillus stearothermophilus] | 78 | 64 |
| 160 | 5 | 2058 | 3050 | gi\|1161061 | dioxygenase [Methylobacterium extorquens] | 78 | 60 |
| 161 | 3 | 1466 | 3295 | gnl\|PID\|e280490 | unknown [Streptococcus pneumoniae] | 78 | 62 |
| 169 | 1 | 2 | 2206 | gi\|1072361 | pyruvate-formate-lyase [Clostridium pasteurianum] | 78 | 61 |
| 171 | 2 | 2833 | 3897 | sp\|P28367\|RF2_BACSU | PROBABLE PEPTIDE CHAIN RELEASE FACTOR 2 (RF-2) (FRAGMENT). | 78 | 64 |
| 180 | 15 | 14851 | 15567 | gi\|1773199 | hypothetical protein [Escherichia coli] | 78 | 67 |
| 185 | 1 | 1142 | 3 | pir\|C33496\|C33496 | hisC homolog - Bacillus subtilis | 78 | 59 |
| 188 | 3 | 1863 | 4178 | gnl\|PID\|e256969 | nifJ gene product [Enterobacter agglomerans] | 78 | 62 |
| 216 | 7 | 5136 | 5600 | gnl\|PID\|e276830 | UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis] | 78 | 60 |
| 216 | 8 | 5531 | 6508 | gnl\|PID\|e276830 | UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis] | 78 | 63 |
| 238 | 26 | 24515 | 25387 | gi\|396681 | rhamnulose-1-phosphate aldolase [Escherichia coli] | 78 | 56 |
| 256 | 6 | 4189 | 6237 | gi\|467427 | methionyl-tRNA synthetase [Bacillus | 78 | 67 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | subtilis] | | |
| 292 | 4 | 2063 | 2353 | gi\|1742823 | Proton/sodium-glutamate symport protein (Glutamate-aspartate carrier protein). [Escherichia coli] | 78 | 62 |
| 305 | 1 | 268 | 1872 | gi\|143582 | spoIIIEA protein [Bacillus subtilis] | 78 | 58 |
| 337 | 2 | 2332 | 1448 | gi\|308861 | GTG start codon [Lactococcus lactis] | 78 | 63 |
| 338 | 2 | 606 | 1466 | gi\|1773142 | similar to the 20.2kd protein in TETB-EXOA region of B. subtilis [Escherichia coli] | 78 | 66 |
| 362 | 1 | 109 | 429 | gi\|150719 | cadmium resistance protein [Plasmid pI258] | 78 | 51 |
| 379 | 3 | 2878 | 1922 | gi\|887824 | ORF_o310 [Escherichia coli] | 78 | 60 |
| 446 | 2 | 962 | 1636 | gi\|537235 | Kenn Rudd identifies as gpmB [Escherichia coli] | 78 | 43 |
| 495 | 5 | 3038 | 3502 | gi\|634107 | kdpB [Escherichia coli] | 78 | 58 |
| 502 | 3 | 3077 | 1470 | gi\|1652592 | peptide-chain-release factor 3 [Synechocystis sp.] | 78 | 58 |
| 523 | 1 | 2 | 616 | gi\|289288 | lexA [Bacillus subtilis] | 78 | 59 |
| 571 | 1 | 99 | 365 | gnl\|PID\|e249644 | YneF [Bacillus subtilis] | 78 | 65 |
| 573 | 3 | 1258 | 1971 | gi\|1731683 | component II of heptaprenyl diphosphate synthase [Bacillus stearothermophilus] | 78 | 50 |
| 575 | 2 | 434 | 168 | gi\|58831 | The experimental evidence that this sequence codes for a complete gag otein is that transfection of the viral genome results in oduction of infectious virus [Cas-Br-E murine leukemia virus] p\|P27460\|GAG_MLVCB GAG POLYPROTEIN (CONTAINS: CORE PROTEIN P15; N | 78 | 47 |
| 607 | 1 | 148 | 708 | gi\|530410 | Ala-tRNA synthetase [Mycoplasma capricolum] | 78 | 63 |
| 655 | 2 | 300 | 899 | gi\|147404 | mannose permease subunit II-M-Man [Escherichia coli] | 78 | 60 |
| 704 | 1 | 181 | 2 | gi\|467430 | unknown [Bacillus subtilis] | 78 | 63 |
| 708 | 1 | 1 | 378 | gi\|443985 | alcohol dehydrogenase [Entamoeba | 78 | 61 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | histolytica] | | |
| 732 | 1 | 661 | 2 | gi\|1064791 | function umknown [Bacillus subtilis] | 78 | 55 |
| 785 | 1 | 2 | 679 | gi\|556014 | UDP-N-acetyl muramate-alanine ligase [Bacillus subtilis] | 78 | 59 |
| 786 | 1 | 2 | 172 | gi\|536992 | SugES [Escherichia coli] | 78 | 60 |
| 820 | 2 | 1602 | 1144 | gi\|153749 | UDPglucose 4-epimerase [Streptococcus thermophilus] pir\|A44509\|A44509 UDPglucose 4-epimerase (EC 5.1.3.2) - treptococcus thermophilus | 78 | 60 |
| 887 | 1 | 337 | 2 | gi\|495046 | tripeptidase [Lactococcus lactis] | 78 | 70 |
| 970 | 2 | 395 | 234 | gi\|1652190 | Fat protein [Synechocystis sp.] | 78 | 51 |
| 4 | 7 | 6069 | 5656 | gi\|1573482 | high affinity ribose transport protein (rbsD) [Haemophilus influenzae] | 77 | 51 |
| 45 | 16 | 12065 | 14047 | gi\|666069 | orf2 gene product [Lactobacillus leichmannii] | 77 | 51 |
| 49 | 13 | 8199 | 9992 | gnl\|PID\|e228615 | homologous to yqcC of the skin element [Bacillus subtilis] | 77 | 59 |
| 60 | 2 | 2895 | 1300 | gi\|143373 | phosphoribosyl aminoimidazole carboxy formyl ormyltransferase/inosine monophosphate cyclohydrolase (PUR-H(J)) Bacillus subtilis] | 77 | 63 |
| 70 | 6 | 5118 | 3874 | gi\|912464 | No definition line found [Escherichia coli] | 77 | 53 |
| 70 | 7 | 5172 | 5756 | gi\|288413 | glutamate dehydrogenase (NADP+) [Corynebacterium glutamicum] pir\|S32227\|S32227 glutamate dehydrogenase (NADP+) (EC 1.4.1.4) - orynebacterium glutamicum | 77 | 65 |
| 74 | 10 | 7303 | 5864 | gi\|289284 | cysteinyl-tRNA synthetase [Bacillus subtilis] | 77 | 62 |
| 74 | 12 | 9559 | 8078 | gi\|289282 | glutamyl-tRNA synthetase [Bacillus subtilis] | 77 | 57 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 88 | 6 | 3013 | 3843 | gi\|535351 | CodY [Bacillus subtilis] | 77 | 57 |
| 89 | 6 | 5749 | 2510 | gi\|1695686 | pyruvate carboxylase [Bacillus stearothermophilus] | 77 | 62 |
| 91 | 1 | 396 | 728 | gi\|1184044 | L-glutamine:D-fructose-6-P amidotransferase precursor [Thermus aquaticus thermophilus] | 77 | 66 |
| 98 | 4 | 3992 | 5710 | gi\|984804 | transmembrane protein [Bacillus subtilis] | 77 | 56 |
| 124 | 1 | 2 | 940 | gnl\|PID\|e199002 | prolidase PepQ [Lactobacillus delbrueckii] | 77 | 60 |
| 158 | 5 | 4845 | 4171 | gi\|435297 | unknown [Lactococcus lactis] | 77 | 48 |
| 162 | 6 | 7426 | 5882 | gi\|142992 | glycerol kinase (glpK) (EC 2.7.1.30) [Bacillus subtilis] pir\|B45868\|B45868 glycerol kinase (EC 2.7.1.30) - Bacillus subtilis sp\|P18157\|GLPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL - PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK). | 77 | 60 |
| 164 | 1 | 179 | 1102 | gi\|882532 | ORF_o294 [Escherichia coli] | 77 | 57 |
| 164 | 22 | 24158 | 23646 | gi\|1573564 | hypothetical [Haemophilus influenzae] | 77 | 36 |
| 171 | 6 | 6656 | 7639 | gi\|1303855 | YqgH [Bacillus subtilis] | 77 | 59 |
| 171 | 9 | 9198 | 9683 | gi\|1591672 | phosphate transport system ATP-binding protein [Methanococcus jannaschii] | 77 | 57 |
| 202 | 4 | 2967 | 3422 | gi\|147782 | ruvA protein (gtg start) [Escherichia coli] | 77 | 50 |
| 202 | 6 | 3662 | 4693 | gi\|147783 | ruvB protein [Escherichia coli] | 77 | 58 |
| 213 | 1 | 3 | 1046 | gi\|1103865 | formyl-tetrahydrofolate synthetase [Streptococcus mutans] | 77 | 63 |
| 217 | 10 | 6870 | 7742 | gi\|414014 | ipa-90d gene product [Bacillus subtilis] | 77 | 50 |
| 223 | 5 | 4171 | 4902 | gnl\|PID\|e254974 | autolysin response regulator [Bacillus subtilis] | 77 | 55 |
| 223 | 7 | 5024 | 5473 | gnl\|PID\|e254975 | hypothetical protein [Bacillus subtilis] | 77 | 58 |
| 228 | 10 | 7747 | 6035 | gi\|467409 | DNA polymerase III subunit [Bacillus subtilis] | 77 | 61 |
| 229 | 15 | 16711 | 14261 | gnl\|PID\|e290286 | priA [Bacillus subtilis] | 77 | 62 |

**Table 2:** ***E. faecalis*** **- Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 232 | 3 | 1742 | 1437 | gi\|142708 | comG3 gene product [Bacillus subtilis] | 77 | 50 |
| 238 | 25 | 23174 | 24511 | pir\|B48649\|B48649 | L-rhamnose isomerase (EC 5.3.1.14) - Escherichia coli | 77 | 59 |
| 238 | 32 | 29472 | 28708 | gi\|451072 | di-tripeptide transporter [Lactococcus lactis] | 77 | 56 |
| 244 | 4 | 3591 | 2809 | gi\|1773173 | similar to M. jannaschii MJ0938 [Escherichia coli] | 77 | 60 |
| 269 | 5 | 3890 | 3522 | gi\|1303793 | YqeL [Bacillus subtilis] | 77 | 55 |
| 276 | 6 | 2840 | 2328 | pir\|PC1127\|PC1127 | hypothetical 110 protein (lytA 5' region) - Lactococcus lactis phage US3 (fragment) | 77 | 50 |
| 291 | 1 | 119 | 916 | gi\|556014 | UDP-N-acetyl muramate-alanine ligase [Bacillus subtilis] | 77 | 63 |
| 304 | 2 | 941 | 2020 | gnl\|PID\|e285001 | CTORF239 [Staphylococcus aureus] | 77 | 62 |
| 305 | 4 | 3618 | 4394 | gi\|709993 | hypothetical protein [Bacillus subtilis] | 77 | 54 |
| 327 | 8 | 5697 | 6005 | gi\|153570 | H+ ATPase [Enterococcus faecalis] | 77 | 61 |
| 341 | 4 | 1206 | 1937 | gi\|1303951 | YqiZ [Bacillus subtilis] | 77 | 62 |
| 360 | 1 | 429 | 4 | gi\|897754 | nonstructural protein NSP3 [Human rotavirus] | 77 | 38 |
| 362 | 3 | 541 | 1239 | gi\|1001826 | cadmium-transporting ATPase [Synechocystis sp.] | 77 | 60 |
| 363 | 9 | 13917 | 12652 | gi\|1574390 | C4-dicarboxylate transport protein [Haemophilus influenzae] | 77 | 55 |
| 367 | 14 | 7218 | 6679 | pir\|A02766\|R5BS0F | ribosomal protein L6 - Bacillus stearothermophilus | 77 | 63 |
| 386 | 8 | 5456 | 5776 | gnl\|PID\|e281578 | hypothetical 12.2 kd protein [Bacillus subtilis] | 77 | 61 |
| 394 | 4 | 3706 | 4167 | pir\|B39096\|B39096 | alkaline phosphatase (EC 3.1.3.1) III precursor - Bacillus subtilis | 77 | 55 |
| 402 | 1 | 710 | 3 | gi\|533105 | unknown [Bacillus subtilis] | 77 | 59 |
| 408 | 2 | 1357 | 584 | gi\|666983 | putative ATP binding subunit [Bacillus subtilis] | 77 | 58 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 460 | 6 | 3562 | 4938 | gi\|1055246 | biotin carboxylase [Bacillus subtilis] | 77 | 60 |
| 466 | 7 | 8657 | 9253 | gi\|147402 | mannose permease subunit III-Man [Escherichia coli] | 77 | 61 |
| 475 | 5 | 3794 | 3234 | gi\|532547 | ORF14 [Enterococcus faecalis] | 77 | 68 |
| 498 | 1 | 1 | 603 | gi\|410137 | ORFX13 [Bacillus subtilis] | 77 | 58 |
| 515 | 1 | 107 | 574 | gi\|1303815 | YqeY [Bacillus subtilis] | 77 | 60 |
| 518 | 6 | 2980 | 4518 | gi\|1402515 | membrane-spanning transporter protein [Clostridium perfringens] | 77 | 56 |
| 523 | 5 | 2527 | 2333 | gi\|149601 | thymidylate synthase (EC 2.1.1.45) [Lactobacillus casei] | 77 | 66 |
| 526 | 2 | 1782 | 436 | gi\|1750124 | xylose isomerase [Bacillus subtilis] | 77 | 62 |
| 552 | 7 | 6809 | 6135 | gi\|534045 | antiterminator [Bacillus subtilis] | 77 | 51 |
| 607 | 3 | 778 | 936 | gi\|1015321 | alanyl-tRNA synthetase [Homo sapiens] | 77 | 51 |
| 624 | 3 | 2289 | 2555 | gnl\|PID\|e187971 | orf121 gene product [Lactococcus lactis] | 77 | 57 |
| 781 | 1 | 15 | 485 | gi\|580883 | ipa-88d gene product [Bacillus subtilis] | 77 | 65 |
| 850 | 2 | 895 | 572 | gi\|142520 | thioredoxin [Bacillus subtilis] | 77 | 59 |
| 853 | 1 | 186 | 4 | gi\|39962 | ribosomal protein L35 (AA 1-66) [Bacillus stearothermophilus] ir\|S05347\|R5BS35 ribosomal protein L35 - Bacillus earothermophilus | 77 | 66 |
| 944 | 1 | 2 | 172 | gi\|425467 | transposase [Lactobacillus helveticus] | 77 | 50 |
| 10 | 1 | 1 | 258 | gnl\|PID\|e234078 | hom [Lactococcus lactis] | 76 | 63 |
| 12 | 4 | 7650 | 5842 | gnl\|PID\|e254877 | unknown [Mycobacterium tuberculosis] | 76 | 57 |
| 17 | 29 | 29022 | 28153 | gi\|1500003 | mutator mutT protein [Methanococcus jannaschii] | 76 | 47 |
| 23 | 15 | 8897 | 10285 | gi\|153960 | ethanolamine ammonia-lyase (eutB) [Salmonella typhimurium] pir\|A36570\|A36570 ethanolamine ammonia-lyase (EC 4.3.1.7) 55K chain Salmonella typhimurium | 76 | 64 |
| 29 | 2 | 1024 | 500 | gi\|40011 | ORF17 (AA 1-161) [Bacillus subtilis] | 76 | 61 |
| 33 | 1 | 14 | 1552 | gi\|148304 | beta-1,4-N-acetylmuramoylhydrolase | 76 | 60 |

Table 2: E. *faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | [Enterococcus hirae] pir\|A42296\|A42296 lysozyme 2 (EC 3.2.1.-) precursor - Enterococcus irae (ATCC 9790) | | |
| 34 | 7 | 7432 | 6965 | gi\|44067 | ORF1 C-terminal [Lactococcus lactis] | 76 | 59 |
| 45 | 8 | 3708 | 4166 | gi\|1303698 | BltD [Bacillus subtilis] | 76 | 56 |
| 47 | 9 | 12849 | 10270 | gi\|1002520 | MutS [Bacillus subtilis] | 76 | 59 |
| 55 | 8 | 3614 | 4105 | gi\|1303915 | YqhZ [Bacillus subtilis] | 76 | 53 |
| 55 | 11 | 6385 | 6642 | gi\|216583 | ORF1 [Escherichia coli] | 76 | 45 |
| 57 | 14 | 17283 | 16597 | gi\|1183887 | integral membrane protein [Bacillus subtilis] | 76 | 56 |
| 59 | 6 | 3112 | 2426 | gi\|392872 | repressor protein [Pasteurella multocida] | 76 | 47 |
| 64 | 1 | 1242 | 46 | gi\|483941 | blt gene product [Bacillus subtilis] | 76 | 55 |
| 67 | 3 | 1370 | 2146 | gnl\|PID\|e199390 | orotate phosphoribosyltransferase [Lactobacillus plantarum] | 76 | 57 |
| 69 | 2 | 837 | 334 | gi\|1377831 | unknown [Bacillus subtilis] | 76 | 57 |
| 70 | 1 | 164 | 1588 | gi\|895751 | putative 6-phospho-beta-glucosidase [Bacillus subtilis] pir\|S57762\|S57762 probable 6-phospho-beta-glucosidase - Bacillus ubtilis | 76 | 60 |
| 74 | 11 | 7826 | 7269 | pir\|E53402\|E53402 | serine O-acetyltransferase (EC 2.3.1.30) - Bacillus stearothermophilus | 76 | 54 |
| 74 | 13 | 10073 | 9588 | gi\|289281 | unknown [Bacillus subtilis] | 76 | 60 |
| 85 | 11 | 7809 | 7102 | gi\|457634 | butyrate kinase [Clostridium acetobutylicum] | 76 | 61 |
| 94 | 8 | 6036 | 4801 | gi\|142538 | aspartate aminotransferase [Bacillus sp.] | 76 | 57 |
| 94 | 14 | 17174 | 12801 | gi\|40060 | DNA polymerase III (AA 1-1437) [Bacillus subtilis] p\|P13267\|DP3A_BACSU DNA POLYMERASE III, ALPHA CHAIN (EC 2.7.7.7). | 76 | 62 |
| 94 | 15 | 19140 | 17407 | gi\|1573733 | prolyl-tRNA synthetase (proS) [Haemophilus influenzae] | 76 | 54 |
| 95 | 1 | 1 | 1290 | gi\|472918 | v-type Na-ATPase [Enterococcus hirae] | 76 | 59 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 95 | 4 | 2367 | 3194 | gi\|487276 | Na+ -ATPase subunit C [Enterococcus hirae] | 76 | 48 |
| 99 | 1 | 1 | 171 | gi\|1353874 | unknown [Rhodobacter capsulatus] | 76 | 52 |
| 100 | 5 | 5414 | 5064 | gi\|1591962 | M. jannaschii predicted coding region MJ1322 [Methanococcus jannaschii] | 76 | 46 |
| 100 | 27 | 23165 | 21198 | gi\|216151 | DNA polymerase (gene L; ttg start codon) [Bacteriophage SPO2] gi\|579197 SPO2 DNA polymerase (aa 1-648) [Bacteriophage SPO2] pir\|A21498\|DJBPS2 DNA-directed DNA polymerase (EC 2.7.7.7) - phage PO2 | 76 | 62 |
| 106 | 1 | 1511 | 264 | gi\|1750108 | YnbA [Bacillus subtilis] | 76 | 61 |
| 116 | 4 | 2480 | 2854 | gi\|755602 | unknown [Bacillus subtilis] | 76 | 60 |
| 116 | 6 | 3299 | 3625 | gi\|1146234 | dihydrodipicolinate reductase [Bacillus subtilis] | 76 | 56 |
| 122 | 5 | 3029 | 3619 | gi\|467436 | unknown [Bacillus subtilis] | 76 | 52 |
| 123 | 10 | 9109 | 10389 | gi\|1773196 | similar to B. stearothermophilus N-carbamyl-L-amino acid amidohydrolase [Escherichia coli] | 76 | 61 |
| 124 | 5 | 4087 | 3182 | gi\|974332 | NAD(P)H-dependent dihydroxyacetone-phosphate reductase [Bacillus ubtilis] | 76 | 58 |
| 130 | 5 | 3341 | 4294 | gi\|308853 | transmembrane protein [Lactococcus lactis] | 76 | 55 |
| 132 | 3 | 2265 | 5117 | gi\|1673889 | (AE000022) Mycoplasma pneumoniae, excinuclease ABC subunit A; similar to Swiss-Prot Accession Number P07671, from E. coli [Mycoplasma pneumoniae] | 76 | 59 |
| 138 | 34 | 25849 | 25409 | gi\|143795 | transfer RNA-Tyr synthetase [Bacillus subtilis] | 76 | 56 |
| 139 | 1 | 3 | 350 | gnl\|PID\|e191395 | mobilisation protein [Lactococcus lactis] | 76 | 65 |
| 141 | 1 | 2 | 544 | gi\|662792 | single-stranded DNA binding protein [unidentified eubacterium] | 76 | 64 |
| 155 | 9 | 7612 | 7058 | gnl\|PID\|e247026 | orf6 [Lactobacillus sake] | 76 | 57 |
| 164 | 4 | 1889 | 2416 | gi\|727436 | putative 20-kDa protein [Lactococcus lactis] | 76 | 55 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 181 | 5 | 3475 | 2288 | gi\|1147744 | PSR [Enterococcus hirae] | 76 | 53 |
| 181 | 8 | 6281 | 4986 | gi\|683583 | 5-enolpyruvylshikimate-3-phosphate synthase [Lactococcus lactis] pir\|S52580\|S52580 3-phosphoshikimate 1-carboxyvinyltransferase (EC .5.1.19) - Lactococcus lactis | 76 | 62 |
| 197 | 7 | 7662 | 8102 | gi\|1783253 | homologous to many ATP-binding transport proteins; hypothetical [Bacillus subtilis] | 76 | 58 |
| 222 | 16 | 10780 | 11298 | gi\|1591856 | hypothetical protein (SP:P15889) [Methanococcus jannaschii] | 76 | 64 |
| 229 | 1 | 1 | 138 | gi\|148316 | NaH-antiporter protein [Enterococcus hirae] | 76 | 47 |
| 233 | 6 | 3946 | 3341 | gi\|1591652 | hypothetical protein (SP:P31065) [Methanococcus jannaschii] | 76 | 60 |
| 238 | 2 | 844 | 1848 | gi\|622991 | mannitol transport protein [Bacillus stearothermophilus] sp\|P50852\|PTMB_BACST PTS SYSTEM, MANNITOL-SPECIFIC IIBC COMPONENT EIIBC-MTL) (MANNITOL- PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE NZYME II, BC COMPONENT) (EC 2.7.1.69) (EII-MTL). | 76 | 64 |
| 238 | 9 | 7235 | 7957 | gi\|1592142 | ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii] | 76 | 49 |
| 249 | 2 | 543 | 1235 | gi\|143156 | membrane bound protein [Bacillus subtilis] | 76 | 45 |
| 262 | 3 | 4131 | 2692 | gnl\|PID\|e281591 | catalase [Bacillus subtilis] | 76 | 65 |
| 265 | 1 | 2 | 400 | gi\|141858 | replication-associated protein [Plasmid pAD1] | 76 | 52 |
| 271 | 13 | 8175 | 10844 | gi\|397973 | Mg2+ transport ATPase [Salmonella typhimurium] | 76 | 57 |
| 323 | 4 | 4128 | 4568 | gnl\|PID\|e249023 | T19B10.3 [Caenorhabditis elegans] | 76 | 60 |
| 329 | 5 | 3270 | 2560 | gi\|310631 | ATP binding protein [Streptococcus gordonii] | 76 | 54 |
| 356 | 1 | 971 | 3 | gi\|971479 | orf3 gene product [Lactobacillus | 76 | 52 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | delbrueckii] | | |
| 371 | 1 | 1564 | 944 | gi\|1750125 | xylulose kinase [Bacillus subtilis] | 76 | 57 |
| 375 | 6 | 5137 | 4238 | gi\|1644202 | unknown [Bacillus subtilis] | 76 | 58 |
| 382 | 2 | 508 | 2769 | gi\|442360 | ClpC adenosine triphosphatase [Bacillus subtilis] | 76 | 60 |
| 399 | 11 | 7811 | 8845 | gi\|1572970 | acetate:SH-citrate lyase ligase (AMP) [Haemophilus influenzae] | 76 | 54 |
| 399 | 13 | 9126 | 10034 | gi\|1572968 | citrate lyase beta chain (acyl lyase subunit) (citE) [Haemophilus influenzae] | 76 | 57 |
| 485 | 1 | 3 | 1262 | gi\|564018 | dihydrofolate synthetase [Streptococcus pneumoniae] | 76 | 54 |
| 486 | 2 | 970 | 344 | gi\|1256617 | adenine phosphoribosyltransferase [Bacillus subtilis] | 76 | 61 |
| 536 | 1 | 220 | 2 | gi\|437389 | transposase [Lactococcus lactis] | 76 | 59 |
| 552 | 3 | 3969 | 2491 | gi\|882609 | 6-phospho-beta-glucosidase [Escherichia coli] | 76 | 63 |
| 634 | 2 | 697 | 918 | gi\|1022725 | unknown [Staphylococcus haemolyticus] | 76 | 52 |
| 684 | 3 | 1191 | 688 | gi\|1256653 | DNA-binding protein [Bacillus subtilis] | 76 | 65 |
| 752 | 1 | 1111 | 929 | gi\|407907 | ORF2 [Staphylococcus xylosus] | 76 | 46 |
| 822 | 1 | 548 | 237 | gi\|144313 | 6.0 kd ORF [Plasmid ColE1] | 76 | 73 |
| 923 | 1 | 2 | 421 | gi\|153843 | trypsin-resistant surface T6 protein (tee6) precursor [Streptococcus yogenes] | 76 | 57 |
| 953 | 2 | 534 | 187 | gi\|1592339 | hypothetical protein (PIR:S52522) [Methanococcus jannaschii] | 76 | 44 |
| 965 | 2 | 564 | 343 | gi\|1098898 | CTRP [Plasmodium falciparum] | 76 | 69 |
| 7 | 4 | 3754 | 4161 | gi\|495046 | tripeptidase [Lactococcus lactis] | 75 | 61 |
| 25 | 1 | 2 | 580 | gi\|1575577 | DNA-binding response regulator [Thermotoga maritima] | 75 | 57 |
| 45 | 7 | 3090 | 3350 | gi\|1673663 | (AE000003) Mycoplasma pneumoniae, E07_orf166 Protein [Mycoplasma pneumoniae] | 75 | 35 |
| 47 | 6 | 7526 | 6957 | gi\|1673843 | (AE000019) Mycoplasma pneumoniae, pilB | 75 | 58 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | homolog; similar to GenBank Accession Number E64124, from H. influenzae [Mycoplasma pneumoniae] | | |
| 51 | 1 | 15 | 1520 | sp\|P39168\|ATMA_ECOLI | MG(2+) TRANSPORT ATPASE, P-TYPE 1 (EC 3.6.1.-). | 75 | 58 |
| 54 | 11 | 3761 | 3579 | gi\|1504026 | similar to C.elegans protein (Z37093) [Homo sapiens] | 75 | 56 |
| 55 | 5 | 1648 | 2562 | gi\|1303901 | YqhT [Bacillus subtilis] | 75 | 58 |
| 56 | 8 | 5873 | 5358 | gi\|895749 | putative cellobiose phosphotransferase enzyme II'' [Bacillus ubtilis] | 75 | 49 |
| 58 | 2 | 2707 | 1916 | gi\|1658403 | formate dehydrogenase alpha subunit [Moorella thermoacetica] | 75 | 58 |
| 71 | 1 | 110 | 1429 | gi\|1304007 | LysA [Bacillus subtilis] | 75 | 58 |
| 74 | 5 | 3436 | 3074 | gi\|467433 | unknown [Bacillus subtilis] | 75 | 61 |
| 74 | 8 | 5491 | 4631 | gi\|467483 | unknown [Bacillus subtilis] | 75 | 60 |
| 77 | 1 | 3 | 992 | gi\|1653966 | 47 kD protein [Synechocystis sp.] | 75 | 34 |
| 81 | 1 | 26 | 862 | gi\|1064809 | homologous to sp:HTRA_ECOLI [Bacillus subtilis] | 75 | 55 |
| 89 | 11 | 11651 | 9801 | gi\|1573881 | hypothetical [Haemophilus influenzae] | 75 | 51 |
| 96 | 3 | 2521 | 1643 | gi\|1531619 | NodB [Rhizobium sp.] | 75 | 54 |
| 98 | 9 | 11494 | 10199 | gi\|1573043 | hypothetical [Haemophilus influenzae] | 75 | 53 |
| 110 | 12 | 11326 | 10283 | gi\|1184121 | auxin-induced protein [Vigna radiata] | 75 | 51 |
| 117 | 13 | 11200 | 9944 | gi\|457635 | vancomycin histidine protein kinase [Enterococcus faecium] gi\|801884 vanS [Transposon Tn1546] | 75 | 51 |
| 122 | 6 | 3812 | 5206 | gi\|467439 | temperature sensitive cell division [Bacillus subtilis] | 75 | 59 |
| 128 | 12 | 8262 | 7921 | gi\|466473 | cellobiose phosphotransferase enzyme II' [Bacillus tearothermophilus] | 75 | 48 |
| 128 | 38 | 31848 | 30733 | gi\|216300 | peptidoglycan synthesis enzyme [Bacillus subtilis] sp\|P37585\|MURG_BACSU MURG PROTEIN UPD-N-ACETYLGLUCOSAMINE--N- | 75 | 56 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | ACETYLMURAMYL-PENTAPEPTIDE)PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE RANSFERASE). | | |
| 129 | 2 | 1916 | 2134 | gnl\|PID\|e267624 | Unknown, highly similar to Pseudomonas putida 4-oxalocrotonate tautomerase [Bacillus subtilis] | 75 | 47 |
| 130 | 4 | 2375 | 3343 | gi\|495179 | transmembrane protein [Lactococcus lactis] | 75 | 55 |
| 133 | 1 | 3 | 1514 | gnl\|PID\|e254877 | unknown [Mycobacterium tuberculosis] | 75 | 54 |
| 158 | 13 | 12326 | 11634 | gi\|809660 | deoxyribose-phosphate aldolase [Bacillus subtilis] pir\|S49455\|S49455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - acillus subtilis | 75 | 66 |
| 162 | 13 | 14285 | 12543 | gi\|1653222 | cation-transporting ATPase PacL [Synechocystis sp.] | 75 | 60 |
| 170 | 2 | 1280 | 921 | sp\|P07999\|DHGB_BACME | GLUCOSE 1-DEHYDROGENASE B (EC 1.1.1.47). | 75 | 62 |
| 171 | 7 | 7618 | 8523 | gi\|1303856 | YqgI [Bacillus subtilis] | 75 | 52 |
| 179 | 14 | 14668 | 15255 | gi\|457177 | alkyl hydroperoxide reductase [Salmonella typhimurium] sp\|P19479\|AHPC_SALTY ALKYL HYDROPEROXIDE REDUCTASE C22 PROTEIN (EC .6.4.-). {SUB 2-187} | 75 | 55 |
| 181 | 6 | 4470 | 3604 | gi\|683585 | prephenate dehydratase [Lactococcus lactis] | 75 | 49 |
| 191 | 1 | 183 | 560 | gnl\|PID\|e261991 | putative orf [Bacillus subtilis] | 75 | 57 |
| 197 | 3 | 2117 | 3592 | gi\|1783250 | homologous to cytochrome d ubiquinol oxidase subunit I; hypothetical [Bacillus subtilis] | 75 | 60 |
| 215 | 3 | 2545 | 2201 | gnl\|PID\|e284996 | ORF136 [Staphylococcus aureus] | 75 | 54 |
| 216 | 1 | 2 | 256 | gi\|153570 | H+ ATPase [Enterococcus faecalis] | 75 | 53 |
| 223 | 4 | 2406 | 4193 | gi\|862312 | lytS gene product [Staphylococcus aureus] | 75 | 56 |
| 227 | 5 | 3004 | 3567 | gi\|144729 | butanol dehydrogenase [Clostridium acetobutylicum] sp\|Q04944\|ADHA_CLOAB NADH- | 75 | 53 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | DEPENDENT BUTANOL DEHYDROGENASE A (EC .1.1.-) (BDH I). | | |
| 228 | 9 | 6032 | 5700 | gi\|467410 | unknown [Bacillus subtilis] | 75 | 59 |
| 229 | 16 | 17081 | 16848 | gi\|207398 | tropomyosin T class IVd alpha-3 [Rattus norvegicus] | 75 | 42 |
| 238 | 8 | 6038 | 7237 | gi\|141927 | czcB gene product [Alcaligenes eutrophus] | 75 | 39 |
| 244 | 10 | 7795 | 7460 | gi\|467419 | unknown [Bacillus subtilis] | 75 | 56 |
| 247 | 1 | 7 | 1431 | gi\|577569 | PepV [Lactobacillus delbrueckii] | 75 | 54 |
| 250 | 5 | 3416 | 3201 | gi\|1580783 | sperm receptor [Strongylocentrotus purpuratus] | 75 | 50 |
| 256 | 1 | 2 | 562 | gi\|709991 | hypothetical protein [Bacillus subtilis] | 75 | 56 |
| 262 | 2 | 1031 | 2479 | gi\|142783 | DNA photolyase [Bacillus firmus] | 75 | 59 |
| 263 | 1 | 222 | 890 | gi\|148304 | beta-1,4-N-acetylmuramoylhydrolase [Enterococcus hirae] pir\|A42296\|A42296 lysozyme 2 (EC 3.2.1.-) precursor - Enterococcus irae (ATCC 9790) | 75 | 60 |
| 266 | 5 | 2224 | 1982 | gnl\|PID\|e253211 | ORF YDL065c [Saccharomyces cerevisiae] | 75 | 50 |
| 269 | 2 | 1477 | 707 | gi\|1736647 | ORF_ID:o347#4; similar to [SwissProt Accession Number P44634] [Escherichia coli] | 75 | 61 |
| 276 | 11 | 7415 | 4593 | gnl\|PID\|e221269 | tail protein [Bacteriophage CP-1] | 75 | 54 |
| 279 | 17 | 14992 | 14651 | gi\|1389549 | ORF3 [Bacillus subtilis] | 75 | 61 |
| 292 | 11 | 7829 | 8470 | gi\|160693 | sporozoite surface protein [Plasmodium yoelii] | 75 | 50 |
| 295 | 2 | 489 | 1157 | gi\|533099 | endonuclease III [Bacillus subtilis] | 75 | 59 |
| 307 | 4 | 3804 | 4889 | gi\|1321625 | exo-alpha-1, 4-glucosidase [Bacillus stearothermophilus] | 75 | 60 |
| 322 | 4 | 1088 | 1996 | gi\|310303 | mosA [Rhizobium meliloti] | 75 | 63 |
| 331 | 1 | 1 | 294 | gi\|1016092 | ribosomal protein S14 [Cyanophora paradoxa] | 75 | 57 |
| 334 | 7 | 6860 | 7969 | gi\|409286 | bmrU [Bacillus subtilis] | 75 | 45 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 340 | 1 | 3 | 743 | gi\|288413 | glutamate dehydrogenase (NADP+) [Corynebacterium glutamicum] pir\|S32227\|S32227 glutamate dehydrogenase (NADP+) (EC 1.4.1.4) - orynebacterium glutamicum | 75 | 60 |
| 343 | 2 | 1497 | 778 | gi\|46602 | putative transposase (AA 1 - 224) [Staphylococcus aureus] ir\|S12093\|S12093 probable IS431mec protein - Staphylococcus aureus p\|P19380\|TRA2_STAAU TRANSPOSASE FOR INSERTION SEQUENCE-LIKE ELEMENT 431MEC. | 75 | 54 |
| 372 | 3 | 865 | 1629 | gi\|146282 | gut operon repressor (gutR) [Escherichia coli] | 75 | 58 |
| 372 | 7 | 6614 | 5307 | gnl\|PID\|e255128 | trigger factor [Bacillus subtilis] | 75 | 62 |
| 387 | 3 | 1721 | 1353 | gi\|580902 | ORF6 gene product [Bacillus subtilis] | 75 | 53 |
| 399 | 30 | 28774 | 29805 | gi\|146278 | glucitol-specific enzyme II (gutA) [Escherichia coli] pir\|A26725\|WQEC2S phosphotransferase system enzyme II (EC .7.1.69), sorbitol-specific, factor II - Escherichia coli sp\|P05705\|PTHB_ECOLI PTS SYSTEM, GLUCITOL/SORBITOL-SPECIFIC IIBC OMPONENT (EIIBC-GUT) | 75 | 61 |
| 399 | 33 | 31077 | 32768 | gi\|517205 | 67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes] | 75 | 59 |
| 404 | 6 | 4994 | 4332 | gi\|1303921 | YqiF [Bacillus subtilis] | 75 | 64 |
| 404 | 7 | 4984 | 4829 | gi\|1303921 | YqiF [Bacillus subtilis] | 75 | 60 |
| 419 | 1 | 320 | 3 | gi\|496283 | lysin [Bacteriophage Tuc2009] | 75 | 67 |
| 431 | 3 | 1139 | 759 | sp\|P46351\|YZGD_BACSU | HYPOTHETICAL 45.4 KD PROTEIN IN THIAMINASE I 5'REGION. | 75 | 60 |
| 473 | 1 | 166 | 2 | gnl\|PID\|e229299 | R04D3.8 [Caenorhabditis elegans] | 75 | 35 |
| 481 | 1 | 1 | 351 | gi\|1573766 | phosphoglyceromutase (gpmA) [Haemophilus influenzae] | 75 | 64 |
| 492 | 1 | 440 | 3 | gi\|806487 | ORF211; putative [Lactococcus lactis] | 75 | 57 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 595 | 1 | 705 | 181 | gi\|147485 | queA [Escherichia coli] | 75 | 51 |
| 619 | 2 | 879 | 319 | gi\|1063246 | low homology to P14 protein of Heamophilus influenzar and 14.2 kDa protein of Escherichia coli [Bacillus subtilis] | 75 | 59 |
| 663 | 1 | 15 | 1544 | gi\|475112 | enzyme IIabc [Pediococcus pentosaceus] | 75 | 54 |
| 701 | 4 | 662 | 946 | gi\|143793 | tyrosyl-tRNA synthetase [Bacillus caldotenax] | 75 | 60 |
| 719 | 1 | 970 | 419 | gi\|727436 | putative 20-kDa protein [Lactococcus lactis] | 75 | 56 |
| 886 | 1 | 101 | 409 | gi\|143150 | levR [Bacillus subtilis] | 75 | 59 |
| 939 | 1 | 403 | 191 | gi\|425467 | transposase [Lactobacillus helveticus] | 75 | 53 |
| 984 | 2 | 66 | 227 | gi\|1652190 | Fat protein [Synechocystis sp.] | 75 | 48 |
| 17 | 2 | 2592 | 2924 | gi\|532556 | ORF23 [Enterococcus faecalis] | 74 | 53 |
| 17 | 25 | 24449 | 25639 | gi\|1458228 | mutY homolog [Homo sapiens] | 74 | 50 |
| 21 | 7 | 4729 | 5229 | gi\|726320 | putative protein of unknown function encoded by the IS200-like lement [Yersinia pestis] | 74 | 57 |
| 32 | 9 | 5819 | 4488 | gi\|1498962 | M. jannaschii predicted coding region MJ0188 [Methanococcus jannaschii] | 74 | 41 |
| 38 | 1 | 707 | 3 | gi\|142152 | sulfate permease (gtg start codon) [Synechococcus PCC6301] pir\|A30301\|GRYCS7 sulfate transport protein - Synechococcus sp. PCC 7942) | 74 | 53 |
| 44 | 1 | 1 | 927 | gi\|1377823 | aminopeptidase [Bacillus subtilis] | 74 | 63 |
| 60 | 8 | 8747 | 8070 | gi\|143368 | phosphoribosylformyl glycinamidine synthetase I (PUR-L; gtg start odon) [Bacillus subtilis] | 74 | 63 |
| 72 | 8 | 7388 | 7119 | gnl\|PID\|e209004 | glutaredoxin-like protein [Lactococcus lactis] | 74 | 53 |
| 91 | 4 | 1031 | 2257 | gi\|726480 | L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus ubtilis] | 74 | 58 |
| 105 | 7 | 5553 | 5855 | gi\|467418 | unknown [Bacillus subtilis] | 74 | 63 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 110 | 18 | 16903 | 15842 | gi\|45288 | arcB (AA 1-336) [Pseudomonas aeruginosa] | 74 | 57 |
| 112 | 3 | 1112 | 636 | gi\|887824 | ORF_o310 [Escherichia coli] | 74 | 53 |
| 123 | 8 | 6105 | 7619 | gi\|1773191 | similar to Pseudomonas sp. ORF5 [Escherichia coli] | 74 | 60 |
| 128 | 1 | 2 | 1315 | gi\|143961 | pyruvate phosphate dikinase [Clostridium symbiosum] pir\|A36231\|KIQAPO pyruvate,orthophosphate dikinase (EC 2.7.9.1) - lostridium symbiosum | 74 | 58 |
| 128 | 26 | 18866 | 20401 | gi\|1303961 | YqjJ [Bacillus subtilis] | 74 | 57 |
| 150 | 5 | 4653 | 5303 | gi\|495046 | tripeptidase [Lactococcus lactis] | 74 | 53 |
| 159 | 8 | 7500 | 6850 | gi\|581098 | GlnQ (AA 1-240); gtg start [Escherichia coli] | 74 | 53 |
| 179 | 1 | 1259 | 57 | gi\|537080 | ribonucleoside triphosphate reductase [Escherichia coli] pir\|A47331\|A47331 oxygen-sensitive ribonucleoside-triphosphate eductase (EC 1.17.4.-) - Escherichia coli | 74 | 62 |
| 183 | 2 | 1669 | 224 | gi\|1146200 | DNA or RNA helicase, DNA-dependent ATPase [Bacillus subtilis] | 74 | 53 |
| 213 | 4 | 2265 | 3200 | gi\|1373157 | orf-X; hypothetical protein; Method: conceptual translation supplied by author [Bacillus subtilis] | 74 | 63 |
| 229 | 13 | 13774 | 12806 | gnl\|PID\|e290288 | Met-tRNAi formyl transferase [Bacillus subtilis] | 74 | 55 |
| 238 | 31 | 28648 | 28052 | gi\|451072 | di-tripeptide transporter [Lactococcus lactis] | 74 | 56 |
| 244 | 8 | 6409 | 5552 | gi\|467422 | unknown [Bacillus subtilis] | 74 | 60 |
| 249 | 1 | 7 | 411 | gi\|1591758 | diaminopimelate epimerase [Methanococcus jannaschii] | 74 | 51 |
| 270 | 3 | 1832 | 3955 | gi\|1303829 | YqfK [Bacillus subtilis] | 74 | 55 |
| 276 | 3 | 1668 | 1357 | gi\|496282 | holin [Bacteriophage Tuc2009] | 74 | 54 |
| 288 | 9 | 5807 | 5076 | gi\|530063 | glycerol uptake facilitator [Streptococcus | 74 | 60 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | pneumoniae] sp\|P52281\|GLPF_STRPN GLYCEROL UPTAKE FACILITATOR PROTEIN. | | |
| 292 | 21 | 16780 | 17547 | gi\|1573646 | Mg(2+) transport ATPase protein C (mgtC) (SP:P22037) [Haemophilus influenzae] | 74 | 42 |
| 297 | 1 | 682 | 11 | gnl\|PID\|e255093 | hypothetical protein [Bacillus subtilis] | 74 | 54 |
| 298 | 3 | 3562 | 3095 | gi\|1303970 | YqjS [Bacillus subtilis] | 74 | 46 |
| 321 | 10 | 5081 | 6028 | pir\|A32950\|A32950 | probable reductase protein - Leishmania major | 74 | 56 |
| 327 | 2 | 904 | 3285 | gi\|1573876 | virulence associated protein homolog (vacB) [Haemophilus influenzae] | 74 | 53 |
| 334 | 5 | 3942 | 5432 | gi\|1652678 | amidase [Synechocystis sp.] | 74 | 57 |
| 341 | 13 | 13007 | 12069 | gi\|39881 | ORF 311 (AA 1-311) [Bacillus subtilis] | 74 | 53 |
| 362 | 7 | 3529 | 5274 | gnl\|PID\|e255093 | hypothetical protein [Bacillus subtilis] | 74 | 58 |
| 376 | 3 | 1282 | 2346 | gi\|1773090 | transfer RNA-guanine transglycosylase [Escherichia coli] | 74 | 59 |
| 421 | 2 | 48 | 1400 | gi\|710632 | beta-glucosidase [Bacillus subtilis] | 74 | 58 |
| 471 | 1 | 815 | 3 | gi\|854234 | cymG gene product [Klebsiella oxytoca] | 74 | 53 |
| 480 | 2 | 263 | 607 | gi\|1303994 | YqkM [Bacillus subtilis] | 74 | 48 |
| 518 | 7 | 4409 | 5002 | gi\|145821 | EBG enzyme alpha subunit [Escherichia coli] | 74 | 47 |
| 539 | 8 | 6607 | 7179 | gi\|1165295 | D3703.8p [Saccharomyces cerevisiae] | 74 | 57 |
| 542 | 1 | 750 | 4 | gi\|1064810 | function unknown [Bacillus subtilis] | 74 | 56 |
| 559 | 1 | 1204 | 5 | gi\|43821 | nifJ protein (AA 1-1171) [Klebsiella pneumoniae] p\|P03833\|NIFJ_KLEPN PYRUVATE-FLAVODOXIN OXIDOREDUCTASE (EC -.-.-). | 74 | 58 |
| 579 | 3 | 1373 | 1624 | gi\|1237013 | ORF2 [Bacillus subtilis] | 74 | 46 |
| 624 | 4 | 2518 | 3669 | gi\|467394 | recombination protein [Bacillus subtilis] | 74 | 56 |
| 688 | 1 | 623 | 3 | gi\|662880 | novel hemolytic factor [Bacillus cereus] | 74 | 48 |
| 763 | 1 | 106 | 441 | gi\|153955 | envM protein [Salmonella typhimurium] | 74 | 46 |
| 811 | 1 | 3 | 158 | gi\|309662 | pheromone binding protein [Plasmid pCF10] | 74 | 57 |
| 852 | 1 | 2 | 601 | gi\|309662 | pheromone binding protein [Plasmid pCF10] | 74 | 53 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 935 | 1 | 976 | 2 | gi\|467403 | seryl-tRNA synthetase [Bacillus subtilis] | 74 | 59 |
| 22 | 2 | 2178 | 2471 | gi\|467460 | unknown [Bacillus subtilis] | 73 | 61 |
| 24 | 2 | 1126 | 3150 | gi\|1303822 | YqfF [Bacillus subtilis] | 73 | 54 |
| 33 | 6 | 6638 | 6970 | gi\|536971 | ORF_o76 [Escherichia coli] | 73 | 56 |
| 48 | 1 | 621 | 1241 | gnl\|PID\|e274111 | aggregation promoting protein [Lactobacillus gasseri] | 73 | 67 |
| 48 | 6 | 5327 | 7225 | gi\|1185289 | 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase [Bacillus subtilis] | 73 | 56 |
| 50 | 2 | 1097 | 2008 | gi\|1498295 | homoserine kinase homolog [Streptococcus pneumoniae] | 73 | 55 |
| 52 | 4 | 2793 | 4334 | gi\|473902 | alpha-acetolactate synthase [Lactococcus lactis] | 73 | 59 |
| 55 | 1 | 1 | 261 | gi\|396365 | alternate name yjbA [Escherichia coli] | 73 | 36 |
| 60 | 6 | 5935 | 5549 | gi\|551881 | amidophosphoribosyltransferase [Lactobacillus casei] pir\|PC1136\|PC1136 purF protein - Lactobacillus casei (fragment) sp\|P35853\|PUR1_LACCA AMIDOPHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.14) GLUTAMINE PHOSPHORIBOSYLPYROPHOSPHATE AMIDOTRANSFERASE) (ATASE) FRAGMENT | 73 | 57 |
| 74 | 2 | 477 | 1355 | gnl\|PID\|e233567 | unknown [Mycobacterium tuberculosis] | 73 | 54 |
| 81 | 19 | 14213 | 13845 | gi\|606073 | ORF_o169 [Escherichia coli] | 73 | 52 |
| 93 | 7 | 2861 | 4075 | gi\|405134 | acetate kinase [Bacillus subtilis] | 73 | 56 |
| 100 | 1 | 1057 | 2 | gi\|1353561 | ORF44 [Bacteriophage r1t] | 73 | 52 |
| 100 | 41 | 28872 | 28627 | gi\|188492 | heat shock-induced protein [Homo sapiens] | 73 | 42 |
| 104 | 4 | 5558 | 5274 | gi\|312440 | aspartate carbamoyltransferase [Bacillus caldolyticus] pir\|S34318\|S34318 aspartate carbamoyltransferase (EC 2.1.3.2) - acillus caldolyticus | 73 | 55 |
| 119 | 5 | 3264 | 3638 | gi\|473707 | positive regulator for virulence factors [Clostridium perfringens] | 73 | 39 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 123 | 17 | 16156 | 15665 | gi\|1303703 | YrkD [Bacillus subtilis] | 73 | 37 |
| 123 | 18 | 16133 | 16465 | gi\|1303893 | YqhL [Bacillus subtilis] | 73 | 43 |
| 124 | 3 | 2165 | 1722 | gi\|486661 | TMnm related protein [Saccharomyces cerevisiae] | 73 | 45 |
| 127 | 6 | 5778 | 5101 | gi\|290561 | o188 [Escherichia coli] | 73 | 48 |
| 128 | 10 | 6896 | 7201 | pir\|S37387\|S37387 | internalin A precursor - Listeria monocytogenes | 73 | 53 |
| 137 | 2 | 980 | 1954 | gi\|1276882 | EpsI [Streptococcus thermophilus] | 73 | 56 |
| 141 | 3 | 942 | 2777 | gi\|467336 | unknown [Bacillus subtilis] | 73 | 49 |
| 146 | 7 | 5611 | 4739 | gi\|149395 | lacC [Lactococcus lactis] | 73 | 56 |
| 154 | 6 | 3566 | 4621 | gi\|1354775 | pfoS/R [Treponema pallidum] | 73 | 46 |
| 155 | 8 | 7136 | 6726 | gnl\|PID\|e247026 | orf6 [Lactobacillus sake] | 73 | 61 |
| 158 | 8 | 8693 | 7119 | gi\|1674275 | (AE000056) Mycoplasma pneumoniae, hypothetical ABC transporter (yjcW) homolog; similar to Swiss-Prot Accession Number P32721, from E. coli [Mycoplasma pneumoniae] | 73 | 45 |
| 162 | 4 | 4039 | 3305 | gi\|142997 | glycerol uptake facilitator [Bacillus subtilis] | 73 | 55 |
| 165 | 4 | 3962 | 3105 | gi\|882736 | ORF_f278 [Escherichia coli] | 73 | 58 |
| 171 | 3 | 3952 | 4689 | gnl\|PID\|e63527 | FtsE [Mycobacterium tuberculosis] | 73 | 56 |
| 171 | 5 | 5673 | 6596 | gi\|1303854 | YqgG [Bacillus subtilis] | 73 | 59 |
| 179 | 9 | 9302 | 10414 | gnl\|PID\|e254984 | hypothetical protein [Bacillus subtilis] | 73 | 55 |
| 180 | 1 | 24 | 1151 | gi\|43985 | nifS-like gene [Lactobacillus delbrueckii] | 73 | 56 |
| 181 | 12 | 10036 | 9674 | gnl\|PID\|e220317 | chorismate mutase [Staphylococcus xylosus] | 73 | 50 |
| 181 | 13 | 10713 | 10003 | gi\|39813 | phospho-2-dehydro-3-deoxyheptonate aldolase [Bacillus subtilis] ir\|S21418\|S21418 phospho-2-dehydro-3-deoxyheptonate aldolase (EC 1.2.15) - Bacillus subtilis | 73 | 56 |
| 183 | 3 | 2716 | 1667 | gi\|1146199 | putative [Bacillus subtilis] | 73 | 36 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 198 | 1 | 869 | 108 | gi\|142854 | homologous to E. coli radC gene product and to unidentified protein rom Staphylococcus aureus [Bacillus subtilis] | 73 | 47 |
| 210 | 1 | 956 | 3 | gnl\|PID\|e281310 | acetyl coenzyme A acetyltransferase (thiolase) [Thermoanaerobacterium thermosaccharolyticum] | 73 | 54 |
| 230 | 1 | 1 | 171 | gi\|304143 | S-layer protein [Bacillus circulans] | 73 | 46 |
| 235 | 1 | 715 | 2 | gi\|1732315 | transport system permease homolog [Listeria monocytogenes] | 73 | 49 |
| 235 | 2 | 888 | 676 | gi\|551726 | sporulation protein [Bacillus subtilis] | 73 | 54 |
| 242 | 4 | 3290 | 3517 | gnl\|PID\|e236570 | orf6 gene product [Enterococcus faecalis] | 73 | 30 |
| 242 | 8 | 5914 | 6492 | gi\|1742340 | HipB protein. [Escherichia coli] | 73 | 49 |
| 250 | 3 | 3037 | 2411 | gi\|1174238 | TipB [Pseudomonas fluorescens] | 73 | 57 |
| 254 | 5 | 1124 | 792 | gi\|580900 | ORF3 gene product [Bacillus subtilis] | 73 | 52 |
| 269 | 9 | 5507 | 5154 | gi\|1303790 | YqeI [Bacillus subtilis] | 73 | 60 |
| 269 | 12 | 7989 | 7345 | gi\|285621 | undefined open reading frame [Bacillus stearothermophilus] | 73 | 54 |
| 284 | 1 | 1 | 915 | gi\|455528 | ORF2 [Streptococcus thermophilus bacteriophage] | 73 | 54 |
| 290 | 3 | 1932 | 2678 | gnl\|PID\|e248883 | unknown [Mycobacterium tuberculosis] | 73 | 57 |
| 295 | 8 | 4521 | 4739 | gi\|145478 | putative [Escherichia coli] | 73 | 56 |
| 296 | 1 | 2 | 1846 | gnl\|PID\|e249642 | transketolase [Bacillus subtilis] | 73 | 59 |
| 310 | 4 | 3488 | 3036 | gi\|1591900 | nucleoside diphosphate kinase [Methanococcus jannaschii] | 73 | 48 |
| 313 | 1 | 17 | 778 | gi\|1658371 | cyclic beta-1,2-glucan modification protein [Rhizobium meliloti] | 73 | 60 |
| 314 | 3 | 2642 | 2067 | gi\|1330343 | C34D4.12 gene product [Caenorhabditis elegans] | 73 | 56 |
| 325 | 1 | 492 | 4 | gi\|407908 | EIIscr [Staphylococcus xylosus] | 73 | 56 |
| 345 | 19 | 20549 | 21901 | gi\|443691 | glutathione reductase [Streptococcus thermophilus] | 73 | 59 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 359 | 4 | 3280 | 2252 | gi\|1001478 | hypothetical protein [Synechocystis sp.] | 73 | 50 |
| 374 | 1 | 884 | 3 | gi\|435123 | PacL [Synechococcus sp.] | 73 | 58 |
| 379 | 6 | 5676 | 4339 | gi\|887822 | possible frameshift at end to join to next ORF? [Escherichia coli] | 73 | 57 |
| 383 | 4 | 3815 | 3387 | gi\|1651732 | mutator MutT protein [Synechocystis sp.] | 73 | 52 |
| 392 | 4 | 3454 | 5202 | gi\|294587 | minimal change nephritis transmembrane glycoprotein [Rattus orvegicus] | 73 | 56 |
| 394 | 5 | 4267 | 5250 | gi\|49011 | amidinotransferase II [Streptomyces griseus] | 73 | 42 |
| 395 | 10 | 4252 | 4608 | gi\|1591139 | M. jannaschii predicted coding region MJ0435 [Methanococcus jannaschii] | 73 | 48 |
| 397 | 1 | 885 | 4 | gnl\|PID\|e249658 | GrlA [Bacillus subtilis] | 73 | 56 |
| 399 | 15 | 10007 | 11569 | gi\|565619 | citrate lyase alpha-subunit [Klebsiella pneumoniae] pir\|S60776\|S60776 citrate (pro-3S)-lyase (EC 4.1.3.6) alpha chain - lebsiella pneumoniae | 73 | 54 |
| 416 | 2 | 660 | 1649 | gi\|475114 | regulatory protein [Pediococcus pentosaceus] | 73 | 50 |
| 436 | 6 | 4124 | 3540 | gi\|727436 | putative 20-kDa protein [Lactococcus lactis] | 73 | 53 |
| 446 | 3 | 1618 | 4260 | gi\|882711 | exonuclease V alpha-subunit [Escherichia coli] | 73 | 48 |
| 462 | 1 | 819 | 43 | gi\|1399011 | immunogenic secreted protein precursor [Streptococcus pyogenes] | 73 | 63 |
| 482 | 5 | 3181 | 2501 | gi\|1072419 | glcB gene product [Staphylococcus carnosus] | 73 | 55 |
| 495 | 4 | 1340 | 3031 | gi\|146547 | kdpA [Escherichia coli] | 73 | 55 |
| 523 | 4 | 2354 | 1821 | pir\|A00392\|RDSODF | dihydrofolate reductase (EC 1.5.1.3) - Enterococcus faecium | 73 | 54 |
| 543 | 5 | 3099 | 2893 | gi\|19743 | nsGRP-2 [Nicotiana sylvestris] | 73 | 53 |
| 567 | 1 | 9 | 740 | gi\|1147601 | cyclophilin isoform 4 [Caenorhabditis elegans] | 73 | 54 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 629 | 1 | 945 | 4 | gi\|1006620 | ABC transporter [Synechocystis sp.] | 73 | 46 |
| 714 | 2 | 344 | 556 | gi\|1045872 | ATP-binding protein [Mycoplasma genitalium] | 73 | 61 |
| 747 | 1 | 320 | 3 | gi\|437389 | transposase [Lactococcus lactis] | 73 | 56 |
| 764 | 1 | 3 | 515 | gi\|532554 | ORF21 [Enterococcus faecalis] | 73 | 50 |
| 766 | 1 | 683 | 3 | gi\|1673788 | (AE000015) Mycoplasma pneumoniae, fructose-bisphosphate aldolase; similar to Swiss-Prot Accession Number P13243, from B. subtilis [Mycoplasma pneumoniae] | 73 | 52 |
| 880 | 1 | 198 | 4 | gi\|309661 | regulatory protein [Plasmid pCF10] | 73 | 50 |
| 897 | 1 | 3 | 170 | gi\|807976 | unknown [Saccharomyces cerevisiae] | 73 | 57 |
| 5 | 1 | 223 | 2 | gnl\|PID\|e255315 | unknown [Mycobacterium tuberculosis] | 72 | 56 |
| 8 | 5 | 4158 | 4799 | gi\|587088 | shikimate kinase [Bacillus subtilis] | 72 | 54 |
| 19 | 6 | 2600 | 2833 | gi\|34844 | embryonic myosin heavy chain (AA 1 - 1940) [Homo sapiens] ir\|S04090\|S04090 myosin heavy chain, skeletal muscle, embryonic - man | 72 | 38 |
| 19 | 25 | 12872 | 14605 | gnl\|PID\|e242896 | orf5 [Bacteriophage A2] | 72 | 52 |
| 21 | 4 | 2777 | 2598 | gi\|54115 | skeletal muscle chloride channel [Mus musculus domesticus] | 72 | 45 |
| 23 | 7 | 3702 | 4847 | gi\|144714 | NADPH-dependent butanol dehydrogenase [Clostridium acetobutylicum] pir\|JU0053\|JU0053 NADPH-dependent butanol dehydrogenase - lostridium acetobutylicum | 72 | 48 |
| 32 | 1 | 1073 | 3 | gi\|1303839 | YqfR [Bacillus subtilis] | 72 | 50 |
| 39 | 8 | 4137 | 3244 | pir\|A32950\|A32950 | probable reductase protein - Leishmania major | 72 | 55 |
| 43 | 3 | 969 | 1919 | gi\|290494 | o287 [Escherichia coli] | 72 | 46 |
| 45 | 2 | 911 | 1567 | gi\|1039479 | ORFU [Lactococcus lactis] | 72 | 50 |
| 55 | 6 | 2549 | 2896 | gi\|755602 | unknown [Bacillus subtilis] | 72 | 51 |
| 55 | 7 | 3178 | 3660 | gi\|1303914 | YqhY [Bacillus subtilis] | 72 | 49 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 60 | 1 | 1302 | 34 | gi\|143374 | phosphoribosyl glycinamide synthetase (PUR-D; gtg start codon) Bacillus subtilis] | 72 | 59 |
| 60 | 3 | 3422 | 2838 | gi\|143372 | phosphoribosyl glycinamide formyltransferase (PUR-N) [Bacillus ubtilis] | 72 | 48 |
| 60 | 10 | 9771 | 9010 | gi\|143367 | phosphoribosyl aminoidazole succinocarboxamide synthetase (PUR-C; tg start codon) [Bacillus subtilis] | 72 | 57 |
| 70 | 5 | 3615 | 3833 | sp\|P43672\|YCBH_ECOLI | HYPOTHETICAL 14.4 KD PROTEIN IN PYRD-PQIA INTERGENIC REGION. | 72 | 48 |
| 79 | 2 | 632 | 841 | gi\|1652343 | ABC transporter [Synechocystis sp.] | 72 | 47 |
| 85 | 2 | 1843 | 770 | gi\|1354775 | pfoS/R [Treponema pallidum] | 72 | 45 |
| 87 | 1 | 2 | 745 | gi\|42029 | ORF1 gene product [Escherichia coli] | 72 | 47 |
| 88 | 1 | 124 | 1047 | gi\|535348 | CodV [Bacillus subtilis] | 72 | 50 |
| 88 | 7 | 3862 | 4752 | gi\|149413 | ORF [Lactococcus lactis] | 72 | 51 |
| 91 | 2 | 611 | 877 | gi\|726480 | L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus ubtilis] | 72 | 57 |
| 98 | 16 | 16302 | 15163 | gi\|147326 | transport protein [Escherichia coli] | 72 | 57 |
| 101 | 6 | 4676 | 4023 | gi\|1109685 | ProW [Bacillus subtilis] | 72 | 53 |
| 104 | 3 | 5331 | 3982 | gi\|312441 | dihydroorotase [Bacillus caldolyticus] | 72 | 58 |
| 114 | 10 | 11165 | 12205 | gi\|556881 | Similar to Saccharomyces cerevisiae SUA5 protein [Bacillus subtilis] pir\|S49358\|S49358 ipc-29d protein - Bacillus subtilis sp\|P39153\|YWLC_BACSU HYPOTHETICAL 37.0 KD PROTEIN IN SPOIIR-GLYC NTERGENIC REGION. | 72 | 60 |
| 128 | 19 | 14325 | 11560 | gi\|143150 | levR [Bacillus subtilis] | 72 | 58 |
| 130 | 2 | 382 | 1437 | gi\|308850 | ATP binding protein [Lactococcus lactis] | 72 | 55 |
| 135 | 4 | 5012 | 3693 | gi\|413940 | ipa-16d gene product [Bacillus subtilis] | 72 | 56 |
| 150 | 6 | 5114 | 5878 | gi\|495046 | tripeptidase [Lactococcus lactis] | 72 | 54 |
| 154 | 9 | 5850 | 5677 | gi\|425467 | transposase [Lactobacillus helveticus] | 72 | 52 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 168 | 4 | 1375 | 1563 | gi\|1652869 | NADH dehydrogenase [Synechocystis sp.] | 72 | 55 |
| 173 | 5 | 2879 | 4024 | gnl\|PID\|e254877 | unknown [Mycobacterium tuberculosis] | 72 | 57 |
| 179 | 2 | 1608 | 2399 | gi\|709993 | hypothetical protein [Bacillus subtilis] | 72 | 45 |
| 179 | 6 | 7584 | 7844 | gi\|1161934 | DltC [Lactobacillus casei] | 72 | 54 |
| 180 | 21 | 19948 | 21105 | gi\|1773197 | similar to M. fervidus malate dehydrogenase [Escherichia coli] | 72 | 55 |
| 182 | 1 | 3 | 413 | gi\|1146182 | putative [Bacillus subtilis] | 72 | 48 |
| 200 | 23 | 13106 | 12789 | gi\|1707358 | polyprotein precurser [Soybean mosaic virus] | 72 | 34 |
| 204 | 6 | 2462 | 2289 | gi\|1200525 | dihydrolipoamide acetyltransferase [Pseudomonas aeruginosa] | 72 | 61 |
| 204 | 9 | 6374 | 5187 | gi\|1732040 | alcohol dehydrogenase [Actinobacillus pleuropneumoniae] | 72 | 56 |
| 205 | 1 | 463 | 71 | gi\|42029 | ORF1 gene product [Escherichia coli] | 72 | 57 |
| 210 | 7 | 6433 | 5279 | gi\|142978 | glycerol dehydrogenase [Bacillus stearothermophilus] pir\|JQ1474\|JQ1474 glycerol dehydrogenase (EC 1.1.1.6) - Bacillus tearothermophilus | 72 | 46 |
| 213 | 6 | 4086 | 5141 | gi\|431231 | uracil permease [Bacillus caldolyticus] | 72 | 51 |
| 223 | 1 | 99 | 833 | gi\|1573615 | ATP-binding protein (abc) [Haemophilus influenzae] | 72 | 47 |
| 227 | 1 | 26 | 886 | gi\|1070015 | protein-dependent [Bacillus subtilis] | 72 | 52 |
| 228 | 4 | 2047 | 2481 | gi\|467339 | unknown [Bacillus subtilis] | 72 | 50 |
| 238 | 17 | 14728 | 15582 | gi\|882736 | ORF_f278 [Escherichia coli] | 72 | 59 |
| 250 | 6 | 4169 | 4765 | gi\|437389 | transposase [Lactococcus lactis] | 72 | 56 |
| 258 | 7 | 5296 | 7089 | gi\|192185 | acid beta-galactosidase [Mus musculus] | 72 | 53 |
| 266 | 3 | 2024 | 1773 | gi\|145149 | ORFd [Escherichia coli] | 72 | 50 |
| 269 | 8 | 5142 | 4477 | gi\|1303791 | YqeJ [Bacillus subtilis] | 72 | 45 |
| 276 | 13 | 9843 | 8152 | gnl\|PID\|e59644 | predicted 86.4kd protein; 52Kd observed [Mycobacteriophage 15] | 72 | 48 |
| 278 | 2 | 965 | 1573 | gi\|425467 | transposase [Lactobacillus helveticus] | 72 | 52 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 279 | 2 | 1305 | 340 | gnl\|PID\|e198981 | ttg start [Campylobacter coli] | 72 | 47 |
| 283 | 4 | 1668 | 2045 | gi\|1353563 | ORF46 [Bacteriophage r1t] | 72 | 48 |
| 286 | 2 | 789 | 2606 | gi\|1651216 | Pz-peptidase [Bacillus licheniformis] | 72 | 52 |
| 290 | 4 | 2676 | 3239 | gi\|1653645 | ribosome releasing factor [Synechocystis sp.] | 72 | 56 |
| 301 | 2 | 1762 | 899 | gi\|606013 | CG Site No. 829 [Escherichia coli] | 72 | 57 |
| 362 | 2 | 377 | 688 | gi\|1001826 | cadmium-transporting ATPase [Synechocystis sp.] | 72 | 53 |
| 369 | 1 | 582 | 142 | gi\|153745 | mannitol-specific enzyme III [Streptococcus mutans] pir\|B44798\|B44798 mannitol-specific factor III, MtlF - treptococcus mutans | 72 | 47 |
| 379 | 2 | 1934 | 1527 | gi\|1055071 | C23G10.2 gene product [Caenorhabditis elegans] | 72 | 51 |
| 384 | 2 | 694 | 1098 | gi\|1208474 | hypothetical protein [Synechocystis sp.] | 72 | 49 |
| 388 | 1 | 291 | 4 | gi\|1673836 | (AE000018) Mycoplasma pneumoniae, osmotically inducible protein; similar to Swiss-Prot Accession Number P23929, from E. coli [Mycoplasma pneumoniae] | 72 | 43 |
| 401 | 6 | 3995 | 5137 | gi\|508242 | ORF 6, putative Galf synthesis pathway protein [Escherichia coli] gi\|510253 orf6 [Escherichia coli] | 72 | 62 |
| 404 | 2 | 2119 | 776 | gi\|466474 | cellobiose phosphotransferase enzyme II'' [Bacillus tearothermophilus] | 72 | 48 |
| 416 | 4 | 3461 | 1980 | gi\|710632 | beta-glucosidase [Bacillus subtilis] | 72 | 55 |
| 416 | 7 | 6285 | 5551 | gnl\|PID\|e269549 | Unknown [Bacillus subtilis] | 72 | 52 |
| 419 | 3 | 759 | 505 | gi\|928830 | ORF75; putative [Lactococcus lactis phage BK5-T] | 72 | 47 |
| 441 | 4 | 3420 | 4676 | gi\|1732195 | beta-cystathionase [Vibrio furnissii] | 72 | 54 |
| 460 | 3 | 1385 | 2641 | gi\|1652389 | beta ketoacyl-acyl carrier protein synthase [Synechocystis sp.] | 72 | 55 |
| 460 | 5 | 3129 | 3560 | gnl\|PID\|e289141 | similar to hydroxymyristoyl-(acyl carrier | 72 | 54 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | protein) dehydratase [Bacillus subtilis] | | |
| 460 | 8 | 5817 | 6023 | gi\|285621 | undefined open reading frame [Bacillus stearothermophilus] | 72 | 57 |
| 462 | 2 | 1591 | 785 | gi\|148304 | beta-1,4-N-acetylmuramoylhydrolase [Enterococcus hirae] pir\|A42296\|A42296 lysozyme 2 (EC 3.2.1.-) precursor - Enterococcus irae (ATCC 9790) | 72 | 51 |
| 467 | 1 | 2 | 706 | gi\|148711 | 6-aminohexanoate-cyclic-dimer hydrolase [Flavobacterium sp.] gi\|488343 6-aminohexanoate-cyclic-dimer hydrolase [Flavobacterium p.] | 72 | 50 |
| 469 | 3 | 1144 | 1419 | gi\|466474 | cellobiose phosphotransferase enzyme II'' [Bacillus tearothermophilus] | 72 | 48 |
| 493 | 1 | 1124 | 240 | sp\|P50848\|YPWA_BACSU | HYPOTHETICAL 58.2 KD PROTEIN IN KDGT-XPT INTERGENIC REGION. | 72 | 58 |
| 536 | 2 | 379 | 218 | gi\|437389 | transposase [Lactococcus lactis] | 72 | 58 |
| 543 | 1 | 574 | 86 | gi\|290513 | f470 [Escherichia coli] | 72 | 47 |
| 592 | 1 | 57 | 680 | gi\|987092 | ABC-transporter [Streptomyces hygroscopicus] | 72 | 55 |
| 666 | 2 | 551 | 967 | gi\|1064786 | function unknown [Bacillus subtilis] | 72 | 48 |
| 762 | 1 | 974 | 273 | gi\|304928 | pantothenate synthetase [Escherichia coli] | 72 | 55 |
| 792 | 1 | 401 | 3 | pir\|A36933\|A36933 | diacylglycerol kinase homolog - Streptococcus mutans | 72 | 50 |
| 873 | 1 | 183 | 4 | gnl\|PID\|e258329 | oxaloacetate decarboxylase alpha-chain [Legionella pneumophila] | 72 | 55 |
| 4 | 4 | 3799 | 3155 | gi\|496943 | ORF [Saccharomyces cerevisiae] | 71 | 45 |
| 10 | 2 | 180 | 977 | gnl\|PID\|e234078 | hom [Lactococcus lactis] | 71 | 49 |
| 16 | 7 | 4922 | 6097 | gi\|534982 | phosphoglucomutase [Spinacia oleracea] | 71 | 54 |
| 21 | 6 | 4148 | 3972 | gi\|1736645 | Proline/betaine transporter (Proline porter II) (PPII). [Escherichia coli] | 71 | 50 |
| 23 | 27 | 16452 | 17459 | gi\|1408503 | yxeR gene product [Bacillus subtilis] | 71 | 52 |
| 25 | 7 | 5812 | 6669 | gi\|413943 | ipa-19d gene product [Bacillus subtilis] | 71 | 58 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 31 | 1 | 80 | 946 | gi\|534045 | antiterminator [Bacillus subtilis] | 71 | 47 |
| 39 | 3 | 755 | 1297 | sp\|P09997\|YIDA_ECOLI | HYPOTHETICAL 29.7 KD PROTEIN IN IBPA-GYRB INTERGENIC REGION. | 71 | 50 |
| 39 | 7 | 2537 | 3193 | pir\|C43748\|C43748 | hypothetical protein (pepX 3' region) - Lactococcus lactis subsp. lactis | 71 | 54 |
| 45 | 10 | 5119 | 5484 | gi\|606044 | ORF_o130; Geneplot suggests frameshift, none found [Escherichia oli] | 71 | 51 |
| 48 | 10 | 11722 | 10148 | gi\|20432 | 4-coumarate:CoA ligase Pc4Cl-1 (AA 1-544) [Petroselinum crispum] ir\|S01667\|S01667 4-coumarate--CoA ligase (EC 6.2.1.12) (clone 4CL-1) - parsley | 71 | 39 |
| 55 | 4 | 1470 | 1709 | gi\|1303901 | YqhT [Bacillus subtilis] | 71 | 54 |
| 57 | 10 | 12899 | 13060 | gi\|40053 | phenylalanyl-tRNA synthetase alpha subunit [Bacillus subtilis] ir\|S11730\|YFBSA phenylalanine--tRNA ligase (EC 6.1.1.20) alpha ain - Bacillus subtilis | 71 | 45 |
| 58 | 3 | 3743 | 2571 | gi\|1658403 | formate dehydrogenase alpha subunit [Moorella thermoacetica] | 71 | 51 |
| 68 | 11 | 8225 | 8602 | gi\|793910 | surface antigen [Homo sapiens] | 71 | 49 |
| 74 | 4 | 2908 | 2042 | gi\|467435 | unknown [Bacillus subtilis] | 71 | 55 |
| 85 | 3 | 3267 | 1966 | gi\|142613 | branched chain alpha-keto acid dehydrogenase E2 [Bacillus subtilis] gi\|1303944 BfmBB [Bacillus subtilis] | 71 | 56 |
| 111 | 8 | 5737 | 4253 | gi\|1256135 | YbbF [Bacillus subtilis] | 71 | 50 |
| 111 | 9 | 6590 | 5730 | gi\|1573762 | glucokinase regulator [Haemophilus influenzae] | 71 | 53 |
| 120 | 1 | 111 | 353 | gnl\|PID\|e235823 | unknown [Schizosaccharomyces pombe] | 71 | 52 |
| 123 | 11 | 10387 | 11196 | gi\|1773195 | hypothetical [Escherichia coli] | 71 | 55 |
| 151 | 3 | 4045 | 3098 | gi\|1256618 | transport protein [Bacillus subtilis] | 71 | 51 |
| 172 | 6 | 3949 | 4806 | gi\|1262288 | CdsA [Brucella abortus] | 71 | 56 |
| 172 | 7 | 5264 | 6448 | gi\|40100 | rodC (tag3) polypeptide (AA 1-746) [Bacillus subtilis] ir\|S06049\|S06049 rodC | 71 | 52 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | protein - Bacillus subtilis p\|P13485\|TAGF_BACSU TEICHOIC ACID BIOSYNTHESIS PROTEIN F. | | |
| 190 | 7 | 3454 | 3122 | gi\|532556 | ORF23 [Enterococcus faecalis] | 71 | 52 |
| 195 | 24 | 9850 | 11871 | gi\|405564 | traE [Plasmid pSK41] | 71 | 45 |
| 215 | 4 | 3361 | 2711 | gi\|1573086 | uridine kinase (uridine monophosphokinase) (udk) [Haemophilus influenzae] | 71 | 51 |
| 218 | 2 | 1456 | 2613 | gnl\|PID\|e254644 | membrane protein [Streptococcus pneumoniae] | 71 | 41 |
| 222 | 3 | 1205 | 2053 | gnl\|PID\|e255114 | glutamate racemase [Bacillus subtilis] | 71 | 56 |
| 222 | 4 | 1611 | 1387 | gi\|1001195 | phosphate transport system permease protein PstA [Synechocystis sp.] | 71 | 57 |
| 222 | 14 | 8852 | 9853 | gi\|466720 | No definition line found [Escherichia coli] | 71 | 53 |
| 238 | 22 | 19256 | 20578 | gi\|595299 | YgiK [Salmonella typhimurium] | 71 | 50 |
| 255 | 3 | 2692 | 1061 | gnl\|PID\|e254877 | unknown [Mycobacterium tuberculosis] | 71 | 55 |
| 265 | 5 | 2960 | 1581 | gi\|1039479 | ORFU [Lactococcus lactis] | 71 | 58 |
| 276 | 2 | 1359 | 538 | gi\|496283 | lysin [Bacteriophage Tuc2009] | 71 | 63 |
| 290 | 5 | 3552 | 4379 | gi\|1016162 | ABC transporter subunit [Cyanophora paradoxa] | 71 | 49 |
| 290 | 7 | 5659 | 6912 | gi\|1001708 | NifS [Synechocystis sp.] | 71 | 56 |
| 292 | 3 | 948 | 2156 | gnl\|PID\|e233874 | hypothetical protein [Bacillus subtilis] | 71 | 55 |
| 318 | 4 | 3229 | 2285 | gi\|1256138 | YbbI [Bacillus subtilis] | 71 | 54 |
| 333 | 1 | 145 | 741 | gi\|293011 | unknown protein [Lactococcus lactis] | 71 | 50 |
| 344 | 1 | 76 | 396 | gi\|853775 | unknown [Bacillus subtilis] | 71 | 53 |
| 350 | 1 | 138 | 1394 | gi\|1652389 | beta ketoacyl-acyl carrier protein synthase [Synechocystis sp.] | 71 | 57 |
| 363 | 4 | 4184 | 5674 | gi\|1657518 | similar to fdrA gene of E. coli [Escherichia coli] | 71 | 54 |
| 364 | 5 | 5319 | 6563 | gi\|1657522 | hypothetical protein [Escherichia coli] | 71 | 46 |
| 367 | 13 | 6539 | 6162 | gi\|44225 | ribosomal protein L18 (AA 1-116) | 71 | 51 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | [Mycoplasma capricolum] ir\|S02847\|R5YM18 ribosomal protein L18 - Mycoplasma capricolum GC3) | | |
| 379 | 7 | 6884 | 5655 | gi\|887821 | ORF_o398 [Escherichia coli] | 71 | 50 |
| 399 | 9 | 6528 | 7664 | gi\|154198 | oxaloacetate decarboxylase [Salmonella typhimurium] pir\|C44465\|C44465 sodium ion pump oxaloacetate decarboxylase ubunit beta - Salmonella typhimurium | 71 | 50 |
| 399 | 18 | 13540 | 14778 | gi\|143165 | malic enzyme (EC 1.1.1.38) [Bacillus stearothermophilus] pir\|A33307\|DEBSXS malate dehydrogenase oxaloacetate-decarboxylating) (EC 1.1.1.38) - Bacillus tearothermophilus | 71 | 46 |
| 404 | 4 | 3769 | 3029 | gi\|143402 | recombination protein (ttg start codon) [Bacillus subtilis] gi\|1303923 RecN [Bacillus subtilis] | 71 | 48 |
| 464 | 1 | 1532 | 216 | gi\|895749 | putative cellobiose phosphotransferase enzyme II'' [Bacillus ubtilis] | 71 | 40 |
| 464 | 3 | 2088 | 2846 | gi\|1486242 | unknown [Bacillus subtilis] | 71 | 39 |
| 481 | 2 | 954 | 409 | gi\|144729 | butanol dehydrogenase [Clostridium acetobutylicum] sp\|Q04944\|ADHA_CLOAB NADH-DEPENDENT BUTANOL DEHYDROGENASE A (EC .1.1.-) (BDH I). | 71 | 58 |
| 482 | 4 | 2503 | 1841 | gi\|1072418 | glcA gene product [Staphylococcus carnosus] | 71 | 58 |
| 496 | 2 | 1636 | 848 | gi\|1001226 | methionine aminopeptidase [Synechocystis sp.] | 71 | 51 |
| 503 | 2 | 1624 | 650 | gi\|39478 | ATP binding protein of transport ATPases [Bacillus firmus] ir\|S15486\|S15486 ATP-binding protein - Bacillus firmus p\|P26946\|YATR_BACFI HYPOTHETICAL ABC TRANSPORTER ATP-BINDING OTEIN. | 71 | 49 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 513 | 2 | 1590 | 982 | gnl\|PID\|e202290 | unknown [Lactobacillus sake] | 71 | 46 |
| 530 | 1 | 2 | 1534 | gi\|1542974 | AbcA [Thermoanaerobacterium thermosulfurigenes] | 71 | 52 |
| 537 | 1 | 706 | 365 | gi\|929972 | ORFB; similar to B. anthracis SterneL element ORFB; putative S150-like transposase [Bacillus anthracis] | 71 | 57 |
| 553 | 1 | 304 | 1287 | gi\|1653479 | regulatory components of sensory transduction system [Synechocystis sp.] | 71 | 48 |
| 573 | 9 | 5560 | 5090 | gi\|143799 | MtrA [Bacillus subtilis] | 71 | 59 |
| 583 | 1 | 21 | 341 | gi\|1064791 | function umknown [Bacillus subtilis] | 71 | 50 |
| 584 | 2 | 638 | 276 | gi\|662792 | single-stranded DNA binding protein [unidentified eubacterium] | 71 | 58 |
| 585 | 1 | 282 | 809 | gi\|666972 | ORF 168 [Synechococcus sp.] | 71 | 46 |
| 611 | 1 | 985 | 2 | gi\|1039479 | ORFU [Lactococcus lactis] | 71 | 55 |
| 616 | 1 | 350 | 3 | gi\|1088272 | nitrogen fixation protein [Bacillus cereus] | 71 | 52 |
| 624 | 1 | 61 | 399 | gi\|40014 | pot. ORF 446 (aa 1-446) [Bacillus subtilis] | 71 | 53 |
| 624 | 2 | 608 | 1732 | gi\|40015 | pot. ORF 378 (aa 1-378) [Bacillus subtilis] | 71 | 51 |
| 659 | 1 | 76 | 582 | gi\|1591045 | hypothetical protein (SP:P31466) [Methanococcus jannaschii] | 71 | 51 |
| 668 | 2 | 836 | 1030 | gi\|467330 | replicative DNA helicase [Bacillus subtilis] | 71 | 60 |
| 683 | 1 | 582 | 118 | gnl\|PID\|e264663 | CinA [Streptococcus pneumoniae] | 71 | 55 |
| 701 | 3 | 411 | 797 | gi\|143795 | transfer RNA-Tyr synthetase [Bacillus subtilis] | 71 | 51 |
| 720 | 1 | 1 | 351 | gi\|1595810 | type-I signal peptidase SpsB [Staphylococcus aureus] | 71 | 55 |
| 724 | 2 | 1020 | 415 | gnl\|PID\|e239621 | ORF YNL218w [Saccharomyces cerevisiae] | 71 | 51 |
| 790 | 2 | 658 | 383 | gi\|1783253 | homologous to many ATP-binding transport proteins; hypothetical [Bacillus subtilis] | 71 | 48 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 799 | 1 | 505 | 906 | gi\|580866 | ipa-12d gene product [Bacillus subtilis] | 71 | 45 |
| 974 | 2 | 139 | 333 | gi\|1778531 | HI0021 homolog [Escherichia coli] | 71 | 42 |
| 980 | 1 | 156 | 497 | gi\|437389 | transposase [Lactococcus lactis] | 71 | 57 |
| 4 | 3 | 3170 | 2418 | gi\|1001805 | hypothetical protein [Synechocystis sp.] | 70 | 55 |
| 17 | 21 | 18642 | 21527 | gi\|145821 | EBG enzyme alpha subunit [Escherichia coli] | 70 | 53 |
| 19 | 8 | 2894 | 3952 | gi\|1353527 | ORF10 [Bacteriophage r1t] | 70 | 58 |
| 23 | 6 | 2640 | 3230 | gi\|699336 | C. freundii orfW homologue [Mycobacterium leprae] sp\|P53523\|Y02Y_MYCLE HYPOTHETICAL 20.9 KD PROTEIN U471A. | 70 | 43 |
| 27 | 3 | 1011 | 493 | gi\|1001644 | regulatory components of sensory transduction system [Synechocystis sp.] | 70 | 44 |
| 31 | 2 | 1095 | 1337 | gi\|1100076 | PTS-dependent enzyme II [Clostridium longisporum] | 70 | 55 |
| 32 | 10 | 6527 | 5817 | gi\|1591789 | M. jannaschii predicted coding region MJ1163 [Methanococcus jannaschii] | 70 | 51 |
| 33 | 7 | 6930 | 7235 | gi\|536972 | ORF_o90a [Escherichia coli] | 70 | 45 |
| 35 | 2 | 500 | 2533 | gi\|43819 | nagE gene product [Klebsiella pneumoniae] | 70 | 50 |
| 47 | 13 | 15837 | 14512 | gi\|150209 | ORF 1 [Mycoplasma mycoides] | 70 | 44 |
| 49 | 15 | 10409 | 11179 | gi\|853751 | N-acetylmuramoyl-L-alanine amidase [Bacteriophage A511] | 70 | 54 |
| 57 | 7 | 8365 | 12189 | gi\|142440 | ATP-dependent nuclease [Bacillus subtilis] | 70 | 48 |
| 57 | 16 | 18656 | 18033 | gi\|388565 | major cell-binding factor [Campylobacter jejuni] | 70 | 52 |
| 59 | 9 | 4985 | 7060 | gnl\|PID\|e254877 | unknown [Mycobacterium tuberculosis] | 70 | 49 |
| 72 | 6 | 6771 | 4600 | gi\|557567 | ribonucleotide reductase R1 subunit [Mycobacterium tuberculosis] sp\|P50640\|RIR1_MYCTU RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE ALPHA HAIN (EC 1.17.4.1) (RIBONUCLEOTIDE REDUCTASE) (R1 SUBUNIT) FRAGMENT). | 70 | 53 |
| 76 | 8 | 5960 | 6343 | gi\|1063251 | no homologous protein [Bacillus subtilis] | 70 | 52 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 81 | 16 | 12529 | 11723 | gi\|1732200 | PTS permease for mannose subunit IIPMan [Vibrio furnissii] | 70 | 52 |
| 98 | 7 | 8974 | 7874 | gi\|1573045 | hypothetical [Haemophilus influenzae] | 70 | 46 |
| 110 | 2 | 1353 | 502 | gi\|1399848 | unknown [Synechococcus PCC7942] | 70 | 52 |
| 123 | 7 | 5009 | 5527 | gi\|143284 | negative regulator pai 1 [Bacillus subtilis] | 70 | 51 |
| 123 | 22 | 19729 | 20412 | gi\|1591493 | glutamine transport ATP-binding protein Q [Methanococcus jannaschii] | 70 | 48 |
| 133 | 6 | 5905 | 6498 | gi\|746399 | transcription elongation factor [Escherichia coli] | 70 | 50 |
| 134 | 1 | 1 | 384 | gi\|1146242 | aspartate 1-decarboxylase [Bacillus subtilis] | 70 | 49 |
| 138 | 10 | 8543 | 7953 | gi\|467371 | LACI family of transcriptional repreesor (probable) [Bacillus ubtilis] | 70 | 50 |
| 160 | 3 | 1263 | 1520 | gi\|1468939 | meso-2,3-butanediol dehydrogenase (D-acetoin forming) [Klebsiella pneumoniae] | 70 | 45 |
| 174 | 3 | 2279 | 1572 | gi\|413931 | ipa-7d gene product [Bacillus subtilis] | 70 | 44 |
| 177 | 2 | 2104 | 1022 | gnl\|PID\|e186242 | D-mannonate hydrolase [Thermotoga neapolitana] | 70 | 52 |
| 178 | 2 | 1320 | 532 | gi\|499659 | K+ channel protein [Panulirus interruptus] | 70 | 51 |
| 180 | 18 | 17770 | 18729 | gi\|887824 | ORF_o310 [Escherichia coli] | 70 | 50 |
| 180 | 22 | 21072 | 22526 | gi\|1573294 | hypothetical [Haemophilus influenzae] | 70 | 40 |
| 181 | 9 | 7409 | 6279 | sp\|P20692\|TYRA_BACSU | PREPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH). | 70 | 49 |
| 197 | 5 | 4529 | 6340 | gi\|1783252 | homologous to many ATP-binding transport proteins including SwissProt:CYDD_ECOLI; hypothetical [Bacillus subtilis] | 70 | 47 |
| 200 | 21 | 12419 | 11820 | gi\|290943 | HindIII modification methyltransferase [Haemophilus influenzae] sp\|P43871\|MTH3_HAEIN MODIFICATION METHYLASE HINDIII (EC 2.1.1.72) ADENINE-SPECIFIC METHYLTRANSFERASE HINDIII) | 70 | 47 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | (M.HINDIII). | | |
| 210 | 4 | 3877 | 3269 | gi\|602683 | orfC [Mycoplasma capricolum] | 70 | 47 |
| 217 | 2 | 405 | 707 | gi\|153767 | ORF [Streptococcus pneumoniae] | 70 | 56 |
| 222 | 8 | 4940 | 6046 | gi\|537033 | ORF_f356 [Escherichia coli] | 70 | 54 |
| 222 | 15 | 9825 | 10553 | gi\|537039 | ORF_o228a [Escherichia coli] | 70 | 56 |
| 227 | 4 | 1871 | 2893 | gi\|1070014 | protein-dependent [Bacillus subtilis] | 70 | 44 |
| 228 | 2 | 1343 | 792 | gi\|1742730 | Protein AraJ precursor. [Escherichia coli] | 70 | 50 |
| 228 | 5 | 3470 | 2574 | gi\|1573390 | hypothetical [Haemophilus influenzae] | 70 | 54 |
| 231 | 2 | 2470 | 1238 | gi\|1574085 | H. influenzae predicted coding region HI1048 [Haemophilus influenzae] | 70 | 48 |
| 235 | 4 | 2779 | 2138 | gi\|309662 | pheromone binding protein [Plasmid pCF10] | 70 | 46 |
| 239 | 4 | 5807 | 6409 | gi\|682765 | mccB gene product [Escherichia coli] | 70 | 41 |
| 248 | 1 | 3 | 350 | gi\|143725 | putative [Bacillus subtilis] | 70 | 52 |
| 254 | 4 | 838 | 497 | gi\|49318 | ORF4 gene product [Bacillus subtilis] | 70 | 48 |
| 256 | 3 | 1737 | 2612 | gi\|596092 | putative multiple membrane domain protein; possible TTG initiation odon at position 1064, near putative RBS at position 1052 Streptococcus pyogenes] | 70 | 51 |
| 279 | 15 | 14547 | 14224 | gi\|1389549 | ORF3 [Bacillus subtilis] | 70 | 50 |
| 283 | 6 | 2279 | 3190 | gi\|853751 | N-acetylmuramoyl-L-alanine amidase [Bacteriophage A511] | 70 | 52 |
| 292 | 8 | 5557 | 6534 | gi\|474195 | This ORF is homologous to a 40.0 kd hypothetical protein in the htrB ' region from E. coli, Accession Number X61000 [Mycoplasma-like rganism] | 70 | 50 |
| 294 | 8 | 2776 | 3375 | gi\|1750126 | YncB [Bacillus subtilis] | 70 | 47 |
| 294 | 10 | 3742 | 4020 | gi\|984581 | YafQ [Escherichia coli] | 70 | 50 |
| 299 | 1 | 905 | 132 | gi\|606309 | ORF_o265; gtg start [Escherichia coli] | 70 | 40 |
| 300 | 3 | 3200 | 2784 | gi\|289260 | comE ORF1 [Bacillus subtilis] | 70 | 50 |
| 301 | 9 | 8564 | 7590 | gi\|1303865 | YqgR [Bacillus subtilis] | 70 | 52 |
| 336 | 2 | 661 | 921 | gi\|202864 | [Rat alternatively spliced mRNA.], gene | 70 | 47 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | product [Rattus norvegicus] | | |
| 339 | 1 | 269 | 3 | gi\|786163 | Ribosomal Protein L10 [Bacillus subtilis] | 70 | 50 |
| 351 | 9 | 4760 | 4359 | gi\|799235 | dTDP-6-deoxy-L-lyxo-4-hexulose reductase [Escherichia coli] | 70 | 45 |
| 399 | 28 | 28203 | 28793 | gi\|146278 | glucitol-specific enzyme II (gutA) [Escherichia coli] pir\|A26725\|WQEC2S phosphotransferase system enzyme II (EC .7.1.69), sorbitol-specific, factor II - Escherichia coli sp\|P05705\|PTHB_ECOLI PTS SYSTEM, GLUCITOL/SORBITOL-SPECIFIC IIBC OMPONENT (EIIBC-GUT) | 70 | 52 |
| 406 | 1 | 1 | 552 | gi\|49315 | ORF1 gene product [Bacillus subtilis] | 70 | 50 |
| 436 | 5 | 2417 | 2193 | gi\|773665 | transposase [Lactococcus lactis] | 70 | 36 |
| 482 | 3 | 1887 | 1660 | gi\|48680 | ptsG-like product [Bacillus subtilis] | 70 | 47 |
| 529 | 3 | 6587 | 7030 | gi\|1022726 | unknown [Staphylococcus haemolyticus] | 70 | 44 |
| 535 | 2 | 1702 | 965 | gi\|1747435 | KdpE [Clostridium acetobutylicum] | 70 | 52 |
| 543 | 2 | 1248 | 547 | gi\|1591045 | hypothetical protein (SP:P31466) [Methanococcus jannaschii] | 70 | 47 |
| 543 | 8 | 4084 | 3878 | gi\|511976 | SERP gene gene product [Plasmodium falciparum] | 70 | 60 |
| 560 | 3 | 1037 | 876 | gi\|558458 | acidic 82 kDa protein [Homo sapiens] | 70 | 40 |
| 573 | 4 | 1920 | 2258 | gi\|336639 | prephytoene pyrophosphate dehydrogenase [Cyanophora paradoxa] gi\|1016130 prenyl transferase [Cyanophora paradoxa] pir\|A40433\|A40433 prephytoene pyrophosphatase dehydrogenase (crtE) omolog - Cyanophora paradoxa | 70 | 32 |
| 599 | 2 | 244 | 573 | gi\|42029 | ORF1 gene product [Escherichia coli] | 70 | 49 |
| 608 | 3 | 867 | 556 | gi\|475032 | formamidopyrimidine-DNA glycosylase [Streptococcus mutans] sp\|P55045\|FPG_STRMU FORMAMIDOPYRIMIDINE-DNA GLYCOSYLASE (EC .2.2.23) (FAPY-DNA GLYCOSYLASE). | 70 | 53 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 636 | 1 | 2 | 628 | gi\|606309 | ORF_o265; gtg start [Escherichia coli] | 70 | 50 |
| 670 | 2 | 2157 | 1828 | gi\|1657698 | hyaluronan receptor [Homo sapiens] | 70 | 41 |
| 702 | 1 | 103 | 870 | gi\|149490 | sucrose-6-phosphate hydrolase [Lactococcus lactis] pir\|JH0754\|JH0754 sucrose-6-phosphate hydrolase (EC 3.2.1.-) - actococcus lactis | 70 | 51 |
| 726 | 2 | 725 | 480 | gnl\|PID\|e240103 | unknown ORF [Saccharomyces cerevisiae] | 70 | 41 |
| 854 | 1 | 1 | 207 | gi\|532653 | thermonuclease [Staphylococcus hyicus] | 70 | 51 |
| 901 | 1 | 238 | 447 | gi\|172022 | myosin 1 isoform (MYO2) [Saccharomyces cerevisiae] | 70 | 20 |
| 940 | 1 | 1 | 318 | gi\|1039479 | ORFU [Lactococcus lactis] | 70 | 56 |
| 1 | 2 | 2112 | 1213 | gi\|413976 | ipa-52r gene product [Bacillus subtilis] | 69 | 51 |
| 8 | 2 | 2196 | 778 | gi\|1510108 | ORF-1 [Agrobacterium tumefaciens] | 69 | 50 |
| 8 | 9 | 7949 | 6654 | gi\|1196907 | daunorubicin resistance protein [Streptomyces peucetius] | 69 | 44 |
| 16 | 3 | 1618 | 2574 | gi\|1109684 | ProV [Bacillus subtilis] | 69 | 53 |
| 17 | 26 | 25781 | 26944 | gi\|485275 | 53.6 kDa protein [Streptococcus pneumoniae] | 69 | 44 |
| 17 | 35 | 32300 | 32770 | gi\|1574146 | pfs protein (pfs) [Haemophilus influenzae] | 69 | 53 |
| 23 | 30 | 18107 | 18538 | gnl\|PID\|e249656 | YneT [Bacillus subtilis] | 69 | 59 |
| 25 | 8 | 6653 | 6994 | gi\|413943 | ipa-19d gene product [Bacillus subtilis] | 69 | 46 |
| 37 | 2 | 2042 | 186 | gi\|143331 | alkaline phosphatase regulatory protein [Bacillus subtilis] pir\|A27650\|A27650 regulatory protein phoR - Bacillus subtilis sp\|P23545\|PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN HOR (EC 2.7.3.-). | 69 | 52 |
| 39 | 2 | 528 | 767 | gi\|1408493 | homologous to SwissProt:YIDA_ECOLI hypothetical protein [Bacillus subtilis] | 69 | 52 |
| 56 | 6 | 4809 | 3457 | gi\|1591610 | probable ATP-dependent helicase [Methanococcus jannaschii] | 69 | 45 |
| 67 | 5 | 3042 | 3938 | gi\|1658188 | oxidative stress transcriptional regulator | 69 | 39 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | [Erwinia carotovora] | | |
| 68 | 3 | 684 | 1529 | gnl\|PID\|e214719 | PlcR protein [Bacillus thuringiensis] | 69 | 45 |
| 72 | 4 | 2099 | 3394 | gi\|882672 | ORF_o313 [Escherichia coli] | 69 | 37 |
| 81 | 15 | 11820 | 10915 | gi\|1732201 | PTS permease for mannose subunit IIBMan [Vibrio furnissii] | 69 | 44 |
| 83 | 20 | 14001 | 15800 | gi\|1230668 | Similar to Arginyl-tRNA synthetase (Swiss Prot. accession number P11875) [Saccharomyces cerevisiae] | 69 | 44 |
| 85 | 6 | 6309 | 5299 | sp\|P54533\|DLD2_BACSU | LIPOAMIDE DEHYDROGENASE COMPONENT (E3) OF BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE COMPLEX (EC 1.8.1.4) (DIHYDROLIPOAMIDE DEHYDROGENASE) (LPD-VAL). | 69 | 46 |
| 86 | 3 | 2084 | 3367 | gi\|143318 | phosphoglycerate kinase [Bacillus megaterium] | 69 | 53 |
| 94 | 2 | 1401 | 751 | gi\|755216 | N-acetylmuramidase [Lactococcus lactis] | 69 | 41 |
| 94 | 16 | 20498 | 19197 | gi\|1208948 | unknown [Escherichia coli] | 69 | 47 |
| 98 | 8 | 10201 | 9029 | gi\|563934 | similar to E. coli hypothetical protein: PIR Accession Number Q0614] [Bacillus subtilis] | 69 | 51 |
| 109 | 4 | 2350 | 1316 | gi\|396501 | aspartyl-tRNA synthetase [Thermus aquaticus thermophilus] pir\|S33743\|S33743 aspartate--tRNA ligase (EC 6.1.1.12) - Thermus quaticus | 69 | 56 |
| 114 | 1 | 83 | 1522 | gi\|1658402 | formate dehydrogenase beta subunit [Moorella thermoacetica] | 69 | 45 |
| 123 | 9 | 7617 | 8984 | gi\|1773192 | similar to S. cerevisiae dal1 [Escherichia coli] | 69 | 50 |
| 128 | 11 | 7940 | 7578 | gi\|895750 | putative cellobiose phosphotransferase enzyme III [Bacillus ubtilis] | 69 | 53 |
| 130 | 10 | 8764 | 9036 | gi\|1641 | put. Na(+)/glucose co-transporter (AA 1-662) [Oryctolagus cuniculus] \|1717 | 69 | 47 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | cortical sodium-D-glucose cotransporter [Oryctolagus iculus] | | |
| 138 | 26 | 16721 | 17545 | pir\|A25805\|A25805 | L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus subtilis | 69 | 55 |
| 139 | 2 | 310 | 1083 | gi\|1408587 | relaxase [Lactococcus lactis lactis] | 69 | 46 |
| 139 | 9 | 5196 | 4984 | gi\|473955 | DNA-binding protein [Lactobacillus sp.] | 69 | 34 |
| 142 | 9 | 5559 | 4564 | gi\|623073 | ORF360; putative [Bacteriophage LL-H] | 69 | 47 |
| 155 | 6 | 4658 | 5818 | gi\|1591260 | endoglucanase [Methanococcus jannaschii] | 69 | 48 |
| 158 | 12 | 11671 | 11201 | gi\|606744 | cytidine deaminase [Bacillus subtilis] | 69 | 52 |
| 162 | 5 | 5888 | 4032 | gi\|142993 | glycerol-3-phosphate dehydrogenase (glpD) (EC 1.1.99.5) [Bacillus ubtilis] | 69 | 54 |
| 180 | 2 | 1901 | 1203 | gi\|1575577 | DNA-binding response regulator [Thermotoga maritima] | 69 | 49 |
| 197 | 4 | 3571 | 4602 | gi\|1783251 | homologous to cytochrome d ubiquinol oxidase subunit II; hypothetical [Bacillus subtilis] | 69 | 46 |
| 197 | 6 | 6283 | 7701 | gi\|1783253 | homologous to many ATP-binding transport proteins; hypothetical [Bacillus subtilis] | 69 | 49 |
| 222 | 1 | 201 | 10 | gi\|149901 | gene codes for a 19 kDa protein [Mycobacterium avium] sp\|P46733\|19KD_MYCAV 19 KD LIPOPROTEIN ANTIGEN PRECURSOR. | 69 | 50 |
| 223 | 28 | 23857 | 24567 | gnl\|PID\|e269548 | Unknown [Bacillus subtilis] | 69 | 53 |
| 228 | 3 | 2031 | 1285 | gi\|1742730 | Protein AraJ precursor. [Escherichia coli] | 69 | 45 |
| 229 | 8 | 7390 | 6698 | gi\|1162980 | ribulose-5-phosphate 3-epimerase [Spinacia oleracea] | 69 | 52 |
| 238 | 27 | 25243 | 25695 | gi\|305005 | ORF_f104 [Escherichia coli] | 69 | 53 |
| 253 | 3 | 1067 | 921 | gi\|1591278 | aspartokinase I [Methanococcus jannaschii] | 69 | 39 |
| 260 | 4 | 2110 | 3105 | gi\|580841 | F1 [Bacillus subtilis] | 69 | 45 |
| 268 | 3 | 2287 | 1910 | gi\|460026 | repressor protein [Streptococcus pneumoniae] | 69 | 48 |
| 269 | 7 | 4532 | 4083 | gi\|1303792 | YqeK [Bacillus subtilis] | 69 | 50 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 271 | 15 | 11040 | 12236 | gi\|1303805 | YqeR [Bacillus subtilis] | 69 | 48 |
| 271 | 16 | 12444 | 12809 | gi\|435490 | orf1 gene product [Lactococcus lactis] | 69 | 46 |
| 281 | 3 | 1277 | 2068 | gi\|1303968 | YqjQ [Bacillus subtilis] | 69 | 50 |
| 281 | 6 | 5004 | 5534 | gi\|1773151 | adenine phosphoribosyltransferase [Escherichia coli] | 69 | 54 |
| 292 | 24 | 19939 | 18398 | gi\|1652664 | glutamine-binding periplasmic protein [Synechocystis sp.] | 69 | 45 |
| 323 | 3 | 2708 | 4243 | gi\|179401 | beta-D-galactosidase precursor (EC 3.2.1.23) [Homo sapiens] gi\|179423 beta-galactosidase precursor (EC 3.2.1.23) [Homo sapiens] pir\|A32688\|A32611 beta-galactosidase (EC 3.2.1.23) precursor - uman | 69 | 56 |
| 330 | 2 | 1388 | 2353 | gi\|1303783 | YqeC [Bacillus subtilis] | 69 | 48 |
| 332 | 1 | 2 | 223 | gi\|1653594 | hemolysin [Synechocystis sp.] | 69 | 50 |
| 338 | 9 | 7035 | 7607 | gi\|467442 | stage V sporulation [Bacillus subtilis] | 69 | 55 |
| 341 | 1 | 1 | 408 | gi\|1477741 | histidine periplasmic binding protein P29 [Campylobacter jejuni] | 69 | 50 |
| 368 | 2 | 972 | 598 | gi\|516826 | rat GCP360 [Rattus rattus] | 69 | 33 |
| 375 | 4 | 3405 | 2599 | gi\|1215693 | putative orf; GT9_orf434 [Mycoplasma pneumoniae] | 69 | 38 |
| 386 | 1 | 2 | 166 | gi\|1549376 | putative protein [Synechococcus PCC7942] | 69 | 42 |
| 396 | 4 | 1248 | 1715 | gi\|410132 | ORFX8 [Bacillus subtilis] | 69 | 50 |
| 398 | 4 | 2763 | 2927 | gi\|466475 | putative phospho-beta-glucosidase [Bacillus stearothermophilus] pir\|D49898\|D49898 cellobiose phosphotransferase system celC - acillus stearothermophilus | 69 | 55 |
| 421 | 5 | 2950 | 3471 | gi\|1574625 | H. influenzae predicted coding region HI1074 [Haemophilus influenzae] | 69 | 45 |
| 423 | 4 | 2408 | 2893 | gnl\|PID\|e163522 | rnhB [Haemophilus influenzae] | 69 | 55 |
| 436 | 3 | 1763 | 1521 | gi\|155032 | ORF B [Plasmid pEa34] | 69 | 37 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 452 | 1 | 3 | 341 | gi\|1591139 | M. jannaschii predicted coding region MJ0435 [Methanococcus jannaschii] | 69 | 52 |
| 470 | 3 | 1816 | 2181 | gi\|437389 | transposase [Lactococcus lactis] | 69 | 56 |
| 471 | 2 | 2003 | 813 | gi\|854233 | cymF gene product [Klebsiella oxytoca] | 69 | 49 |
| 478 | 1 | 822 | 4 | gi\|142521 | deoxyribodipyrimidine photolyase [Bacillus subtilis] gnl\|PID\|e255102 deoxyribodipyrimidine photolyase [Bacillus ubtilis] | 69 | 63 |
| 490 | 4 | 1447 | 1289 | gi\|699379 | glvr-1 protein [Mycobacterium leprae] | 69 | 41 |
| 518 | 2 | 213 | 605 | pir\|S00076\|R5BS12 | ribosomal protein L12 - Bacillus stearothermophilus | 69 | 59 |
| 536 | 4 | 1471 | 1653 | gi\|1146240 | ketopantoate hydroxymethyltransferase [Bacillus subtilis] | 69 | 53 |
| 539 | 5 | 3796 | 5091 | gi\|973231 | gamma-glutamyl phosphate reductase [Lycopersicon esculentum] | 69 | 54 |
| 566 | 1 | 1 | 231 | gi\|45741 | ORFE [Enterococcus faecalis] | 69 | 50 |
| 579 | 5 | 2729 | 3595 | gi\|145887 | malonyl coenzyme A-acyl carrier protein transacylase [Escherichia oli] | 69 | 49 |
| 583 | 2 | 373 | 912 | gi\|1064791 | function umknown [Bacillus subtilis] | 69 | 55 |
| 605 | 1 | 254 | 3 | pir\|S39743\|S39743 | hypothetical protein - Bacillus subtilis | 69 | 37 |
| 630 | 2 | 1659 | 1231 | gi\|153672 | lactose repressor [Streptococcus mutans] | 69 | 47 |
| 634 | 1 | 36 | 731 | gi\|1022725 | unknown [Staphylococcus haemolyticus] | 69 | 53 |
| 662 | 1 | 486 | 73 | gi\|467431 | high level kasgamycin resistance [Bacillus subtilis] sp\|P37468\|KSGA_BACSU DIMETHYLADENOSINE TRANSFERASE (EC 2.1.1.-) S-ADENOSYLMETHIONINE-6-N', N'-ADENOSYL(RRNA) DIMETHYLTRANSFERASE) 16S RRNA DIMETHYLASE) (HIGH LEVEL KASUGAMYCIN RESISTANCE PROTEIN SGA) (K | 69 | 55 |
| 689 | 1 | 340 | 26 | gi\|1017817 | membrane spanning protein [Streptomyces coelicolor] | 69 | 41 |
| 756 | 2 | 300 | 500 | gi\|520596 | Mre2 protein [Saccharomyces cerevisiae] | 69 | 46 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 792 | 2 | 855 | 460 | gi\|1303823 | YqfG [Bacillus subtilis] | 69 | 55 |
| 916 | 1 | 4 | 789 | gnl\|PID\|e253114 | ornithine carbamoyltransferase [Pyrococcus furiosus] | 69 | 57 |
| 7 | 3 | 2609 | 3748 | gi\|1303836 | YqfO [Bacillus subtilis] | 68 | 50 |
| 16 | 5 | 4165 | 4689 | gi\|142450 | ahrC protein [Bacillus subtilis] | 68 | 46 |
| 17 | 16 | 12826 | 13071 | gi\|222681 | RNA polymerase [Tomato spotted wilt virus] | 68 | 50 |
| 17 | 32 | 31402 | 31572 | gi\|1303984 | YqkG [Bacillus subtilis] | 68 | 44 |
| 17 | 33 | 31509 | 32009 | gi\|1303984 | YqkG [Bacillus subtilis] | 68 | 50 |
| 29 | 1 | 19 | 282 | gi\|1234787 | up-regulated by thyroid hormone in tadpoles; expressed specifically in the tail and only at metamorphosis; membrane bound or extracellular protein; C-terminal basic region [Xenopus laevis] | 68 | 37 |
| 29 | 3 | 1087 | 1950 | gi\|407878 | leucine rich protein [Streptococcus equisimilis] | 68 | 45 |
| 45 | 1 | 204 | 959 | gi\|1039479 | ORFU [Lactococcus lactis] | 68 | 50 |
| 47 | 7 | 8108 | 7527 | gi\|142853 | homologous to unidentified E. coli protein [Bacillus subtilis] gi\|143161 maf [Bacillus subtilis] | 68 | 46 |
| 52 | 6 | 4304 | 5050 | gnl\|PID\|e124050 | alpha-acetolactate decarboxylase [Lactococcus lactis] | 68 | 53 |
| 58 | 5 | 5961 | 4807 | gi\|466365 | potential NAD-reducing hydrogenase subunit [Desulfovibrio ructosovorans] | 68 | 49 |
| 68 | 8 | 4036 | 4743 | gi\|1673727 | (AE000009) Mycoplasma pneumoniae, glutamine transport ATP-binding protein; similar to Swiss-Prot Accession Number P10346, from E. coli [Mycoplasma pneumoniae] | 68 | 44 |
| 72 | 5 | 4441 | 3434 | gi\|1395209 | ribonucleotide reductase R2-2 small subunit [Mycobacterium tuberculosis] | 68 | 52 |
| 80 | 1 | 836 | 3 | gi\|474176 | regulator protein [Staphylococcus xylosus] | 68 | 48 |
| 81 | 2 | 793 | 1359 | gi\|1064809 | homologous to sp:HTRA_ECOLI [Bacillus | 68 | 48 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | subtilis] | | |
| 85 | 9 | 6911 | 6711 | gi\|144893 | butyrate kinase [Clostridium acetobutylicum] | 68 | 55 |
| 89 | 8 | 7184 | 5970 | gi\|1469784 | putative cell division protein ftsW [Enterococcus hirae] | 68 | 44 |
| 91 | 3 | 828 | 1076 | gi\|726480 | L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus ubtilis] | 68 | 53 |
| 103 | 1 | 1019 | 3 | gi\|143365 | phosphoribosyl aminoimidazole carboxylase II (PUR-K; ttg start odon) [Bacillus subtilis] | 68 | 50 |
| 106 | 2 | 2441 | 1509 | gi\|146860 | delta-2-isopentenyl pyrophosphate transferase [Escherichia coli] gi\|537012 tRNA delta-2-isopentenylpyrophosphate (IPP) transferase Escherichia coli] | 68 | 47 |
| 112 | 1 | 558 | 100 | gnl\|PID\|e242290 | carbamate kinase [Clostridium perfringens] | 68 | 50 |
| 116 | 3 | 2383 | 1496 | gi\|755601 | unknown [Bacillus subtilis] | 68 | 42 |
| 119 | 3 | 2136 | 1201 | gi\|1171125 | thioredoxin reductase [Clostridium litorale] | 68 | 49 |
| 121 | 4 | 3697 | 4650 | gi\|790945 | aryl-alcohol dehydrogenase [Bacillus subtilis] | 68 | 48 |
| 123 | 26 | 24262 | 24801 | gi\|537235 | Kenn Rudd identifies as gpmB [Escherichia coli] | 68 | 51 |
| 123 | 27 | 24887 | 25888 | gi\|143150 | levR [Bacillus subtilis] | 68 | 51 |
| 126 | 4 | 2773 | 1844 | gi\|551854 | ORF2 [Erwinia herbicola] | 68 | 54 |
| 131 | 1 | 150 | 1058 | gi\|1387979 | 44% identity over 302 residues with hypothetical protein from Synechocystis sp, accession D64006_CD; expression induced by environmental stress; some similarity to glycosyl transferases; two potential membrane-spanning helices [Bacillus subtil | 68 | 44 |
| 134 | 3 | 2154 | 1804 | sp\|P39213\|YI91_SHI | INSERTION ELEMENT IS911 HYPOTHETICAL 12.7 | 68 | 43 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | DY | KD PROTEIN. | | |
| 138 | 19 | 12285 | 12656 | gi\|1438847 | homologue of hypothetical 17.6 kDa protein in rplI-cpdB intergenic region of E. coli [Bacillus subtilis] | 68 | 43 |
| 151 | 2 | 2784 | 1654 | gi\|143365 | phosphoribosyl aminoimidazole carboxylase II (PUR-K; ttg start odon) [Bacillus subtilis] | 68 | 45 |
| 164 | 23 | 24352 | 24119 | gi\|1573564 | hypothetical [Haemophilus influenzae] | 68 | 40 |
| 166 | 2 | 970 | 1260 | gi\|151968 | nifS [Rhodobacter sphaeroides] | 68 | 41 |
| 172 | 2 | 1320 | 2015 | gi\|1208965 | hypothetical 23.3 kd protein [Escherichia coli] | 68 | 46 |
| 175 | 1 | 900 | 451 | gi\|468207 | Submitter comments: A Mg2+ transporting P-type ATPase highly omologous with mgtB ATPase at 80 min on Salmonella chromosome. ediates the influx of Mg2+ only. Transcription regulated by xtracellular Mg2+ [Salmonella typhimurium] | 68 | 47 |
| 180 | 14 | 12551 | 14956 | gi\|565641 | FdrA protein [Escherichia coli] | 68 | 49 |
| 186 | 1 | 3 | 686 | gi\|405804 | transposase [Streptococcus thermophilus] | 68 | 51 |
| 200 | 1 | 239 | 3 | gi\|468016 | immunoglobulin heavy chain binding protein [Giardia intestinalis] | 68 | 42 |
| 201 | 4 | 4468 | 3686 | gi\|304013 | abcA [Aeromonas salmonicida] | 68 | 50 |
| 204 | 10 | 6833 | 6468 | gi\|488430 | alcohol dehydrogenase 2 [Entamoeba histolytica] | 68 | 51 |
| 214 | 3 | 3360 | 2491 | gi\|928834 | integrase [Lactococcus lactis phage BK5-T] | 68 | 50 |
| 229 | 9 | 8277 | 7375 | gi\|1574569 | hypothetical [Haemophilus influenzae] | 68 | 41 |
| 229 | 14 | 14288 | 13740 | gnl\|PID\|e290287 | polypeptide deformylase [Bacillus subtilis] | 68 | 50 |
| 230 | 5 | 4593 | 3532 | gi\|143002 | proton glutamate symport protein [Bacillus caldotenax] pir\|S26246\|S26246 glutamate/aspartate transport protein - Bacillus aldotenax | 68 | 29 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 244 | 1 | 1 | 891 | gi\|537080 | ribonucleoside triphosphate reductase [Escherichia coli] pir\|A47331\|A47331 oxygen-sensitive ribonucleoside-triphosphate eductase (EC 1.17.4.-) - Escherichia coli | 68 | 54 |
| 244 | 5 | 4249 | 3551 | gi\|1773172 | hypothetical protein [Escherichia coli] | 68 | 46 |
| 244 | 7 | 5670 | 5212 | gi\|467423 | unknown [Bacillus subtilis] | 68 | 43 |
| 264 | 9 | 3925 | 3734 | gi\|914991 | Similar to hemoglobinase [Saccharomyces cerevisiae] pir\|S59796\|S59796 hypothetical protein D9798.2 - yeast Saccharomyces cerevisiae) | 68 | 44 |
| 271 | 7 | 3484 | 4686 | gi\|1469784 | putative cell division protein ftsW [Enterococcus hirae] | 68 | 50 |
| 271 | 11 | 6817 | 6548 | gi\|413948 | ipa-24d gene product [Bacillus subtilis] | 68 | 50 |
| 288 | 3 | 1638 | 1333 | gi\|562039 | NADH dehydrogenase, subunit 2 [Acanthamoeba castellanii] pir\|S53835\|S53835 *NADH dehydrogenase* chain 2 - Acanthamoeba astellanii mitochondrion (SGC6) | 68 | 50 |
| 295 | 6 | 3537 | 4472 | gi\|555668 | glycosylasparaginase precursor [Flavobacterium meningosepticum] | 68 | 41 |
| 296 | 2 | 3143 | 1950 | gi\|1742630 | Bicyclomycin resistance protein (Sulfonamide resistance protein). [Escherichia coli] | 68 | 34 |
| 301 | 3 | 3271 | 1760 | gi\|413960 | ipa-36d galT gene product [Bacillus subtilis] | 68 | 53 |
| 315 | 3 | 2230 | 905 | gi\|1653498 | ABC transporter [Synechocystis sp.] | 68 | 47 |
| 318 | 2 | 1285 | 854 | gi\|43940 | EIII-F Sor PTS [Klebsiella pneumoniae] | 68 | 39 |
| 320 | 2 | 1178 | 621 | gi\|664842 | sister of P-glycoprotein [Sus scrofa domestica] | 68 | 46 |
| 331 | 2 | 342 | 566 | pir\|B48396\|B48396 | ribosomal protein L33 - Bacillus stearothermophilus | 68 | 59 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 336 | 1 | 1 | 663 | gi\|1006591 | cation-transporting ATPase PacL [Synechocystis sp.] | 68 | 44 |
| 338 | 6 | 4004 | 5035 | gi\|155276 | aldehyde dehydrogenase [Vibrio cholerae] | 68 | 51 |
| 338 | 12 | 10404 | 11165 | gi\|467444 | transcription-repair coupling factor [Bacillus subtilis] sp\|P37474\|MFD_BACSU TRANSCRIPTION-REPAIR COUPLING FACTOR (TRCF). | 68 | 46 |
| 341 | 3 | 743 | 1222 | gi\|1183886 | integral membrane protein [Bacillus subtilis] | 68 | 45 |
| 351 | 6 | 2992 | 2561 | gi\|580881 | ipa-73d gene product [Bacillus subtilis] | 68 | 53 |
| 363 | 8 | 12517 | 9950 | gi\|1652980 | H(+)-transporting ATPase [Synechocystis sp.] | 68 | 46 |
| 368 | 3 | 1269 | 1736 | gnl\|PID\|e209005 | homologous to ORF2 in nrdEF operons of E.coli and S.typhimurium [Lactococcus lactis] | 68 | 37 |
| 386 | 11 | 6564 | 6115 | gi\|765072 | ORF3 [Staphylococcus aureus] | 68 | 46 |
| 395 | 3 | 935 | 729 | gi\|15521 | ORF 3 (AA 1-90) [Bacteriophage phi-105] | 68 | 34 |
| 399 | 8 | 6073 | 6519 | gi\|153584 | biotin carboxyl carrier protein [Streptococcus mutans] sp\|P29337\|BCCP_STRMU BIOTIN CARBOXYL CARRIER PROTEIN (BCCP). | 68 | 53 |
| 408 | 3 | 2289 | 1336 | gi\|41572 | GlnP (AA 1-219) [Escherichia coli] | 68 | 40 |
| 420 | 1 | 559 | 2 | gi\|1592142 | ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii] | 68 | 51 |
| 423 | 2 | 254 | 1294 | gi\|1773109 | similar to S. typhimurium apbA [Escherichia coli] | 68 | 47 |
| 423 | 3 | 1465 | 2421 | gi\|1653032 | hypothetical protein [Synechocystis sp.] | 68 | 40 |
| 428 | 1 | 859 | 2 | gi\|1652454 | hypothetical protein [Synechocystis sp.] | 68 | 48 |
| 432 | 7 | 4626 | 3901 | gi\|1573285 | hypothetical [Haemophilus influenzae] | 68 | 55 |
| 434 | 1 | 90 | 1889 | gi\|1542975 | AbcB [Thermoanaerobacterium thermosulfurigenes] | 68 | 50 |
| 441 | 5 | 4674 | 5156 | gi\|467437 | unknown [Bacillus subtilis] | 68 | 48 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 455 | 4 | 3835 | 4080 | gi\|19815 | luminal binding protein (BiP) [Nicotiana tabacum] | 68 | 40 |
| 530 | 2 | 394 | 546 | gi\|763326 | unknown [Saccharomyces cerevisiae] | 68 | 42 |
| 531 | 2 | 810 | 622 | gi\|1146183 | putative [Bacillus subtilis] | 68 | 51 |
| 537 | 3 | 1353 | 1192 | gi\|929968 | ORFA; similar to B. anthracis WeyAR element ORFA; putative ransposase [Bacillus anthracis] | 68 | 56 |
| 539 | 3 | 2725 | 2231 | gi\|1353537 | dUTPase [Bacteriophage r1t] | 68 | 53 |
| 569 | 1 | 3 | 446 | gi\|146544 | 18 kD protein [Escherichia coli] | 68 | 47 |
| 591 | 2 | 656 | 174 | gi\|1039479 | ORFU [Lactococcus lactis] | 68 | 42 |
| 652 | 2 | 739 | 1032 | gi\|1303715 | YrkP [Bacillus subtilis] | 68 | 50 |
| 671 | 2 | 436 | 1617 | gi\|413959 | ipa-35d galK gene product [Bacillus subtilis] | 68 | 50 |
| 684 | 1 | 466 | 2 | gnl\|PID\|e248400 | orfRM1 gene product [Bacillus subtilis] | 68 | 40 |
| 693 | 1 | 2 | 787 | gi\|405804 | transposase [Streptococcus thermophilus] | 68 | 46 |
| 700 | 2 | 772 | 596 | gi\|153801 | enzyme scr-II [Streptococcus mutans] | 68 | 50 |
| 735 | 1 | 118 | 609 | gi\|969027 | gamma-aminobutyrate permease [Bacillus subtilis] sp\|P46349\|GABP_BACSU GABA PERMEASE (4-AMINO BUTYRATE TRANSPORT ARRIER) (GAMA-AMINOBUTYRATE PERMEASE). | 68 | 40 |
| 750 | 1 | 2 | 529 | gi\|893358 | PgsA [Bacillus subtilis] | 68 | 54 |
| 762 | 2 | 1588 | 950 | gi\|1146240 | ketopantoate hydroxymethyltransferase [Bacillus subtilis] | 68 | 49 |
| 790 | 1 | 407 | 3 | gi\|142224 | attachment protein ChvA (ttg strart codon) [Agrobacterium umefaciens] | 68 | 55 |
| 882 | 1 | 3 | 278 | gi\|57572 | glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating) attus rattus] | 68 | 48 |
| 950 | 1 | 140 | 568 | gi\|882736 | ORF_f278 [Escherichia coli] | 68 | 53 |
| 969 | 2 | 554 | 339 | gi\|1118031 | similar to neural cell adhesion molecules and neuroglians in their IG-like C2-type domains [Caenorhabditis elegans] | 68 | 47 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 970 | 1 | 297 | 73 | gi\|474404 | cyclophilin [Tolypocladium inflatum] | 68 | 40 |
| 1 | 1 | 1103 | 3 | gi\|48790 | ORF 3 [Pseudomonas putida] | 67 | 50 |
| 29 | 10 | 7156 | 6614 | sp\|P36672\|PTTB_ECOLI | PTS SYSTEM, TREHALOSE-SPECIFIC IIBC COMPONENT (EIIBC-TRE) (TREHALOSE- PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT) (EC 2.7.1.69) (EII-TRE). | 67 | 52 |
| 48 | 8 | 8035 | 9141 | gi\|975627 | N-acylamino acid racemase [Amycolatopsis sp.] | 67 | 48 |
| 55 | 12 | 6621 | 7439 | gi\|391610 | farnesyl diphosphate synthase [Bacillus stearothermophilus] pir\|JX0257\|JX0257 geranyltranstransferase (EC 2.5.1.10) - Bacillus tearothermophilus | 67 | 47 |
| 57 | 13 | 13972 | 16401 | gnl\|PID\|e255138 | phenylalanyl-tRNA synthetase beta subunit [Bacillus subtilis] | 67 | 47 |
| 63 | 4 | 1917 | 2729 | gi\|1321629 | MIP related protein of E. coli [Escherichia coli] | 67 | 47 |
| 68 | 12 | 8600 | 8923 | gi\|793910 | surface antigen [Homo sapiens] | 67 | 43 |
| 72 | 7 | 7138 | 6740 | gnl\|PID\|e209005 | homologous to ORF2 in nrdEF operons of E.coli and S.typhimurium [Lactococcus lactis] | 67 | 39 |
| 72 | 10 | 8309 | 9433 | gi\|1199515 | ferrous iron transport protein B [Escherichia coli] | 67 | 41 |
| 85 | 5 | 5315 | 4296 | gi\|142611 | branched chain alpha-keto acid dehydrogenase E1-alpha [Bacillus ubtilis] | 67 | 52 |
| 101 | 5 | 4149 | 3100 | gi\|1109686 | ProX [Bacillus subtilis] | 67 | 48 |
| 110 | 4 | 2335 | 1292 | gi\|1066343 | mu-crystallin [Homo sapiens] | 67 | 48 |
| 114 | 12 | 12936 | 13520 | gi\|146218 | serine hydroxymethyltransferase [Escherichia coli] | 67 | 50 |
| 115 | 5 | 3137 | 2010 | gi\|1256150 | YbaR [Bacillus subtilis] | 67 | 47 |
| 115 | 6 | 3199 | 2792 | gi\|1652593 | hypothetical protein [Synechocystis sp.] | 67 | 45 |
| 123 | 25 | 22739 | 24208 | gi\|148711 | 6-aminohexanoate-cyclic-dimer hydrolase [Flavobacterium sp.] gi\|488343 6- | 67 | 50 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | aminohexanoate-cyclic-dimer hydrolase [Flavobacterium p.] | | |
| 124 | 6 | 5139 | 4267 | gi\|1016770 | prolipoprotein diacylglyceryl transferase [Staphylococcus aureus] | 67 | 50 |
| 125 | 2 | 1306 | 221 | gi\|853743 | L-alanoyl-D-glutamate peptidase [Bacteriophage A118] | 67 | 50 |
| 128 | 36 | 29462 | 28737 | gi\|142940 | ftsA [Bacillus subtilis] | 67 | 46 |
| 138 | 27 | 17602 | 18183 | gi\|1256639 | putative [Bacillus subtilis] | 67 | 50 |
| 138 | 31 | 21578 | 20097 | gi\|143245 | Na+/H+ antiporter [Bacillus firmus] | 67 | 42 |
| 138 | 33 | 25165 | 23249 | gi\|1498811 | M. jannaschii predicted coding region MJ0050 [Methanococcus jannaschii] | 67 | 45 |
| 138 | 36 | 28690 | 27362 | gnl\|PID\|e269549 | Unknown [Bacillus subtilis] | 67 | 47 |
| 144 | 4 | 3271 | 3717 | gi\|1753229 | PKCI [Borrelia burgdorferi] | 67 | 52 |
| 145 | 3 | 1435 | 2511 | gi\|1573615 | ATP-binding protein (abc) [Haemophilus influenzae] | 67 | 47 |
| 146 | 5 | 4657 | 2804 | gi\|1045034 | beta-galactosidase [Xanthomonas campestris pv. manihotis] | 67 | 51 |
| 149 | 3 | 1978 | 1367 | gi\|806536 | membrane protein [Bacillus acidopullulyticus] | 67 | 51 |
| 156 | 1 | 3 | 365 | gnl\|PID\|e265539 | ClpB-homologue [Thermus aquaticus thermophilus] | 67 | 42 |
| 158 | 15 | 14863 | 13766 | gi\|1573487 | rbs repressor (rbsR) [Haemophilus influenzae] | 67 | 40 |
| 158 | 17 | 16483 | 15959 | gi\|677850 | hypothetical protein [Staphylococcus aureus] | 67 | 51 |
| 159 | 7 | 6872 | 6006 | gi\|1303949 | YqiX [Bacillus subtilis] | 67 | 41 |
| 159 | 9 | 8103 | 7498 | gi\|1303950 | YqiY [Bacillus subtilis] | 67 | 41 |
| 165 | 11 | 9846 | 9004 | gi\|606079 | ORF_o267 [Escherichia coli] | 67 | 36 |
| 169 | 2 | 2151 | 3047 | gi\|42371 | pyruvate formate-lyase activating enzyme (AA 1-246) [Escherichia li] | 67 | 44 |
| 179 | 13 | 13648 | 14451 | gnl\|PID\|e257631 | methyltransferase [Lactococcus lactis] | 67 | 45 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 180 | 28 | 28656 | 29801 | gi\|666005 | hypothetical protein [Bacillus subtilis] | 67 | 48 |
| 194 | 6 | 2774 | 4231 | gi\|143245 | Na+/H+ antiporter [Bacillus firmus] | 67 | 41 |
| 194 | 10 | 6472 | 8259 | gi\|622991 | mannitol transport protein [Bacillus stearothermophilus] sp\|P50852\|PTMB_BACST PTS SYSTEM, MANNITOL-SPECIFIC IIBC COMPONENT EIIBC-MTL) (MANNITOL- PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE NZYME II, BC COMPONENT) (EC 2.7.1.69) (EII-MTL). | 67 | 50 |
| 204 | 5 | 1924 | 3006 | gi\|1235684 | mevalonate pyrophosphate decarboxylase [Saccharomyces cerevisiae] | 67 | 50 |
| 214 | 1 | 42 | 1196 | gi\|606013 | CG Site No. 829 [Escherichia coli] | 67 | 36 |
| 219 | 2 | 524 | 850 | gnl\|PID\|e257628 | ORF [Lactococcus lactis] | 67 | 42 |
| 223 | 15 | 13640 | 14407 | gi\|496520 | orf iota [Streptococcus pyogenes] | 67 | 54 |
| 227 | 3 | 1011 | 1892 | gi\|1070013 | protein-dependent [Bacillus subtilis] | 67 | 37 |
| 233 | 12 | 9340 | 8339 | gi\|507880 | xanthine dehydrogenase [Gallus gallus] | 67 | 50 |
| 238 | 10 | 7951 | 9183 | gi\|1653948 | hypothetical protein [Synechocystis sp.] | 67 | 45 |
| 246 | 3 | 783 | 1430 | gnl\|PID\|e233869 | hypothetical protein [Bacillus subtilis] | 67 | 47 |
| 256 | 2 | 570 | 1601 | gi\|709992 | hypothetical protein [Bacillus subtilis] | 67 | 36 |
| 266 | 2 | 1266 | 835 | gi\|963038 | ArpU [Enterococcus hirae] | 67 | 42 |
| 285 | 1 | 3 | 809 | gi\|40014 | pot. ORF 446 (aa 1-446) [Bacillus subtilis] | 67 | 53 |
| 288 | 10 | 6838 | 5801 | gi\|1651806 | hypothetical protein [Synechocystis sp.] | 67 | 45 |
| 301 | 10 | 8822 | 8562 | gi\|1303864 | YqgQ [Bacillus subtilis] | 67 | 43 |
| 312 | 5 | 2377 | 2595 | gi\|709991 | hypothetical protein [Bacillus subtilis] | 67 | 52 |
| 353 | 1 | 3 | 1472 | gi\|151259 | HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii] pir\|A44756\|A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp. | 67 | 48 |
| 359 | 2 | 984 | 439 | gi\|1773190 | similar to E. coli yhaE [Escherichia coli] | 67 | 45 |
| 359 | 3 | 2244 | 982 | gi\|1001478 | hypothetical protein [Synechocystis sp.] | 67 | 30 |
| 364 | 8 | 8469 | 7816 | gi\|496943 | ORF [Saccharomyces cerevisiae] | 67 | 50 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 386 | 12 | 6625 | 7833 | gnl\|PID\|e254644 | membrane protein [Streptococcus pneumoniae] | 67 | 36 |
| 394 | 2 | 497 | 2635 | gnl\|PID\|e255093 | hypothetical protein [Bacillus subtilis] | 67 | 45 |
| 399 | 6 | 5410 | 3971 | gi\|665994 | hypothetical protein [Bacillus subtilis] | 67 | 45 |
| 414 | 1 | 1 | 1227 | gi\|1621027 | high affinity potassium transporter [Debaryomyces occidentalis] | 67 | 40 |
| 453 | 2 | 618 | 391 | gi\|537189 | ORF_f132 [Escherichia coli] | 67 | 45 |
| 458 | 1 | 825 | 226 | gnl\|PID\|e189917 | ORF 28.5 [Escherichia coli] | 67 | 45 |
| 460 | 2 | 644 | 1387 | gi\|1502421 | 3-ketoacyl-acyl carrier protein reductase [Bacillus subtilis] | 67 | 48 |
| 460 | 4 | 2622 | 3131 | gi\|1399830 | biotin carboxyl carrier protein [Synechococcus PCC7942] | 67 | 53 |
| 474 | 1 | 1456 | 77 | gi\|495277 | histidine kinase [Streptococcus pneumoniae] | 67 | 54 |
| 488 | 6 | 3892 | 3032 | gi\|437389 | transposase [Lactococcus lactis] | 67 | 47 |
| 490 | 1 | 460 | 2 | gi\|1742830 | ORF_ID:o326#2; similar to [SwissProt Accession Number P37794] [Escherichia coli] | 67 | 43 |
| 582 | 1 | 2 | 787 | gi\|1408485 | yxdM gene product [Bacillus subtilis] | 67 | 38 |
| 629 | 2 | 1280 | 915 | gi\|1006620 | ABC transporter [Synechocystis sp.] | 67 | 50 |
| 633 | 2 | 941 | 390 | gnl\|PID\|e221400 | tex gene product [Bordetella pertussis] | 67 | 54 |
| 655 | 1 | 47 | 313 | gi\|147403 | mannose permease subunit II-P-Man [Escherichia coli] | 67 | 48 |
| 671 | 3 | 1630 | 2415 | sp\|P13226\|GALE_STRLI | UDP-GLUCOSE 4-EPIMERASE (EC 5.1.3.2) (GALACTOWALDENASE). | 67 | 52 |
| 682 | 2 | 1428 | 595 | gi\|147404 | mannose permease subunit II-M-Man [Escherichia coli] | 67 | 42 |
| 704 | 3 | 977 | 411 | gi\|467428 | unknown [Bacillus subtilis] | 67 | 45 |
| 711 | 1 | 590 | 168 | gi\|471236 | orf3 [Haemophilus influenzae] | 67 | 37 |
| 784 | 1 | 253 | 2 | gnl\|PID\|e236287 | site-specific DNA-methyltransferase [Bacillus stearothermophilus] | 67 | 44 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 907 | 1 | 209 | 3 | gi\|5119 | topoisomerase I [Schizosaccharomyces pombe] | 67 | 42 |
| 908 | 1 | 275 | 96 | gi\|1591045 | hypothetical protein (SP:P31466) [Methanococcus jannaschii] | 67 | 46 |
| 960 | 1 | 499 | 98 | gi\|405804 | transposase [Streptococcus thermophilus] | 67 | 50 |
| 963 | 1 | 259 | 2 | pir\|S34632\|S34632 | dnaJ protein homolog - human | 67 | 54 |
| 964 | 1 | 164 | 628 | bbs\|173803 | CD4+ T cell-stimulating antigen [Listeria monocytogenes, 85EO-1167, Peptide Partial, 268 aa] [Listeria monocytogenes] | 67 | 49 |
| 5 | 4 | 1438 | 2403 | gi\|1303810 | YqeT [Bacillus subtilis] | 66 | 50 |
| 7 | 1 | 24 | 1727 | gi\|145220 | alanyl-tRNA synthetase [Escherichia coli] | 66 | 50 |
| 7 | 2 | 1858 | 2646 | gi\|687599 | orfA1; transposon insertion into orfA1 impairs growth and virulence f L. monocytogenes [Listeria monocytogenes] | 66 | 58 |
| 8 | 1 | 3 | 707 | gi\|1303830 | YqfL [Bacillus subtilis] | 66 | 45 |
| 9 | 1 | 182 | 1051 | gi\|467399 | IMP dehydrogenase [Bacillus subtilis] | 66 | 51 |
| 17 | 11 | 8383 | 8598 | gi\|457336 | Pv200 [Plasmodium vivax] | 66 | 42 |
| 18 | 14 | 5903 | 6136 | gi\|294706 | trfA [Plasmid RK2] | 66 | 50 |
| 23 | 12 | 5951 | 6895 | gi\|1652472 | ethylene response sensor protein [Synechocystis sp.] | 66 | 51 |
| 23 | 17 | 11198 | 11881 | gi\|466517 | pduB [Salmonella typhimurium] | 66 | 44 |
| 23 | 19 | 12395 | 13501 | gi\|145206 | alcohol dehydrogenase (adhE) [Escherichia coli] | 66 | 47 |
| 34 | 5 | 5987 | 6232 | gi\|397360 | yNucR endo-exonuclease [Saccharomyces cerevisiae] | 66 | 46 |
| 43 | 2 | 782 | 1018 | gi\|513417 | non-structural polyprotein of pSP6-SFV4 [unidentified] | 66 | 46 |
| 43 | 5 | 3757 | 2324 | gnl\|PID\|e154145 | penicillin binding protein 4 [Staphylococcus aureus] | 66 | 44 |
| 56 | 4 | 2351 | 1662 | gi\|49272 | Asparaginase [Bacillus licheniformis] | 66 | 44 |
| 57 | 2 | 950 | 1735 | gi\|1657505 | hypothetical protein [Escherichia coli] | 66 | 46 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 57 | 4 | 3117 | 3932 | gi\|1657507 | hypothetical protein [Escherichia coli] | 66 | 41 |
| 57 | 8 | 12269 | 12646 | gi\|1622733 | orf108; unknown function [Butyrivibrio fibrisolvens] | 66 | 44 |
| 62 | 2 | 547 | 1302 | gi\|413967 | ipa-43d gene product [Bacillus subtilis] | 66 | 50 |
| 62 | 5 | 2633 | 1905 | gi\|475110 | fructokinase [Pediococcus pentosaceus] | 66 | 51 |
| 74 | 7 | 4661 | 4086 | gi\|467484 | unknown [Bacillus subtilis] | 66 | 47 |
| 81 | 18 | 13878 | 13717 | gi\|146724 | enzyme III-Man function protein (manX (ptsL)) [Escherichia coli] gi\|41976 manX gene product (AA 1-315) [Escherichia coli] | 66 | 35 |
| 94 | 17 | 20780 | 21253 | gi\|142955 | glucose dehydrogenase (EC 1.1.1.47) [Bacillus subtilis] pir\|S36090\|S36090 glucose 1-dehydrogenase (EC 1.1.1.47) - Bacillus ubtilis | 66 | 47 |
| 98 | 15 | 15165 | 14338 | gi\|147327 | transport protein [Escherichia coli] | 66 | 34 |
| 105 | 3 | 1726 | 3183 | gnl\|PID\|e205173 | orf1 gene product [Lactobacillus helveticus] | 66 | 45 |
| 110 | 17 | 15811 | 14804 | gi\|887824 | ORF_o310 [Escherichia coli] | 66 | 52 |
| 112 | 2 | 712 | 443 | gnl\|PID\|e242290 | carbamate kinase [Clostridium perfringens] | 66 | 51 |
| 123 | 1 | 1 | 540 | gi\|1573538 | H. influenzae predicted coding region HI0552 [Haemophilus influenzae] | 66 | 39 |
| 123 | 33 | 30312 | 31460 | gi\|1498930 | M. jannaschii predicted coding region MJ0158 [Methanococcus jannaschii] | 66 | 48 |
| 125 | 8 | 4914 | 4474 | gi\|1736749 | Exopolysaccharide production protein PSS. [Escherichia coli] | 66 | 54 |
| 128 | 25 | 18201 | 18878 | gnl\|PID\|e255543 | putative iron dependant repressor [Staphylococcus epidermidis] | 66 | 48 |
| 131 | 3 | 2311 | 3213 | gi\|38969 | lacF gene product [Agrobacterium radiobacter] | 66 | 37 |
| 131 | 5 | 3588 | 3394 | gi\|1303823 | YqfG [Bacillus subtilis] | 66 | 29 |
| 135 | 1 | 1214 | 45 | gi\|1498930 | M. jannaschii predicted coding region MJ0158 [Methanococcus jannaschii] | 66 | 48 |
| 135 | 10 | 7764 | 7405 | gi\|530825 | OVT1 [Onchocerca volvulus] | 66 | 47 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 144 | 13 | 12859 | 10739 | pir\|A40614\|A40614 | penicillin-binding protein pbpF - Bacillus subtilis | 66 | 47 |
| 145 | 5 | 3224 | 4063 | gi\|349531 | lipoprotein [Pasteurella haemolytica] | 66 | 45 |
| 146 | 2 | 1497 | 619 | gi\|147404 | mannose permease subunit II-M-Man [Escherichia coli] | 66 | 38 |
| 149 | 2 | 1097 | 1282 | gi\|1762962 | FemA [Staphylococcus simulans] | 66 | 38 |
| 150 | 3 | 1443 | 2417 | gnl\|PID\|e185374 | ceuE gene product [Campylobacter coli] | 66 | 46 |
| 150 | 8 | 6487 | 6903 | gi\|1377842 | unknown [Bacillus subtilis] | 66 | 43 |
| 164 | 20 | 21846 | 22646 | gi\|1279769 | FdhC [Methanobacterium thermoformicicum] | 66 | 57 |
| 164 | 25 | 24555 | 25688 | pir\|A43577\|A43577 | regulatory protein pfoR - Clostridium perfringens | 66 | 47 |
| 178 | 1 | 383 | 3 | gi\|763052 | integrase [Bacteriophage T270] | 66 | 47 |
| 195 | 19 | 8698 | 8516 | bbs\|169008 | homeobox gene [Drosophila sp.] | 66 | 55 |
| 207 | 1 | 166 | 1554 | gi\|619724 | MgtE [Bacillus firmus] | 66 | 39 |
| 207 | 3 | 2312 | 2010 | gi\|1204258 | soluble protein [Escherichia coli] | 66 | 44 |
| 211 | 3 | 1523 | 1729 | gi\|289932 | MHC class II beta chain [Cyphotilapia frontosa] | 66 | 66 |
| 213 | 3 | 1811 | 2308 | gi\|153045 | prolipoprotein signal peptidase [Staphylococcus aureus] pir\|S20433\|S20433 lsp protein - Staphylococcus aureus sp\|P31024\|LSPA_STAAU LIPOPROTEIN SIGNAL PEPTIDASE (EC 3.4.23.36) PROLIPOPROTEIN SIGNAL PEPTIDASE) (SIGNAL PEPTIDASE II) (SPASE II). | 66 | 40 |
| 221 | 7 | 2524 | 3468 | gi\|1353527 | ORF10 [Bacteriophage r1t] | 66 | 44 |
| 222 | 13 | 8272 | 8988 | gi\|466719 | No definition line found [Escherichia coli] | 66 | 48 |
| 223 | 18 | 15210 | 15971 | gi\|496520 | orf iota [Streptococcus pyogenes] | 66 | 57 |
| 232 | 5 | 3494 | 2715 | gi\|142706 | comG1 gene product [Bacillus subtilis] | 66 | 41 |
| 235 | 3 | 1774 | 734 | gi\|580897 | OppB gene product [Bacillus subtilis] | 66 | 47 |
| 244 | 2 | 906 | 1520 | gi\|15354 | ORF 55.9 [Bacteriophage T4] | 66 | 46 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 259 | 3 | 2355 | 1867 | gi\|56312 | Gephyrin [Rattus norvegicus] | 66 | 55 |
| 271 | 1 | 1 | 675 | gi\|1574748 | tRNA pseudouridine 55 synthase (truB) [Haemophilus influenzae] | 66 | 53 |
| 277 | 1 | 1 | 927 | gi\|1303799 | YqeN [Bacillus subtilis] | 66 | 45 |
| 291 | 5 | 4587 | 3547 | gnl\|PID\|e257609 | sugar-binding transport protein [Anaerocellum thermophilum] | 66 | 46 |
| 292 | 25 | 20451 | 19912 | gi\|1649035 | high-affinity periplasmic glutamine binding protein [Salmonella typhimurium] | 66 | 50 |
| 300 | 1 | 2302 | 77 | gi\|289262 | comE ORF3 [Bacillus subtilis] | 66 | 46 |
| 301 | 4 | 4290 | 3265 | sp\|P13226\|GALE_STRLI | UDP-GLUCOSE 4-EPIMERASE (EC 5.1.3.2) (GALACTOWALDENASE). | 66 | 51 |
| 301 | 5 | 4516 | 4689 | gnl\|PID\|e212164 | PSII, protein N [Odontella sinensis] | 66 | 58 |
| 314 | 1 | 360 | 4 | gi\|467452 | unknown [Bacillus subtilis] | 66 | 43 |
| 315 | 4 | 2559 | 2209 | gi\|1653498 | ABC transporter [Synechocystis sp.] | 66 | 44 |
| 320 | 3 | 2406 | 1081 | gnl\|PID\|e250352 | unknown [Mycobacterium tuberculosis] | 66 | 35 |
| 332 | 2 | 157 | 921 | gi\|1303875 | YqhB [Bacillus subtilis] | 66 | 44 |
| 334 | 2 | 1001 | 3076 | gi\|1651660 | DNA ligase [Synechocystis sp.] | 66 | 48 |
| 338 | 1 | 2 | 616 | gi\|845686 | ORF-27 [Staphylococcus aureus] | 66 | 54 |
| 338 | 7 | 5011 | 5496 | gi\|912476 | No definition line found [Escherichia coli] | 66 | 48 |
| 341 | 5 | 1935 | 3107 | gi\|142538 | aspartate aminotransferase [Bacillus sp.] | 66 | 44 |
| 343 | 3 | 2548 | 2045 | gnl\|PID\|e289147 | similar to single strand binding protein [Bacillus subtilis] | 66 | 44 |
| 345 | 20 | 22093 | 22461 | gi\|1657795 | dihydroneopterin aldolase [Methylobacterium extorquens] | 66 | 45 |
| 353 | 3 | 2621 | 2379 | gnl\|PID\|e257628 | ORF [Lactococcus lactis] | 66 | 52 |
| 365 | 4 | 5117 | 4779 | gi\|1742868 | Mutator MutT protein (7,8-dihydro-8-oxoguanine-triphosphatase) (8-oxo-dgtpase) (EC 3.6.1.-) (DGTP pyrophosphohydrolase). [Escherichia coli] | 66 | 54 |
| 376 | 1 | 3 | 1076 | gi\|1778517 | glycerol dehydrogenase homolog | 66 | 45 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | [Escherichia coli] | | |
| 394 | 7 | 5980 | 5648 | gi\|486358 | ORF YKL202w [Saccharomyces cerevisiae] | 66 | 38 |
| 421 | 4 | 1469 | 2539 | gi\|606375 | ORF_f345 [Escherichia coli] | 66 | 48 |
| 475 | 6 | 3978 | 3763 | gi\|532547 | ORF14 [Enterococcus faecalis] | 66 | 48 |
| 491 | 8 | 7710 | 7081 | gi\|1000453 | TreR [Bacillus subtilis] | 66 | 49 |
| 526 | 1 | 392 | 3 | gi\|1750125 | xylulose kinase [Bacillus subtilis] | 66 | 49 |
| 552 | 6 | 6147 | 5917 | gi\|1432152 | PTS antiterminator [Klebsiella oxytoca] | 66 | 37 |
| 571 | 2 | 560 | 1153 | gi\|1773132 | multidrug resistance-like ATP-binding protein Mdl [Escherichia coli] | 66 | 38 |
| 575 | 3 | 1075 | 539 | gi\|1651722 | guanylate kinase [Synechocystis sp.] | 66 | 48 |
| 608 | 2 | 631 | 113 | gi\|1213334 | OrfX; hypothetical 22.5 KD protein downstream of type IV prepilin leader peptidase gene; Method: conceptual translation supplied by author [Vibrio vulnificus] | 66 | 41 |
| 640 | 1 | 877 | 2 | sp\|P50487\|YCPX_CLOPE | HYPOTHETICAL PROTEIN IN CPE 5'REGION (FRAGMENT). | 66 | 36 |
| 734 | 1 | 2 | 343 | gi\|1653602 | hypothetical protein [Synechocystis sp.] | 66 | 43 |
| 802 | 1 | 2 | 292 | gnl\|PID\|e280516 | voltage-gated sodium channel [Mus musculus] | 66 | 58 |
| 812 | 2 | 343 | 531 | gi\|511075 | ORF2 [Streptococcus agalactiae] | 66 | 51 |
| 823 | 1 | 1 | 393 | gi\|1303843 | YqfV [Bacillus subtilis] | 66 | 42 |
| 891 | 1 | 82 | 402 | gi\|567769 | ORF5; predicted protein shows similarity to ATP-binding transport roteins AmiE and AmiF of Streptococcus pneumoniae; disruptuion of RF5 leads to aminopterin resistance [Streptococcus parasanguis] | 66 | 52 |
| 5 | 6 | 2630 | 3154 | gi\|1303811 | YqeU [Bacillus subtilis] | 65 | 50 |
| 6 | 1 | 2 | 628 | gi\|1742303 | Acyl carrier protein phosphodiesterase (ACP phosphodiesterase) (fragment). [Escherichia coli] | 65 | 43 |
| 18 | 6 | 3360 | 2518 | gi\|601880 | rep protein [Bacillus borstelensis] | 65 | 40 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 21 | 11 | 7933 | 7706 | gi\|1500521 | M. jannaschii predicted coding region MJ1623 [Methanococcus jannaschii] | 65 | 32 |
| 23 | 20 | 13459 | 13881 | gi\|488430 | alcohol dehydrogenase 2 [Entamoeba histolytica] | 65 | 43 |
| 23 | 25 | 15987 | 16178 | gnl\|PID\|e248966 | F32D8.5 [Caenorhabditis elegans] | 65 | 50 |
| 27 | 2 | 526 | 302 | gi\|1001644 | regulatory components of sensory transduction system [Synechocystis sp.] | 65 | 44 |
| 29 | 9 | 6770 | 5727 | sp\|P36672\|PTTB_ECOLI | PTS SYSTEM, TREHALOSE-SPECIFIC IIBC COMPONENT (EIIBC-TRE) (TREHALOSE- PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT) (EC 2.7.1.69) (EII-TRE). | 65 | 45 |
| 31 | 5 | 4611 | 5207 | gi\|171625 | guanylate kinase [Saccharomyces cerevisiae] | 65 | 39 |
| 32 | 7 | 4085 | 3915 | gi\|150158 | 29 kD protein [Mycoplasma genitalium] | 65 | 51 |
| 33 | 8 | 7396 | 7638 | gi\|1573421 | protein translocation protein, low temperature (secG) [Haemophilus influenzae] | 65 | 26 |
| 35 | 1 | 2 | 499 | gi\|1737500 | transcription antiterminator [Bacillus stearothermophilus] | 65 | 40 |
| 45 | 6 | 2537 | 3037 | gi\|511455 | unknown [Coxiella burnetii] | 65 | 37 |
| 46 | 3 | 1028 | 2254 | gi\|1001642 | dGTP triphosphohydrolase [Synechocystis sp.] | 65 | 43 |
| 47 | 12 | 14524 | 14264 | gi\|150209 | ORF 1 [Mycoplasma mycoides] | 65 | 34 |
| 50 | 3 | 2866 | 2051 | gi\|1303830 | YqfL [Bacillus subtilis] | 65 | 40 |
| 57 | 11 | 12955 | 13332 | gnl\|PID\|e254999 | phenylalany-tRNA synthetase beta subunit [Bacillus subtilis] | 65 | 51 |
| 62 | 1 | 2 | 484 | gi\|1573470 | H. influenzae predicted coding region HI0491 [Haemophilus influenzae] | 65 | 57 |
| 68 | 1 | 49 | 282 | gi\|1573250 | aspartate aminotransferase (aspC) [Haemophilus influenzae] | 65 | 52 |
| 72 | 2 | 567 | 1325 | gi\|466645 | alternate name yhiD [Escherichia coli] | 65 | 40 |
| 81 | 5 | 3711 | 2938 | gi\|1732200 | PTS permease for mannose subunit IIPMan | 65 | 43 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | [Vibrio furnissii] | | |
| 83 | 18 | 12506 | 12745 | pir\|D64042\|D64042 | ribosomal-protein-alanine acetyltransferase (rimI) homolog - Haemophilus influenzae (strain Rd KW20) | 65 | 50 |
| 100 | 38 | 28229 | 28032 | gi\|183075 | glial fibrillary acidic protein [Homo sapiens] | 65 | 43 |
| 105 | 1 | 912 | 106 | pir\|S15248\|YQBZCD | fimC protein - Dichelobacter nodosus (serotype D) | 65 | 46 |
| 106 | 5 | 6097 | 5102 | gi\|1143204 | ORF2; Method: conceptual translation supplied by author [Shigella sonnei] | 65 | 44 |
| 109 | 3 | 1165 | 899 | gi\|1573390 | hypothetical [Haemophilus influenzae] | 65 | 55 |
| 110 | 7 | 5579 | 4257 | pir\|B44514\|B44514 | hypothetical protein 1 (vnfA 5' region) - Azotobacter vinelandii | 65 | 43 |
| 120 | 3 | 1249 | 1632 | sp\|P54746\|YBGB_ECOLI | HYPOTHETICAL PROTEIN IN HRSA 3'REGION (FRAGMENT). | 65 | 48 |
| 122 | 2 | 896 | 1654 | gi\|1335913 | unknown [Erysipelothrix rhusiopathiae] | 65 | 48 |
| 145 | 4 | 2509 | 3210 | gi\|1208965 | hypothetical 23.3 kd protein [Escherichia coli] | 65 | 40 |
| 149 | 7 | 4407 | 3502 | gi\|145173 | 35 kDa protein [Escherichia coli] | 65 | 46 |
| 154 | 8 | 5738 | 4926 | gi\|405804 | transposase [Streptococcus thermophilus] | 65 | 47 |
| 155 | 1 | 306 | 512 | gi\|285627 | E.coli SecE homologous protein [Bacillus subtilis] pir\|S39858\|S39858 secE protein homolog - Bacillus subtilis sp\|Q06799\|SECE_BACSU PREPROTEIN TRANSLOCASE SECE SUBUNIT. | 65 | 48 |
| 158 | 1 | 150 | 1103 | gi\|289272 | ferrichrome-binding protein [Bacillus subtilis] | 65 | 40 |
| 158 | 16 | 14885 | 15946 | gi\|467172 | add; L308_C2_206 [Mycobacterium leprae] | 65 | 36 |
| 173 | 4 | 2103 | 2912 | gnl\|PID\|e254877 | unknown [Mycobacterium tuberculosis] | 65 | 41 |
| 173 | 12 | 9749 | 9054 | gi\|1652864 | hypothetical protein [Synechocystis sp.] | 65 | 50 |
| 179 | 16 | 15674 | 17035 | gi\|1171125 | thioredoxin reductase [Clostridium litorale] | 65 | 41 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 180 | 26 | 26911 | 28266 | sp\|P13692\|P54_ENTFC | P54 PROTEIN PRECURSOR. | 65 | 39 |
| 193 | 6 | 2893 | 3795 | gi\|39787 | adaA [Bacillus subtilis] | 65 | 45 |
| 194 | 5 | 1843 | 2238 | gi\|47394 | 5-oxoprolyl-peptidase [Streptococcus pyogenes] | 65 | 48 |
| 199 | 1 | 894 | 82 | gi\|1591118 | nitrate transport ATP-binding protein [Methanococcus jannaschii] | 65 | 46 |
| 200 | 24 | 13441 | 13136 | gi\|144926 | toxin A [Clostridium difficile] | 65 | 39 |
| 202 | 3 | 2925 | 1846 | gi\|413968 | ipa-44d gene product [Bacillus subtilis] | 65 | 46 |
| 203 | 1 | 797 | 3 | gi\|1377832 | unknown [Bacillus subtilis] | 65 | 45 |
| 204 | 3 | 1065 | 1472 | gi\|1008996 | unknown [Schizosaccharomyces pombe] | 65 | 51 |
| 205 | 4 | 1029 | 1685 | gi\|148989 | truncated tetracycline resistance repressor (non-functional) Haemophilus parainfluenzae] | 65 | 42 |
| 206 | 8 | 5037 | 4807 | pir\|D60110\|D60110 | repetitive protein antigen 3 - Trypanosoma cruzi (fragment) | 65 | 41 |
| 217 | 1 | 411 | 4 | gi\|1146181 | putative [Bacillus subtilis] | 65 | 43 |
| 217 | 4 | 1092 | 3065 | gi\|984229 | penicillin-binding protein 1a [Streptococcus pneumoniae] | 65 | 48 |
| 223 | 27 | 23445 | 23879 | gnl\|PID\|e269486 | Unknown [Bacillus subtilis] | 65 | 47 |
| 225 | 6 | 5138 | 3984 | gi\|39956 | IIGlc [Bacillus subtilis] | 65 | 47 |
| 229 | 5 | 5528 | 5130 | gi\|1303914 | YqhY [Bacillus subtilis] | 65 | 33 |
| 229 | 10 | 10697 | 8517 | gnl\|PID\|e266933 | unknown [Mycobacterium tuberculosis] | 65 | 46 |
| 233 | 3 | 2413 | 1526 | gi\|887825 | ORF_f541 [Escherichia coli] | 65 | 46 |
| 236 | 4 | 6975 | 4789 | gi\|405863 | yohA [Escherichia coli] | 65 | 43 |
| 237 | 4 | 1460 | 1816 | gi\|305080 | myosin heavy chain [Entamoeba histolytica] | 65 | 42 |
| 238 | 24 | 21690 | 23228 | gi\|305008 | rhamnulokinase [Escherichia coli] | 65 | 49 |
| 242 | 3 | 2192 | 3280 | gnl\|PID\|e221269 | tail protein [Bacteriophage CP-1] | 65 | 37 |
| 244 | 6 | 5172 | 4228 | gi\|1653197 | hypothetical protein [Synechocystis sp.] | 65 | 51 |
| 259 | 5 | 3684 | 2779 | gi\|559900 | F49E2.1 [Caenorhabditis elegans] | 65 | 39 |
| 259 | 6 | 4243 | 3749 | gi\|1743887 | molybdopterin cofactor biosynthesis enzyme | 65 | 50 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | [Bradyrhizobium japonicum] | | |
| 260 | 1 | 140 | 478 | gi\|895748 | putative cellobiose phosphotransferase enzyme II' [Bacillus ubtilis] | 65 | 55 |
| 269 | 6 | 4113 | 3907 | gi\|1303792 | YqeK [Bacillus subtilis] | 65 | 39 |
| 271 | 12 | 7731 | 6772 | gi\|1657534 | cyn operon transcriptional activator [Escherichia coli] | 65 | 45 |
| 275 | 9 | 6413 | 5361 | gi\|1773132 | multidrug resistance-like ATP-binding protein Mdl [Escherichia coli] | 65 | 48 |
| 276 | 4 | 1813 | 1583 | gi\|1504014 | similar to myosin heavy chain: Containing ATP/GTP-binding site motif A(P-loop) [Homo sapiens] | 65 | 34 |
| 279 | 14 | 14254 | 10625 | gi\|1237015 | ORF4 [Bacillus subtilis] | 65 | 45 |
| 281 | 2 | 692 | 1279 | gi\|1303962 | YqjK [Bacillus subtilis] | 65 | 50 |
| 295 | 5 | 2279 | 3388 | gi\|436965 | [malA] gene products [Bacillus stearothermophilus] pir\|S43914\|S43914 hypothetical protein 1 - Bacillus tearothermophilus | 65 | 41 |
| 298 | 1 | 63 | 1142 | gi\|928834 | integrase [Lactococcus lactis phage BK5-T] | 65 | 44 |
| 301 | 8 | 7592 | 7176 | gi\|1303893 | YqhL [Bacillus subtilis] | 65 | 50 |
| 311 | 3 | 4658 | 5701 | gnl\|PID\|e221269 | tail protein [Bacteriophage CP-1] | 65 | 40 |
| 326 | 1 | 2 | 247 | gi\|466520 | pocR [Salmonella typhimurium] | 65 | 38 |
| 329 | 1 | 789 | 523 | gi\|1303895 | YqhN [Bacillus subtilis] | 65 | 36 |
| 345 | 5 | 3363 | 3641 | gi\|895749 | putative cellobiose phosphotransferase enzyme II'' [Bacillus ubtilis] | 65 | 51 |
| 369 | 3 | 1635 | 1207 | gi\|1480429 | putative transcriptional regulator [Bacillus stearothermophilus] | 65 | 45 |
| 373 | 2 | 815 | 1630 | gi\|1277032 | unknown [Bacillus subtilis] | 65 | 41 |
| 379 | 9 | 11301 | 8275 | gi\|887828 | was o492p and o826p before splice [Escherichia coli] | 65 | 49 |
| 386 | 13 | 7903 | 8145 | gnl\|PID\|e217382 | M7.9 [Caenorhabditis elegans] | 65 | 39 |
| 395 | 4 | 1028 | 1231 | gi\|1592033 | M. jannaschii predicted coding region | 65 | 30 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | MJ1387 [Methanococcus jannaschii] | | |
| 396 | 3 | 1000 | 1272 | gi\|1045900 | hypothetical protein (GB:L09228_17) [Mycoplasma genitalium] | 65 | 44 |
| 422 | 3 | 2050 | 1262 | gi\|405907 | yejD [Escherichia coli] | 65 | 50 |
| 438 | 1 | 44 | 358 | gi\|530798 | LysB [Bacteriophage phi-LC3] | 65 | 39 |
| 460 | 1 | 119 | 646 | gi\|1502420 | malonyl-CoA:Acyl carrier protein transacylase [Bacillus subtilis] | 65 | 46 |
| 463 | 1 | 870 | 121 | gi\|1651917 | tRNA(m1G37)methyltransferase [Synechocystis sp.] | 65 | 47 |
| 468 | 1 | 2 | 823 | gi\|216457 | ORF [Escherichia coli] | 65 | 46 |
| 470 | 1 | 34 | 816 | gi\|530798 | LysB [Bacteriophage phi-LC3] | 65 | 47 |
| 476 | 1 | 21 | 830 | gi\|1006591 | cation-transporting ATPase PacL [Synechocystis sp.] | 65 | 46 |
| 510 | 7 | 4875 | 6092 | gi\|143150 | levR [Bacillus subtilis] | 65 | 46 |
| 565 | 2 | 686 | 339 | gi\|143833 | PBSX repressor [Bacillus subtilis] | 65 | 51 |
| 566 | 2 | 198 | 743 | gi\|496501 | RepS [Streptococcus pyogenes] | 65 | 34 |
| 604 | 5 | 1875 | 2078 | gi\|1590997 | M. jannaschii predicted coding region MJ0272 [Methanococcus jannaschii] | 65 | 49 |
| 608 | 1 | 194 | 3 | gnl\|PID\|e290940 | unknown [Mycobacterium tuberculosis] | 65 | 35 |
| 648 | 1 | 60 | 953 | gi\|1591145 | hypothetical protein (HI0902) [Methanococcus jannaschii] | 65 | 31 |
| 657 | 4 | 2531 | 1620 | gi\|1500015 | amidase [Methanococcus jannaschii] | 65 | 46 |
| 691 | 1 | 2 | 718 | gnl\|PID\|e248400 | orfRM1 gene product [Bacillus subtilis] | 65 | 48 |
| 704 | 2 | 474 | 175 | gi\|467428 | unknown [Bacillus subtilis] | 65 | 50 |
| 758 | 2 | 408 | 683 | gi\|451201 | ORF1 [Bacillus subtilis] | 65 | 44 |
| 778 | 1 | 833 | 3 | gi\|410137 | ORFX13 [Bacillus subtilis] | 65 | 40 |
| 793 | 1 | 1 | 564 | gi\|912436 | oligo-1,6-glucosidase [Bacillus thermoglucosidasius] pir\|A41707\|A41707 oligo-1,6-glucosidase (EC 3.2.1.10) - Bacillus hermoglucosidasius | 65 | 40 |
| 827 | 1 | 364 | 2 | gi\|852076 | MrgA [Bacillus subtilis] | 65 | 33 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 856 | 1 | 209 | 3 | gi\|1575605 | 4-methyl-5-nitrocatechol oxygenase [Burkholderia sp.] | 65 | 45 |
| 890 | 1 | 966 | 745 | pir\|A44803\|A44803 | pG1 protein - human (fragment) | 65 | 63 |
| 4 | 1 | 2 | 958 | gnl\|PID\|e265530 | yorfE [Streptococcus pneumoniae] | 64 | 43 |
| 5 | 8 | 4212 | 5579 | gi\|407881 | stringent response-like protein [Streptococcus equisimilis] pir\|S39975\|S39975 stringent response-like protein - Streptococcus quisimilis | 64 | 47 |
| 8 | 4 | 4047 | 3304 | gi\|1573150 | dihydrolipoamide acetyltransferase (acoC) [Haemophilus influenzae] | 64 | 37 |
| 17 | 14 | 11709 | 10393 | gi\|155109 | ORF 1B [Thermus aquaticus thermophilus] | 64 | 37 |
| 19 | 12 | 6499 | 6801 | gi\|1303755 | YqbO [Bacillus subtilis] | 64 | 32 |
| 23 | 1 | 1 | 303 | gi\|1022963 | dextransucrase [Leuconostoc mesenteroides] | 64 | 50 |
| 28 | 4 | 7059 | 6505 | gi\|1568609 | 18kDA protein [Streptococcus pneumoniae] | 64 | 45 |
| 31 | 3 | 1316 | 2986 | gi\|1100076 | PTS-dependent enzyme II [Clostridium longisporum] | 64 | 47 |
| 47 | 2 | 2665 | 3408 | gi\|1742154 | Phosphoglycolate phosphatase (EC 3.1.3.18). [Escherichia coli] | 64 | 52 |
| 48 | 2 | 1699 | 1310 | gi\|142702 | A competence protein 2 [Bacillus subtilis] | 64 | 41 |
| 54 | 8 | 2750 | 2352 | gi\|951052 | ORF9, putative [Streptococcus pneumoniae] | 64 | 31 |
| 57 | 15 | 18035 | 17274 | gi\|1183886 | integral membrane protein [Bacillus subtilis] | 64 | 40 |
| 62 | 4 | 1968 | 1699 | gi\|475110 | fructokinase [Pediococcus pentosaceus] | 64 | 52 |
| 100 | 42 | 29329 | 29039 | gi\|951048 | excisionase [Streptococcus pneumoniae] | 64 | 37 |
| 102 | 4 | 3726 | 4805 | gi\|215331 | morphogenesis protein [Bacteriophage phi-29] | 64 | 43 |
| 106 | 3 | 3296 | 2439 | gi\|1303930 | YqiK [Bacillus subtilis] | 64 | 44 |
| 123 | 12 | 12960 | 11314 | sp\|P37047\|YAEG_ECOLI | HYPOTHETICAL 44.3 KD PROTEIN IN HTRA-DAPD INTERGENIC REGION. | 64 | 40 |
| 128 | 2 | 1285 | 1614 | gi\|143961 | pyruvate phosphate dikinase [Clostridium symbiosum] pir\|A36231\|KIQAPO | 64 | 52 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | pyruvate,orthophosphate dikinase (EC 2.7.9.1) - lostridium symbiosum | | |
| 128 | 8 | 6178 | 4757 | gi\|40665 | beta-glucosidase [Clostridium thermocellum] | 64 | 41 |
| 133 | 2 | 1748 | 2248 | gi\|1591027 | ferripyochelin binding protein [Methanococcus jannaschii] | 64 | 46 |
| 150 | 1 | 35 | 673 | gnl\|PID\|e185372 | ceuC gene product [Campylobacter coli] | 64 | 38 |
| 158 | 6 | 6038 | 5040 | gi\|1045801 | hypothetical protein (SP:P32720) [Mycoplasma genitalium] | 64 | 35 |
| 164 | 7 | 3620 | 4903 | gnl\|PID\|e283116 | unknown similar to quinolon resistance protein NorA [Bacillus subtilis] | 64 | 41 |
| 171 | 11 | 10107 | 10784 | gi\|1591668 | phosphate transport system regulatory protein [Methanococcus jannaschii] | 64 | 40 |
| 179 | 4 | 4826 | 6373 | gi\|149535 | D-alanine activating enzyme [Lactobacillus casei] | 64 | 51 |
| 181 | 4 | 2251 | 1364 | gi\|671632 | unknown [Staphylococcus aureus] | 64 | 38 |
| 190 | 11 | 11302 | 10355 | gi\|599850 | orf1 gene product [Lactobacillus sake] | 64 | 33 |
| 195 | 37 | 15344 | 16033 | gi\|1736499 | Lysostaphin precursor (EC 3.5.1.-). [Escherichia coli] | 64 | 49 |
| 199 | 4 | 4000 | 5631 | gi\|746574 | similar to M. musculus transport system membrane protein, Nramp PIR:A40739) and S. cerevisiae SMF1 protein (PIR:A45154) Caenorhabditis elegans] | 64 | 37 |
| 202 | 1 | 1 | 1560 | gi\|309662 | pheromone binding protein [Plasmid pCF10] | 64 | 45 |
| 204 | 7 | 3000 | 4115 | gi\|1591731 | melvalonate kinase [Methanococcus jannaschii] | 64 | 41 |
| 208 | 1 | 308 | 1090 | gi\|473821 | 'tetrahydrodipicolinate N-succinyltransferase' [Escherichia coli] gi\|1552743 tetrahydrodipicolinate N-succinyltransferase Escherichia coli] | 64 | 42 |
| 216 | 9 | 6501 | 6698 | gi\|47373 | 7 kDa protein [Streptococcus pneumoniae] | 64 | 35 |
| 221 | 18 | 8268 | 8513 | gi\|1389837 | complement regulatory protein [Trypanosoma | 64 | 28 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | cruzi] | | |
| 231 | 4 | 2964 | 2632 | gnl\|PID\|e279941 | muconate cycloisomerase [Rhodococcus erythropolis] | 64 | 37 |
| 234 | 2 | 751 | 302 | gnl\|PID\|e194709 | N-terminal part of a protein of unknown function [Chlamydia psittaci] | 64 | 42 |
| 238 | 18 | 15580 | 16392 | gi\|537108 | ORF_f254 [Escherichia coli] | 64 | 44 |
| 245 | 1 | 14 | 868 | gi\|153247 | endo-beta-N-acetylglucosaminidase H [Streptomyces plicatus] pir\|A00903\|RBSMHP mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) H precursor - treptomyces plicatus | 64 | 51 |
| 272 | 2 | 584 | 1144 | gi\|580781 | signal peptidase [Bacillus licheniformis] | 64 | 47 |
| 281 | 5 | 2659 | 5019 | gi\|147550 | recJ [Escherichia coli] | 64 | 46 |
| 290 | 12 | 9496 | 10371 | gi\|45713 | P.putida genes rpmH, rnpA, 9k, 60k, 50k, gidA, gidB, uncI and uncB seudomonas putida] | 64 | 42 |
| 298 | 4 | 4029 | 3466 | gi\|147780 | rts gene product [Escherichia coli] | 64 | 43 |
| 301 | 20 | 16216 | 15977 | gi\|170482 | prosystemin [Solanum lycopersicum] | 64 | 57 |
| 301 | 21 | 17732 | 17391 | gi\|405804 | transposase [Streptococcus thermophilus] | 64 | 52 |
| 307 | 1 | 198 | 1964 | gi\|1255196 | BSMA [Bacillus stearothermophilus] | 64 | 48 |
| 320 | 5 | 3441 | 3070 | gi\|972900 | ArtP [Haemophilus influenzae] | 64 | 38 |
| 341 | 9 | 7690 | 6413 | gi\|1161380 | IcaA [Staphylococcus epidermidis] | 64 | 30 |
| 345 | 6 | 3589 | 4848 | gi\|902932 | L-methionine gamma-lyase [Pseudomonas putida] | 64 | 45 |
| 348 | 1 | 453 | 22 | gi\|1591957 | M. jannaschii predicted coding region MJ1318 [Methanococcus jannaschii] | 64 | 32 |
| 350 | 2 | 1372 | 1830 | gnl\|PID\|e289141 | similar to hydroxymyristoyl-(acyl carrier protein) dehydratase [Bacillus subtilis] | 64 | 44 |
| 351 | 7 | 3291 | 2917 | gi\|49013 | dTDP-dihydrostreptose synthase [Streptomyces griseus] ir\|S18618\|SYSMPG dTDP-dihydrostreptose synthase - Streptomyces iseus | 64 | 46 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 352 | 2 | 780 | 1028 | gi\|173431 | H+-ATPase [Schizosaccharomyces pombe] | 64 | 38 |
| 386 | 10 | 5952 | 6161 | gnl\|PID\|e243284 | ORF YGL056c [Saccharomyces cerevisiae] | 64 | 50 |
| 398 | 2 | 1233 | 1808 | gi\|147920 | 3-methyladenine-DNA glycosylase I (tag) [Escherichia coli] | 64 | 47 |
| 399 | 12 | 8761 | 9159 | gi\|1778534 | HI0024 homolog [Escherichia coli] | 64 | 40 |
| 409 | 1 | 657 | 1607 | gi\|1773157 | ferrochelatase [Escherichia coli] | 64 | 41 |
| 446 | 1 | 266 | 775 | gi\|563845 | orf gene product [Bacillus circulans] | 64 | 53 |
| 462 | 4 | 1714 | 1959 | gi\|169461 | serine proteinase inhibitor [Populus trichocarpa x Populus eltoides] | 64 | 50 |
| 466 | 6 | 5621 | 8539 | gi\|143150 | levR [Bacillus subtilis] | 64 | 43 |
| 501 | 2 | 891 | 1469 | gi\|467109 | rim; 30S Ribosomal protein S18 alanine acetyltransferase; 229_C1_170 [Mycobacterium leprae] | 64 | 44 |
| 512 | 1 | 1 | 279 | gi\|1651948 | hypothetical protein [Synechocystis sp.] | 64 | 35 |
| 516 | 1 | 466 | 2 | gi\|155027 | 6'-N-acetyltransferase [Transposon Tn2426] | 64 | 35 |
| 516 | 2 | 556 | 759 | gi\|1653387 | nitrogen assimilation regulatory protein [Synechocystis sp.] | 64 | 58 |
| 523 | 2 | 904 | 662 | gi\|159464 | armadillo protein [Musca domestica] | 64 | 45 |
| 537 | 2 | 1083 | 844 | gi\|929966 | truncated ORFB due to a basepair deletion; similar to B. anthracis terneR element ORFB [Bacillus anthracis] | 64 | 42 |
| 549 | 1 | 309 | 4 | gi\|1279769 | FdhC [Methanobacterium thermoformicicum] | 64 | 48 |
| 552 | 4 | 5960 | 3945 | gi\|1100076 | PTS-dependent enzyme II [Clostridium longisporum] | 64 | 47 |
| 556 | 1 | 3 | 224 | gi\|727437 | putative 37-kDa protein [Lactococcus lactis] | 64 | 49 |
| 557 | 2 | 767 | 1120 | gnl\|PID\|e257629 | transcription factor [Lactococcus lactis] | 64 | 44 |
| 602 | 1 | 428 | 156 | gi\|520407 | orf2; GTG start codon [Bacillus thuringiensis] | 64 | 50 |
| 603 | 1 | 1 | 165 | gi\|1621445 | sporulation protein Cse15 [Bacillus subtilis] | 64 | 32 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 626 | 1 | 3 | 992 | gi\|1574715 | thioredoxin reductase (trxB) [Haemophilus influenzae] | 64 | 40 |
| 628 | 2 | 240 | 446 | gi\|1165281 | Smg [Borrelia burgdorferi] | 64 | 41 |
| 723 | 1 | 23 | 829 | gi\|1620648 | surface protein Rib [Streptococcus agalactiae] | 64 | 50 |
| 739 | 1 | 4 | 378 | gi\|143835 | PBSX repressor [Bacillus subtilis] | 64 | 37 |
| 748 | 1 | 139 | 765 | gi\|498816 | ORF7; homology to regions 4.1 and 4.2 of sigma factors [Bacillus ubtilis] | 64 | 35 |
| 758 | 1 | 3 | 410 | gi\|451201 | ORF1 [Bacillus subtilis] | 64 | 34 |
| 808 | 1 | 368 | 3 | gi\|142833 | ORF2 [Bacillus subtilis] | 64 | 47 |
| 818 | 2 | 415 | 663 | gi\|854020 | U41, major DNA binding protein [Human herpesvirus 6] | 64 | 40 |
| 906 | 1 | 2 | 433 | gi\|1303865 | YqgR [Bacillus subtilis] | 64 | 44 |
| 17 | 28 | 28175 | 27612 | gi\|151824 | ORF5 [Plasmid R46] | 63 | 34 |
| 19 | 18 | 9546 | 9722 | gi\|288661 | ORF5 product [Bacteriophage P2] | 63 | 45 |
| 39 | 5 | 1841 | 2329 | gi\|1573292 | hypothetical [Haemophilus influenzae] | 63 | 47 |
| 41 | 1 | 1531 | 2 | gi\|580896 | nodB protein (aa 1-219) [Bradyrhizobium sp.] | 63 | 43 |
| 55 | 10 | 5052 | 6410 | gi\|1303917 | YqiB [Bacillus subtilis] | 63 | 42 |
| 80 | 2 | 1852 | 824 | gi\|38722 | precursor (aa -20 to 381) [Acinetobacter calcoaceticus] ir\|A29277\|A29277 aldose 1-epimerase (EC 5.1.3.3) - Acinetobacter lcoaceticus | 63 | 42 |
| 81 | 10 | 6724 | 6221 | gi\|1591234 | hypothetical protein (SP:P42297) [Methanococcus jannaschii] | 63 | 40 |
| 81 | 14 | 9175 | 10848 | gi\|309662 | pheromone binding protein [Plasmid pCF10] | 63 | 44 |
| 86 | 1 | 2 | 1006 | gi\|143316 | [gap] gene products [Bacillus megaterium] | 63 | 43 |
| 89 | 13 | 12929 | 12639 | gi\|1377841 | unknown [Bacillus subtilis] | 63 | 44 |
| 98 | 14 | 14365 | 13502 | sp\|P45169\|POTC_HAEIN | SPERMIDINE/PUTRESCINE TRANSPORT SYSTEM PERMEASE PROTEIN POTC. | 63 | 37 |
| 100 | 24 | 20444 | 17985 | gi\|563258 | virulence-associated protein E | 63 | 44 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | [Dichelobacter nodosus] | | |
| 102 | 2 | 2441 | 2599 | gi\|1619835 | MOB [Bacillus thuringiensis israelensis] | 63 | 28 |
| 110 | 22 | 19725 | 20705 | gi\|1763011 | lysophospholipase homolog [Homo sapiens] | 63 | 48 |
| 115 | 1 | 481 | 92 | gi\|467360 | unknown [Bacillus subtilis] | 63 | 38 |
| 128 | 30 | 25257 | 24397 | gi\|1518679 | orf [Bacillus subtilis] | 63 | 39 |
| 138 | 18 | 12236 | 11580 | gi\|405516 | This ORF is homologous to nitroreductase from Enterobacter cloacae, ccession Number A38686, and Salmonella, Accession Number P15888 Mycoplasma-like organism] | 63 | 39 |
| 143 | 2 | 167 | 1096 | pir\|S39416\|S39416 | metallothionein 10-I - blue mussel | 63 | 63 |
| 158 | 9 | 10023 | 8893 | bbs\|173803 | CD4+ T cell-stimulating antigen [Listeria monocytogenes, 85EO-1167, Peptide Partial, 268 aa] [Listeria monocytogenes] | 63 | 48 |
| 164 | 6 | 3041 | 3301 | gi\|1573583 | H. influenzae predicted coding region HI0594 [Haemophilus influenzae] | 63 | 31 |
| 164 | 18 | 18502 | 21708 | gi\|1015903 | ORF YJR151c [Saccharomyces cerevisiae] | 63 | 45 |
| 165 | 3 | 3084 | 2278 | gi\|537108 | ORF_f254 [Escherichia coli] | 63 | 45 |
| 166 | 1 | 83 | 1045 | gi\|762778 | NifS gene product [Anabaena azollae] | 63 | 49 |
| 168 | 3 | 638 | 1489 | gi\|805022 | Ndi1p [Saccharomyces cerevisiae] | 63 | 32 |
| 171 | 12 | 10655 | 10810 | gi\|152403 | phosphate regulatory protein [Rhizobium meliloti] | 63 | 50 |
| 172 | 1 | 242 | 1336 | gi\|1552775 | ATP-binding protein [Escherichia coli] | 63 | 45 |
| 179 | 11 | 11236 | 12111 | gnl\|PID\|e245033 | unknown [Mycobacterium tuberculosis] | 63 | 42 |
| 179 | 15 | 15289 | 15765 | gi\|1353197 | thioredoxin reductase [Eubacterium acidaminophilum] | 63 | 44 |
| 180 | 3 | 3412 | 1892 | gi\|1064813 | homologous to sp:PHOR_BACSU [Bacillus subtilis] | 63 | 40 |
| 180 | 7 | 7063 | 7926 | gi\|1657516 | hypothetical protein [Escherichia coli] | 63 | 41 |
| 187 | 1 | 1 | 729 | gi\|1651957 | hypothetical protein [Synechocystis sp.] | 63 | 34 |
| 195 | 17 | 7717 | 8280 | gi\|431928 | MunI methyltransferase [Mycoplasma sp.] | 63 | 44 |
| 202 | 8 | 5311 | 6165 | gi\|606162 | ORF_f229 [Escherichia coli] | 63 | 48 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 202 | 10 | 7848 | 8681 | gi\|606018 | ORF_o783 [Escherichia coli] | 63 | 47 |
| 208 | 3 | 2979 | 2341 | gi\|1006613 | hypothetical protein [Synechocystis sp.] | 63 | 40 |
| 221 | 3 | 874 | 1146 | gnl\|PID\|e265530 | yorfE [Streptococcus pneumoniae] | 63 | 42 |
| 227 | 2 | 856 | 1254 | gi\|438459 | homologous to E. coli hydrophobic Fe-uptake components FepD, FecD; utative [Bacillus subtilis] | 63 | 41 |
| 231 | 3 | 2618 | 2448 | gi\|606248 | 30S ribosomal subunit protein S3 [Escherichia coli] | 63 | 42 |
| 233 | 9 | 6773 | 6144 | gi\|887827 | ORF_o192 [Escherichia coli] | 63 | 41 |
| 234 | 1 | 348 | 70 | gi\|494958 | ExpZ [Bacillus subtilis] | 63 | 32 |
| 240 | 2 | 1230 | 721 | gnl\|PID\|e252616 | DcuC protein [Escherichia coli] | 63 | 38 |
| 244 | 9 | 7512 | 6508 | gi\|467421 | similar to B. subtilis DnaH [Bacillus subtilis] sp\|P37540\|YAAS_BACSU HYPOTHETICAL 37.6 KD PROTEIN IN XPAC-ABRB NTERGENIC REGION. | 63 | 43 |
| 255 | 5 | 3600 | 2818 | gi\|1486244 | unknown [Bacillus subtilis] | 63 | 47 |
| 258 | 1 | 3 | 449 | gi\|1041115 | TRAC [Plasmid pPD1] | 63 | 38 |
| 259 | 4 | 2842 | 2342 | gnl\|PID\|e290788 | unknown [Mycobacterium tuberculosis] | 63 | 42 |
| 265 | 8 | 3313 | 3480 | gi\|694074 | emml gene product [Streptococcus pyogenes] | 63 | 42 |
| 276 | 18 | 12505 | 11654 | gi\|601878 | beta-1,3-glucanase bglH [Bacillus circulans] | 63 | 36 |
| 294 | 5 | 2012 | 2275 | gi\|288661 | ORF5 product [Bacteriophage P2] | 63 | 40 |
| 301 | 7 | 7063 | 6704 | gnl\|PID\|e290998 | unknown [Mycobacterium tuberculosis] | 63 | 41 |
| 345 | 2 | 2279 | 2725 | gi\|413940 | ipa-16d gene product [Bacillus subtilis] | 63 | 39 |
| 351 | 8 | 4361 | 3306 | gi\|398120 | TDP-glucose oxireductase [Xanthomonas campestris] | 63 | 47 |
| 359 | 1 | 526 | 14 | gi\|1001605 | 3-hydroxyisobutyrate dehydrogenase [Synechocystis sp.] | 63 | 36 |
| 364 | 6 | 6741 | 7277 | gi\|1736473 | ORF_ID:o335#13; similar to [SwissProt Accession Number P36088] [Escherichia coli] | 63 | 42 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 378 | 2 | 683 | 1414 | gi\|529016 | aminoglycoside 6-adenylyltransferase [Bacillus subtilis] pir\|JU0059\|XXBSG aminoglycoside 6-adenylyltransferase (EC 2.7.7.-) Bacillus subtilis | 63 | 41 |
| 392 | 2 | 783 | 1646 | gi\|1772644 | orfR gene product [Bacillus subtilis] | 63 | 34 |
| 399 | 2 | 574 | 1407 | gi\|40023 | B.subtilis genes rpmH, rnpA, 50kd, gidA and gidB [Bacillus subtilis] i\|467388 stage III sporulation [Bacillus subtilis] ir\|S18073\|S18073 spoIIIJ protein - Bacillus subtilis | 63 | 42 |
| 403 | 1 | 754 | 2 | gi\|1303938 | YqiS [Bacillus subtilis] | 63 | 52 |
| 404 | 5 | 4149 | 3745 | gi\|142450 | ahrC protein [Bacillus subtilis] | 63 | 42 |
| 430 | 1 | 2 | 1222 | gi\|1046082 | M. genitalium predicted coding region MG372 [Mycoplasma genitalium] | 63 | 40 |
| 432 | 1 | 3 | 1241 | gi\|1001328 | UDP-MurNac-tripeptide synthetase [Synechocystis sp.] | 63 | 33 |
| 432 | 4 | 1970 | 3016 | gi\|1161061 | dioxygenase [Methylobacterium extorquens] | 63 | 41 |
| 463 | 2 | 1324 | 851 | gi\|1573163 | hypothetical [Haemophilus influenzae] | 63 | 40 |
| 466 | 4 | 2843 | 3730 | gnl\|PID\|e261988 | putative ORF [Bacillus subtilis] | 63 | 41 |
| 472 | 1 | 527 | 3 | gi\|556885 | Unknown [Bacillus subtilis] | 63 | 50 |
| 517 | 3 | 2803 | 1646 | gi\|531265 | lipophilic protein which affects bacterial lysis rate and ethicillin resistance level [Staphylococcus aureus] pir\|A55856\|A55856 llm protein - Staphylococcus aureus | 63 | 38 |
| 538 | 1 | 206 | 3 | gi\|172657 | serine-protein kinase [Saccharomyces cerevisiae] | 63 | 47 |
| 539 | 4 | 2997 | 3851 | gi\|973230 | gamma-glutamyl kinase [Lycopersicon esculentum] | 63 | 43 |
| 565 | 3 | 756 | 1010 | gi\|1303724 | YqaF [Bacillus subtilis] | 63 | 51 |
| 573 | 7 | 4518 | 3709 | gi\|1652352 | dihydropteroate pyrophosphorylase [Synechocystis sp.] | 63 | 45 |
| 579 | 2 | 361 | 1344 | gi\|1573114 | beta-ketoacyl-acyl carrier protein | 63 | 41 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | synthase III (fabH) [Haemophilus influenzae] | | |
| 593 | 2 | 390 | 1037 | gi\|409286 | bmrU [Bacillus subtilis] | 63 | 33 |
| 707 | 1 | 647 | 171 | gi\|511596 | interleukin-2 [Canis familiaris] | 63 | 33 |
| 714 | 1 | 2 | 268 | gnl\|PID\|e213832 | putative inner membrane protein [Bacillus licheniformis] | 63 | 38 |
| 724 | 1 | 562 | 239 | gnl\|PID\|e255315 | unknown [Mycobacterium tuberculosis] | 63 | 49 |
| 759 | 1 | 681 | 4 | gi\|437639 | [Plasmodium falciparum 3' end.], gene product [Plasmodium alciparum] | 63 | 28 |
| 794 | 1 | 981 | 313 | gi\|451201 | ORF1 [Bacillus subtilis] | 63 | 37 |
| 811 | 2 | 609 | 184 | gi\|150553 | regulatory protein [Plasmid pCF10] | 63 | 30 |
| 835 | 1 | 2 | 262 | gi\|1736496 | RpiR protein. [Escherichia coli] | 63 | 41 |
| 11 | 1 | 2 | 1144 | gi\|143150 | levR [Bacillus subtilis] | 62 | 48 |
| 12 | 5 | 8710 | 7673 | gi\|1486244 | unknown [Bacillus subtilis] | 62 | 43 |
| 15 | 3 | 1167 | 2957 | gi\|1592101 | adenine deaminase [Methanococcus jannaschii] | 62 | 40 |
| 16 | 4 | 2572 | 4092 | gi\|1109685 | ProW [Bacillus subtilis] | 62 | 37 |
| 23 | 4 | 1279 | 2067 | gi\|41432 | fepC gene product [Escherichia coli] | 62 | 35 |
| 23 | 26 | 16176 | 16454 | gi\|154499 | carbon dioxide concentrating mechanism protein [Synechococcus sp.] pir\|C36904\|C36904 carbon dioxide concentrating mechanism protein cmL - Synechococcus sp. (PCC 7942) | 62 | 41 |
| 31 | 6 | 5322 | 5774 | gi\|532309 | 25 kDa protein [Escherichia coli] | 62 | 38 |
| 68 | 4 | 1606 | 2778 | gi\|1732203 | GlcNAc 6-P deacetylase [Vibrio furnissii] | 62 | 44 |
| 72 | 1 | 1 | 540 | gi\|1573097 | glucosamine-6-phosphate deaminase protein (nagB) [Haemophilus influenzae] | 62 | 26 |
| 76 | 3 | 1937 | 2227 | gi\|928830 | ORF75; putative [Lactococcus lactis phage BK5-T] | 62 | 34 |
| 83 | 16 | 11700 | 12272 | gi\|1592161 | N-terminal acetyltransferase complex, subunit ARD1 [Methanococcus jannaschii] | 62 | 33 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 83 | 19 | 12685 | 13737 | gi\|1653193 | sialoglycoprotease [Synechocystis sp.] | 62 | 42 |
| 91 | 6 | 3232 | 3789 | gi\|1762962 | FemA [Staphylococcus simulans] | 62 | 37 |
| 100 | 43 | 29676 | 29317 | gi\|963033 | orf1 gene product [Enterococcus hirae] | 62 | 45 |
| 101 | 8 | 7410 | 6481 | gi\|1161061 | dioxygenase [Methylobacterium extorquens] | 62 | 45 |
| 110 | 3 | 653 | 871 | gi\|992683 | mdm2-D [Homo sapiens] | 62 | 37 |
| 110 | 8 | 8440 | 5810 | gi\|784897 | beta-N-acetylhexosaminidase [Streptococcus pneumoniae] pir\|A56390\|A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - treptococcus pneumoniae | 62 | 46 |
| 111 | 2 | 1057 | 287 | gnl\|PID\|e253280 | ORF YDL238c [Saccharomyces cerevisiae] | 62 | 45 |
| 114 | 5 | 6886 | 7662 | gi\|152719 | flavocytochrome c [Shewanella putrefaciens] | 62 | 37 |
| 115 | 4 | 1401 | 1994 | gi\|1303978 | YqkA [Bacillus subtilis] | 62 | 46 |
| 118 | 1 | 545 | 225 | gi\|39431 | oligo-1,6-glucosidase [Bacillus cereus] | 62 | 40 |
| 119 | 8 | 4625 | 4356 | gi\|1522673 | type I restriction enzyme [Methanococcus jannaschii] | 62 | 33 |
| 120 | 2 | 257 | 1270 | gnl\|PID\|e235823 | unknown [Schizosaccharomyces pombe] | 62 | 41 |
| 121 | 8 | 7543 | 8034 | gi\|39475 | formamidopyrimidine-DNA glycosylase [Bacillus firmus] ir\|S11489\|S11489 formamidopyrimidine-DNA glycosidase (EC 3.2.2.23) Bacillus firmus | 62 | 48 |
| 123 | 2 | 1677 | 592 | gi\|882252 | conjugated bile acid hydrolase [Clostridium perfringens] sp\|P54965\|CBH_CLOPE CHOLOYLGLYCINE HYDROLASE (EC 3.5.1.24) CONJUGATED BILE ACID HYDROLASE) (CBAH) (BILE SALT HYDROLASE). | 62 | 40 |
| 128 | 16 | 10895 | 9408 | gi\|1742834 | PTS system, cellobiose-specific IIC component (EIIC-CEL) (Cellobiose- permease IIC component) (Phosphotransferase enzyme II, C component). [Escherichia coli] | 62 | 43 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 128 | 29 | 24254 | 23544 | gi\|1518680 | minicell-associated protein DivIVA [Bacillus subtilis] | 62 | 37 |
| 128 | 35 | 28843 | 28103 | gi\|142940 | ftsA [Bacillus subtilis] | 62 | 42 |
| 133 | 4 | 3434 | 4165 | gnl\|PID\|e235174 | unknown [Mycobacterium tuberculosis] | 62 | 38 |
| 134 | 2 | 1679 | 933 | gi\|155032 | ORF B [Plasmid pEa34] | 62 | 36 |
| 146 | 6 | 4923 | 4651 | gi\|153675 | tagatose 6-P kinase [Streptococcus mutans] | 62 | 48 |
| 149 | 5 | 3318 | 2527 | gi\|1591587 | pantothenate metabolism flavoprotein [Methanococcus jannaschii] | 62 | 35 |
| 152 | 9 | 4830 | 5747 | gi\|1652461 | lactose transport system permease protein LacF [Synechocystis sp.] | 62 | 39 |
| 163 | 2 | 1341 | 544 | gi\|533098 | DnaD protein [Bacillus subtilis] | 62 | 41 |
| 164 | 14 | 9567 | 9322 | gi\|1118060 | coded for by C. elegans cDNA yk3d11.5; coded for by C. elegans cDNA yk5f4.5 [Caenorhabditis elegans] | 62 | 27 |
| 172 | 8 | 6613 | 7146 | gi\|915199 | ggaB [Bacillus subtilis] | 62 | 33 |
| 173 | 13 | 11127 | 9736 | gi\|1653484 | hypothetical protein [Synechocystis sp.] | 62 | 44 |
| 177 | 1 | 1077 | 364 | gi\|1572994 | 2-keto-3-deoxy-6-phosphogluconate aldolase (eda) [Haemophilus influenzae] | 62 | 38 |
| 178 | 4 | 1683 | 1318 | gnl\|PID\|e155310 | Orf2 [Bacteriophage TP901-1] | 62 | 51 |
| 179 | 5 | 6425 | 7576 | gi\|1161933 | DltB [Lactobacillus casei] | 62 | 44 |
| 180 | 13 | 12470 | 10842 | sp\|P37047\|YAEG_ECOLI | HYPOTHETICAL 44.3 KD PROTEIN IN HTRA-DAPD INTERGENIC REGION. | 62 | 38 |
| 181 | 14 | 11649 | 10735 | gi\|1742758 | Shikimate 5-dehydrogenase (EC 1.1.1.25). [Escherichia coli] | 62 | 41 |
| 197 | 2 | 516 | 1442 | gi\|623476 | transcriptional activator [Providencia stuartii] sp\|P43463\|AARP_PROST TRANSCRIPTIONAL ACTIVATOR AARP. | 62 | 34 |
| 206 | 5 | 2728 | 1790 | gnl\|PID\|e265638 | unknown [Mycobacterium tuberculosis] | 62 | 37 |
| 210 | 2 | 938 | 2290 | gi\|528991 | unknown [Bacillus subtilis] | 62 | 41 |
| 221 | 15 | 7083 | 7280 | gnl\|PID\|e219154 | K08F4.5 [Caenorhabditis elegans] | 62 | 44 |
| 222 | 11 | 7141 | 8022 | gi\|537034 | ORF_o488 [Escherichia coli] | 62 | 39 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 223 | 9 | 6924 | 6358 | gnl\|PID\|e283128 | unknown, highly similar to E. coli YecD hypothtical 21.8 KD protein in aspS 5'region and to isochorismatase [Bacillus subtilis] | 62 | 42 |
| 225 | 4 | 2055 | 2885 | gi\|18724 | pyrroline-5-carboxylate reductase (AA 1-274) [Glycine max] ir\|S10186\|S10186 pyrroline-5-carboxylate reductase (EC 1.5.1.2) - ybean | 62 | 39 |
| 229 | 11 | 11428 | 10670 | gnl\|PID\|e235745 | hypothetical protein [Mycobacterium leprae] | 62 | 36 |
| 231 | 1 | 1244 | 3 | gi\|48808 | dciAE gene product [Bacillus subtilis] | 62 | 45 |
| 233 | 1 | 801 | 4 | gi\|143391 | ORF2 [Bacillus subtilis] | 62 | 42 |
| 233 | 13 | 10471 | 9431 | gi\|887825 | ORF_f541 [Escherichia coli] | 62 | 35 |
| 242 | 1 | 3 | 149 | gi\|532549 | ORF16 [Enterococcus faecalis] | 62 | 44 |
| 255 | 2 | 443 | 1009 | gi\|639789 | ORF9 [Mycoplasma pneumoniae] | 62 | 44 |
| 266 | 6 | 2349 | 2158 | gnl\|PID\|e194945 | yeast sds22 homolog [Homo sapiens] | 62 | 37 |
| 270 | 1 | 3 | 314 | gi\|1303827 | YqfI [Bacillus subtilis] | 62 | 35 |
| 270 | 7 | 5136 | 4447 | gi\|1303958 | YqjG [Bacillus subtilis] | 62 | 41 |
| 279 | 1 | 271 | 2 | gnl\|PID\|e185372 | ceuC gene product [Campylobacter coli] | 62 | 44 |
| 301 | 11 | 9598 | 8798 | gi\|1303863 | YqgP [Bacillus subtilis] | 62 | 45 |
| 306 | 2 | 750 | 1202 | gi\|148771 | ribosomal protein HmaS4 [Haloarcula marismortui] | 62 | 41 |
| 308 | 3 | 2328 | 1684 | gnl\|PID\|e238666 | hypothetical protein [Bacillus subtilis] | 62 | 40 |
| 309 | 5 | 8806 | 8573 | gi\|1591861 | M. jannaschii predicted coding region MJ1230 [Methanococcus jannaschii] | 62 | 37 |
| 318 | 3 | 2278 | 1283 | gi\|1256134 | YbbE [Bacillus subtilis] | 62 | 37 |
| 321 | 3 | 1433 | 1792 | gi\|606080 | ORF_o290; Geneplot suggests frameshift linking to o267, not found Escherichia coli] | 62 | 37 |
| 338 | 13 | 11175 | 12770 | gi\|467446 | similar to SpoVB [Bacillus subtilis] | 62 | 38 |
| 345 | 11 | 10519 | 11793 | gi\|1736789 | Collagenase precursor (EC 3.4.-.-). | 62 | 40 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | [Escherichia coli] | | |
| 345 | 21 | 22459 | 22947 | gi\|1657794 | 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase [Methylobacterium extorquens] | 62 | 47 |
| 358 | 1 | 902 | 36 | gi\|409241 | penicillin-binding protein 2 [Staphylococcus aureus] | 62 | 44 |
| 362 | 6 | 2930 | 3493 | gnl\|PID\|e255091 | hypothetical protein [Bacillus subtilis] | 62 | 37 |
| 363 | 2 | 3242 | 1581 | gnl\|PID\|e254997 | hypothetical protein [Bacillus subtilis] | 62 | 40 |
| 365 | 2 | 400 | 1770 | gi\|143150 | levR [Bacillus subtilis] | 62 | 42 |
| 372 | 5 | 2525 | 4489 | gi\|1045736 | fructose-permease IIBC component [Mycoplasma genitalium] | 62 | 43 |
| 373 | 1 | 3 | 851 | gi\|438462 | transmembrane protein [Bacillus subtilis] | 62 | 36 |
| 375 | 1 | 2 | 1336 | gi\|732813 | branched-chain amino acid carrier [Lactobacillus delbrueckii] pir\|S60180\|S60180 branched-chain amino acid carrier brnQ - actobacillus delbrueckii | 62 | 43 |
| 375 | 3 | 2592 | 1831 | gi\|1644206 | unknown [Bacillus subtilis] | 62 | 43 |
| 391 | 2 | 142 | 510 | gi\|151776 | ORF3 [Escherichia coli] | 62 | 31 |
| 396 | 2 | 254 | 1051 | gi\|410131 | ORFX7 [Bacillus subtilis] | 62 | 41 |
| 423 | 1 | 197 | 6 | pir\|A33592\|A33592 | repressor protein catM - Acinetobacter calcoaceticus | 62 | 38 |
| 436 | 1 | 704 | 3 | gi\|455376 | unidentified reading frame L (ORFL) (putative); putative [Transposon n10] | 62 | 32 |
| 466 | 8 | 9320 | 10480 | gi\|147402 | mannose permease subunit III-Man [Escherichia coli] | 62 | 44 |
| 488 | 5 | 2175 | 2927 | gi\|532546 | ORF13 [Enterococcus faecalis] | 62 | 40 |
| 510 | 4 | 2572 | 3078 | gi\|43941 | EIII-B Sor PTS [Klebsiella pneumoniae] | 62 | 35 |
| 517 | 2 | 1533 | 736 | gi\|559388 | epsX gene product [Acinetobacter calcoaceticus] | 62 | 53 |
| 519 | 1 | 2 | 1084 | gi\|1652876 | hypothetical protein [Synechocystis sp.] | 62 | 41 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 535 | 1 | 353 | 69 | gi\|1196922 | unknown protein [Insertion sequence IS861] | 62 | 33 |
| 579 | 1 | 1 | 363 | gi\|535052 | involved in protein secretion [Bacillus subtilis] | 62 | 22 |
| 656 | 5 | 5351 | 5956 | gnl\|PID\|e290931 | unknown [Mycobacterium tuberculosis] | 62 | 40 |
| 666 | 1 | 445 | 128 | gi\|483940 | transcription regulator [Bacillus subtilis] | 62 | 42 |
| 682 | 1 | 597 | 172 | gi\|146724 | enzyme III-Man function protein (manX (ptsL)) [Escherichia coli] gi\|41976 manX gene product (AA 1-315) [Escherichia coli] | 62 | 37 |
| 771 | 1 | 3 | 365 | gi\|1773086 | similar to S. typhimurium ProY [Escherichia coli] | 62 | 44 |
| 831 | 1 | 390 | 94 | gnl\|PID\|e255000 | hypothetical protein [Bacillus subtilis] | 62 | 55 |
| 15 | 5 | 4421 | 5260 | gnl\|PID\|e214719 | PlcR protein [Bacillus thuringiensis] | 61 | 38 |
| 16 | 6 | 4705 | 4938 | gi\|758425 | complement component C3 [Xenopus laevis/gilli] | 61 | 44 |
| 23 | 16 | 10279 | 11214 | sp\|P19265\|EUTC_SALTY | ETHANOLAMINE AMMONIA-LYASE LIGHT CHAIN (EC 4.3.1.7). | 61 | 46 |
| 33 | 2 | 1789 | 2205 | gi\|413958 | ipa-34d gene product [Bacillus subtilis] | 61 | 36 |
| 33 | 5 | 4756 | 6594 | gi\|1001823 | cadmium-transporting ATPase [Synechocystis sp.] | 61 | 38 |
| 37 | 4 | 2813 | 3295 | gi\|1256140 | YbbK [Bacillus subtilis] | 61 | 51 |
| 37 | 7 | 5973 | 5215 | gnl\|PID\|e269488 | Unknown [Bacillus subtilis] | 61 | 33 |
| 49 | 4 | 1567 | 1839 | gnl\|PID\|e139445 | major tail protein [Bacteriophage B1] | 61 | 43 |
| 56 | 1 | 108 | 641 | gi\|1574067 | H. influenzae predicted coding region HI1034 [Haemophilus influenzae] | 61 | 35 |
| 59 | 1 | 1 | 1002 | gi\|763513 | ORF4; putative [Streptomyces violaceoruber] | 61 | 37 |
| 69 | 7 | 4837 | 5523 | gnl\|PID\|e254877 | unknown [Mycobacterium tuberculosis] | 61 | 34 |
| 72 | 11 | 9262 | 10476 | gi\|1591272 | ferrous iron transport protein B [Methanococcus jannaschii] | 61 | 45 |
| 83 | 2 | 731 | 1549 | gi\|755152 | highly hydrophobic integral membrane | 61 | 41 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | protein [Bacillus subtilis] sp\|P42953\|TAGG_BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGG. | | |
| 87 | 2 | 2067 | 925 | gi\|1573129 | hypothetical [Haemophilus influenzae] | 61 | 46 |
| 103 | 5 | 2689 | 3495 | gi\|1685111 | orf1091 [Streptococcus thermophilus] | 61 | 45 |
| 110 | 13 | 11455 | 11820 | gi\|1001825 | transcriptional repressor SmtB [Synechocystis sp.] | 61 | 42 |
| 110 | 15 | 14048 | 12588 | gi\|1573583 | H. influenzae predicted coding region HI0594 [Haemophilus influenzae] | 61 | 38 |
| 111 | 3 | 1675 | 1055 | gnl\|PID\|e253280 | ORF YDL238c [Saccharomyces cerevisiae] | 61 | 34 |
| 111 | 4 | 1838 | 2518 | gi\|1574513 | hypothetical [Haemophilus influenzae] | 61 | 50 |
| 111 | 5 | 2535 | 3158 | gi\|537235 | Kenn Rudd identifies as gpmB [Escherichia coli] | 61 | 40 |
| 121 | 1 | 3 | 1397 | gi\|290643 | ATPase [Enterococcus hirae] | 61 | 50 |
| 123 | 28 | 25608 | 27734 | gi\|143150 | levR [Bacillus subtilis] | 61 | 39 |
| 125 | 5 | 3455 | 2589 | gi\|148921 | LicD protein [Haemophilus influenzae] | 61 | 47 |
| 128 | 14 | 9382 | 9146 | gi\|575361 | protein kinase PkpA [Phycomyces blakesleeanus] | 61 | 38 |
| 138 | 32 | 23151 | 21628 | gi\|1184262 | GadC [Shigella flexneri] | 61 | 34 |
| 144 | 8 | 6311 | 5325 | gi\|710422 | cmp-binding-factor 1 [Staphylococcus aureus] | 61 | 39 |
| 171 | 4 | 4601 | 5566 | gi\|41500 | ORF 3 (AA 1-352); 38 kD (put. ftsX) [Escherichia coli] | 61 | 31 |
| 172 | 3 | 2006 | 2848 | gi\|303560 | ORF271 [Escherichia coli] | 61 | 42 |
| 173 | 7 | 5146 | 6228 | gi\|1256134 | YbbE [Bacillus subtilis] | 61 | 31 |
| 197 | 8 | 9183 | 8182 | gi\|143803 | GerC3 [Bacillus subtilis] | 61 | 33 |
| 217 | 5 | 3007 | 3462 | gi\|1749414 | unnamed protein product [Schizosaccharomyces pombe] | 61 | 43 |
| 217 | 8 | 6099 | 5464 | gi\|143456 | rpoE protein (ttg start codon) [Bacillus subtilis] | 61 | 37 |
| 222 | 6 | 3400 | 3927 | gnl\|PID\|e255118 | hypothetical protein [Bacillus subtilis] | 61 | 41 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 225 | 3 | 1946 | 981 | gi\|1574660 | xylose operon regluatory protein (xylR) [Haemophilus influenzae] | 61 | 43 |
| 237 | 2 | 203 | 952 | gi\|1019108 | alternate start at bp 59; ORF [Bacteriophage phi-80] | 61 | 52 |
| 237 | 7 | 3058 | 3279 | gnl\|PID\|e246904 | ORF YPL169c [Saccharomyces cerevisiae] | 61 | 32 |
| 262 | 1 | 20 | 913 | gnl\|PID\|e214719 | PlcR protein [Bacillus thuringiensis] | 61 | 35 |
| 271 | 17 | 12725 | 13504 | gi\|143057 | ORF39 [Bacillus subtilis] | 61 | 31 |
| 275 | 8 | 5370 | 3697 | gi\|1542975 | AbcB [Thermoanaerobacterium thermosulfurigenes] | 61 | 41 |
| 280 | 2 | 692 | 3079 | gi\|1001352 | ABC transporter [Synechocystis sp.] | 61 | 42 |
| 294 | 7 | 2276 | 2767 | gi\|662792 | single-stranded DNA binding protein [unidentified eubacterium] | 61 | 44 |
| 301 | 12 | 9965 | 9519 | gi\|1303861 | YqgN [Bacillus subtilis] | 61 | 41 |
| 308 | 1 | 1471 | 26 | gi\|1276882 | EpsI [Streptococcus thermophilus] | 61 | 36 |
| 314 | 2 | 475 | 1662 | gi\|975351 | PatB [Bacillus subtilis] | 61 | 42 |
| 321 | 9 | 3762 | 4193 | gi\|1732202 | PTS permease for mannose subunit IIIMan N terminal domain [Vibrio furnissii] | 61 | 40 |
| 323 | 5 | 5118 | 5537 | gi\|532540 | ORF7 [Enterococcus faecalis] | 61 | 28 |
| 324 | 7 | 4800 | 5156 | gi\|146122 | H-protein [Escherichia coli] | 61 | 39 |
| 338 | 3 | 1456 | 1989 | pir\|A47071\|A47071 | orf1 immediately 5' of nifS - Bacillus subtilis | 61 | 43 |
| 341 | 2 | 342 | 947 | gi\|1736577 | Octopine transport system permease protein OccM. [Escherichia coli] | 61 | 41 |
| 349 | 3 | 1788 | 1363 | pir\|G64143\|G64143 | hypothetical protein HI0143 - Haemophilus influenzae (strain Rd KW20) | 61 | 38 |
| 369 | 2 | 1261 | 587 | gi\|153744 | ORF X; putative [Streptococcus mutans] | 61 | 33 |
| 371 | 2 | 1801 | 1562 | gi\|48836 | xylulokinase [Staphylococcus xylosus] | 61 | 40 |
| 372 | 4 | 1575 | 2543 | gi\|149395 | lacC [Lactococcus lactis] | 61 | 43 |
| 379 | 11 | 12683 | 11727 | gi\|887829 | D21141 uses 2nd start; frame determined by Lac fusion [Escherichia oli] | 61 | 40 |
| 383 | 5 | 5625 | 3820 | gi\|624072 | similar to Escherichia coli | 61 | 36 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | glycerophosphoryl diester hosphodiesterase, Swiss-Prot Accession Number P10908 [Paramecium ursaria Chlorella virus 1] | | |
| 395 | 2 | 771 | 517 | gnl\|PID\|e276251 | T23G11.6 [Caenorhabditis elegans] | 61 | 42 |
| 399 | 20 | 15621 | 15812 | gi\|472527 | protein phosphatase 1 [Schizosaccharomyces pombe] | 61 | 44 |
| 413 | 1 | 3 | 749 | gnl\|PID\|e289144 | ywpE [Bacillus subtilis] | 61 | 42 |
| 427 | 1 | 1079 | 288 | gi\|403373 | glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pir\|S37251\|S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis | 61 | 42 |
| 436 | 4 | 2045 | 1761 | gi\|48669 | pot. ORF B [Shigella sonnei] | 61 | 38 |
| 437 | 1 | 1158 | 244 | gi\|580866 | ipa-12d gene product [Bacillus subtilis] | 61 | 47 |
| 482 | 2 | 1676 | 1167 | bbs\|158786 | 4A11 antigen, sperm tail membrane antigen=putative sucrose-specific phosphotransferase enzyme II homolog [mice, testis, Peptide Partial, 172 aa] [Mus sp.] | 61 | 42 |
| 490 | 3 | 1291 | 1094 | gnl\|PID\|e248473 | putative phosphate permease [Arabidopsis thaliana] | 61 | 35 |
| 514 | 1 | 687 | 142 | gi\|1742775 | msm operon regulatory protein. [Escherichia coli] | 61 | 36 |
| 541 | 1 | 758 | 3 | gi\|1591732 | cobalt transport ATP-binding protein O [Methanococcus jannaschii] | 61 | 39 |
| 551 | 3 | 2163 | 1600 | gi\|671632 | unknown [Staphylococcus aureus] | 61 | 38 |
| 603 | 2 | 163 | 564 | gi\|1408587 | relaxase [Lactococcus lactis lactis] | 61 | 39 |
| 637 | 8 | 4539 | 4769 | gi\|143559 | subtilin [Bacillus subtilis] | 61 | 38 |
| 765 | 1 | 34 | 681 | gi\|408888 | orfA 5' of intG [Lactobacillus bacteriophage phi adh] pir\|PN0468\|PN0468 hypothetical protein 106 - Lactobacillus | 61 | 40 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | gasseri fragment) | | |
| 773 | 1 | 53 | 1207 | gi\|143841 | xylose repressor [Bacillus subtilis] | 61 | 36 |
| 798 | 1 | 175 | 381 | gi\|187572 | located at OATL1 [Homo sapiens] | 61 | 32 |
| 5 | 2 | 303 | 998 | gi\|1783264 | homologous to DNA glycosylases; hypothetical [Bacillus subtilis} | 60 | 50 |
| 8 | 8 | 5891 | 6550 | gi\|1777939 | Pfs [Treponema pallidum] | 60 | 40 |
| 11 | 7 | 4096 | 4935 | gi\|147404 | mannose permease subunit II-M-Man [Escherichia coli] | 60 | 41 |
| 11 | 8 | 4919 | 5254 | gi\|467125 | glmS; L-Glucosamine:D-fructose-6-phosphate aminotransferase; 229_C3_238 [Mycobacterium leprae] | 60 | 30 |
| 17 | 9 | 7736 | 8203 | gi\|496514 | orf zeta [Streptococcus pyogenes] | 60 | 42 |
| 20 | 1 | 3 | 443 | gi\|861137 | chitin binding protein [Streptomyces olivaceoviridis] pir\|S55001\|S55001 CHB1 protein - Streptomyces olivaceoviridis {SUB -30} | 60 | 40 |
| 21 | 3 | 1970 | 684 | gi\|1778520 | hypothetical protein [Escherichia coli] | 60 | 43 |
| 23 | 11 | 5357 | 5953 | gi\|619066 | NAST [Azotobacter vinelandii] | 60 | 31 |
| 34 | 4 | 6662 | 3279 | gi\|153952 | polymerase III polymerase subunit (dnaE) [Salmonella typhimurium] pir\|A45915\|A45915 DNA-directed DNA polymerase (EC 2.7.7.7) III lpha chain - Salmonella typhimurium | 60 | 37 |
| 39 | 1 | 47 | 466 | gi\|1561567 | Unknown [Bacillus subtilis] | 60 | 35 |
| 39 | 4 | 1855 | 1361 | gi\|298045 | Orf154 [Streptomyces ambofaciens] | 60 | 41 |
| 48 | 4 | 2554 | 4128 | gi\|1255259 | o-succinylbenzoic acid (OSB) CoA ligase [Staphylococcus aureus] | 60 | 40 |
| 56 | 9 | 6682 | 5795 | gi\|413940 | ipa-16d gene product [Bacillus subtilis] | 60 | 40 |
| 65 | 3 | 2105 | 2593 | gi\|1573061 | hypothetical [Haemophilus influenzae] | 60 | 34 |
| 72 | 9 | 7854 | 8330 | gi\|606343 | CG Site No. 28964 [Escherichia coli] | 60 | 39 |
| 81 | 3 | 2053 | 1406 | gi\|1574770 | phenylalanyl-tRNA synthetase beta-subunit (pheT) [Haemophilus influenzae] | 60 | 46 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 81 | 4 | 2987 | 2130 | gi\|147404 | mannose permease subunit II-M-Man [Escherichia coli] | 60 | 34 |
| 81 | 12 | 8280 | 7150 | gnl\|PID\|e254984 | hypothetical protein [Bacillus subtilis] | 60 | 44 |
| 83 | 22 | 16887 | 16537 | gi\|509672 | repressor protein [Bacteriophage Tuc2009] | 60 | 33 |
| 89 | 1 | 698 | 60 | gi\|840838 | hypothetical 21.7 kDa protein in ftsY 5' region [Pseudomonas eruginosa] | 60 | 36 |
| 89 | 12 | 12641 | 11856 | gi\|1377843 | unknown [Bacillus subtilis] | 60 | 40 |
| 89 | 17 | 18879 | 15844 | gi\|666069 | orf2 gene product [Lactobacillus leichmannii] | 60 | 37 |
| 94 | 6 | 2281 | 3384 | gi\|468760 | ORF334 [Rhizobium meliloti] | 60 | 36 |
| 98 | 1 | 12 | 1970 | gi\|1652892 | ABC transporter [Synechocystis sp.] | 60 | 38 |
| 99 | 3 | 978 | 1460 | gi\|473955 | DNA-binding protein [Lactobacillus sp.] | 60 | 31 |
| 100 | 35 | 26818 | 26333 | gi\|347851 | junctional sarcoplasmic reticulum glycoprotein [Oryctolagus uniculus] | 60 | 48 |
| 100 | 45 | 30072 | 30449 | gi\|143547 | Sin regulatory protein (ttg start codon) [Bacillus subtilis] gi\|1303886 SinR [Bacillus subtilis] | 60 | 43 |
| 102 | 8 | 5923 | 6561 | gi\|1633572 | Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus] | 60 | 25 |
| 109 | 1 | 362 | 3 | pir\|S10655\|S10655 | hypothetical protein X - Pyrococcus woesei (fragment) | 60 | 33 |
| 110 | 16 | 14806 | 14087 | pir\|JH0364\|JH0364 | hypothetical protein 176 (SAGP 5' region) - Streptococcus pyogenes | 60 | 35 |
| 110 | 20 | 18929 | 18414 | gi\|142450 | ahrC protein [Bacillus subtilis] | 60 | 39 |
| 110 | 21 | 19124 | 19624 | gi\|142450 | ahrC protein [Bacillus subtilis] | 60 | 40 |
| 111 | 1 | 289 | 2 | gi\|1256618 | transport protein [Bacillus subtilis] | 60 | 31 |
| 122 | 7 | 5627 | 9589 | gi\|217191 | 5'-nucleotidase precursor [Vibrio parahaemolyticus] | 60 | 39 |
| 123 | 5 | 4390 | 3659 | gi\|1197667 | vitellogenin [Anolis pulchellus] | 60 | 27 |
| 123 | 20 | 18102 | 18407 | gi\|1303705 | YrkF [Bacillus subtilis] | 60 | 34 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 128 | 32 | 26229 | 25492 | gi\|1652485 | hypothetical protein [Synechocystis sp.] | 60 | 29 |
| 129 | 5 | 4421 | 6259 | gi\|1303853 | YqgF [Bacillus subtilis] | 60 | 36 |
| 131 | 2 | 1112 | 2338 | gi\|699112 | ugpC gene product [Mycobacterium leprae] | 60 | 41 |
| 131 | 4 | 3194 | 4036 | gi\|296356 | putative membrane transport protein [Clostridium perfringens]<br>pir\|A56641\|A56641 probable membrane transport protein - Clostridium erfringens | 60 | 32 |
| 131 | 8 | 6669 | 7901 | gi\|537054 | 2',3'-cyclic-nucleotide 2'-phosphodiesterase [Escherichia coli]<br>pir\|S56438\|S56438 2',3'-cyclic-nucleotide 2'-phosphodiesterase (EC .1.4.16) - Escherichia coli | 60 | 40 |
| 133 | 11 | 9854 | 10240 | gnl\|PID\|e249654 | YneR [Bacillus subtilis] | 60 | 37 |
| 138 | 7 | 6793 | 6263 | gi\|1486247 | unknown [Bacillus subtilis] | 60 | 48 |
| 146 | 4 | 2831 | 2328 | gi\|39979 | P18 [Bacillus subtilis] | 60 | 38 |
| 149 | 6 | 3504 | 3316 | gi\|145173 | 35 kDa protein [Escherichia coli] | 60 | 47 |
| 154 | 5 | 2599 | 3558 | gi\|1773109 | similar to S. typhimurium apbA [Escherichia coli] | 60 | 41 |
| 155 | 5 | 3061 | 4701 | gi\|388269 | traC [Plasmid pAD1] | 60 | 38 |
| 155 | 11 | 8565 | 8927 | gi\|1197460 | MtfB [Escherichia coli] | 60 | 39 |
| 158 | 10 | 11123 | 10032 | gi\|581809 | tmbC gene product [Treponema pallidum] | 60 | 39 |
| 165 | 7 | 6131 | 5700 | gi\|1439527 | EIIA-man [Lactobacillus curvatus] | 60 | 35 |
| 172 | 4 | 3169 | 3810 | gi\|1001342 | hypothetical protein [Synechocystis sp.] | 60 | 42 |
| 174 | 2 | 1574 | 762 | gi\|1045808 | hypothetical protein (GB:U00021_19) [Mycoplasma genitalium] | 60 | 35 |
| 181 | 7 | 4975 | 4460 | gi\|683584 | shikimate kinase [Lactococcus lactis] | 60 | 33 |
| 183 | 6 | 2719 | 2955 | gi\|1146198 | ferredoxin [Bacillus subtilis] | 60 | 37 |
| 189 | 2 | 3528 | 2221 | gi\|396301 | matches PS00041: Bacterial regulatory proteins, araC family ignature [Escherichia coli] | 60 | 35 |
| 193 | 5 | 3121 | 2600 | gi\|39788 | adaB [Bacillus subtilis] | 60 | 49 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 195 | 11 | 4623 | 6569 | gnl\|PID\|e250887 | potential coding region [Clostridium difficile] | 60 | 39 |
| 202 | 2 | 1837 | 1607 | gi\|693939 | membrane ATPase [Haloferax volcanii] | 60 | 32 |
| 206 | 7 | 4794 | 3754 | gi\|1574702 | hypothetical [Haemophilus influenzae] | 60 | 42 |
| 209 | 2 | 1308 | 433 | pir\|A38587\|A38587 | collagen, corneal - chicken (fragment) | 60 | 51 |
| 220 | 3 | 4263 | 1213 | gi\|437706 | alternative truncated translation product from E.coli [Streptococcus neumoniae] | 60 | 41 |
| 222 | 9 | 6019 | 6522 | gi\|882463 | protein-N(pi)-phosphohistidine-sugar phosphotransferase [Escherichia oli] | 60 | 47 |
| 222 | 12 | 8001 | 8336 | gi\|537035 | ORF_o101 [Escherichia coli] | 60 | 33 |
| 233 | 2 | 1294 | 827 | gi\|145091 | flavodoxin [Desulfovibrio salexigens] | 60 | 39 |
| 242 | 11 | 7370 | 7627 | gi\|1353404 | cytochrome oxidase subunit I [Metridium senile] | 60 | 28 |
| 249 | 3 | 1109 | 1768 | gi\|143156 | membrane bound protein [Bacillus subtilis] | 60 | 41 |
| 251 | 3 | 4053 | 1933 | gi\|1235662 | RfbC [Myxococcus xanthus] | 60 | 42 |
| 256 | 4 | 2614 | 3867 | gi\|532612 | ecotropic retrovirus receptor [Mus musculus] | 60 | 37 |
| 260 | 2 | 1539 | 802 | gi\|1208447 | metalloprotease transporter [Serratia marcescens] | 60 | 35 |
| 261 | 5 | 4528 | 3179 | gnl\|PID\|e246728 | histidine kinase [Streptococcus gordonii] | 60 | 25 |
| 269 | 3 | 2723 | 1563 | gi\|1591618 | M. jannaschii predicted coding region MJ0951 [Methanococcus jannaschii] | 60 | 39 |
| 269 | 4 | 3541 | 2780 | gi\|1303794 | YqeM [Bacillus subtilis] | 60 | 36 |
| 269 | 11 | 7164 | 6595 | gi\|1303787 | YqeG [Bacillus subtilis] | 60 | 38 |
| 271 | 2 | 677 | 1651 | gnl\|PID\|e269877 | riboflavin kinase [Bacillus subtilis] | 60 | 43 |
| 271 | 3 | 1639 | 2247 | gi\|537148 | ORF_f181 [Escherichia coli] | 60 | 41 |
| 271 | 18 | 13502 | 13762 | pir\|S39341\|S39341 | grpE protein - Lactococcus lactis | 60 | 40 |
| 277 | 2 | 1662 | 979 | gi\|1773109 | similar to S. typhimurium apbA [Escherichia coli] | 60 | 41 |
| 279 | 13 | 10627 | 9773 | gi\|290545 | f270 [Escherichia coli] | 60 | 41 |
| 290 | 2 | 790 | 1695 | gi\|152886 | elongation factor Ts (tsf) [Spiroplasma | 60 | 38 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | citri] | | |
| 291 | 4 | 3571 | 2612 | gnl\|PID\|e257610 | sugar-binding transport protein [Anaerocellum thermophilum] | 60 | 40 |
| 295 | 3 | 1309 | 2094 | gi\|1000453 | TreR [Bacillus subtilis] | 60 | 37 |
| 301 | 15 | 11063 | 11344 | gi\|535274 | ORF1 [Streptococcus thermophilus] | 60 | 36 |
| 310 | 3 | 2903 | 1266 | gi\|809765 | aspartate aminotransferase (AA 1-402) [Sulfolobus solfataricus] pir\|S07088\|S07088 aspartate transaminase (EC 2.6.1.1) - Sulfolobus olfataricus | 60 | 44 |
| 316 | 2 | 319 | 119 | bbs\|115298 | polyprotein(coat protein) [raspberry ringspot virus RRV, Peptide, 1107 aa] [Raspberry ringspot virus] | 60 | 28 |
| 320 | 4 | 3085 | 2483 | gi\|143002 | proton glutamate symport protein [Bacillus caldotenax] pir\|S26246\|S26246 glutamate/aspartate transport protein - Bacillus aldotenax | 60 | 26 |
| 323 | 1 | 1 | 681 | gi\|1477486 | transposase [Burkholderia cepacia] | 60 | 44 |
| 330 | 4 | 3361 | 4488 | gi\|1778517 | glycerol dehydrogenase homolog [Escherichia coli] | 60 | 48 |
| 356 | 3 | 2471 | 2205 | gi\|57633 | neuronal myosin heavy chain [Rattus rattus] | 60 | 40 |
| 362 | 5 | 2458 | 2925 | gnl\|PID\|e255090 | hypothetical protein [Bacillus subtilis] | 60 | 36 |
| 364 | 4 | 4096 | 5349 | gi\|1657522 | hypothetical protein [Escherichia coli] | 60 | 41 |
| 383 | 1 | 654 | 4 | gnl\|PID\|e288399 | F56H6.k [Caenorhabditis elegans] | 60 | 39 |
| 383 | 2 | 2208 | 853 | gi\|143536 | sigma factor 54 [Bacillus subtilis] | 60 | 37 |
| 386 | 2 | 130 | 510 | gi\|1046053 | hypothetical protein (SP:P32049) [Mycoplasma genitalium] | 60 | 42 |
| 399 | 26 | 25892 | 27757 | gi\|895747 | putative cel operon regulator [Bacillus subtilis] | 60 | 30 |
| 399 | 27 | 27721 | 28239 | gi\|146281 | gut operon activator (gutM) [Escherichia coli] | 60 | 35 |
| 401 | 4 | 2081 | 3523 | gi\|142833 | ORF2 [Bacillus subtilis] | 60 | 36 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 405 | 2 | 1353 | 763 | gi\|633113 | ORF3 [Streptococcus sobrinus] | 60 | 42 |
| 407 | 7 | 4380 | 4589 | gi\|1674126 | (AE000043) Mycoplasma pneumoniae, MG280 homolog, from M. genitalium [Mycoplasma pneumoniae] | 60 | 39 |
| 408 | 1 | 12 | 539 | gi\|455006 | orf6 [Rhodococcus fascians] | 60 | 42 |
| 421 | 7 | 4113 | 3925 | gi\|60020 | ORF31 (AA1-868) [Human herpesvirus 3] | 60 | 43 |
| 452 | 3 | 712 | 2223 | gi\|532554 | ORF21 [Enterococcus faecalis] | 60 | 38 |
| 462 | 3 | 2066 | 1551 | gi\|1015903 | ORF YJR151c [Saccharomyces cerevisiae] | 60 | 37 |
| 480 | 1 | 12 | 272 | gi\|468715 | sss gene product [Pseudomonas aeruginosa] | 60 | 34 |
| 487 | 1 | 1091 | 3 | gi\|388269 | traC [Plasmid pAD1] | 60 | 39 |
| 490 | 5 | 2108 | 1479 | gi\|699379 | glvr-1 protein [Mycobacterium leprae] | 60 | 29 |
| 507 | 1 | 221 | 751 | gi\|1303952 | YqjA [Bacillus subtilis] | 60 | 37 |
| 511 | 1 | 449 | 63 | gi\|391610 | farnesyl diphosphate synthase [Bacillus stearothermophilus] pir\|JX0257\|JX0257 geranyltranstransferase (EC 2.5.1.10) - Bacillus tearothermophilus | 60 | 42 |
| 551 | 2 | 1521 | 604 | gi\|1256648 | putative [Bacillus subtilis] | 60 | 37 |
| 552 | 1 | 887 | 63 | gi\|537235 | Kenn Rudd identifies as gpmB [Escherichia coli] | 60 | 40 |
| 610 | 1 | 1 | 792 | gi\|1321625 | exo-alpha-1, 4-glucosidase [Bacillus stearothermophilus] | 60 | 45 |
| 642 | 1 | 402 | 214 | gi\|992964 | thioredoxin [Arabidopsis thaliana] | 60 | 36 |
| 646 | 1 | 642 | 265 | gi\|1041115 | TRAC [Plasmid pPD1] | 60 | 32 |
| 661 | 2 | 305 | 943 | gi\|1651536 | 3-oxoacyl-[acyl-carrier-protein] reductase [Escherichia coli] | 60 | 37 |
| 678 | 1 | 536 | 3 | gi\|532554 | ORF21 [Enterococcus faecalis] | 60 | 39 |
| 716 | 1 | 799 | 305 | gi\|886040 | ORFtxe1 [Clostridium difficile] | 60 | 38 |
| 717 | 1 | 2 | 472 | gi\|1402529 | ORF8 [Enterococcus faecalis] | 60 | 31 |
| 727 | 1 | 516 | 82 | gi\|471283 | ORF [Synechococcus PCC6301] | 60 | 41 |
| 770 | 1 | 327 | 4 | gi\|467451 | unknown [Bacillus subtilis] | 60 | 33 |
| 843 | 1 | 234 | 4 | gi\|2819 | transferase (GAL10) (AA 1 - 687) | 60 | 37 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | [Kluyveromyces lactis] r\|S01407\|XUVKG UDPglucose 4-epimerase (EC 5.1.3.2) - yeast uyveromyces marxianus var. lactis) | | |
| 21 | 1 | 341 | 3 | gi\|1778519 | hypothetical protein [Escherichia coli] | 59 | 47 |
| 23 | 2 | 290 | 1303 | gi\|1407800 | ABC-type permease [Yersinia pestis] | 59 | 36 |
| 23 | 13 | 6720 | 7388 | gi\|1652472 | ethylene response sensor protein [Synechocystis sp.] | 59 | 37 |
| 23 | 18 | 11892 | 12413 | gi\|825627 | major carboxysome shell protein [Thiobacillus neapolitanus] pir\|S60136\|S60136 major carboxysome shell protein - Thiobacillus eapolitanus | 59 | 42 |
| 29 | 4 | 1989 | 2852 | gi\|1742383 | ORF_ID:o276#3; similar to [PIR Accession Number S11432] [Escherichia coli] | 59 | 48 |
| 32 | 8 | 4504 | 4064 | gi\|1046081 | hypothetical protein (GB:D26185_10) [Mycoplasma genitalium] | 59 | 33 |
| 37 | 9 | 6670 | 6284 | gi\|290561 | o188 [Escherichia coli] | 59 | 44 |
| 47 | 1 | 2 | 2743 | gnl\|PID\|e248792 | unknown [Mycobacterium tuberculosis] | 59 | 46 |
| 48 | 5 | 4017 | 5492 | gi\|1185288 | isochorismate synthase [Bacillus subtilis] | 59 | 40 |
| 49 | 5 | 1797 | 2093 | gi\|496280 | structural protein [Bacteriophage Tuc2009] | 59 | 41 |
| 59 | 8 | 3324 | 5057 | gi\|1486244 | unknown [Bacillus subtilis] | 59 | 35 |
| 72 | 14 | 13937 | 13434 | gi\|532540 | ORF7 [Enterococcus faecalis] | 59 | 25 |
| 81 | 20 | 14659 | 14219 | gi\|39978 | P16 [Bacillus subtilis] | 59 | 38 |
| 98 | 2 | 1961 | 2617 | gi\|41519 | P30 protein (AA 1-240) [Escherichia coli] | 59 | 39 |
| 102 | 3 | 2542 | 3774 | gi\|1674376 | (AE000062) Mycoplasma pneumoniae, MG148 homolog, from M. genitalium [Mycoplasma pneumoniae] | 59 | 30 |
| 116 | 2 | 907 | 1458 | gi\|1146225 | putative [Bacillus subtilis] | 59 | 37 |
| 116 | 7 | 3532 | 4842 | gi\|1146238 | poly(A) polymerase [Bacillus subtilis] | 59 | 41 |
| 128 | 20 | 15626 | 14310 | gi\|1001719 | ATP-dependent RNA helicase DeaD [Synechocystis sp.] | 59 | 34 |
| 134 | 4 | 3158 | 3850 | gi\|1477486 | transposase [Burkholderia cepacia] | 59 | 40 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 137 | 1 | 1 | 999 | gi\|1065948 | similar to thymidine diphosphoglucose 4,6-dehydratase [Caenorhabditis elegans] | 59 | 40 |
| 138 | 8 | 7489 | 6827 | gnl\|PID\|e264435 | Putative orf YCLX8c, len:192 [Saccharomyces cerevisiae] | 59 | 36 |
| 140 | 1 | 3 | 656 | gnl\|PID\|e254943 | unknown [Mycobacterium tuberculosis] | 59 | 32 |
| 165 | 13 | 10427 | 9849 | gi\|1732199 | PTS permease for mannose subunit IIIMan C terminal domain [Vibrio furnissii] | 59 | 37 |
| 167 | 1 | 2 | 1045 | gi\|1573128 | hypothetical [Haemophilus influenzae] | 59 | 38 |
| 173 | 2 | 430 | 2160 | gi\|1486244 | unknown [Bacillus subtilis] | 59 | 31 |
| 179 | 10 | 10432 | 11199 | gi\|288299 | ORF1 gene product [Bacillus megaterium] | 59 | 34 |
| 179 | 12 | 12117 | 13148 | gi\|1045964 | hypothetical protein (GB:U14003_297) [Mycoplasma genitalium] | 59 | 41 |
| 181 | 11 | 9684 | 8575 | gi\|1653152 | 3-dehydroquinate synthase [Synechocystis sp.] | 59 | 41 |
| 223 | 24 | 20736 | 21974 | gi\|1573051 | succinyl-diaminopimelate desuccinylase (dapE) [Haemophilus influenzae] | 59 | 48 |
| 229 | 12 | 12818 | 11421 | gi\|1652035 | fmu and fmv protein [Synechocystis sp.] | 59 | 39 |
| 244 | 3 | 2836 | 1565 | gi\|1303959 | YqjH [Bacillus subtilis] | 59 | 45 |
| 265 | 9 | 4116 | 3868 | gi\|311100 | translational activator [Saccharomyces cerevisiae] | 59 | 28 |
| 272 | 1 | 1 | 546 | gi\|490320 | Y gene product [unidentified] | 59 | 41 |
| 279 | 16 | 14774 | 14370 | gi\|1389549 | ORF3 [Bacillus subtilis] | 59 | 46 |
| 283 | 8 | 3222 | 3401 | gi\|153047 | lysostaphin (ttg start codon) [Staphylococcus simulans] pir\|A25881\|A25881 lysostaphin precursor - Staphylococcus simulans sp\|P10547\|LSTP_STASI LYSOSTAPHIN PRECURSOR (EC 3.5.1.-). | 59 | 43 |
| 288 | 5 | 2617 | 3144 | gi\|1142714 | phosphoenolpyruvate:mannose phosphotransferase element IIB [Lactobacillus curvatus] | 59 | 45 |
| 292 | 19 | 14837 | 16792 | gi\|495646 | ATPase [Transposon Tn5422] | 59 | 40 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 295 | 1 | 49 | 495 | gi\|533098 | DnaD protein [Bacillus subtilis] | 59 | 39 |
| 315 | 2 | 907 | 653 | gi\|1574802 | hypothetical [Haemophilus influenzae] | 59 | 38 |
| 318 | 6 | 4549 | 4058 | gi\|43941 | EIII-B Sor PTS [Klebsiella pneumoniae] | 59 | 35 |
| 345 | 3 | 2707 | 3507 | gi\|895749 | putative cellobiose phosphotransferase enzyme II'' [Bacillus ubtilis] | 59 | 38 |
| 351 | 5 | 2646 | 2371 | gi\|1666506 | RfbC [Leptospira interrogans] | 59 | 30 |
| 355 | 21 | 15237 | 17222 | gi\|515738 | ORF2; putative [Oenococcus oeni] | 59 | 35 |
| 384 | 1 | 14 | 754 | gi\|1162959 | homologous to HI0365 in Haemophilus influenzae; ORF1 [Pseudomonas aeruginosa] | 59 | 34 |
| 385 | 1 | 3 | 533 | gi\|1146197 | putative [Bacillus subtilis] | 59 | 36 |
| 394 | 13 | 13137 | 12160 | gnl\|PID\|e243582 | ORF YGR263c [Saccharomyces cerevisiae] | 59 | 36 |
| 399 | 1 | 224 | 580 | gi\|580904 | homologous to E.coli rnpA [Bacillus subtilis] | 59 | 38 |
| 412 | 1 | 3 | 2927 | gi\|1620648 | surface protein Rib [Streptococcus agalactiae] | 59 | 43 |
| 412 | 2 | 2918 | 3559 | gi\|1620648 | surface protein Rib [Streptococcus agalactiae] | 59 | 43 |
| 416 | 6 | 5283 | 3940 | gi\|1100076 | PTS-dependent enzyme II [Clostridium longisporum] | 59 | 38 |
| 437 | 2 | 1561 | 1136 | gi\|580866 | ipa-12d gene product [Bacillus subtilis] | 59 | 44 |
| 495 | 2 | 438 | 614 | gi\|1500472 | M. jannaschii predicted coding region MJ1577 [Methanococcus jannaschii] | 59 | 45 |
| 502 | 1 | 853 | 188 | gi\|1063248 | No homologous protein [Bacillus subtilis] | 59 | 25 |
| 573 | 8 | 5092 | 4493 | gi\|1573226 | hypothetical [Haemophilus influenzae] | 59 | 39 |
| 579 | 4 | 1716 | 2717 | gnl\|PID\|e280724 | unknown [Mycobacterium tuberculosis] | 59 | 41 |
| 600 | 1 | 1 | 504 | gi\|49386 | internal region of the penicillin-binding protein 2B gene treptococcus pneumoniae] | 59 | 40 |
| 616 | 3 | 904 | 533 | gi\|289265 | [Bacillus sp. (KSM 64) endo-1,4-beta-glucanase gene, complete cds.], ene products [Bacillus sp.] | 59 | 44 |
| 657 | 1 | 432 | 4 | gi\|1651338 | PnuC protein [Escherichia coli] | 59 | 37 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 699 | 1 | 416 | 165 | gnl\|PID\|e199096 | PepR1 [Lactobacillus delbrueckii] | 59 | 23 |
| 713 | 4 | 3709 | 2660 | gi\|515738 | ORF2; putative [Oenococcus oeni] | 59 | 37 |
| 715 | 1 | 698 | 84 | gi\|1176399 | EpiF [Staphylococcus epidermidis] | 59 | 42 |
| 737 | 2 | 660 | 199 | gi\|666000 | hypothetical protein [Bacillus subtilis] | 59 | 43 |
| 744 | 1 | 395 | 3 | gi\|1732057 | MUC.CL-1 [Trypanosoma cruzi] | 59 | 45 |
| 746 | 1 | 3 | 554 | gi\|141858 | replication-associated protein [Plasmid pAD1] | 59 | 36 |
| 869 | 1 | 2 | 250 | gi\|1432153 | cellobiose-specific PTS permease [Klebsiella oxytoca] | 59 | 40 |
| 4 | 8 | 6948 | 6067 | gi\|147516 | ribokinase [Escherichia coli] | 58 | 42 |
| 11 | 6 | 3312 | 4121 | gi\|1732200 | PTS permease for mannose subunit IIPMan [Vibrio furnissii] | 58 | 35 |
| 16 | 9 | 7684 | 6932 | gnl\|PID\|e233879 | hypothetical protein [Bacillus subtilis] | 58 | 48 |
| 23 | 14 | 7440 | 8903 | gi\|142940 | ftsA [Bacillus subtilis] | 58 | 39 |
| 30 | 2 | 570 | 1283 | gi\|1644202 | unknown [Bacillus subtilis] | 58 | 37 |
| 48 | 7 | 7186 | 8037 | gi\|1573247 | hypothetical [Haemophilus influenzae] | 58 | 35 |
| 49 | 7 | 2395 | 2871 | gnl\|PID\|e210884 | c2 gene product [Bacteriophage B1] | 58 | 34 |
| 54 | 1 | 1014 | 91 | gi\|46645 | ORF (rlx) [Staphylococcus aureus] | 58 | 46 |
| 55 | 3 | 1221 | 511 | gi\|726443 | No definition line found [Caenorhabditis elegans] | 58 | 41 |
| 58 | 1 | 1904 | 696 | gi\|1591564 | molybdenum cofactor biosynthesis moeA protein [Methanococcus jannaschii] | 58 | 39 |
| 58 | 8 | 7238 | 6996 | gi\|1279769 | FdhC [Methanobacterium thermoformicicum] | 58 | 54 |
| 72 | 12 | 12117 | 10897 | gi\|763052 | integrase [Bacteriophage T270] | 58 | 37 |
| 77 | 2 | 1155 | 1910 | gi\|1245464 | YfeA [Yersinia pestis] | 58 | 34 |
| 78 | 1 | 2589 | 49 | gi\|40663 | sialidase [Clostridium septicum] | 58 | 40 |
| 88 | 9 | 5854 | 6528 | gi\|1619623 | hemin binding protein [Yersinia enterocolitica] | 58 | 37 |
| 93 | 6 | 2639 | 2863 | gi\|405133 | putative [Bacillus subtilis] | 58 | 33 |
| 98 | 13 | 13523 | 12432 | gi\|147329 | transport protein [Escherichia coli] | 58 | 41 |
| 100 | 12 | 8550 | 8224 | gi\|1736642 | Invasin. [Escherichia coli] | 58 | 47 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 102 | 7 | 5688 | 5969 | gi\|808869 | human gcp372 [Homo sapiens] | 58 | 30 |
| 105 | 5 | 3716 | 4501 | gi\|143729 | transcription activator [Bacillus subtilis] | 58 | 40 |
| 107 | 1 | 511 | 2 | gi\|1303827 | YqfI [Bacillus subtilis] | 58 | 34 |
| 108 | 2 | 1040 | 1732 | gi\|1592142 | ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii] | 58 | 37 |
| 114 | 6 | 7608 | 8444 | gi\|152719 | flavocytochrome c [Shewanella putrefaciens] | 58 | 40 |
| 117 | 14 | 11813 | 11115 | gi\|1575577 | DNA-binding response regulator [Thermotoga maritima] | 58 | 42 |
| 122 | 1 | 1 | 936 | gi\|393269 | adhesion protein [Streptococcus pneumoniae] | 58 | 38 |
| 123 | 23 | 20379 | 21617 | gi\|1653948 | hypothetical protein [Synechocystis sp.] | 58 | 38 |
| 133 | 8 | 7362 | 8480 | gi\|143498 | degS protein [Bacillus subtilis] | 58 | 38 |
| 133 | 9 | 8437 | 9087 | gi\|143089 | iep protein [Bacillus subtilis] | 58 | 31 |
| 138 | 3 | 3551 | 2898 | gi\|216114 | DNA polymerase [Bacteriophage SPO1] | 58 | 41 |
| 138 | 5 | 5819 | 5049 | gnl\|PID\|e289148 | highly similar to phosphotransferase system regulator [Bacillus subtilis] | 58 | 38 |
| 138 | 17 | 11419 | 10379 | gi\|1674137 | (AE000044) Mycoplasma pneumoniae, lipoate protein ligase; similar to Swiss-Prot Accession Number P32099, from E. coli [Mycoplasma pneumoniae] | 58 | 37 |
| 139 | 8 | 5002 | 4808 | gi\|153607 | dpnD gene product [Streptococcus pneumoniae] | 58 | 43 |
| 146 | 9 | 7817 | 6627 | gi\|606076 | ORF_o384 [Escherichia coli] | 58 | 43 |
| 150 | 10 | 7529 | 7894 | gi\|141852 | sialidase [Actinomyces viscosus] | 58 | 28 |
| 152 | 10 | 5717 | 6637 | gi\|296356 | putative membrane transport protein [Clostridium perfringens] pir\|A56641\|A56641 probable membrane transport protein - Clostridium erfringens | 58 | 36 |
| 162 | 10 | 11009 | 11185 | gi\|42655 | pi protein [Escherichia coli] | 58 | 37 |
| 164 | 3 | 1793 | 1608 | gi\|881499 | parathion hydrolase (phosphotriesterase)- | 58 | 41 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | related protein [Mus usculus] | | |
| 165 | 6 | 5640 | 4975 | gi\|1146190 | 2-keto-3-deoxy-6-phosphogluconate aldolase [Bacillus subtilis] | 58 | 39 |
| 165 | 10 | 9038 | 8199 | gi\|606080 | ORF_o290; Geneplot suggests frameshift linking to o267, not found Escherichia coli] | 58 | 35 |
| 168 | 1 | 1 | 657 | gi\|413930 | ipa-6d gene product [Bacillus subtilis] | 58 | 41 |
| 170 | 1 | 923 | 234 | gi\|1573505 | hypothetical [Haemophilus influenzae] | 58 | 30 |
| 176 | 1 | 1 | 1101 | gi\|1652379 | cation-transporting P-ATPase [Synechocystis sp.] | 58 | 30 |
| 180 | 12 | 10237 | 10410 | gi\|408123 | V-ATPase 14kD subunit peptide [Drosophila melanogaster] pir\|S38436\|S38436 H+-transporting ATPase (EC 3.6.1.35) 14K chain - ruit fly (Drosophila melanogaster) | 58 | 33 |
| 193 | 3 | 2077 | 1388 | gi\|1256633 | putative [Bacillus subtilis] | 58 | 39 |
| 193 | 4 | 2602 | 2075 | gi\|147920 | 3-methyladenine-DNA glycosylase I (tag) [Escherichia coli] | 58 | 33 |
| 194 | 9 | 6492 | 5500 | sp\|P09997\|YIDA_ECOLI | HYPOTHETICAL 29.7 KD PROTEIN IN IBPA-GYRB INTERGENIC REGION. | 58 | 38 |
| 201 | 5 | 5152 | 4466 | gi\|755152 | highly hydrophobic integral membrane protein [Bacillus subtilis] sp\|P42953\|TAGG_BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGG. | 58 | 28 |
| 210 | 9 | 6546 | 7265 | gi\|466520 | pocR [Salmonella typhimurium] | 58 | 36 |
| 220 | 1 | 3 | 569 | gi\|467441 | expressed at the end of exponential growyh under condtions in which he enzymes of the TCA cycle are repressed [Bacillus subtilis] sp\|P14194\|CTC_BACSU GENERAL STRESS PROTEIN CTC. {SUB 2-204} gi\|40219 partial ctc gene product (AA 1-186) [Bacillus subtilis] | 58 | 38 |
| 222 | 10 | 6520 | 7143 | gi\|1674024 | (AE000033) Mycoplasma pneumoniae, | 58 | 41 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | hypothetical protein (yjfS) homolog; similar to Swiss-Prot Accession Number P39301, from E. coli [Mycoplasma pneumoniae] | | |
| 233 | 7 | 4984 | 3944 | gi\|147806 | selenium metabolism protein [Escherichia coli] | 58 | 45 |
| 238 | 14 | 12128 | 12910 | gi\|1736468 | Pectin degradation repressor protein KdgR. [Escherichia coli] | 58 | 37 |
| 244 | 11 | 8102 | 7809 | gi\|467418 | unknown [Bacillus subtilis] | 58 | 37 |
| 246 | 1 | 1 | 276 | gi\|65291 | receptor tyrosine kiase preprotein [Xiphophorus sp.} ir\|S06142\|S06142 kinase-related transforming protein (Tu) (EC 7.1.-) precursor - southern platyfish | 58 | 32 |
| 255 | 4 | 2927 | 2559 | gi\|1652384 | ABC transporter [Synechocystis sp.] | 58 | 41 |
| 258 | 9 | 8025 | 8966 | gi\|147402 | mannose permease subunit III-Man [Escherichia coli] | 58 | 35 |
| 259 | 2 | 1801 | 893 | gi\|1591564 | molybdenum cofactor biosynthesis moeA protein [Methanococcus jannaschii] | 58 | 39 |
| 260 | 3 | 1754 | 2254 | gi\|580841 | F1 [Bacillus subtilis] | 58 | 38 |
| 271 | 4 | 2382 | 2738 | gi\|40067 | X gene product [Bacillus sphaericus] | 58 | 37 |
| 279 | 8 | 6237 | 6536 | gi\|1783243 | homologous to jojC gene product (B. subtilis; prf:2111327a); hypothetical [Bacillus subtilis] | 58 | 34 |
| 301 | 1 | 753 | 175 | gi\|499196 | ORF1 [Streptomyces lincolnensis] | 58 | 37 |
| 304 | 1 | 100 | 849 | gi\|1653322 | hypothetical protein [Synechocystis sp.] | 58 | 41 |
| 313 | 2 | 748 | 1650 | gi\|1658371 | cyclic beta-1,2-glucan modification protein [Rhizobium meliloti] | 58 | 36 |
| 321 | 11 | 6033 | 6533 | gi\|1573292 | hypothetical [Haemophilus influenzae] | 58 | 34 |
| 322 | 6 | 3819 | 5069 | gi\|23897 | 5'-nucleotidase [Homo sapiens] | 58 | 34 |
| 324 | 5 | 3259 | 4452 | gi\|1469784 | putative cell division protein ftsW [Enterococcus hirae] | 58 | 37 |
| 328 | 1 | 1 | 270 | gi\|882579 | CG Site No. 29739 [Escherichia coli] | 58 | 43 |

Table 2: E. *faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 330 | 8 | 6228 | 6758 | gi\|43941 | EIII-B Sor PTS [Klebsiella pneumoniae] | 58 | 37 |
| 334 | 4 | 3634 | 3963 | gi\|1001306 | hypothetical protein [Synechocystis sp.] | 58 | 34 |
| 345 | 17 | 18899 | 20044 | gi\|853809 | ORF3 [Clostridium perfringens] | 58 | 30 |
| 363 | 7 | 8475 | 9944 | gi\|348056 | trans-acting positive regulator [Bacillus anthracis] | 58 | 33 |
| 375 | 7 | 6472 | 5279 | gi\|1408501 | homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis] | 58 | 42 |
| 394 | 12 | 10689 | 12095 | gi\|537034 | ORF_o488 [Escherichia coli] | 58 | 32 |
| 399 | 3 | 1383 | 2198 | gi\|580905 | B.subtilis genes rpmH, rnpA, 50kd, gidA and gidB [Bacillus subtilis] gi\|580919 Jag [Bacillus subtilis] | 58 | 36 |
| 399 | 16 | 11544 | 12098 | gi\|1572965 | hypothetical [Haemophilus influenzae] | 58 | 39 |
| 399 | 19 | 14776 | 15654 | gi\|1778530 | CitG homolog [Escherichia coli] | 58 | 40 |
| 407 | 2 | 738 | 553 | gi\|170553 | pyruvate kinase [Trichoderma reesei] | 58 | 38 |
| 416 | 5 | 4045 | 3389 | gi\|475112 | enzyme IIabc [Pediococcus pentosaceus] | 58 | 41 |
| 449 | 4 | 1421 | 879 | gi\|928834 | integrase [Lactococcus lactis phage BK5-T] | 58 | 32 |
| 497 | 1 | 3 | 458 | gi\|160628 | reticulocyte binding protein 2 [Plasmodium vivax] | 58 | 30 |
| 594 | 1 | 285 | 4 | gi\|1353874 | unknown [Rhodobacter capsulatus] | 58 | 39 |
| 637 | 6 | 3451 | 2765 | pir\|D61615\|D61615 | sericin MG-1 - greater wax moth (fragment) | 58 | 52 |
| 653 | 1 | 595 | 245 | gi\|1408585 | LtrD [Lactococcus lactis lactis] | 58 | 41 |
| 656 | 4 | 3713 | 5209 | sp\|P13692\|P54_ENTFC | P54 PROTEIN PRECURSOR. | 58 | 37 |
| 656 | 6 | 5988 | 6467 | gi\|1017818 | phosphotyrosine protein phosphatase [Streptomyces coelicolor] | 58 | 48 |
| 667 | 1 | 88 | 1467 | bbs\|177441 | OsNramp1=Nramp1 homolog/Bcg product homolog [Oryza sativa, indica, cv. IR 36, etiolated shoots, Peptide, 517 aa] [Oryza sativa] | 58 | 40 |
| 686 | 1 | 892 | 233 | pir\|A24255\|A24255 | chorion class A protein L11 precursor - | 58 | 38 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | silkworm | | |
| 706 | 1 | 1002 | 607 | gi\|1001762 | hypothetical protein [Synechocystis sp.] | 58 | 32 |
| 801 | 1 | 254 | 12 | gnl\|PID\|e243641 | unknown [Mycobacterium tuberculosis] | 58 | 29 |
| 848 | 1 | 212 | 3 | gnl\|PID\|e254644 | membrane protein [Streptococcus pneumoniae] | 58 | 37 |
| 975 | 1 | 3 | 422 | gi\|290545 | f270 [Escherichia coli] | 58 | 35 |
| 11 | 4 | 2345 | 2833 | gi\|1439527 | EIIA-man [Lactobacillus curvatus] | 57 | 46 |
| 16 | 2 | 1426 | 365 | gi\|780550 | acetyl transferase [Rhizobium loti] | 57 | 35 |
| 18 | 3 | 1593 | 925 | gnl\|PID\|e137594 | xerC recombinase [Lactobacillus leichmannii] | 57 | 36 |
| 19 | 15 | 8058 | 8267 | gi\|1590922 | cell division inhibitor [Methanococcus jannaschii] | 57 | 42 |
| 19 | 23 | 11938 | 12318 | gi\|1294760 | structural protein; orfL3; putative [Bacteriophage phi-41] | 57 | 46 |
| 25 | 9 | 7743 | 6958 | gnl\|PID\|e255000 | hypothetical protein [Bacillus subtilis] | 57 | 40 |
| 47 | 3 | 3857 | 4462 | gi\|1353540 | ORF23 [Bacteriophage r1t] | 57 | 35 |
| 65 | 10 | 7180 | 8919 | gi\|496254 | fibronectin/fibrinogen-binding protein [Streptococcus pyogenes] | 57 | 40 |
| 68 | 7 | 3923 | 3705 | gi\|336656 | ribosomal protein secY [Cyanophora paradoxa] | 57 | 28 |
| 70 | 4 | 2317 | 3645 | pir\|S11158\|YESAEE | erythromycin resistance protein - Staphylococcus epidermidis plasmid pUL5050 | 57 | 40 |
| 76 | 1 | 55 | 1095 | gi\|1353562 | Structural protein [Bacteriophage r1t] | 57 | 41 |
| 91 | 11 | 9070 | 8849 | gi\|550321 | beta-fructofuranosidase [Chenopodium rubrum] | 57 | 30 |
| 94 | 4 | 1740 | 1495 | gi\|47406 | penicillin-binding protein 1a [Streptococcus pneumoniae] ir\|S28031\|S28031 penicillin-binding protein 1a - Streptococcus eumoniae (strain 456) (fragment) | 57 | 30 |
| 98 | 6 | 7766 | 6849 | gi\|409286 | bmrU [Bacillus subtilis] | 57 | 31 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 100 | 22 | 17294 | 15912 | gnl\|PID\|e289150 | member of the SNF2 helicase family [Bacillus subtilis] | 57 | 30 |
| 102 | 1 | 66 | 2465 | gi\|405564 | traE [Plasmid pSK41] | 57 | 28 |
| 110 | 14 | 11757 | 12497 | gi\|854601 | unknown [Schizosaccharomyces pombe] | 57 | 38 |
| 114 | 9 | 10291 | 11139 | gi\|853777 | product similar to E.coli PRFA2 protein [Bacillus subtilis] pir\|S55438\|S55438 ywkE protein - Bacillus subtilis sp\|P45873\|HEMK_BACSU POSSIBLE PROTOPORPHYRINOGEN OXIDASE (EC .3.3.-). | 57 | 38 |
| 115 | 3 | 955 | 1461 | gi\|396347 | alternate name yjaB [Escherichia coli] | 57 | 33 |
| 123 | 3 | 1925 | 2932 | gi\|1001731 | low affinity sulfate transporter [Synechocystis sp.] | 57 | 39 |
| 124 | 7 | 6026 | 5118 | gi\|1674310 | (AE000058) Mycoplasma pneumoniae, MG085 homolog, from M. genitalium [Mycoplasma pneumoniae] | 57 | 30 |
| 128 | 9 | 7530 | 6235 | gi\|413940 | ipa-16d gene product [Bacillus subtilis] | 57 | 36 |
| 128 | 31 | 25487 | 25206 | gi\|1651915 | hypothetical protein [Synechocystis sp.] | 57 | 42 |
| 128 | 33 | 26878 | 26150 | gi\|1001387 | hypothetical protein [Synechocystis sp.] | 57 | 30 |
| 128 | 37 | 30730 | 29600 | gi\|406877 | DivIB protein [Bacillus licheniformis] | 57 | 35 |
| 130 | 9 | 7408 | 8556 | gi\|343539 | NADH dehydrogenase subunit 4 [Trypanosoma brucei] | 57 | 27 |
| 144 | 1 | 1013 | 219 | gi\|1652518 | hypothetical protein [Synechocystis sp.] | 57 | 45 |
| 144 | 6 | 4145 | 5254 | gi\|149581 | maturation protein [Lactobacillus paracasei] | 57 | 38 |
| 146 | 1 | 617 | 192 | gi\|147402 | mannose permease subunit III-Man [Escherichia coli] | 57 | 33 |
| 153 | 1 | 83 | 991 | gi\|147336 | transmembrane protein [Escherichia coli] | 57 | 33 |
| 160 | 8 | 4718 | 4134 | gi\|305333 | zeta-crystallin [Cavia porcellus] | 57 | 39 |
| 167 | 8 | 14891 | 14688 | gi\|206354 | protein kinase C, zeta subspecies [Rattus norvegicus] pir\|A30314\|A30314 protein kinase C (EC 2.7.1.-) zeta - rat sp\|P09217\|KPCZ_RAT PROTEIN KINASE C, ZETA | 57 | 39 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | TYPE (EC 2.7.1.-) NPKC-ZETA). | | |
| 174 | 1 | 760 | 2 | gnl\|PID\|e191403 | ORFA gene product [Chloroflexus aurantiacus] | 57 | 42 |
| 176 | 4 | 3347 | 3568 | gi\|1236529 | cyclomaltodextrinase [Bacillus sp.] | 57 | 46 |
| 194 | 8 | 4786 | 5457 | gi\|405516 | This ORF is homologous to nitroreductase from Enterobacter cloacae, ccession Number A38686, and Salmonella, Accession Number P15888 Mycoplasma-like organism] | 57 | 26 |
| 199 | 3 | 3207 | 3764 | gi\|216350 | ORF [Bacillus subtilis] | 57 | 38 |
| 202 | 5 | 3356 | 3664 | gi\|1183841 | Holliday junction binding protein [Pseudomonas aeruginosa] | 57 | 34 |
| 202 | 12 | 10911 | 10192 | gi\|971338 | anaerobic regulatory protein [Bacillus subtilis] | 57 | 27 |
| 205 | 3 | 1022 | 468 | gi\|1783240 | hypothetical [Bacillus subtilis] | 57 | 38 |
| 223 | 2 | 779 | 1501 | gi\|1208965 | hypothetical 23.3 kd protein [Escherichia coli] | 57 | 32 |
| 223 | 3 | 1499 | 2332 | gi\|303560 | ORF271 [Escherichia coli] | 57 | 35 |
| 223 | 11 | 8404 | 12198 | gi\|158079 | period protein [Drosophila serrata] | 57 | 40 |
| 237 | 9 | 3685 | 3906 | gi\|514919 | phosphofructokinase [Drosophila melanogaster] | 57 | 31 |
| 242 | 7 | 5760 | 5020 | gi\|1574596 | H. influenzae predicted coding region HI1738 [Haemophilus influenzae] | 57 | 33 |
| 250 | 2 | 1243 | 1485 | gnl\|PID\|e275819 | K08G2.8 [Caenorhabditis elegans] | 57 | 47 |
| 276 | 28 | 16565 | 16332 | gi\|886375 | variant-specific surface protein [Plasmodium falciparum] | 57 | 47 |
| 288 | 6 | 3157 | 3363 | gi\|147403 | mannose permease subunit II-P-Man [Escherichia coli] | 57 | 39 |
| 289 | 1 | 141 | 818 | gi\|1742822 | Phosphoglycolate phosphatase (EC 3.1.3.18). [Escherichia coli] | 57 | 40 |
| 292 | 20 | 15930 | 15721 | gi\|854201 | putative polymerase [Infectious bursal disease virus] | 57 | 47 |
| 294 | 4 | 1454 | 2014 | gi\|454303 | LDJ2 gene product [Allium porrum] | 57 | 41 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 295 | 4 | 2052 | 2342 | pir\|S48588\|S48588 | hypothetical protein - Mycoplasma capricolum (SGC3) (fragment) | 57 | 39 |
| 301 | 14 | 10921 | 10148 | gnl\|PID\|e262045 | putative orf [Bacillus subtilis] | 57 | 38 |
| 306 | 1 | 2 | 793 | gi\|216715 | HpaI methyltransferase [Haemophilus parainfluenzae] pir\|S28681\|S28681 site-specific DNA-methyltransferase adenine-specific) (EC 2.1.1.72) HpaI - Haemophilus parainfluenzae sp\|P29538\|MTH1_HAEPA MODIFICATION METHYLASE HPAI (EC 2.1.1.72) ADENINE-SPECIFIC MET | 57 | 36 |
| 306 | 8 | 5418 | 5663 | gi\|1591542 | M. jannaschii predicted coding region MJ0857 [Methanococcus jannaschii] | 57 | 42 |
| 308 | 2 | 1732 | 1487 | gi\|1518045 | FlbF protein [Borrelia burgdorferi] | 57 | 28 |
| 321 | 2 | 1030 | 1458 | gi\|606080 | ORF_o290; Geneplot suggests frameshift linking to o267, not found Escherichia coli] | 57 | 30 |
| 351 | 4 | 2342 | 1587 | gi\|1591853 | M. jannaschii predicted coding region MJ1222 [Methanococcus jannaschii] | 57 | 37 |
| 355 | 30 | 20619 | 20861 | gi\|1136394 | There are three putative hydrophobic domains in the central region. [Homo sapiens] | 57 | 42 |
| 364 | 10 | 9415 | 8852 | gi\|38722 | precursor (aa -20 to 381) [Acinetobacter calcoaceticus] ir\|A29277\|A29277 aldose 1-epimerase (EC 5.1.3.3) - Acinetobacter lcoaceticus | 57 | 32 |
| 365 | 3 | 4715 | 1812 | gi\|914990 | Similar to DEAD box family helicases [Saccharomyces cerevisiae] pir\|S59797\|S59797 hypothetical protein D9798.1 - yeast Saccharomyces cerevisiae) | 57 | 35 |
| 378 | 1 | 615 | 10 | gi\|1652989 | hypothetical protein [Synechocystis sp.] | 57 | 35 |
| 379 | 1 | 1457 | 114 | gi\|1256618 | transport protein [Bacillus subtilis] | 57 | 36 |
| 390 | 1 | 1426 | 2 | gi\|387880 | collagen adhesin [Staphylococcus aureus] | 57 | 37 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 422 | 1 | 2 | 409 | gi\|1591837 | M. jannaschii predicted coding region MJ1207 [Methanococcus jannaschii] | 57 | 37 |
| 447 | 1 | 397 | 131 | gi\|214566 | keratin protein XK81 [Xenopus laevis] | 57 | 33 |
| 454 | 2 | 1095 | 889 | gi\|1783256 | sigma factor [Bacillus subtilis] | 57 | 28 |
| 504 | 2 | 641 | 1426 | gi\|42081 | nagD gene product (AA 1-250) [Escherichia coli] | 57 | 32 |
| 524 | 2 | 963 | 577 | gi\|143724 | putative [Bacillus subtilis] | 57 | 43 |
| 535 | 4 | 4862 | 4305 | gi\|146549 | kdpC [Escherichia coli] | 57 | 40 |
| 547 | 2 | 426 | 719 | gi\|533098 | DnaD protein [Bacillus subtilis] | 57 | 33 |
| 548 | 1 | 316 | 717 | gi\|397973 | Mg2+ transport ATPase [Salmonella typhimurium] | 57 | 33 |
| 639 | 2 | 359 | 105 | gnl\|PID\|e247390 | P-type ATPase [Dictyostelium discoideum] | 57 | 31 |
| 641 | 1 | 941 | 180 | gnl\|PID\|e261990 | putative orf [Bacillus subtilis] | 57 | 36 |
| 686 | 3 | 1298 | 3259 | gi\|496506 | orf gamma [Streptococcus pyogenes] | 57 | 37 |
| 686 | 6 | 2200 | 2847 | gi\|404800 | putative [Saccharopolyspora erythraea] | 57 | 47 |
| 782 | 2 | 591 | 860 | gi\|1591270 | alanyl-tRNA synthetase [Methanococcus jannaschii] | 57 | 32 |
| 844 | 1 | 3 | 182 | gi\|849217 | Weak similarity to Streptococcus Protein V, a type-II IgG receptor PIR accession number S17354) and Giardia lamblia median body rotein (PIR accession number S33821) [Saccharomyces cerevisiae] pir\|S61181\|S61181 hypothetical protein D9740.10 - yeast Sacchar | 57 | 34 |
| 859 | 1 | 174 | 4 | gi\|1762584 | polygalacturonase isoenzyme 1 beta subunit homolog [Arabidopsis thaliana] | 57 | 28 |
| 967 | 1 | 381 | 4 | gi\|309662 | pheromone binding protein [Plasmid pCF10] | 57 | 40 |
| 11 | 5 | 2817 | 3314 | gi\|43941 | EIII-B Sor PTS [Klebsiella pneumoniae] | 56 | 30 |
| 15 | 1 | 80 | 892 | gi\|1574803 | spermidine/putrescine-binding periplasmic protein precursor (potD) [Haemophilus influenzae] | 56 | 32 |
| 37 | 8 | 6327 | 6088 | gi\|290561 | o188 [Escherichia coli] | 56 | 41 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 44 | 2 | 1169 | 1360 | gi\|16096 | peroxidase [Armoracia rusticana] | 56 | 37 |
| 56 | 3 | 1881 | 1363 | gi\|49272 | Asparaginase [Bacillus licheniformis] | 56 | 33 |
| 65 | 1 | 102 | 887 | gi\|1377832 | unknown [Bacillus subtilis] | 56 | 41 |
| 75 | 9 | 5817 | 4306 | gi\|1235712 | polyprotein [Infectious pancreatic necrosis virus] | 56 | 30 |
| 83 | 7 | 3260 | 4051 | gi\|1652645 | phosphoglycolate phosphatase [Synechocystis sp.] | 56 | 30 |
| 95 | 3 | 1793 | 2389 | pir\|C53610\|C53610 | ntpE protein - Enterococcus hirae | 56 | 28 |
| 100 | 3 | 5076 | 1915 | gi\|1353559 | ORF42 [Bacteriophage r1t] | 56 | 35 |
| 100 | 16 | 10581 | 10369 | gi\|868224 | No definition line found [Caenorhabditis elegans] | 56 | 35 |
| 100 | 48 | 31841 | 32770 | gi\|460025 | ORF2, putative [Streptococcus pneumoniae] | 56 | 38 |
| 108 | 5 | 4007 | 3336 | gi\|288301 | ORF2 gene product [Bacillus megaterium] | 56 | 34 |
| 109 | 2 | 1032 | 325 | gi\|413976 | ipa-52r gene product [Bacillus subtilis] | 56 | 36 |
| 119 | 7 | 3958 | 5304 | gi\|498842 | VirS [Clostridium perfringens] | 56 | 35 |
| 123 | 32 | 29479 | 30345 | gi\|39981 | P30 [Bacillus subtilis] | 56 | 38 |
| 126 | 1 | 521 | 3 | gi\|147403 | mannose permease subunit II-P-Man [Escherichia coli] | 56 | 29 |
| 130 | 6 | 4296 | 6104 | gi\|308854 | oligopeptide binding protein [Lactococcus lactis] | 56 | 33 |
| 131 | 7 | 5267 | 6613 | gi\|466589 | CG Site No. 39 [Escherichia coli] | 56 | 32 |
| 133 | 5 | 4358 | 5758 | gi\|1573431 | aminodeoxychorismate lyase (pabC) [Haemophilus influenzae] | 56 | 40 |
| 138 | 20 | 13680 | 12670 | gi\|1590951 | UDP-glucose 4-epimerase [Methanococcus jannaschii] | 56 | 40 |
| 138 | 29 | 19764 | 18823 | gi\|44864 | H.8 outer membrane protein (AA -17 to 71) [Neisseria gonorrhoeae] ir\|S02720\|S02720 outer membrane protein H.8 precursor - Neisseria norrhoeae | 56 | 33 |
| 145 | 7 | 5611 | 7179 | gi\|1652892 | ABC transporter [Synechocystis sp.] | 56 | 33 |
| 146 | 10 | 8545 | 7811 | gi\|41519 | P30 protein (AA 1-240) [Escherichia coli] | 56 | 28 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 150 | 4 | 2979 | 4637 | gi\|309662 | pheromone binding protein [Plasmid pCF10] | 56 | 32 |
| 159 | 5 | 5362 | 5066 | gi\|576733 | apocytochrome b [Trypanoplasma borreli] | 56 | 43 |
| 164 | 13 | 8864 | 15031 | gi\|1654116 | protein F2 [Streptococcus pyogenes] | 56 | 43 |
| 179 | 7 | 7790 | 9118 | gi\|413926 | ipa-2r gene product [Bacillus subtilis] | 56 | 33 |
| 187 | 4 | 2239 | 1667 | gi\|1573061 | hypothetical [Haemophilus influenzae] | 56 | 18 |
| 200 | 19 | 11473 | 10724 | gi\|498817 | ORF8; homologous to small subunit of phage terminases [Bacillus ubtilis] | 56 | 35 |
| 206 | 6 | 3766 | 2759 | gi\|474837 | ORF1 [Thermoanaerobacterium thermosulfurigenes] sp\|P38541\|YAMB_THETU HYPOTHETICAL 35.6 KD PROTEIN IN AMYB 5'REGION ORF1). | 56 | 34 |
| 207 | 2 | 2091 | 1672 | gi\|1204258 | soluble protein [Escherichia coli] | 56 | 40 |
| 217 | 9 | 6661 | 6158 | gi\|1017427 | elastic titin [Homo sapiens] | 56 | 28 |
| 225 | 7 | 6007 | 5099 | gi\|1742675 | Phosphotransferase system enzyme II (EC 2.7.1.69) MalX [Escherichia coli] | 56 | 46 |
| 230 | 3 | 595 | 3153 | gi\|437706 | alternative truncated translation product from E.coli [Streptococcus neumoniae] | 56 | 34 |
| 236 | 2 | 1486 | 515 | gi\|415664 | catabolite control protein [Bacillus megaterium] sp\|P46828\|CCPA_BACME GLUCOSE-RESISTANCE AMYLASE REGULATOR CATABOLITE CONTROL PROTEIN). | 56 | 35 |
| 236 | 7 | 9255 | 8599 | gi\|343544 | ATPase 6 [Trypanosoma brucei] | 56 | 48 |
| 238 | 15 | 13059 | 13718 | gi\|1146190 | 2-keto-3-deoxy-6-phosphogluconate aldolase [Bacillus subtilis] | 56 | 37 |
| 238 | 20 | 17734 | 18756 | gi\|1574060 | hypothetical [Haemophilus influenzae] | 56 | 32 |
| 238 | 23 | 21613 | 20726 | gi\|151361 | member of the AraC/XylS family of transcriptional regulators Pseudomonas aeruginosa] | 56 | 36 |
| 242 | 6 | 4103 | 4477 | gi\|886858 | nicotinic acetylcholine receptor [Caenorhabditis elegans] pir\|S57648\|S57648 nicotinic acetylcholine receptor - Caenorhabditis legans | 56 | 35 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 260 | 5 | 3170 | 3781 | gnl\|PID\|e58151 | F3 [Bacillus subtilis] | 56 | 43 |
| 279 | 6 | 5140 | 2831 | gi\|581100 | gamma-glutamylcysteine synthetase (aa 1-518) [Escherichia coli] pir\|A24136\|SYECEC glutamate--cysteine ligase (EC 6.3.2.2) - scherichia coli | 56 | 42 |
| 279 | 9 | 6434 | 7228 | gi\|1783243 | homologous to jojC gene product (B. subtilis; prf:2111327a); hypothetical [Bacillus subtilis] | 56 | 29 |
| 292 | 14 | 10719 | 11504 | gi\|45738 | ORFC [Enterococcus faecalis] | 56 | 37 |
| 313 | 3 | 3039 | 1831 | gi\|474915 | orf 337; translated orf similarity to SW: BCR_ECOLI bicyclomycin esistance protein of Escherichia coli [Coxiella burnetii] pir\|S44207\|S44207 hypothetical protein 337 - Coxiella burnetii (SUB -338) | 56 | 31 |
| 313 | 5 | 4233 | 3589 | gi\|405883 | yeiL [Escherichia coli] | 56 | 30 |
| 322 | 5 | 1994 | 3715 | gi\|1377831 | unknown [Bacillus subtilis] | 56 | 34 |
| 353 | 2 | 2353 | 1310 | gnl\|PID\|e254644 | membrane protein [Streptococcus pneumoniae] | 56 | 26 |
| 394 | 14 | 13289 | 14143 | gi\|142836 | repressor protein [Bacillus subtilis] | 56 | 30 |
| 399 | 32 | 30208 | 30891 | gi\|396293 | similar to Bacillus subtilis hypoth. 20 kDa protein, in tsr 3' egion [Escherichia coli] | 56 | 38 |
| 402 | 2 | 1267 | 914 | gi\|170710 | alpha-type gliadin precursor protein [Triticum aestivum] | 56 | 45 |
| 408 | 4 | 2825 | 2220 | gnl\|PID\|e257696 | collagen binding protein [Lactobacillus reuteri] | 56 | 36 |
| 432 | 5 | 3105 | 3302 | gi\|11678 | atpE gene product [Marchantia polymorpha] | 56 | 33 |
| 443 | 2 | 844 | 1089 | gi\|1256138 | YbbI [Bacillus subtilis] | 56 | 36 |
| 499 | 2 | 875 | 1666 | gi\|1499876 | magnesium and cobalt transport protein [Methanococcus jannaschii] | 56 | 30 |
| 510 | 6 | 3864 | 4733 | gi\|147404 | mannose permease subunit II-M-Man [Escherichia coli] | 56 | 34 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 543 | 6 | 3706 | 3113 | gi\|563812 | XCAP-C [Xenopus laevis] | 56 | 32 |
| 609 | 2 | 390 | 653 | gi\|48745 | principal sigma subunit (AA 1-442) [Streptomyces coelicolor] ir\|S11712\|S11712 translation initiation factor sigma hrdB - reptomyces coelicolor | 56 | 37 |
| 626 | 2 | 1124 | 2104 | gi\|950197 | unknown [Corynebacterium glutamicum] | 56 | 40 |
| 787 | 1 | 2 | 634 | gnl\|PID\|e283826 | orf c04012 [Sulfolobus solfataricus] | 56 | 26 |
| 820 | 1 | 1220 | 3 | gi\|44001 | galactose-1-P-uridyl transferase [Lactobacillus helveticus] ir\|B47032\|B47032 galactose-1-phosphate uridyl transferase - ctobacillus helveticus | 56 | 35 |
| 875 | 1 | 1 | 144 | gi\|455178 | 16K protein [Escherichia coli] | 56 | 46 |
| 906 | 2 | 307 | 846 | gi\|144858 | ORF A [Clostridium perfringens] | 56 | 34 |
| 941 | 1 | 3 | 335 | gi\|160299 | glutamic acid-rich protein [Plasmodium falciparum] pir\|A54514\|A54514 glutamic acid-rich protein precursor - Plasmodium alciparum | 56 | 23 |
| 5 | 5 | 2451 | 2951 | gi\|1303811 | YqeU [Bacillus subtilis] | 55 | 39 |
| 8 | 10 | 8312 | 7947 | gi\|1196907 | daunorubicin resistance protein [Streptomyces peucetius] | 55 | 29 |
| 17 | 24 | 23626 | 24465 | gnl\|PID\|e285322 | RecX protein [Mycobacterium smegmatis] | 55 | 28 |
| 17 | 31 | 31027 | 30344 | gi\|143830 | xpaC [Bacillus subtilis] | 55 | 22 |
| 17 | 34 | 31991 | 32302 | gnl\|PID\|e229183 | C11G6.3 [Caenorhabditis elegans] | 55 | 34 |
| 30 | 1 | 2 | 478 | pir\|S10655\|S10655 | hypothetical protein X - Pyrococcus woesei (fragment) | 55 | 34 |
| 49 | 14 | 9998 | 10411 | gi\|455154 | ORF D [Clostridium perfringens] | 55 | 36 |
| 54 | 3 | 955 | 1332 | gnl\|PID\|e238660 | hypothetical protein [Bacillus subtilis] | 55 | 32 |
| 54 | 10 | 3527 | 3231 | pir\|JQ0405\|JQ0405 | hypothetical 119.5K protein (uvrA region) - Micrococcus luteus | 55 | 45 |
| 67 | 4 | 2313 | 3044 | gi\|555750 | unknown [Neisseria gonorrhoeae] | 55 | 42 |
| 69 | 4 | 2250 | 2020 | gnl\|PID\|e259955 | K04G11.5 [Caenorhabditis elegans] | 55 | 33 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 77 | 5 | 3954 | 2938 | gi\|1001634 | hypothetical protein [Synechocystis sp.] | 55 | 34 |
| 80 | 4 | 4806 | 2482 | gi\|466952 | B1620_F1_30 [Mycobacterium leprae] | 55 | 35 |
| 81 | 6 | 4212 | 3730 | gi\|606073 | ORF_o169 [Escherichia coli] | 55 | 34 |
| 83 | 1 | 66 | 737 | gi\|216064 | morphogenesis protein B [Bacteriophage PZA] | 55 | 36 |
| 89 | 10 | 9486 | 7714 | gi\|148221 | DNA-dependent ATPase, DNA helicase [Escherichia coli] pir\|JS0137\|BVECRQ recQ protein - Escherichia coli | 55 | 35 |
| 91 | 5 | 2507 | 3289 | gi\|153015 | FemA protein [Staphylococcus aureus] | 55 | 35 |
| 100 | 14 | 9974 | 9393 | gi\|558603 | synaptonemal complex protein 1 [Mus musculus] | 55 | 30 |
| 116 | 1 | 1 | 909 | gi\|473901 | ORF1 [Lactococcus lactis] | 55 | 33 |
| 122 | 3 | 1801 | 2655 | gi\|1016216 | putative protein of 299 amino acids [Cyanophora paradoxa] | 55 | 28 |
| 123 | 30 | 28191 | 28721 | gi\|1142714 | phosphoenolpyruvate:mannose phosphotransferase element IIB [Lactobacillus curvatus] | 55 | 29 |
| 128 | 22 | 16664 | 16029 | gi\|606025 | ORF_o221 [Escherichia coli] | 55 | 42 |
| 150 | 7 | 5949 | 6521 | gi\|39573 | P20 (AA 1-178) [Bacillus licheniformis] | 55 | 32 |
| 155 | 7 | 5767 | 6660 | gi\|1763974 | DPPA [Bacillus methanolicus] | 55 | 31 |
| 157 | 1 | 867 | 70 | gi\|1067010 | M153.1 [Caenorhabditis elegans] | 55 | 34 |
| 160 | 9 | 6090 | 4804 | gi\|1592141 | M. jannaschii predicted coding region MJ1507 [Methanococcus jannaschii] | 55 | 31 |
| 176 | 3 | 2060 | 3349 | gi\|153858 | wall-associated protein [Streptococcus mutans] | 55 | 37 |
| 201 | 2 | 3277 | 413 | gi\|1235662 | RfbC [Myxococcus xanthus] | 55 | 36 |
| 202 | 9 | 6199 | 8001 | gi\|606018 | ORF_o783 [Escherichia coli] | 55 | 42 |
| 222 | 7 | 4803 | 4021 | gnl\|PID\|e289148 | highly similar to phosphotransferase system regulator [Bacillus subtilis] | 55 | 40 |
| 238 | 12 | 11465 | 9942 | gnl\|PID\|e266573 | unknown [Mycobacterium tuberculosis] | 55 | 27 |
| 238 | 13 | 11527 | 12027 | gi\|1129093 | unknown protein [Bacillus sp.] | 55 | 36 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 240 | 4 | 1988 | 1215 | gnl\|PID\|e252616 | DcuC protein [Escherichia coli] | 55 | 34 |
| 246 | 2 | 433 | 792 | gnl\|PID\|e233868 | hypothetical protein [Bacillus subtilis] | 55 | 25 |
| 253 | 5 | 1827 | 1549 | gi\|142540 | aspartokinase II [Bacillus sp.] | 55 | 48 |
| 259 | 1 | 895 | 74 | gi\|1006621 | molybdate-binding periplasmic protein [Synechocystis sp.] | 55 | 37 |
| 267 | 1 | 1183 | 2 | gi\|882672 | ORF_o313 [Escherichia coli] | 55 | 27 |
| 292 | 16 | 12843 | 13325 | gi\|561746 | cyclin-dependent protein kinase [Mus musculus] | 55 | 26 |
| 294 | 9 | 3390 | 3752 | gi\|984582 | DinJ [Escherichia coli] | 55 | 26 |
| 300 | 5 | 3914 | 3582 | gi\|1591957 | M. jannaschii predicted coding region MJ1318 [Methanococcus jannaschii] | 55 | 38 |
| 305 | 3 | 2769 | 3527 | gi\|606309 | ORF_o265; gtg start [Escherichia coli] | 55 | 36 |
| 320 | 6 | 4479 | 3475 | gi\|1591732 | cobalt transport ATP-binding protein O [Methanococcus jannaschii] | 55 | 32 |
| 355 | 24 | 18149 | 18322 | gi\|344751 | MDV TK gene product [unidentified] | 55 | 40 |
| 364 | 2 | 2083 | 386 | gi\|1573045 | hypothetical [Haemophilus influenzae] | 55 | 40 |
| 364 | 9 | 8796 | 8575 | gnl\|PID\|e252108 | ORF YOR255w [Saccharomyces cerevisiae] | 55 | 27 |
| 379 | 8 | 8248 | 6872 | gi\|1330236 | dihydropyrimidinase [Homo sapiens] | 55 | 37 |
| 386 | 6 | 3847 | 4332 | gi\|976025 | HrsA [Escherichia coli] | 55 | 27 |
| 441 | 2 | 939 | 1730 | gi\|144859 | ORF B [Clostridium perfringens] | 55 | 28 |
| 482 | 6 | 3515 | 3156 | gi\|606162 | ORF_f229 [Escherichia coli] | 55 | 39 |
| 497 | 9 | 4885 | 5937 | gi\|1041637 | replication initiator protein [Staphylococcus xylosus] | 55 | 33 |
| 546 | 1 | 1 | 1104 | gi\|467446 | similar to SpoVB [Bacillus subtilis] | 55 | 36 |
| 634 | 4 | 2132 | 1524 | gi\|431950 | similar to a B.subtilis gene (GB: BACHEMEHY_5) [Clostridium asteurianum] | 55 | 27 |
| 660 | 2 | 249 | 401 | gnl\|PID\|e254995 | hypothetical protein [Bacillus subtilis] | 55 | 35 |
| 671 | 1 | 288 | 58 | gi\|38722 | precursor (aa -20 to 381) [Acinetobacter calcoaceticus] ir\|A29277\|A29277 aldose 1-epimerase (EC 5.1.3.3) - Acinetobacter lcoaceticus | 55 | 33 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 686 | 2 | 245 | 1141 | gi\|1633572 | Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus] | 55 | 36 |
| 713 | 3 | 2742 | 1438 | gnl\|PID\|e8901 | RESA NF7 Ag13 [Plasmodium falciparum] | 55 | 25 |
| 815 | 1 | 2 | 226 | gi\|1113815 | histidine kinase [Borrelia burgdorferi] | 55 | 36 |
| 857 | 1 | 2 | 520 | gi\|143024 | glucose-resistance amylase regulator [Bacillus subtilis] pir\|S15318\|S15318 ccpA protein - Bacillus subtilis sp\|P25144\|CCPA_BACSU GLUCOSE-RESISTANCE AMYLASE REGULATOR CATABOLITE CONTROL PROTEIN). | 55 | 31 |
| 931 | 1 | 3 | 557 | gi\|1098508 | putative spore germination apparatus protein [Bacillus megaterium] | 55 | 32 |
| 17 | 7 | 6379 | 7218 | gnl\|PID\|e250887 | potential coding region [Clostridium difficile] | 54 | 35 |
| 21 | 9 | 7265 | 6348 | gi\|13441 | NADH dehydrogenase subunit 4L [Phoca vitulina] | 54 | 29 |
| 28 | 2 | 2727 | 3425 | gi\|1001792 | hypothetical protein [Synechocystis sp.] | 54 | 29 |
| 32 | 6 | 4044 | 3523 | gi\|1673660 | (AE000002) Mycoplasma pneumoniae, hypothetical 28K protein; similar to GenBank Accession Number JS0068, from M. pneumoniae [Mycoplasma pneumoniae] | 54 | 36 |
| 33 | 3 | 2274 | 3767 | gnl\|PID\|e245024 | unknown [Mycobacterium tuberculosis] | 54 | 36 |
| 40 | 1 | 1 | 915 | gi\|773349 | BirA protein [Bacillus subtilis] | 54 | 32 |
| 49 | 6 | 2120 | 2485 | gnl\|PID\|e139446 | a2 gene product [Bacteriophage B1] | 54 | 38 |
| 54 | 17 | 8969 | 8661 | gi\|334068 | ORF2 [Suid herpesvirus 1] | 54 | 51 |
| 65 | 2 | 1311 | 2120 | gi\|537207 | ORF_f277 [Escherichia coli] | 54 | 27 |
| 72 | 20 | 21986 | 22435 | gi\|928848 | ORF70'; putative [Lactococcus lactis phage BK5-T] | 54 | 34 |
| 105 | 4 | 3039 | 3827 | gnl\|PID\|e205174 | orf2 gene product [Lactobacillus helveticus] | 54 | 30 |
| 127 | 1 | 884 | 150 | gi\|726443 | No definition line found [Caenorhabditis | 54 | 31 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | elegans] | | |
| 148 | 1 | 1204 | 62 | gi\|467456 | unknown [Bacillus subtilis] | 54 | 37 |
| 156 | 4 | 4360 | 3167 | gi\|1032483 | unidentified ORF downstream of hydrogenase cluster; ORF5 [Anabaena variabilis] | 54 | 30 |
| 160 | 4 | 1523 | 2077 | gnl\|PID\|e255111 | hypothetical protein [Bacillus subtilis] | 54 | 27 |
| 160 | 7 | 4260 | 3745 | gi\|1184121 | auxin-induced protein [Vigna radiata] | 54 | 30 |
| 165 | 5 | 4996 | 3971 | gi\|1772652 | 2-keto-3-deoxygluconate kinase [Haloferax alicantei] | 54 | 36 |
| 176 | 2 | 1044 | 1937 | gi\|162201 | P-type ATPase [Trypanosoma brucei] | 54 | 38 |
| 180 | 29 | 30833 | 29853 | gnl\|PID\|e254644 | membrane protein [Streptococcus pneumoniae] | 54 | 29 |
| 200 | 16 | 7933 | 6656 | gi\|1574238 | traN protein (traN) [Haemophilus influenzae] | 54 | 31 |
| 206 | 1 | 232 | 2 | gi\|1220501 | Rickettsia tsutsugamushi (strain Kp47) gene, complete cds [Rickettsia tsutsugamushi] | 54 | 31 |
| 220 | 4 | 5235 | 4342 | gi\|606080 | ORF_o290; Geneplot suggests frameshift linking to o267, not found Escherichia coli] | 54 | 31 |
| 220 | 5 | 5821 | 5135 | gi\|43942 | first subunit of EII-Sor [Klebsiella pneumoniae] | 54 | 36 |
| 223 | 20 | 17253 | 17747 | gi\|47932 | tonB protein [Salmonella typhimurium] | 54 | 38 |
| 228 | 7 | 4866 | 4033 | gi\|1736828 | Thi4 protein [Escherichia coli] | 54 | 34 |
| 229 | 4 | 5050 | 3371 | gi\|1046078 | M. genitalium predicted coding region MG369 [Mycoplasma genitalium] | 54 | 42 |
| 236 | 3 | 4777 | 1496 | gi\|152271 | 319-kDA protein [Rhizobium meliloti] | 54 | 28 |
| 236 | 5 | 7822 | 6944 | gnl\|PID\|e285031 | Hyp1 protein [Hydra vulgaris] | 54 | 20 |
| 238 | 30 | 27964 | 27746 | gnl\|PID\|e217586 | PlnM [Lactobacillus plantarum] | 54 | 42 |
| 242 | 5 | 3508 | 4050 | gi\|149502 | beta-lactamase [Lactococcus lactis] | 54 | 35 |
| 257 | 1 | 296 | 120 | gi\|1498064 | AtE1 [Arabidopsis thaliana] | 54 | 50 |
| 257 | 6 | 6745 | 5633 | gi\|343949 | var1(40.0) [Saccharomyces cerevisiae] | 54 | 42 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 258 | 8 | 7839 | 7114 | gi\|41519 | P30 protein (AA 1-240) [Escherichia coli] | 54 | 31 |
| 276 | 20 | 13101 | 12880 | gi\|155322 | icsB gene product [Plasmid pWR100] | 54 | 37 |
| 280 | 1 | 618 | 106 | gi\|467356 | unknown [Bacillus subtilis] | 54 | 21 |
| 288 | 4 | 2183 | 2632 | gi\|39978 | P16 [Bacillus subtilis] | 54 | 39 |
| 316 | 1 | 3 | 767 | gi\|143264 | membrane-associated protein [Bacillus subtilis] | 54 | 34 |
| 318 | 7 | 5035 | 4565 | gi\|606080 | ORF_o290; Geneplot suggests frameshift linking to o267, not found Escherichia coli] | 54 | 28 |
| 319 | 3 | 1393 | 2163 | gi\|148327 | vancomycin response regulator [Enterococcus faecium] | 54 | 34 |
| 323 | 2 | 1256 | 2560 | gi\|413940 | ipa-16d gene product [Bacillus subtilis] | 54 | 26 |
| 364 | 7 | 7335 | 7724 | gnl\|PID\|e250171 | F18C12.1 [Caenorhabditis elegans] | 54 | 31 |
| 386 | 5 | 2399 | 3844 | gi\|155369 | PTS enzyme-II fructose [Xanthomonas campestris] | 54 | 37 |
| 392 | 3 | 2004 | 3353 | gi\|872306 | integral membrane protein [Streptomyces pristinaespiralis] pir\|S57509\|S57509 integral membrane protein - Streptomyces ristinaespiralis | 54 | 32 |
| 424 | 5 | 1553 | 1371 | gi\|160316 | major merozoite surface antigen [Plasmodium falciparum] sp\|P50495\|MSP1_PLAFP MEROZOITE SURFACE PROTEIN 1 PRECURSOR MEROZOITE SURFACE ANTIGENS) (PMMSA) (GP195). | 54 | 37 |
| 445 | 2 | 1897 | 1178 | gi\|1781503 | MigA [Pseudomonas aeruginosa] | 54 | 31 |
| 452 | 5 | 2506 | 2805 | gi\|216292 | neopullulanase [Bacillus sp.] | 54 | 34 |
| 457 | 2 | 2178 | 1024 | gi\|405570 | TraK protein shares sequence similarity with a family of proteins ncoded on Gram-negative gene transfer systems such as TraD from the plasmid [Plasmid pSK41] | 54 | 35 |
| 461 | 3 | 627 | 1418 | gi\|797332 | MocD [Agrobacterium tumefaciens] | 54 | 38 |
| 466 | 5 | 5419 | 3770 | gi\|1652892 | ABC transporter [Synechocystis sp.] | 54 | 29 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 475 | 3 | 2745 | 1990 | gi\|532546 | ORF13 [Enterococcus faecalis] | 54 | 35 |
| 495 | 1 | 2 | 295 | gi\|304990 | ORF_o290 [Escherichia coli] | 54 | 21 |
| 502 | 4 | 3518 | 3216 | gi\|1573270 | hemolysin (tlyC) [Haemophilus influenzae] | 54 | 33 |
| 510 | 5 | 3089 | 3931 | gi\|1732200 | PTS permease for mannose subunit IIPMan [Vibrio furnissii] | 54 | 29 |
| 570 | 1 | 1 | 930 | gi\|1001582 | penicillin-binding protein 1A [Synechocystis sp.] | 54 | 31 |
| 573 | 6 | 2763 | 3164 | gi\|416197 | homologous to plasmid R100 pemK gene [Escherichia coli] | 54 | 35 |
| 590 | 1 | 433 | 2 | gi\|532309 | 25 kDa protein [Escherichia coli] | 54 | 33 |
| 643 | 2 | 1202 | 1477 | gnl\|PID\|e125689 | 256 kD golgin [Homo sapiens] | 54 | 29 |
| 705 | 1 | 2 | 682 | gi\|148921 | LicD protein [Haemophilus influenzae] | 54 | 39 |
| 730 | 1 | 370 | 167 | gnl\|PID\|e245531 | ORF YLR068w [Saccharomyces cerevisiae] | 54 | 29 |
| 745 | 1 | 502 | 209 | gi\|581140 | NADH dehydrogenase [Escherichia coli] | 54 | 37 |
| 749 | 1 | 413 | 3 | gi\|664840 | TagB [Dictyostelium discoideum] | 54 | 44 |
| 932 | 1 | 3 | 320 | gi\|537207 | ORF_f277 [Escherichia coli] | 54 | 27 |
| 4 | 6 | 5671 | 4748 | gi\|216267 | ORF2 [Bacillus megaterium] | 53 | 34 |
| 16 | 8 | 6231 | 6806 | gi\|517105 | spermidine acetyltransferase [Escherichia coli] | 53 | 35 |
| 17 | 1 | 2 | 2497 | gi\|387880 | collagen adhesin [Staphylococcus aureus] | 53 | 35 |
| 42 | 4 | 2942 | 3529 | gi\|1633572 | Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus] | 53 | 20 |
| 69 | 6 | 3149 | 4879 | gi\|1486244 | unknown [Bacillus subtilis] | 53 | 30 |
| 72 | 3 | 1455 | 2063 | gi\|1592197 | M. jannaschii predicted coding region MJ1576 [Methanococcus jannaschii] | 53 | 32 |
| 79 | 1 | 83 | 592 | gi\|633757 | pr2 [Mycoplasma hyopneumoniae] | 53 | 28 |
| 83 | 8 | 5179 | 4412 | gi\|496100 | unknown function; putative [Bacteriophage phi-LC3] | 53 | 39 |
| 85 | 10 | 7180 | 6764 | gi\|1303940 | YqiU [Bacillus subtilis] | 53 | 35 |
| 92 | 2 | 789 | 986 | gi\|1372996 | Rho [Borrelia burgdorferi] | 53 | 28 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 95 | 10 | 7546 | 7734 | gi\|162379 | variant surface glycoprotein [Trypanosoma brucei] | 53 | 28 |
| 99 | 4 | 1391 | 1861 | gi\|1499620 | M. jannaschii predicted coding region MJ0798 [Methanococcus jannaschii] | 53 | 34 |
| 100 | 44 | 29982 | 29749 | gi\|1590997 | M. jannaschii predicted coding region MJ0272 [Methanococcus jannaschii] | 53 | 35 |
| 102 | 5 | 4787 | 5089 | gi\|1399011 | immunogenic secreted protein precursor [Streptococcus pyogenes] | 53 | 40 |
| 113 | 1 | 825 | 4 | gnl\|PID\|e264148 | unknown [Mycobacterium tuberculosis] | 53 | 24 |
| 114 | 4 | 6555 | 5113 | gi\|487282 | Na+ -ATPase subunit J [Enterococcus hirae] | 53 | 33 |
| 119 | 6 | 3581 | 3994 | gi\|473707 | positive regulator for virulence factors [Clostridium perfringens] | 53 | 31 |
| 123 | 19 | 16463 | 18115 | gi\|1591361 | NADH oxidase [Methanococcus jannaschii] | 53 | 33 |
| 136 | 1 | 381 | 4 | gi\|152744 | IpaD protein [Shigella flexneri] | 53 | 32 |
| 138 | 9 | 8079 | 7594 | gi\|467371 | LACI family of transcriptional repreesor (probable) [Bacillus ubtilis] | 53 | 29 |
| 142 | 8 | 4594 | 4007 | gi\|755216 | N-acetylmuramidase [Lactococcus lactis] | 53 | 38 |
| 162 | 12 | 12482 | 11937 | gi\|1063250 | low homology to P20 protein of Bacillus lichiniformis and bleomycin acetyltransferase of Streptomyces verticillus [Bacillus subtilis] | 53 | 36 |
| 163 | 1 | 546 | 31 | gi\|153767 | ORF [Streptococcus pneumoniae] | 53 | 34 |
| 163 | 7 | 4973 | 3453 | gi\|29468 | beta-myosin heavy chain (1151 AA) [Homo sapiens] | 53 | 36 |
| 167 | 2 | 1038 | 2006 | gi\|413930 | ipa-6d gene product [Bacillus subtilis] | 53 | 27 |
| 173 | 11 | 8865 | 7843 | gi\|1778569 | YaaF homolog [Escherichia coli] | 53 | 39 |
| 190 | 8 | 6842 | 3549 | gi\|387880 | collagen adhesin [Staphylococcus aureus] | 53 | 38 |
| 199 | 2 | 2725 | 950 | gi\|1652570 | nitrate transport protein NrtB [Synechocystis sp.] | 53 | 32 |
| 200 | 13 | 6184 | 5954 | gi\|1652679 | hypothetical protein [Synechocystis sp.] | 53 | 40 |
| 200 | 17 | 9287 | 7890 | gi\|1574246 | H. influenzae predicted coding region | 53 | 35 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | HI1409 [Haemophilus influenzae] | | |
| 205 | 6 | 2048 | 3229 | gi\|148026 | topoisomerase III [Escherichia coli] | 53 | 32 |
| 211 | 2 | 270 | 1052 | gi\|483940 | transcription regulator [Bacillus subtilis] | 53 | 30 |
| 221 | 10 | 5119 | 5994 | gi\|1353529 | ORF12 [Bacteriophage r1t] | 53 | 44 |
| 232 | 7 | 4344 | 3925 | gi\|1665759 | Similar to Schistosoma mansoni amino acid permease (L25068). [Homo sapiens] | 53 | 35 |
| 238 | 21 | 18705 | 19247 | gi\|1574062 | hypothetical [Haemophilus influenzae] | 53 | 30 |
| 239 | 1 | 2 | 1636 | gi\|433932 | activator of (R)-hydroxyglutaryl-CoA dehydratase [Acidaminococcus ermentans] | 53 | 35 |
| 250 | 1 | 1469 | 318 | gi\|987094 | membrane transport protein [Streptomyces hygroscopicus] | 53 | 22 |
| 253 | 4 | 1759 | 1028 | gi\|537245 | aspartokinase I-homoserine dehydrogenase I [Escherichia coli] pir\|S56629\|S56629 aspartate kinase (EC 2.7.2.4) / homoserine ehydrogenase (EC 1.1.1.3) - Escherichia coli | 53 | 35 |
| 271 | 8 | 4649 | 5800 | gi\|413966 | ipa-42d gene product [Bacillus subtilis] | 53 | 27 |
| 276 | 26 | 15786 | 15112 | gi\|1699017 | ErpB2 [Borrelia burgdorferi] | 53 | 26 |
| 279 | 11 | 8309 | 7797 | gi\|1651934 | hypothetical protein [Synechocystis sp.] | 53 | 35 |
| 288 | 8 | 3997 | 4872 | gi\|43943 | second subunit of EII-Sor [Klebsiella pneumoniae] | 53 | 32 |
| 290 | 6 | 4391 | 5680 | gi\|466882 | pps1; B1496_C2_189 [Mycobacterium leprae] | 53 | 29 |
| 294 | 3 | 1197 | 1481 | gi\|173004 | topoisomerase I [Saccharomyces cerevisiae] | 53 | 40 |
| 330 | 3 | 2351 | 3367 | gi\|466691 | No definition line found [Escherichia coli] | 53 | 34 |
| 334 | 8 | 8172 | 9182 | gi\|1652483 | hypothetical protein [Synechocystis sp.] | 53 | 29 |
| 368 | 1 | 620 | 102 | gi\|487273 | Na+ -ATPase subunit I [Enterococcus hirae] | 53 | 29 |
| 377 | 4 | 2424 | 2260 | gi\|221407 | FP5 [Fowlpox virus] | 53 | 35 |
| 382 | 1 | 257 | 36 | gi\|1592016 | M. jannaschii predicted coding region MJ1371 [Methanococcus jannaschii] | 53 | 32 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 387 | 1 | 2 | 460 | gi\|1574317 | repressor protein (GP:L22692_1) [Haemophilus influenzae] | 53 | 30 |
| 394 | 10 | 8379 | 10412 | gi\|882463 | protein-N(pi)-phosphohistidine-sugar phosphotransferase [Escherichia oli] | 53 | 34 |
| 399 | 4 | 2349 | 3098 | gi\|453287 | OmpR protein [Escherichia coli] | 53 | 27 |
| 420 | 2 | 1378 | 719 | gi\|1437473 | nitrate transporter [Bacillus subtilis] | 53 | 28 |
| 441 | 6 | 5361 | 7937 | gi\|1592205 | M. jannaschii predicted coding region MJ1595 [Methanococcus jannaschii] | 53 | 38 |
| 461 | 1 | 6 | 512 | gi\|1651800 | L-glutamine:D-fructose-6-P amidotransferase [Synechocystis sp.] | 53 | 29 |
| 497 | 3 | 1700 | 1960 | gi\|4328 | RIF1 gene product [Saccharomyces cerevisiae] | 53 | 33 |
| 503 | 1 | 669 | 4 | gnl\|PID\|e202290 | unknown [Lactobacillus sake] | 53 | 30 |
| 538 | 2 | 1053 | 262 | gi\|1613769 | response regulator [Streptococcus pneumoniae] | 53 | 30 |
| 539 | 6 | 6172 | 5183 | gi\|567887 | putative repressor [Streptomyces peucetius] | 53 | 32 |
| 551 | 1 | 629 | 162 | gi\|1256649 | putative [Bacillus subtilis] | 53 | 26 |
| 557 | 1 | 9 | 695 | gi\|143177 | putative [Bacillus subtilis] | 53 | 31 |
| 569 | 2 | 418 | 1158 | gi\|1184684 | MucD [Pseudomonas aeruginosa] | 53 | 26 |
| 614 | 1 | 99 | 581 | gi\|485280 | 28.2 kDa protein [Streptococcus pneumoniae] | 53 | 32 |
| 660 | 1 | 1 | 279 | gnl\|PID\|e288480 | R10E8.f [Caenorhabditis elegans] | 53 | 34 |
| 776 | 1 | 3 | 635 | gi\|151352 | mandelate racemase (EC 5.1.2.2) [Pseudomonas putida] | 53 | 33 |
| 11 | 2 | 1117 | 1656 | gi\|143150 | levR [Bacillus subtilis] | 52 | 29 |
| 17 | 6 | 5327 | 6559 | gnl\|PID\|e250887 | potential coding region [Clostridium difficile] | 52 | 37 |
| 19 | 31 | 17760 | 17978 | gi\|1079556 | dShc [Drosophila melanogaster] | 52 | 42 |
| 19 | 38 | 20306 | 22627 | gnl\|PID\|e139448 | host interacting protein [Bacteriophage B1] | 52 | 32 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 25 | 4 | 2662 | 2087 | gi\|1072067 | PepF [Rhodobacter sphaeroides] | 52 | 23 |
| 25 | 6 | 5596 | 3407 | gi\|1303866 | YqgS [Bacillus subtilis] | 52 | 34 |
| 49 | 3 | 1135 | 1569 | gi\|496279 | putative [Bacteriophage Tuc2009] | 52 | 25 |
| 53 | 1 | 850 | 2 | sp\|P52697\|YBHE_ECOLI | HYPOTHETICAL 30.2 KD PROTEIN IN MODC 3'REGION. | 52 | 35 |
| 54 | 9 | 10909 | 2687 | gi\|1633572 | Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus] | 52 | 30 |
| 57 | 6 | 4779 | 8402 | gi\|142439 | ATP-dependent nuclease [Bacillus subtilis] | 52 | 31 |
| 58 | 6 | 6446 | 5949 | gnl\|PID\|e255921 | F53F4.10 [Caenorhabditis elegans] | 52 | 31 |
| 72 | 13 | 13446 | 13195 | gi\|532541 | ORF8 [Enterococcus faecalis] | 52 | 37 |
| 81 | 17 | 13692 | 12520 | gi\|1732203 | GlcNAc 6-P deacetylase [Vibrio furnissii] | 52 | 35 |
| 84 | 1 | 3 | 1355 | gi\|64288 | fast skeletal muscle Ca-ATPase [Rana esculenta] | 52 | 34 |
| 100 | 2 | 1917 | 1027 | gi\|1353560 | ORF43 [Bacteriophage r1t] | 52 | 34 |
| 101 | 1 | 30 | 1862 | gi\|405957 | yeeF [Escherichia coli] | 52 | 24 |
| 106 | 8 | 8517 | 7600 | gi\|454904 | rfbG gene product [Shigella flexneri] | 52 | 41 |
| 108 | 1 | 1 | 1059 | gnl\|PID\|e255337 | unknown [Mycobacterium tuberculosis] | 52 | 29 |
| 123 | 4 | 2899 | 3495 | gi\|1305720 | prs-associated putative membrane protein [Escherichia coli] | 52 | 24 |
| 128 | 23 | 17561 | 16740 | gi\|473805 | 'regulatory protein sfs1 involved in maltose metabolism' Escherichia coli] | 52 | 32 |
| 130 | 8 | 6693 | 7481 | gi\|1552775 | ATP-binding protein [Escherichia coli] | 52 | 30 |
| 138 | 1 | 40 | 1359 | gi\|1045867 | oligoendopeptidase F [Mycoplasma genitalium] | 52 | 31 |
| 138 | 2 | 2757 | 1384 | gi\|1591425 | hypothetical protein (GP:X91006_2) [Methanococcus jannaschii] | 52 | 26 |
| 138 | 6 | 6317 | 5940 | gi\|1486247 | unknown [Bacillus subtilis] | 52 | 36 |
| 142 | 10 | 7337 | 5466 | gi\|1151158 | repeat organellar protein [Plasmodium chabaudi] | 52 | 34 |
| 149 | 1 | 33 | 1133 | gi\|1762962 | FemA [Staphylococcus simulans] | 52 | 31 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 161 | 1 | 3 | 245 | gi\|151276 | histidine utilization genes repressor protein (hut) [Pseudomonas utida] | 52 | 35 |
| 163 | 4 | 2048 | 1320 | gi\|1064810 | function unknown [Bacillus subtilis] | 52 | 27 |
| 164 | 8 | 4882 | 5103 | gi\|57251 | precursor (AA -35 to 1766) [Rattus norvegicus] | 52 | 38 |
| 165 | 9 | 7247 | 7474 | gi\|1652671 | hypothetical protein [Synechocystis sp.] | 52 | 28 |
| 178 | 5 | 1887 | 1681 | gi\|220704 | cAMP-dependent protein kinase catalytic subunit-beta [Rattus sp.] gi\|191177 cAMP-dependent protein kinase beta-catalytic subunit Cricetulus sp.] | 52 | 36 |
| 180 | 24 | 22536 | 23774 | gi\|581052 | cytosine deaminase [Escherichia coli] | 52 | 28 |
| 190 | 9 | 8891 | 7056 | gi\|1592079 | M. jannaschii predicted coding region MJ1429 [Methanococcus jannaschii] | 52 | 39 |
| 195 | 8 | 2000 | 2272 | gi\|868024 | HIC-1 gene product [Homo sapiens] | 52 | 52 |
| 202 | 11 | 9189 | 10145 | gi\|141861 | traA gene product [Plasmid pAD1] | 52 | 33 |
| 204 | 4 | 1361 | 2011 | gi\|1184118 | mevalonate kinase [Methanobacterium thermoautotrophicum] | 52 | 33 |
| 204 | 8 | 4018 | 5142 | gnl\|PID\|e283860 | carotenoid biosynthetic gene ERWCRTS homolog [Sulfolobus solfataricus] | 52 | 31 |
| 208 | 2 | 1112 | 2296 | gi\|1408501 | homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis] | 52 | 35 |
| 215 | 1 | 772 | 2 | gi\|1480429 | putative transcriptional regulator [Bacillus stearothermophilus] | 52 | 26 |
| 218 | 4 | 4072 | 3425 | gi\|862630 | glyceraldehyde-3-phosphate dehydrogenase [Buchnera aphidicola] sp\|Q07234\|G3P_BUCAP GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC .2.1.12) (GAPDH). | 52 | 35 |
| 228 | 1 | 1 | 741 | gnl\|PID\|e264148 | unknown [Mycobacterium tuberculosis] | 52 | 29 |
| 230 | 2 | 149 | 634 | gi\|437705 | hyaluronidase [Streptococcus pneumoniae] | 52 | 28 |
| 233 | 8 | 6166 | 4982 | gi\|1001708 | NifS [Synechocystis sp.] | 52 | 31 |
| 240 | 3 | 725 | 967 | gi\|399655 | Ca2+ regulatory protein [Saccharomyces | 52 | 21 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | cerevisiae] sp\|P35206\|CSG2_YEAST CSG2 PROTEIN PRECURSOR. | | |
| 288 | 7 | 3171 | 4028 | gi\|147403 | mannose permease subunit II-P-Man [Escherichia coli] | 52 | 27 |
| 318 | 1 | 7 | 819 | gi\|1303849 | YqgB [Bacillus subtilis] | 52 | 33 |
| 330 | 1 | 1062 | 154 | gi\|144859 | ORF B [Clostridium perfringens] | 52 | 29 |
| 330 | 9 | 6815 | 7213 | gi\|1439527 | EIIA-man [Lactobacillus curvatus] | 52 | 31 |
| 345 | 9 | 8348 | 9397 | gi\|606292 | ORF_o696 [Escherichia coli] | 52 | 27 |
| 398 | 3 | 2671 | 1877 | gi\|144859 | ORF B [Clostridium perfringens] | 52 | 29 |
| 411 | 1 | 992 | 3 | gnl\|PID\|e283950 | daunorubicin resistance ATP-binding protein DrrA [Sulfolobus solfataricus] | 52 | 27 |
| 422 | 2 | 1292 | 585 | gi\|537214 | yjjG gene product [Escherichia coli] | 52 | 32 |
| 436 | 2 | 1669 | 1205 | gi\|507323 | ORF1 [Bacillus stearothermophilus] | 52 | 29 |
| 450 | 1 | 119 | 754 | gi\|1573916 | multidrug resistance protein (emrB) [Haemophilus influenzae] | 52 | 32 |
| 453 | 1 | 190 | 381 | gi\|182021 | elastin [Homo sapiens] | 52 | 40 |
| 455 | 7 | 5767 | 4634 | gnl\|PID\|e155312 | integrase [Bacteriophage TP901-1] | 52 | 34 |
| 479 | 1 | 138 | 758 | gi\|1742859 | ORF_ID:o327#7; similar to [SwissProt Accession Number P54449] [Escherichia coli] | 52 | 27 |
| 517 | 1 | 763 | 2 | gi\|152780 | rhamnosyl transferase II [Shigella dysenteriae] | 52 | 29 |
| 518 | 3 | 1735 | 848 | gi\|153858 | wall-associated protein [Streptococcus mutans] | 52 | 20 |
| 526 | 3 | 2297 | 1848 | gi\|147402 | mannose permease subunit III-Man [Escherichia coli] | 52 | 27 |
| 617 | 1 | 1 | 462 | gi\|142863 | replication initiation protein [Bacillus subtilis] | 52 | 35 |
| 639 | 3 | 1068 | 259 | gi\|1591153 | hypothetical protein (SP:P46348) [Methanococcus jannaschii] | 52 | 30 |
| 703 | 1 | 773 | 81 | gi\|793910 | surface antigen [Homo sapiens] | 52 | 31 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 737 | 1 | 235 | 2 | gi\|666000 | hypothetical protein [Bacillus subtilis] | 52 | 29 |
| 791 | 4 | 1368 | 1802 | gnl\|PID\|e269549 | Unknown [Bacillus subtilis] | 52 | 28 |
| 825 | 1 | 1 | 300 | gi\|732538 | No definition line found [Caenorhabditis elegans] | 52 | 28 |
| 981 | 1 | 226 | 2 | gi\|951100 | P45016a-ms1 [Mus spretus] | 52 | 36 |
| 17 | 23 | 23542 | 22163 | gi\|1652483 | hypothetical protein [Synechocystis sp.] | 51 | 32 |
| 65 | 6 | 4302 | 3691 | gi\|397498 | Membrane Ribose Binding Protein [Bacillus subtilis] pir\|S42714\|S42714 membrane ribose-binding protein - Bacillus ubtilis | 51 | 31 |
| 69 | 5 | 2926 | 2537 | gi\|1773150 | hypothetical 14.8kd protein [Escherichia coli] | 51 | 30 |
| 92 | 1 | 973 | 44 | gnl\|PID\|e243523 | ORF YGR130c [Saccharomyces cerevisiae] | 51 | 29 |
| 103 | 6 | 5272 | 3593 | gi\|312940 | threonine kinase [Streptococcus equisimilis] | 51 | 32 |
| 111 | 7 | 4195 | 3317 | pir\|G64143\|G64143 | hypothetical protein HI0143 - Haemophilus influenzae (strain Rd KW20) | 51 | 29 |
| 115 | 7 | 4526 | 3414 | gi\|405879 | yeiH [Escherichia coli] | 51 | 27 |
| 123 | 29 | 27788 | 28207 | gi\|147402 | mannose permease subunit III-Man [Escherichia coli] | 51 | 27 |
| 125 | 1 | 223 | 2 | gi\|4482 | SLY1 gene product [Saccharomyces cerevisiae] | 51 | 37 |
| 128 | 21 | 16156 | 15638 | gi\|606026 | ORF_o115 [Escherichia coli] | 51 | 27 |
| 137 | 4 | 3207 | 5369 | gi\|1673692 | (AE000005) Mycoplasma pneumoniae, C09_orf422 Protein [Mycoplasma pneumoniae] | 51 | 26 |
| 138 | 28 | 18295 | 18771 | gi\|149647 | ORFZ [Listeria monocytogenes] | 51 | 31 |
| 145 | 6 | 4054 | 5271 | gi\|1653860 | N-acyl-L-amino acid amidohydrolase [Synechocystis sp.] | 51 | 41 |
| 155 | 4 | 3019 | 2273 | gi\|1486242 | unknown [Bacillus subtilis] | 51 | 41 |
| 180 | 8 | 7951 | 9189 | gi\|1657522 | hypothetical protein [Escherichia coli] | 51 | 32 |
| 186 | 2 | 859 | 1620 | gi\|511497 | oleoyl-acyl carrier protein thioesterase [Coriandrum sativum] | 51 | 29 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 186 | 3 | 1644 | 2060 | sp\|P37348\|YECE_ECOLI | HYPOTHETICAL PROTEIN IN ASPS 5'REGION (FRAGMENT). | 51 | 38 |
| 194 | 3 | 1521 | 1276 | gi\|332697 | fusion protein [Human parainfluenza virus 2] | 51 | 32 |
| 195 | 7 | 1986 | 3767 | gi\|405570 | TraK protein shares sequence similarity with a family of proteins ncoded on Gram-negative gene transfer systems such as TraD from the plasmid [Plasmid pSK41] | 51 | 28 |
| 197 | 1 | 3 | 494 | gi\|1592234 | DNA topoisomerase I [Methanococcus jannaschii] | 51 | 32 |
| 198 | 2 | 1521 | 862 | gi\|1196483 | unknown protein [Lactobacillus casei] | 51 | 32 |
| 238 | 16 | 13630 | 14730 | gi\|1772652 | 2-keto-3-deoxygluconate kinase [Haloferax alicantei] | 51 | 36 |
| 257 | 5 | 5646 | 4513 | pir\|S43367\|S43367 | metallothionein - Green crab, common shore crab | 51 | 38 |
| 261 | 6 | 4950 | 4519 | gi\|581545 | orf 4 [Staphylococcus aureus] | 51 | 26 |
| 270 | 5 | 4480 | 4220 | gi\|1066975 | F49E2.5a [Caenorhabditis elegans] | 51 | 28 |
| 306 | 10 | 5928 | 6905 | gi\|1752736 | gene required for phosphoylation of oligosaccharides/ has high homology with YJR061w [Saccharomyces cerevisiae] | 51 | 28 |
| 324 | 3 | 1590 | 2405 | gi\|409925 | VirR positive regulator [Streptococcus pyogenes] | 51 | 25 |
| 328 | 2 | 632 | 309 | gi\|466475 | putative phospho-beta-glucosidase [Bacillus stearothermophilus] pir\|D49898\|D49898 cellobiose phosphotransferase system celC - acillus stearothermophilus | 51 | 30 |
| 340 | 2 | 898 | 1152 | gi\|40046 | phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus] ir\|S15936\|NUBSSA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus | 51 | 39 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 340 | 4 | 3617 | 2445 | gi\|763052 | integrase [Bacteriophage T270] | 51 | 33 |
| 379 | 10 | 11742 | 11311 | gi\|887829 | D21141 uses 2nd start; frame determined by Lac fusion [Escherichia oli] | 51 | 34 |
| 380 | 1 | 2 | 1123 | gi\|309662 | pheromone binding protein [Plasmid pCF10] | 51 | 34 |
| 395 | 1 | 526 | 95 | gi\|490986 | phi 105 repessor orf2 [unidentified] | 51 | 27 |
| 424 | 4 | 2512 | 995 | gi\|1633572 | Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus] | 51 | 31 |
| 444 | 1 | 737 | 483 | gi\|1245376 | cardiac ryanodine receptor [Oryctolagus cuniculus] | 51 | 34 |
| 483 | 1 | 1 | 642 | gi\|1303981 | YqkD [Bacillus subtilis] | 51 | 29 |
| 500 | 1 | 2 | 550 | gi\|987094 | membrane transport protein [Streptomyces hygroscopicus] | 51 | 23 |
| 525 | 3 | 492 | 983 | pir\|A57438\|A57438 | tryptophan-rich sensory protein - Rhodobacter sphaeroides (strain 2.4.1) | 51 | 38 |
| 534 | 1 | 2 | 1165 | gi\|147516 | ribokinase [Escherichia coli] | 51 | 33 |
| 547 | 1 | 1 | 387 | gi\|1353528 | ORF11 [Bacteriophage r1t] | 51 | 33 |
| 553 | 2 | 1728 | 1330 | pir\|B55124\|B55124 | thioredoxin - Chlorobium sp. | 51 | 27 |
| 574 | 1 | 2291 | 2476 | bbs\|129435 | RprX=inner membrane signal-transducing protein [Bacteroides fragilis, Peptide, 519 aa] [Bacteroides fragilis] | 51 | 36 |
| 574 | 2 | 3145 | 3420 | gi\|1732202 | PTS permease for mannose subunit IIIMan N terminal domain [Vibrio furnissii] | 51 | 29 |
| 594 | 2 | 530 | 225 | gi\|1657696 | tryptophan hydroxylase [Gallus gallus] | 51 | 40 |
| 605 | 3 | 1220 | 1936 | gnl\|PID\|e289149 | similar to B. subtilis YcsE hypothetical protein [Bacillus subtilis] | 51 | 32 |
| 609 | 1 | 1027 | 74 | gi\|1226279 | strong similarity to Schistosoma amino acid permease (GB:L25068) [Caenorhabditis elegans] | 51 | 26 |
| 656 | 2 | 2033 | 2950 | gi\|143213 | putative [Bacillus subtilis] | 51 | 26 |
| 670 | 1 | 1508 | 369 | gi\|1652222 | hypothetical protein [Synechocystis sp.] | 51 | 25 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 673 | 1 | 2 | 1135 | gi\|532553 | ORF20 [Enterococcus faecalis] | 51 | 27 |
| 674 | 2 | 1158 | 778 | gi\|467451 | unknown [Bacillus subtilis] | 51 | 26 |
| 735 | 2 | 477 | 725 | gi\|757791 | aromatic amino acid permease [Corynebacterium glutamicum] pir\|S52754\|S52754 aromatic amino acid permease - Corynebacterium lutamicum | 51 | 38 |
| 924 | 1 | 794 | 3 | gi\|40663 | sialidase [Clostridium septicum] | 51 | 35 |
| 4 | 5 | 3811 | 4728 | gi\|413948 | ipa-24d gene product [Bacillus subtilis] | 50 | 29 |
| 8 | 3 | 3310 | 2180 | gi\|1592205 | M. jannaschii predicted coding region MJ1595 [Methanococcus jannaschii] | 50 | 28 |
| 11 | 9 | 5269 | 5520 | gi\|1651800 | L-glutamine:D-fructose-6-P amidotransferase [Synechocystis sp.] | 50 | 25 |
| 12 | 6 | 9045 | 8662 | gnl\|PID\|e254943 | unknown [Mycobacterium tuberculosis] | 50 | 23 |
| 15 | 4 | 2911 | 4269 | gi\|1592173 | N-ethylammeline chlorohydrolase [Methanococcus jannaschii] | 50 | 28 |
| 19 | 10 | 4934 | 5530 | gi\|825569 | unknown [Saccharomyces cerevisiae] | 50 | 20 |
| 28 | 5 | 7515 | 7057 | gi\|1230586 | orf10; Method: conceptual translation supplied by author [Vibrio cholerae O139] | 50 | 38 |
| 45 | 9 | 4279 | 5019 | gi\|1591029 | thioredoxin/glutaredoxin [Methanococcus jannaschii] | 50 | 32 |
| 54 | 16 | 7739 | 7590 | gi\|1589837 | cuticle preprocollagen [Meloidogyne incognita] | 50 | 46 |
| 59 | 5 | 1551 | 2345 | gi\|144297 | acetyl esterase (XynC) [Caldocellum saccharolyticum] pir\|B37202\|B37202 acetylesterase (EC 3.1.1.6) (XynC) - Caldocellum accharolyticum | 50 | 34 |
| 62 | 3 | 1650 | 1360 | gnl\|PID\|e205266 | LEA76 homologue type2 [Arabidopsis thaliana] | 50 | 31 |
| 91 | 10 | 8858 | 7521 | gi\|758229 | integrase [Bacteriophage phi-13] | 50 | 31 |
| 112 | 5 | 3548 | 2133 | gi\|1184262 | GadC [Shigella flexneri] | 50 | 25 |
| 123 | 13 | 13099 | 14319 | gi\|178273 | alanine:glyoxylate aminotransferase [Homo sapiens] | 50 | 31 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 123 | 15 | 14395 | 15675 | gi\|467342 | unknown [Bacillus subtilis] | 50 | 28 |
| 123 | 31 | 28700 | 29494 | gi\|43942 | first subunit of EII-Sor [Klebsiella pneumoniae] | 50 | 27 |
| 124 | 2 | 1666 | 1061 | gi\|556016 | similar to plant water stress proteins; ORF2 [Bacillus subtilis] gi\|556016 similar to plant water stress proteins; ORF2 [Bacillus ubtilis] | 50 | 34 |
| 128 | 39 | 32767 | 31829 | gi\|39993 | UDP-N-acetylmuramoylalanine--D-glutamate ligase [Bacillus subtilis] | 50 | 33 |
| 135 | 11 | 8803 | 7694 | gi\|895747 | putative cel operon regulator [Bacillus subtilis] | 50 | 26 |
| 138 | 21 | 14648 | 13653 | gi\|1591472 | malic acid transport protein [Methanococcus jannaschii] | 50 | 26 |
| 146 | 3 | 2338 | 1415 | gi\|1732200 | PTS permease for mannose subunit IIPMan [Vibrio furnissii] | 50 | 27 |
| 160 | 2 | 724 | 1302 | gnl\|PID\|e264218 | F54F3.4 [Caenorhabditis elegans] | 50 | 30 |
| 164 | 15 | 15432 | 16364 | gi\|409286 | bmrU [Bacillus subtilis] | 50 | 27 |
| 167 | 9 | 17082 | 15394 | gi\|143156 | membrane bound protein [Bacillus subtilis] | 50 | 30 |
| 179 | 3 | 2350 | 4485 | gi\|1408485 | yxdM gene product [Bacillus subtilis] | 50 | 24 |
| 180 | 30 | 31056 | 30643 | gnl\|PID\|e254644 | membrane protein [Streptococcus pneumoniae] | 50 | 27 |
| 184 | 1 | 2 | 1015 | gi\|854232 | cymE gene product [Klebsiella oxytoca] | 50 | 24 |
| 194 | 7 | 4335 | 4817 | gi\|1256652 | 25% identity to the E.coli regulatory protein MprA; putative [Bacillus subtilis] | 50 | 30 |
| 195 | 29 | 11712 | 12422 | gi\|662263 | ORF5 [Plasmid pIP501] | 50 | 25 |
| 204 | 1 | 2 | 166 | gi\|328656 | envelope polyprotein [Human immunodeficiency virus type 1] | 50 | 45 |
| 205 | 7 | 3118 | 3861 | gi\|437697 | traE [Plasmid RP4] | 50 | 31 |
| 216 | 11 | 7181 | 7750 | gnl\|PID\|e254644 | membrane protein [Streptococcus pneumoniae] | 50 | 30 |
| 223 | 10 | 7036 | 8082 | gi\|606423 | T09B9.1 [Caenorhabditis elegans] | 50 | 30 |

**Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins**

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 223 | 22 | 19257 | 19799 | gi\|1256141 | YbbL [Bacillus subtilis] | 50 | 29 |
| 233 | 4 | 3102 | 2320 | gi\|887826 | GUG start [Escherichia coli] | 50 | 32 |
| 238 | 6 | 5102 | 3906 | gi\|1161219 | homolgous to D-amino acid dehydrogenase enzyme [Pseudomonas aeruginosa] | 50 | 29 |
| 239 | 3 | 4449 | 5159 | gi\|41519 | P30 protein (AA 1-240) [Escherichia coli] | 50 | 31 |
| 242 | 2 | 147 | 2210 | gi\|160299 | glutamic acid-rich protein [Plasmodium falciparum] pir\|A54514\|A54514 glutamic acid-rich protein precursor - Plasmodium alciparum | 50 | 30 |
| 248 | 2 | 263 | 712 | gi\|143725 | putative [Bacillus subtilis] | 50 | 32 |
| 256 | 8 | 8531 | 7395 | gnl\|PID\|e250452 | C44H9.4 [Caenorhabditis elegans] | 50 | 38 |
| 265 | 3 | 1150 | 893 | gi\|1402527 | ORF6 [Enterococcus faecalis] | 50 | 39 |
| 276 | 24 | 14203 | 14000 | gi\|1591019 | M. jannaschii predicted coding region MJ0297 [Methanococcus jannaschii] | 50 | 33 |
| 276 | 32 | 20601 | 19924 | gi\|1334905 | BXLF2 late reading frame, encodes gp85; homologous to RF 37 VZV and glycoprotein H of HSV (gpIII of VZV) [Human herpesvirus 4] | 50 | 29 |
| 286 | 1 | 1 | 747 | gnl\|PID\|e257895 | homology with truncated ORF2 of pepF2 [Lactococcus lactis] | 50 | 32 |
| 301 | 17 | 11706 | 13313 | gi\|562039 | NADH dehydrogenase, subunit 2 [Acanthamoeba castellanii] pir\|S53835\|S53835 NADH dehydrogenase chain 2 - Acanthamoeba astellanii mitochondrion (SGC6) | 50 | 26 |
| 338 | 5 | 2206 | 3729 | gi\|829194 | bacterial cell wall hydrolase [Enterococcus faecalis] pir\|A38109\|A38109 autolysin - Enterococcus faecalis sp\|P37710\|ALYS_ENTFA AUTOLYSIN (EC 3.5.1.28) N-ACETYLMURAMOYL-L-ALANINE AMIDASE). | 50 | 34 |
| 345 | 12 | 11781 | 13379 | gnl\|PID\|e235181 | unknown [Mycobacterium tuberculosis] | 50 | 32 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 360 | 2 | 2879 | 408 | gi\|40782 | bps2 gene product [Desulfurolobus ambivalens] | 50 | 25 |
| 372 | 1 | 6 | 440 | gi\|1552733 | similar to voltage-gated chloride channel protein [Escherichia coli] | 50 | 31 |
| 372 | 2 | 391 | 738 | gi\|1591749 | TRK system potassium uptake protein A [Methanococcus jannaschii] | 50 | 23 |
| 377 | 3 | 2262 | 1846 | gi\|52797 | kinesin heavy chain [Mus musculus] | 50 | 22 |
| 392 | 1 | 433 | 2 | gi\|147213 | phnP protein [Escherichia coli] | 50 | 33 |
| 399 | 31 | 29803 | 30186 | gi\|146288 | PTS enzyme III glucitol [Escherichia coli] | 50 | 30 |
| 518 | 4 | 2885 | 2040 | gi\|475107 | regulatory protein [Pediococcus pentosaceus] | 50 | 29 |
| 528 | 1 | 3 | 665 | gi\|215098 | excisionase [Bacteriophage 154a] | 50 | 38 |
| 562 | 1 | 631 | 107 | gi\|1592205 | M. jannaschii predicted coding region MJ1595 [Methanococcus jannaschii] | 50 | 28 |
| 596 | 1 | 227 | 1153 | gi\|963039 | orf gene product [Enterococcus hirae] | 50 | 26 |
| 680 | 1 | 2 | 1090 | gi\|1050297 | product p150Glued [Neurospora crassa] | 50 | 27 |
| 755 | 1 | 2 | 430 | gi\|1736469 | Tetracenomycin C resistance and export protein. [Escherichia coli] | 50 | 33 |
| 838 | 1 | 428 | 3 | gi\|530424 | 50S ribosomal protein [Mycoplasma capricolum] | 50 | 30 |
| 14 | 2 | 3453 | 538 | gi\|47049 | asa1 gene product (AA 1-1296) [Enterococcus faecalis] ir\|S10223\|HMS01F aggregation protein asa1 - Enterococcus faecalis asmid pAD1 | 49 | 25 |
| 56 | 7 | 5367 | 4822 | gi\|924754 | glycine reductase complex selenoprotein B [Clostridium litorale] | 49 | 31 |
| 68 | 9 | 4741 | 7389 | gi\|1591494 | M. jannaschii predicted coding region MJ0797 [Methanococcus jannaschii] | 49 | 21 |
| 94 | 10 | 9425 | 6633 | gi\|1146243 | 22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative [Bacillus subtilis] | 49 | 30 |
| 98 | 12 | 12306 | 11701 | gi\|1303784 | YqeD [Bacillus subtilis] | 49 | 26 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 117 | 7 | 4789 | 6228 | gi\|435493 | orf4 gene product [Lactococcus lactis] | 49 | 26 |
| 123 | 21 | 18576 | 19745 | gi\|298032 | EF [Streptococcus suis] | 49 | 29 |
| 125 | 4 | 2358 | 1594 | gnl\|PID\|e237295 | unknown [Saccharomyces cerevisiae] | 49 | 27 |
| 125 | 6 | 4235 | 3453 | gi\|1573885 | glycosyl transferase (lgtD) [Haemophilus influenzae] | 49 | 32 |
| 144 | 5 | 3715 | 4062 | gi\|507130 | emm64 gene product [Streptococcus pyogenes] | 49 | 30 |
| 162 | 8 | 10472 | 9120 | gi\|47045 | NADH oxidase [Enterococcus faecalis] | 49 | 34 |
| 179 | 18 | 18426 | 17848 | gi\|40060 | DNA polymerase III (AA 1-1437) [Bacillus subtilis] p\|P13267\|DP3A_BACSU DNA POLYMERASE III, ALPHA CHAIN (EC 2.7.7.7). | 49 | 27 |
| 180 | 19 | 18727 | 19917 | gi\|143000 | proton glutamate symport protein [Bacillus stearothermophilus] pir\|S26247\|S26247 glutamate/aspartate transport protein - Bacillus tearothermophilus | 49 | 31 |
| 224 | 1 | 145 | 1371 | gi\|1103862 | TolA [Pseudomonas aeruginosa] | 49 | 32 |
| 236 | 8 | 10955 | 9249 | gi\|431272 | lysis protein [Bacillus subtilis] | 49 | 28 |
| 278 | 1 | 757 | 2 | gi\|467478 | unknown [Bacillus subtilis] | 49 | 29 |
| 290 | 8 | 6860 | 7366 | gi\|466875 | nifU; B1496_C1_157 [Mycobacterium leprae] | 49 | 35 |
| 318 | 5 | 4065 | 3190 | gi\|144859 | ORF B [Clostridium perfringens] | 49 | 25 |
| 318 | 8 | 6052 | 5033 | gi\|1439528 | EIIC-man [Lactobacillus curvatus] | 49 | 30 |
| 335 | 1 | 534 | 40 | gi\|216861 | 24K membrane protein [Pseudomonas aeruginosa] | 49 | 24 |
| 338 | 4 | 2861 | 2169 | gnl\|PID\|e288536 | F37H8.a [Caenorhabditis elegans] | 49 | 30 |
| 346 | 4 | 1257 | 2273 | gi\|536970 | ORF_f543 [Escherichia coli] | 49 | 25 |
| 355 | 20 | 12902 | 15262 | gi\|292836 | trichohyalin [Homo sapiens] | 49 | 20 |
| 366 | 1 | 1 | 1437 | gi\|405857 | yehU [Escherichia coli] | 49 | 26 |
| 375 | 8 | 7663 | 6470 | gi\|1573546 | H. influenzae predicted coding region HI0561 [Haemophilus influenzae] | 49 | 30 |
| 377 | 2 | 1624 | 392 | gi\|532553 | ORF20 [Enterococcus faecalis] | 49 | 27 |
| 399 | 5 | 3960 | 3142 | gi\|1742362 | nta operon transcriptional regulator. | 49 | 29 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | [Escherichia coli] | | |
| 456 | 1 | 1070 | 342 | gi\|290533 | similar to E. coli ORF adjacent to suc operon; similar to gntR class f regulatory proteins [Escherichia coli] | 49 | 27 |
| 619 | 1 | 2 | 232 | gi\|665956 | ribosomal protein S20 homolog [Aeromonas sobria] sp\|P45786\|RS20_AERHY 30S RIBOSOMAL PROTEIN S20 (FRAGMENT). sp\|P45788\|RS20_AERSO 30S RIBOSOMAL PROTEIN S20 (FRAGMENT). | 49 | 41 |
| 621 | 1 | 319 | 942 | gi\|149456 | nisin-resistance protein [Lactococcus lactis] | 49 | 29 |
| 630 | 1 | 3 | 1190 | gi\|537145 | ORF_f437 [Escherichia coli] | 49 | 34 |
| 736 | 1 | 859 | 2 | gi\|1592020 | hypothetical protein (SP:P37555) [Methanococcus jannaschii] | 49 | 27 |
| 849 | 1 | 232 | 11 | gi\|145514 | cyclopropane fatty acid synthase [Escherichia coli] | 49 | 35 |
| 47 | 11 | 14140 | 13307 | gi\|1045937 | M. genitalium predicted coding region MG246 [Mycoplasma genitalium] | 48 | 34 |
| 103 | 4 | 2492 | 1605 | gi\|1591514 | membrane protein [Methanococcus jannaschii] | 48 | 19 |
| 127 | 7 | 6836 | 5736 | gi\|1573128 | hypothetical [Haemophilus influenzae] | 48 | 24 |
| 138 | 22 | 14742 | 15590 | gi\|580884 | ipa-89d gene product [Bacillus subtilis] | 48 | 33 |
| 160 | 6 | 3048 | 3665 | gi\|1652295 | serine esterase [Synechocystis sp.] | 48 | 28 |
| 162 | 3 | 3048 | 2491 | gi\|143830 | xpaC [Bacillus subtilis] | 48 | 13 |
| 193 | 2 | 1257 | 310 | gi\|1591153 | hypothetical protein (SP:P46348) [Methanococcus jannaschii] | 48 | 24 |
| 219 | 1 | 61 | 573 | gnl\|PID\|e257628 | ORF [Lactococcus lactis] | 48 | 32 |
| 221 | 11 | 5952 | 6428 | gi\|1303733 | YqaN [Bacillus subtilis] | 48 | 31 |
| 232 | 4 | 2776 | 1712 | gi\|142707 | comG2 gene product [Bacillus subtilis] | 48 | 24 |
| 236 | 6 | 8618 | 7689 | gi\|550075 | cephalosporin-C deacetylase [Bacillus subtilis] | 48 | 26 |
| 238 | 28 | 25896 | 26825 | gi\|47906 | rha regulatory protein [Salmonella | 48 | 31 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
|  |  |  |  |  | typhimurium] |  |  |
| 251 | 2 | 1935 | 640 | gi\|1143026 | ORF10 [Spiroplasma virus] | 48 | 30 |
| 252 | 1 | 2036 | 3 | gnl\|PID\|e228699 | homologous to yqbO of the skin element [Bacillus subtilis] | 48 | 37 |
| 269 | 1 | 481 | 2 | gi\|1045975 | sensory rhodopsin II transducer [Mycoplasma genitalium] | 48 | 28 |
| 315 | 5 | 4604 | 2649 | gi\|396400 | similar to eukaryotic Na+/H+ exchangers [Escherichia coli] sp\|P32703\|YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOXR-ACS NTERGENIC REGION (O549). | 48 | 30 |
| 327 | 1 | 128 | 916 | gi\|216314 | esterase [Bacillus stearothermophilus] | 48 | 30 |
| 330 | 6 | 4486 | 5337 | gi\|43942 | first subunit of EII-Sor [Klebsiella pneumoniae] | 48 | 21 |
| 330 | 7 | 5325 | 6230 | gi\|147404 | mannose permease subunit II-M-Man [Escherichia coli] | 48 | 33 |
| 345 | 10 | 9571 | 10521 | gi\|1736789 | Collagenase precursor (EC 3.4.-.-). [Escherichia coli] | 48 | 26 |
| 509 | 1 | 1 | 444 | gi\|606376 | ORF_o162 [Escherichia coli] | 48 | 33 |
| 531 | 1 | 624 | 109 | sp\|P50848\|YPWA_BACSU | HYPOTHETICAL 58.2 KD PROTEIN IN KDGT-XPT INTERGENIC REGION. | 48 | 33 |
| 549 | 3 | 962 | 369 | gi\|1001212 | molybdenum cofactor biosynthesis protein C [Synechocystis sp.] | 48 | 32 |
| 725 | 1 | 3 | 500 | gi\|1151158 | repeat organellar protein [Plasmodium chabaudi] | 48 | 25 |
| 789 | 1 | 133 | 717 | gi\|42724 | rhaS (AA 1-278) [Escherichia coli] | 48 | 39 |
| 936 | 1 | 32 | 316 | gi\|532549 | ORF16 [Enterococcus faecalis] | 48 | 45 |
| 2 | 2 | 2662 | 449 | gi\|929878 | J1027 gene product [Saccharomyces cerevisiae] | 47 | 20 |
| 4 | 2 | 1002 | 2192 | gi\|763052 | integrase [Bacteriophage T270] | 47 | 29 |
| 21 | 8 | 6350 | 5355 | gi\|1066343 | mu-crystallin [Homo sapiens] | 47 | 29 |
| 25 | 3 | 915 | 2048 | gi\|1064813 | homologous to sp:PHOR_BACSU [Bacillus subtilis] | 47 | 21 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 59 | 2 | 953 | 1378 | gi\|872306 | integral membrane protein [Streptomyces pristinaespiralis] pir\|S57509\|S57509 integral membrane protein - Streptomyces ristinaespiralis | 47 | 26 |
| 81 | 7 | 4970 | 4206 | gi\|1591754 | hypothetical protein (SP:P39364) [Methanococcus jannaschii] | 47 | 22 |
| 82 | 3 | 1534 | 866 | gi\|397526 | clumping factor [Staphylococcus aureus] | 47 | 21 |
| 110 | 5 | 2313 | 3767 | gi\|151928 | 48 kDa protein [Rhodobacter sphaeroides] | 47 | 26 |
| 150 | 11 | 7839 | 9107 | gnl\|PID\|e275490 | C30H6.k [Caenorhabditis elegans] | 47 | 16 |
| 161 | 2 | 116 | 1450 | gnl\|PID\|e283830 | aminotransferase [Sulfolobus solfataricus] | 47 | 23 |
| 165 | 8 | 8081 | 6129 | gi\|924925 | heparinase III protein [Cytophaga heparina] | 47 | 29 |
| 180 | 31 | 31515 | 31054 | gi\|1591753 | N-acetylglucosamine-1-phosphate transferase [Methanococcus jannaschii] | 47 | 29 |
| 194 | 11 | 8247 | 9236 | gi\|1480429 | putative transcriptional regulator [Bacillus stearothermophilus] | 47 | 26 |
| 225 | 2 | 1039 | 701 | gi\|1212992 | Protein sequence and annotation available soon via Swiss-Prot; available at present via e-mail from LABEIT@EMBL-Heidelberg.DE [Homo sapiens] | 47 | 33 |
| 232 | 1 | 196 | 969 | gi\|293033 | integrase [Bacteriophage phi-LC3] | 47 | 30 |
| 232 | 6 | 3687 | 3340 | gi\|142706 | comG1 gene product [Bacillus subtilis] | 47 | 28 |
| 233 | 10 | 8424 | 6739 | gi\|887816 | possible start 13 codons upstream, for o765 [Escherichia coli] | 47 | 35 |
| 346 | 2 | 706 | 1083 | gi\|536970 | ORF_f543 [Escherichia coli] | 47 | 27 |
| 352 | 1 | 112 | 843 | gi\|1591857 | H+-transporting ATPase [Methanococcus jannaschii] | 47 | 28 |
| 410 | 1 | 3 | 980 | gi\|1652869 | NADH dehydrogenase [Synechocystis sp.] | 47 | 30 |
| 465 | 2 | 1976 | 1749 | gi\|211659 | p68 protein; c-rel proto-oncogene [Gallus gallus] | 47 | 30 |
| 491 | 3 | 3752 | 2466 | gi\|881434 | ORFP [Bacillus subtilis] | 47 | 24 |
| 501 | 1 | 48 | 809 | gi\|467429 | unknown [Bacillus subtilis] | 47 | 33 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 532 | 1 | 3 | 287 | gi\|755724 | alpha-toxin [Clostridium novyi] | 47 | 32 |
| 578 | 1 | 707 | 81 | gi\|532547 | ORF14 [Enterococcus faecalis] | 47 | 30 |
| 605 | 4 | 2051 | 2470 | gi\|1783233 | hypothetical [Bacillus subtilis] | 47 | 22 |
| 626 | 3 | 2459 | 2169 | gi\|1573573 | 2',3'-cyclic-nucleotide 2'-phosphodiesterase (cpdB) [Haemophilus influenzae] | 47 | 44 |
| 650 | 1 | 1042 | 341 | gi\|404802 | integrase [Saccharopolyspora erythraea] | 47 | 26 |
| 665 | 1 | 714 | 1175 | gi\|143655 | sporulation protein [Bacillus subtilis] | 47 | 22 |
| 754 | 2 | 1086 | 736 | gi\|143835 | PBSX repressor [Bacillus subtilis] | 47 | 27 |
| 845 | 1 | 2 | 241 | gi\|1303952 | YqjA [Bacillus subtilis] | 47 | 26 |
| 911 | 1 | 1 | 456 | gi\|1019640 | ORFX (a homolog to the prgX gene of the pheromone response plasmid pCF10); putative [Plasmid pHKK701] | 47 | 26 |
| 933 | 1 | 16 | 303 | gi\|331002 | first methionine codon in the ECLF1 ORF [Saimiriine herpesvirus 2] gi\|60394 ORF 73; ECLF1 [Saimiriine herpesvirus 2] | 47 | 29 |
| 17 | 17 | 13073 | 13675 | gi\|1304597 | abortive phage resistance protein [Lactococcus lactis] | 46 | 27 |
| 19 | 11 | 5515 | 6393 | gi\|1353529 | ORF12 [Bacteriophage r1t] | 46 | 28 |
| 42 | 3 | 2460 | 3011 | gi\|1064814 | homologous to sp:PHOP_BACSUB [Bacillus subtilis] | 46 | 33 |
| 49 | 9 | 4042 | 5793 | gnl\|PID\|e59644 | predicted 86.4kd protein; 52Kd observed [Mycobacteriophage 15] | 46 | 22 |
| 74 | 6 | 4039 | 3434 | gi\|143542 | RNA polymerase sigma-30 factor [Bacillus licheniformis] pir\|B28625\|SZBSSL transcription initiation factor sigma H - acillus licheniformis | 46 | 27 |
| 89 | 14 | 14259 | 12967 | gi\|1499089 | M. jannaschii predicted coding region MJ0305 [Methanococcus jannaschii] | 46 | 32 |
| 89 | 15 | 15737 | 14427 | gi\|1653339 | hypothetical protein [Synechocystis sp.] | 46 | 22 |
| 94 | 13 | 12634 | 11132 | gi\|1402515 | membrane-spanning transporter protein [Clostridium perfringens] | 46 | 23 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 100 | 18 | 13493 | 11958 | gi\|15470 | portal protein [Bacteriophage SPP1] | 46 | 31 |
| 144 | 2 | 2364 | 1126 | gnl\|PID\|e183450 | hypothetical EcsB protein [Bacillus subtilis] | 46 | 25 |
| 144 | 9 | 8977 | 6236 | gi\|710421 | unknown [Staphylococcus aureus] | 46 | 24 |
| 152 | 7 | 3397 | 4557 | gnl\|PID\|e254991 | hypothetical protein [Bacillus subtilis] | 46 | 25 |
| 158 | 7 | 7144 | 5993 | gi\|1045800 | ribose transport system permease protein [Mycoplasma genitalium] | 46 | 28 |
| 180 | 11 | 10882 | 10055 | gi\|303953 | esterase [Acinetobacter calcoaceticus] | 46 | 23 |
| 181 | 3 | 1173 | 976 | gi\|1591638 | M. jannaschii predicted coding region MJ0975 [Methanococcus jannaschii] | 46 | 36 |
| 240 | 1 | 715 | 221 | gi\|1766062 | Ats1 [Schizosaccharomyces pombe] | 46 | 28 |
| 254 | 2 | 499 | 2 | gi\|153661 | translational initiation factor IF2 [Enterococcus faecium] sp\|P18311\|IF2_ENTFC INITIATION FACTOR IF-2. | 46 | 32 |
| 262 | 4 | 5276 | 4431 | pir\|A45605\|A45605 | mature-parasite-infected erythrocyte surface antigen MESA - Plasmodium falciparum | 46 | 20 |
| 309 | 1 | 2 | 673 | gi\|1651714 | type 4 prepilin peptidase [Synechocystis sp.] | 46 | 40 |
| 312 | 1 | 18 | 872 | gi\|580884 | ipa-89d gene product [Bacillus subtilis] | 46 | 32 |
| 324 | 6 | 4450 | 4836 | gi\|1061418 | ArsC [Plasmid R46] | 46 | 28 |
| 345 | 1 | 2241 | 1333 | gi\|144859 | ORF B [Clostridium perfringens] | 46 | 24 |
| 386 | 4 | 1438 | 2421 | gi\|405894 | 1-phosphofructokinase [Escherichia coli] | 46 | 31 |
| 395 | 8 | 3584 | 3853 | gnl\|PID\|e120267 | sucrose-phosphate synthase [Beta vulgaris] | 46 | 25 |
| 491 | 2 | 2527 | 1169 | gnl\|PID\|e267595 | Unknown, similar to peptidases [Bacillus subtilis] | 46 | 29 |
| 495 | 3 | 612 | 869 | gi\|406286 | triose phosphate/phosphate translocator [Flaveria pringlei] pir\|S37553\|S37553 triose phosphate/3-phosphoglycerate/phosphate ranslocator - Flaveria pringlei | 46 | 27 |
| 513 | 1 | 2 | 946 | gi\|143024 | glucose-resistance amylase regulator | 46 | 26 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | [Bacillus subtilis] pir\|S15318\|S15318 ccpA protein - Bacillus subtilis sp\|P25144\|CCPA_BACSU GLUCOSE-RESISTANCE AMYLASE REGULATOR CATABOLITE CONTROL PROTEIN). | | |
| 520 | 3 | 914 | 2674 | gi\|1163086 | microfilarial sheath protein SHP3 [Brugia malayi] | 46 | 27 |
| 554 | 1 | 3 | 788 | gi\|413972 | ipa-48r gene product [Bacillus subtilis] | 46 | 27 |
| 568 | 1 | 1574 | 3 | gi\|532549 | ORF16 [Enterococcus faecalis] | 46 | 28 |
| 809 | 1 | 506 | 135 | gi\|49021 | surface exclusion protein (SEA1) [Enterococcus faecalis] ir\|S22452\|S22452 surface exclusion protein seal precursor - terococcus faecalis plasmid pAD1 | 46 | 28 |
| 813 | 1 | 2 | 1090 | gi\|150556 | surface protein [Plasmid pCF10] | 46 | 34 |
| 78 | 2 | 4915 | 2516 | gi\|577295 | The ha1225 gene product is related to human alpha-glucosidase. [Homo apiens] | 45 | 20 |
| 81 | 9 | 6123 | 5386 | gi\|147200 | phnF protein [Escherichia coli] | 45 | 28 |
| 85 | 1 | 120 | 761 | gi\|457514 | gltC [Bacillus subtilis] | 45 | 19 |
| 94 | 11 | 10681 | 9668 | gi\|289753 | homology with nucleolin protein; putative [Caenorhabditis elegans] pir\|S44897\|S44897 ZK1236.2 protein - Caenorhabditis elegans sp\|P34618\|YO82_CAEEL HYPOTHETICAL 33.8 KD PROTEIN ZK1236.2 IN HROMOSOME III. | 45 | 23 |
| 108 | 3 | 2427 | 1789 | gnl\|PID\|e263931 | OrfD [Streptococcus pneumoniae] | 45 | 27 |
| 108 | 4 | 3338 | 2352 | gi\|606150 | ORF_f309 [Escherichia coli] | 45 | 25 |
| 131 | 6 | 3981 | 5309 | gi\|1590845 | hypothetical protein (PIR:S51413) [Methanococcus jannaschii] | 45 | 36 |
| 144 | 11 | 10215 | 8944 | gi\|1001554 | hypothetical protein [Synechocystis sp.] | 45 | 30 |
| 164 | 11 | 8247 | 6736 | gi\|409925 | VirR positive regulator [Streptococcus pyogenes] | 45 | 22 |
| 192 | 1 | 1598 | 591 | gi\|1736826 | Lysozyme M1 precursor (EC 3.2.1.17) (1,4-b-N-acetylmuramidase M1). [Escherichia | 45 | 27 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | coli] | | |
| 223 | 16 | 14409 | 15212 | gi\|1651958 | hypothetical protein [Synechocystis sp.] | 45 | 32 |
| 279 | 7 | 5236 | 5772 | gi\|1736514 | Isochorismatase (EC 3.3.2.1) (2,3 dihydro-2,3 dihydroxybenzoate synthase). [Escherichia coli] | 45 | 29 |
| 364 | 3 | 2419 | 4098 | gi\|309662 | pheromone binding protein [Plasmid pCF10] | 45 | 26 |
| 459 | 1 | 2 | 307 | gi\|1679640 | ORFA [Mycoplasma mycoides mycoides SC] | 45 | 27 |
| 491 | 1 | 1022 | 135 | sp\|P27434\|YFGA_ECOLI | HYPOTHETICAL 36.2 KD PROTEIN IN NDK-GCPE INTERGENIC REGION. | 45 | 20 |
| 496 | 1 | 847 | 2 | gi\|1208489 | serum resistance locus BrkB [Synechocystis sp.] | 45 | 19 |
| 542 | 2 | 1169 | 804 | gi\|1064811 | function unknown [Bacillus subtilis] | 45 | 28 |
| 63 | 3 | 1047 | 1919 | gi\|39848 | U3 [Bacillus subtilis] | 44 | 26 |
| 93 | 3 | 1108 | 1374 | sp\|Q04747\|SRF2_BACSU | SURFACTIN SYNTHETASE SUBUNIT 2. | 44 | 27 |
| 155 | 10 | 8354 | 7620 | sp\|P35136\|SERA_BACSU | D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) (PGDH). | 44 | 29 |
| 215 | 2 | 2192 | 1134 | gi\|468760 | ORF334 [Rhizobium meliloti] | 44 | 31 |
| 303 | 1 | 466 | 2 | gi\|431950 | similar to a B.subtilis gene (GB: BACHEMEHY_5) [Clostridium asteurianum] | 44 | 22 |
| 310 | 1 | 284 | 39 | pir\|S01294\|S01294 | intermediate filament protein B - Roman snail | 44 | 26 |
| 311 | 1 | 122 | 2668 | gi\|532549 | ORF16 [Enterococcus faecalis] | 44 | 27 |
| 320 | 1 | 709 | 2 | gi\|290801 | member of super-family of ABC proteins [Francisella tularensis (var. ovicida)] | 44 | 23 |
| 341 | 14 | 13882 | 12998 | gi\|142863 | replication initiation protein [Bacillus subtilis] | 44 | 16 |
| 345 | 15 | 16445 | 18001 | gi\|151282 | DL-hydantoinase [Pseudomonas sp.] | 44 | 34 |
| 386 | 3 | 1340 | 570 | sp\|P46117\|YARA_PROST | HYPOTHETICAL 31.5 KD PROTEIN IN AARA 3'REGION. | 44 | 19 |
| 862 | 1 | 483 | 4 | gi\|929796 | precursor of the major merozoite surface | 44 | 26 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | antigens [Plasmodium alciparum] | | |
| 19 | 3 | 1695 | 1372 | gi\|603263 | Yel055p [Saccharomyces cerevisiae] | 43 | 31 |
| 45 | 17 | 14045 | 14995 | gnl\|PID\|e233895 | hypothetical protein [Bacillus subtilis] | 43 | 32 |
| 57 | 1 | 667 | 317 | gi\|664840 | TagB [Dictyostelium discoideum] | 43 | 22 |
| 71 | 2 | 1537 | 2568 | gi\|1303981 | YqkD [Bacillus subtilis] | 43 | 26 |
| 72 | 18 | 20511 | 20164 | gi\|349045 | merozoite surface antigen 2 [Plasmodium falciparum] | 43 | 36 |
| 94 | 9 | 6581 | 6039 | gi\|1146245 | putative [Bacillus subtilis] | 43 | 28 |
| 180 | 17 | 16391 | 17656 | gi\|290540 | f445 [Escherichia coli] | 43 | 24 |
| 252 | 2 | 2407 | 1829 | gi\|154381 | chemoreceptor [Salmonella typhimurium] | 43 | 19 |
| 276 | 30 | 19091 | 18480 | gi\|15470 | portal protein [Bacteriophage SPP1] | 43 | 23 |
| 311 | 2 | 2666 | 4639 | gi\|160299 | glutamic acid-rich protein [Plasmodium falciparum] pir\|A54514\|A54514 glutamic acid-rich protein precursor - Plasmodium alciparum | 43 | 28 |
| 631 | 2 | 1126 | 2328 | gi\|1519696 | coded for by C. elegans cDNA yk126f9.5; coded for by C. elegans cDNA yk159h6.3; coded for by C. elegans cDNA yk126f9.3; coded for by C. elegans cDNA yk159h6.5 [Caenorhabditis elegans] | 43 | 27 |
| 11 | 3 | 1509 | 2342 | gi\|143150 | levR [Bacillus subtilis] | 42 | 21 |
| 45 | 14 | 10730 | 12028 | gi\|666069 | orf2 gene product [Lactobacillus leichmannii] | 42 | 23 |
| 72 | 19 | 21070 | 21981 | gnl\|PID\|e236595 | orf7 gene product [Enterococcus faecalis] | 42 | 23 |
| 123 | 35 | 32205 | 32768 | gi\|1772652 | 2-keto-3-deoxygluconate kinase [Haloferax alicantei] | 42 | 27 |
| 136 | 5 | 2737 | 2375 | gi\|153858 | wall-associated protein [Streptococcus mutans] | 42 | 27 |
| 167 | 4 | 2701 | 6540 | gi\|1519696 | coded for by C. elegans cDNA yk126f9.5; coded for by C. elegans cDNA yk159h6.3; coded for by C. elegans cDNA yk126f9.3; coded for by C. elegans cDNA yk159h6.5 | 42 | 27 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| | | | | | [Caenorhabditis elegans] | | |
| 195 | 31 | 12430 | 13155 | pir\|S33124\|S33124 | tpr protein - human | 42 | 24 |
| 211 | 1 | 187 | 2 | gi\|1653346 | GDP-mannose pyrophosphorylase [Synechocystis sp.] | 42 | 33 |
| 242 | 13 | 8089 | 12447 | gi\|951460 | FIM-C.1 gene product [Xenopus laevis] | 42 | 31 |
| 305 | 5 | 4354 | 5340 | gi\|1408485 | yxdM gene product [Bacillus subtilis] | 42 | 25 |
| 355 | 18 | 9964 | 12549 | gi\|532549 | ORF16 [Enterococcus faecalis] | 42 | 30 |
| 446 | 4 | 4428 | 5261 | gi\|47528 | glucosyltransferase S [Streptococcus salivarius] | 42 | 25 |
| 656 | 3 | 2866 | 3456 | gi\|142857 | MreD protein [Bacillus subtilis] | 42 | 25 |
| 686 | 11 | 3646 | 3921 | pir\|A44805\|A44805 | eggshell protein - fluke (Schistosoma haematobium) (subclone SH.E 2-1) | 42 | 42 |
| 920 | 1 | 41 | 316 | gi\|532549 | ORF16 [Enterococcus faecalis] | 42 | 40 |
| 23 | 3 | 729 | 487 | gi\|414525 | meiotin-1 [Lilium longiflorum] | 41 | 41 |
| 56 | 5 | 3511 | 2324 | gi\|1591610 | probable ATP-dependent helicase [Methanococcus jannaschii] | 41 | 21 |
| 98 | 17 | 16843 | 16274 | gi\|1742129 | Immunity repressor protein. [Escherichia coli] | 41 | 23 |
| 167 | 6 | 6734 | 9811 | gnl\|PID\|e249616 | F56H9.1 [Caenorhabditis elegans] | 41 | 37 |
| 171 | 13 | 10879 | 11871 | gi\|331002 | first methionine codon in the ECLF1 ORF [Saimiriine herpesvirus 2] gi\|60394 ORF 73; ECLF1 [Saimiriine herpesvirus 2] | 41 | 23 |
| 181 | 2 | 1012 | 500 | gi\|455315 | ORF 4 [Plasmid pIP404] | 41 | 24 |
| 230 | 4 | 3664 | 3224 | gi\|498251 | glutamate/aspartate transporter II [Homo sapiens] | 41 | 22 |
| 718 | 1 | 2 | 613 | gi\|984656 | ORF3 [Salmonella typhimurium] | 41 | 22 |
| 19 | 30 | 16391 | 17770 | gi\|806704 | Upf2p [Saccharomyces cerevisiae] | 40 | 21 |
| 164 | 16 | 16440 | 17951 | gi\|348056 | trans-acting positive regulator [Bacillus anthracis] | 40 | 22 |
| 200 | 12 | 5956 | 4841 | gi\|1574243 | H. influenzae predicted coding region HI1405 [Haemophilus influenzae] | 40 | 24 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 216 | 10 | 6799 | 7194 | gi\|146279 | glucitol-specific enzyme III (gutB) [Escherichia coli] | 40 | 27 |
| 292 | 13 | 8633 | 10741 | gi\|1008233 | ORF YJL076w [Saccharomyces cerevisiae] | 40 | 18 |
| 345 | 13 | 14050 | 15333 | gi\|581051 | cytosine permease [Escherichia coli] | 40 | 25 |
| 521 | 1 | 177 | 1466 | gi\|289614 | homology with glucose induced repressor, GRR1; putative Caenorhabditis elegans] | 40 | 18 |
| 64 | 3 | 2646 | 1855 | gi\|154924 | spectinomycin adenyltransferase [Transposon Tn554] | 39 | 27 |
| 100 | 17 | 12037 | 10565 | gi\|1052806 | product required for head morphogenesis [Bacteriophage SPP1] | 39 | 24 |
| 529 | 1 | 326 | 4939 | gi\|295671 | selected as a weak suppressor of a mutant of the subunit AC40 of DNA ependant RNA polymerase I and III [Saccharomyces cerevisiae] | 39 | 19 |
| 49 | 2 | 518 | 931 | gi\|166162 | Bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11] | 38 | 19 |
| 54 | 19 | 11264 | 10854 | gi\|160186 | circumsporozoite protein [Plasmodium vivax] | 38 | 31 |
| 164 | 21 | 22793 | 23587 | gi\|603857 | secreted acid phosphatase 2 (SAP2) [Leishmania mexicana] | 38 | 18 |
| 167 | 3 | 2322 | 2756 | gi\|435039 | proline-rich cell wall protein [Gossypium hirsutum] | 38 | 36 |
| 204 | 2 | 133 | 798 | gi\|396401 | No definition line found [Escherichia coli] | 38 | 25 |
| 475 | 2 | 761 | 1792 | gi\|1574532 | H. influenzae predicted coding region HI1680 [Haemophilus influenzae] | 38 | 27 |
| 164 | 19 | 20738 | 21385 | gi\|165704 | [Rabbit smooth muscle myosin light chain kinase mRNA, complete DS.], gene product [Oryctolagus cuniculus] | 37 | 20 |
| 394 | 6 | 5649 | 6395 | gi\|603857 | secreted acid phosphatase 2 (SAP2) [Leishmania mexicana] | 36 | 16 |
| 958 | 1 | 1 | 459 | gi\|951460 | FIM-C.1 gene product [Xenopus laevis] | 36 | 28 |

Table 2: *E. faecalis* - Putative coding regions of novel proteins similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) | Match accession | Match gene name | % Sim | % Ident |
|---|---|---|---|---|---|---|---|
| 399 | 21 | 16383 | 21359 | gi\|1707247 | partial CDS [Caenorhabditis elegans] | 34 | 13 |
| 150 | 12 | 9056 | 11740 | gi\|1015903 | ORF YJR151c [Saccharomyces cerevisiae] | 33 | 19 |
| 195 | 34 | 13017 | 15512 | gi\|632549 | NF-180 [Petromyzon marinus] | 33 | 18 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 2 | 1 | 458 | 3 |
| 2 | 3 | 2208 | 2624 |
| 5 | 3 | 928 | 1440 |
| 8 | 6 | 4792 | 5877 |
| 8 | 7 | 5480 | 5262 |
| 12 | 1 | 2 | 832 |
| 12 | 2 | 771 | 4622 |
| 13 | 1 | 2 | 1684 |
| 14 | 1 | 531 | 130 |
| 15 | 2 | 862 | 1197 |
| 16 | 1 | 51 | 200 |
| 17 | 4 | 3309 | 3665 |
| 17 | 13 | 10079 | 10261 |
| 17 | 18 | 14431 | 13682 |
| 17 | 22 | 21525 | 21956 |
| 17 | 27 | 27055 | 27567 |
| 18 | 4 | 2172 | 1591 |
| 18 | 5 | 2524 | 2249 |
| 18 | 7 | 3467 | 3715 |
| 18 | 8 | 4082 | 3555 |
| 18 | 9 | 4333 | 4055 |
| 18 | 10 | 4395 | 4204 |
| 18 | 11 | 4498 | 4677 |
| 18 | 12 | 4656 | 5393 |
| 18 | 13 | 5878 | 5492 |
| 18 | 15 | 6296 | 6931 |
| 19 | 1 | 1047 | 676 |
| 19 | 2 | 1068 | 1247 |
| 19 | 4 | 1747 | 2031 |
| 19 | 5 | 2244 | 2612 |
| 19 | 7 | 2797 | 2943 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 19 | 9 | 3873 | 4730 |
| 19 | 13 | 6884 | 7420 |
| 19 | 14 | 7428 | 8042 |
| 19 | 16 | 9246 | 8425 |
| 19 | 17 | 9412 | 9615 |
| 19 | 19 | 9733 | 9918 |
| 19 | 20 | 10032 | 10334 |
| 19 | 21 | 10422 | 11009 |
| 19 | 22 | 11516 | 11944 |
| 19 | 24 | 12423 | 12881 |
| 19 | 26 | 14606 | 15427 |
| 19 | 27 | 15414 | 15848 |
| 19 | 28 | 15802 | 16134 |
| 19 | 29 | 16064 | 16393 |
| 19 | 32 | 17846 | 18052 |
| 19 | 33 | 18021 | 18356 |
| 19 | 34 | 18334 | 18684 |
| 19 | 35 | 18659 | 19036 |
| 19 | 36 | 18991 | 19677 |
| 19 | 37 | 19671 | 20132 |
| 19 | 39 | 22603 | 23337 |
| 19 | 40 | 23319 | 25580 |
| 21 | 2 | 762 | 262 |
| 21 | 5 | 3440 | 2925 |
| 21 | 10 | 7684 | 7241 |
| 23 | 5 | 2098 | 2652 |
| 23 | 8 | 4912 | 4709 |
| 23 | 9 | 4911 | 5246 |
| 23 | 10 | 5087 | 5353 |
| 23 | 22 | 14318 | 14926 |
| 23 | 23 | 14924 | 15565 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 23 | 24 | 15559 | 16083 |
| 23 | 29 | 17567 | 18022 |
| 25 | 2 | 553 | 1005 |
| 25 | 5 | 3363 | 2653 |
| 26 | 2 | 1220 | 1654 |
| 27 | 1 | 297 | 4 |
| 28 | 1 | 239 | 2833 |
| 29 | 5 | 3244 | 2822 |
| 29 | 6 | 4014 | 3301 |
| 29 | 7 | 4168 | 4557 |
| 29 | 8 | 5620 | 4595 |
| 32 | 3 | 2646 | 1375 |
| 32 | 4 | 2573 | 3010 |
| 39 | 9 | 4636 | 4986 |
| 40 | 2 | 1346 | 981 |
| 43 | 1 | 120 | 620 |
| 43 | 4 | 1972 | 2280 |
| 45 | 3 | 1557 | 1961 |
| 45 | 4 | 2012 | 2230 |
| 45 | 5 | 2218 | 2553 |
| 45 | 11 | 7226 | 5670 |
| 45 | 12 | 7270 | 10113 |
| 45 | 13 | 10013 | 10732 |
| 46 | 1 | 42 | 872 |
| 46 | 2 | 886 | 1125 |
| 46 | 4 | 2807 | 3100 |
| 47 | 4 | 5101 | 5625 |
| 47 | 10 | 13239 | 12847 |
| 49 | 1 | 106 | 504 |
| 49 | 8 | 2858 | 4132 |
| 49 | 10 | 5777 | 6193 |

Table 3: E. *faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 49 | 11 | 6166 | 6720 |
| 52 | 5 | 3505 | 3110 |
| 52 | 7 | 5160 | 5603 |
| 52 | 8 | 5662 | 5459 |
| 54 | 2 | 400 | 729 |
| 54 | 4 | 1326 | 1610 |
| 54 | 5 | 2354 | 1335 |
| 54 | 6 | 1676 | 2080 |
| 54 | 7 | 2151 | 2576 |
| 54 | 12 | 4181 | 3954 |
| 54 | 13 | 5975 | 6289 |
| 54 | 14 | 6869 | 7144 |
| 54 | 15 | 7433 | 7107 |
| 54 | 18 | 9764 | 11086 |
| 55 | 2 | 252 | 440 |
| 56 | 2 | 1344 | 658 |
| 57 | 9 | 12450 | 12605 |
| 58 | 7 | 7066 | 6425 |
| 59 | 3 | 1350 | 952 |
| 59 | 4 | 1225 | 1515 |
| 59 | 7 | 2958 | 3200 |
| 62 | 6 | 4116 | 3007 |
| 63 | 1 | 77 | 364 |
| 63 | 2 | 455 | 1060 |
| 63 | 7 | 5422 | 5910 |
| 63 | 8 | 5870 | 6751 |
| 63 | 9 | 6688 | 7296 |
| 64 | 2 | 1849 | 1523 |
| 64 | 4 | 3183 | 2644 |
| 64 | 5 | 3422 | 3213 |
| 65 | 5 | 3787 | 3389 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 65 | 7 | 5043 | 4300 |
| 65 | 8 | 5354 | 4959 |
| 65 | 9 | 7005 | 6328 |
| 67 | 6 | 3719 | 4060 |
| 68 | 2 | 569 | 348 |
| 68 | 5 | 3234 | 2821 |
| 68 | 6 | 3808 | 3221 |
| 68 | 10 | 7495 | 8106 |
| 70 | 2 | 2102 | 1614 |
| 70 | 3 | 2019 | 2231 |
| 71 | 3 | 3362 | 3787 |
| 72 | 21 | 22464 | 22709 |
| 72 | 22 | 22690 | 23019 |
| 72 | 23 | 23013 | 23834 |
| 73 | 1 | 154 | 2 |
| 74 | 1 | 61 | 486 |
| 74 | 3 | 1334 | 1981 |
| 75 | 4 | 3227 | 2136 |
| 75 | 5 | 3994 | 3251 |
| 75 | 6 | 3348 | 3632 |
| 75 | 7 | 4519 | 4043 |
| 75 | 8 | 4296 | 4529 |
| 75 | 10 | 6518 | 5769 |
| 76 | 2 | 1079 | 1897 |
| 76 | 4 | 2113 | 2436 |
| 76 | 6 | 4737 | 4105 |
| 77 | 3 | 1874 | 2704 |
| 77 | 4 | 2665 | 2459 |
| 78 | 3 | 5814 | 5398 |
| 79 | 3 | 848 | 1645 |
| 79 | 4 | 2121 | 1642 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 81 | 8 | 5392 | 4961 |
| 81 | 13 | 8428 | 8874 |
| 81 | 21 | 15746 | 14802 |
| 82 | 1 | 858 | 4 |
| 82 | 2 | 198 | 383 |
| 83 | 3 | 2194 | 2604 |
| 83 | 4 | 2728 | 2405 |
| 83 | 6 | 2855 | 3172 |
| 83 | 10 | 7188 | 6184 |
| 83 | 11 | 7415 | 7065 |
| 83 | 17 | 12259 | 12561 |
| 83 | 21 | 15890 | 16456 |
| 83 | 23 | 16946 | 17251 |
| 84 | 5 | 7071 | 7949 |
| 85 | 7 | 6518 | 6174 |
| 89 | 2 | 1012 | 599 |
| 89 | 3 | 1382 | 939 |
| 89 | 4 | 2350 | 1370 |
| 89 | 5 | 2523 | 2314 |
| 89 | 9 | 7505 | 7182 |
| 89 | 16 | 15846 | 15673 |
| 89 | 19 | 20070 | 19045 |
| 90 | 1 | 3 | 689 |
| 91 | 7 | 3834 | 4127 |
| 91 | 8 | 4288 | 5268 |
| 91 | 9 | 7259 | 5748 |
| 91 | 12 | 9737 | 8973 |
| 91 | 13 | 10162 | 9731 |
| 92 | 3 | 1458 | 958 |
| 92 | 4 | 1934 | 1287 |
| 93 | 2 | 479 | 949 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 93 | 4 | 1344 | 1727 |
| 94 | 1 | 770 | 45 |
| 94 | 3 | 1460 | 1618 |
| 94 | 5 | 2279 | 1734 |
| 94 | 12 | 11000 | 10641 |
| 95 | 11 | 7674 | 7907 |
| 95 | 12 | 8604 | 8056 |
| 95 | 13 | 8725 | 8546 |
| 96 | 1 | 758 | 1018 |
| 96 | 2 | 1038 | 1469 |
| 98 | 5 | 6809 | 5994 |
| 98 | 10 | 10338 | 10652 |
| 98 | 11 | 10650 | 11558 |
| 99 | 2 | 232 | 513 |
| 100 | 4 | 3728 | 4048 |
| 100 | 6 | 5866 | 5378 |
| 100 | 7 | 6574 | 5921 |
| 100 | 8 | 6923 | 6534 |
| 100 | 9 | 7355 | 6921 |
| 100 | 10 | 7698 | 7339 |
| 100 | 11 | 8226 | 7744 |
| 100 | 13 | 9395 | 8514 |
| 100 | 15 | 10368 | 10102 |
| 100 | 19 | 14770 | 13505 |
| 100 | 20 | 15300 | 14758 |
| 100 | 21 | 15783 | 15298 |
| 100 | 23 | 17699 | 17292 |
| 100 | 25 | 20933 | 20625 |
| 100 | 26 | 21200 | 20946 |
| 100 | 28 | 23713 | 23156 |
| 100 | 29 | 23948 | 23691 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 100 | 30 | 24312 | 23965 |
| 100 | 31 | 24550 | 24287 |
| 100 | 32 | 24912 | 24565 |
| 100 | 33 | 25173 | 24910 |
| 100 | 34 | 26339 | 25158 |
| 100 | 36 | 27251 | 26994 |
| 100 | 37 | 27945 | 27232 |
| 100 | 39 | 28442 | 28227 |
| 100 | 40 | 28657 | 28403 |
| 100 | 46 | 30439 | 31146 |
| 100 | 47 | 31158 | 31712 |
| 101 | 2 | 850 | 464 |
| 101 | 3 | 2453 | 1899 |
| 102 | 6 | 5023 | 5616 |
| 102 | 9 | 6704 | 7111 |
| 103 | 7 | 5454 | 5296 |
| 105 | 2 | 1244 | 1828 |
| 106 | 4 | 5114 | 3294 |
| 106 | 6 | 7622 | 6168 |
| 106 | 7 | 6577 | 6867 |
| 108 | 6 | 5192 | 4158 |
| 110 | 1 | 2 | 454 |
| 110 | 6 | 3689 | 4207 |
| 110 | 9 | 9374 | 8553 |
| 110 | 10 | 9903 | 9361 |
| 110 | 11 | 10175 | 9843 |
| 111 | 6 | 3118 | 3267 |
| 112 | 4 | 2170 | 1043 |
| 114 | 2 | 1347 | 1135 |
| 116 | 8 | 4782 | 5147 |
| 117 | 4 | 2437 | 2670 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 117 | 6 | 3876 | 4640 |
| 117 | 8 | 5643 | 5927 |
| 117 | 9 | 6195 | 6488 |
| 117 | 12 | 9655 | 9837 |
| 119 | 1 | 3 | 500 |
| 119 | 2 | 670 | 1158 |
| 119 | 4 | 2730 | 2284 |
| 121 | 3 | 2276 | 3670 |
| 123 | 14 | 14304 | 14555 |
| 123 | 16 | 15305 | 15147 |
| 123 | 24 | 21896 | 22663 |
| 123 | 34 | 31458 | 32207 |
| 125 | 3 | 1581 | 1300 |
| 125 | 7 | 4516 | 4346 |
| 126 | 2 | 85 | 312 |
| 127 | 2 | 1047 | 787 |
| 127 | 3 | 2006 | 1299 |
| 127 | 4 | 3432 | 1924 |
| 128 | 4 | 3094 | 2747 |
| 128 | 5 | 3466 | 3305 |
| 128 | 6 | 4625 | 3507 |
| 128 | 7 | 4726 | 4550 |
| 128 | 13 | 8947 | 8522 |
| 128 | 15 | 9325 | 9582 |
| 128 | 17 | 10126 | 10380 |
| 128 | 24 | 17649 | 18038 |
| 129 | 1 | 276 | 1769 |
| 130 | 7 | 6478 | 6702 |
| 130 | 11 | 9386 | 9769 |
| 133 | 7 | 6622 | 7380 |
| 135 | 2 | 2289 | 1153 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 135 | 3 | 3380 | 2271 |
| 135 | 5 | 3778 | 3930 |
| 135 | 6 | 5835 | 5137 |
| 135 | 7 | 6649 | 5852 |
| 135 | 8 | 7021 | 6647 |
| 135 | 9 | 7420 | 7034 |
| 136 | 2 | 963 | 379 |
| 136 | 3 | 2009 | 939 |
| 136 | 4 | 2344 | 1973 |
| 138 | 4 | 5051 | 3636 |
| 138 | 11 | 8499 | 8753 |
| 138 | 12 | 8682 | 8536 |
| 138 | 13 | 8923 | 9270 |
| 138 | 14 | 9333 | 9887 |
| 138 | 15 | 9628 | 10308 |
| 138 | 16 | 10422 | 10216 |
| 138 | 23 | 15980 | 15678 |
| 138 | 24 | 16437 | 16063 |
| 138 | 30 | 19388 | 19828 |
| 139 | 3 | 1068 | 1466 |
| 139 | 4 | 3338 | 1983 |
| 139 | 5 | 3769 | 3317 |
| 139 | 6 | 4114 | 3818 |
| 139 | 7 | 4838 | 4236 |
| 139 | 10 | 5639 | 5175 |
| 142 | 1 | 369 | 106 |
| 142 | 2 | 1005 | 367 |
| 142 | 3 | 2140 | 980 |
| 142 | 4 | 2504 | 2127 |
| 142 | 5 | 2821 | 2474 |
| 142 | 6 | 3294 | 2806 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 142 | 7 | 4000 | 3635 |
| 143 | 1 | 650 | 3 |
| 143 | 3 | 1090 | 173 |
| 143 | 4 | 1044 | 433 |
| 144 | 10 | 7570 | 8403 |
| 144 | 12 | 10727 | 10335 |
| 145 | 1 | 188 | 30 |
| 145 | 2 | 775 | 978 |
| 150 | 9 | 6876 | 7166 |
| 150 | 13 | 11538 | 11242 |
| 152 | 1 | 35 | 445 |
| 152 | 2 | 405 | 914 |
| 152 | 3 | 912 | 1430 |
| 152 | 4 | 1349 | 2212 |
| 152 | 5 | 2210 | 2896 |
| 152 | 6 | 2739 | 3368 |
| 152 | 8 | 4479 | 4694 |
| 152 | 11 | 6647 | 7321 |
| 154 | 7 | 4557 | 4195 |
| 155 | 3 | 1227 | 2180 |
| 155 | 12 | 8726 | 9022 |
| 156 | 3 | 3179 | 2664 |
| 158 | 11 | 10876 | 11220 |
| 160 | 1 | 545 | 3 |
| 162 | 1 | 228 | 1349 |
| 162 | 2 | 2513 | 1653 |
| 162 | 7 | 9163 | 7664 |
| 162 | 9 | 10619 | 10990 |
| 162 | 11 | 11891 | 11427 |
| 163 | 3 | 1043 | 1234 |
| 163 | 5 | 3217 | 2021 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 163 | 6 | 3455 | 3198 |
| 163 | 8 | 5611 | 4931 |
| 163 | 9 | 5969 | 5580 |
| 163 | 10 | 6144 | 5926 |
| 164 | 2 | 1100 | 1687 |
| 164 | 9 | 5729 | 5259 |
| 164 | 10 | 6778 | 5639 |
| 164 | 12 | 8277 | 8450 |
| 164 | 17 | 18224 | 18526 |
| 164 | 24 | 24751 | 24536 |
| 164 | 27 | 25764 | 26369 |
| 165 | 1 | 17 | 481 |
| 165 | 2 | 2213 | 1389 |
| 165 | 12 | 9871 | 9689 |
| 165 | 14 | 11416 | 10367 |
| 166 | 3 | 1250 | 1669 |
| 167 | 5 | 3774 | 3439 |
| 167 | 7 | 10479 | 14498 |
| 167 | 10 | 17476 | 18768 |
| 168 | 2 | 665 | 393 |
| 172 | 9 | 7018 | 6701 |
| 172 | 10 | 7097 | 7930 |
| 173 | 1 | 2 | 412 |
| 173 | 3 | 2341 | 2024 |
| 173 | 6 | 4234 | 5055 |
| 173 | 9 | 7882 | 7295 |
| 173 | 10 | 7413 | 7571 |
| 173 | 14 | 12308 | 11748 |
| 174 | 4 | 2350 | 3021 |
| 174 | 5 | 3082 | 3498 |
| 178 | 3 | 866 | 1105 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 179 | 8 | 8115 | 7816 |
| 179 | 17 | 17407 | 17135 |
| 180 | 4 | 3524 | 4537 |
| 180 | 5 | 4686 | 5687 |
| 180 | 6 | 5897 | 6949 |
| 180 | 9 | 9721 | 9299 |
| 180 | 10 | 9996 | 9715 |
| 180 | 20 | 19805 | 19954 |
| 180 | 23 | 21808 | 21509 |
| 180 | 25 | 24127 | 26460 |
| 180 | 27 | 27977 | 27474 |
| 181 | 1 | 381 | 82 |
| 183 | 1 | 190 | 2 |
| 183 | 4 | 1849 | 2211 |
| 183 | 5 | 2350 | 2568 |
| 183 | 7 | 3592 | 2978 |
| 183 | 8 | 4176 | 3571 |
| 185 | 2 | 1260 | 1424 |
| 185 | 3 | 2722 | 1301 |
| 185 | 4 | 3612 | 2671 |
| 187 | 2 | 727 | 1302 |
| 187 | 3 | 1293 | 1745 |
| 187 | 5 | 2592 | 2173 |
| 189 | 1 | 18 | 2180 |
| 190 | 1 | 466 | 68 |
| 190 | 2 | 896 | 411 |
| 190 | 4 | 1878 | 2165 |
| 190 | 5 | 2740 | 2384 |
| 190 | 10 | 10281 | 8875 |
| 191 | 2 | 861 | 658 |
| 191 | 3 | 1096 | 827 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 192 | 2 | 1881 | 1564 |
| 193 | 1 | 316 | 2 |
| 193 | 7 | 4667 | 3813 |
| 194 | 1 | 30 | 641 |
| 194 | 2 | 608 | 1582 |
| 195 | 1 | 2 | 433 |
| 195 | 2 | 431 | 943 |
| 195 | 3 | 1055 | 465 |
| 195 | 4 | 972 | 1487 |
| 195 | 5 | 1507 | 1995 |
| 195 | 6 | 3314 | 1851 |
| 195 | 9 | 3089 | 3529 |
| 195 | 10 | 3521 | 3312 |
| 195 | 12 | 6604 | 6837 |
| 195 | 13 | 7049 | 6786 |
| 195 | 14 | 6825 | 7700 |
| 195 | 15 | 7682 | 7047 |
| 195 | 16 | 7202 | 7417 |
| 195 | 18 | 8278 | 9036 |
| 195 | 20 | 8583 | 8837 |
| 195 | 21 | 8871 | 9602 |
| 195 | 22 | 9251 | 9403 |
| 195 | 23 | 9600 | 10022 |
| 195 | 25 | 10020 | 10226 |
| 195 | 26 | 11229 | 10024 |
| 195 | 27 | 10659 | 10946 |
| 195 | 28 | 10944 | 11318 |
| 195 | 30 | 12449 | 12246 |
| 195 | 32 | 13212 | 12505 |
| 195 | 33 | 12558 | 12773 |
| 195 | 35 | 13673 | 14011 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 195 | 36 | 14811 | 14143 |
| 195 | 38 | 16061 | 16363 |
| 195 | 39 | 16320 | 16799 |
| 195 | 40 | 16515 | 16333 |
| 196 | 1 | 608 | 1411 |
| 197 | 9 | 9269 | 9553 |
| 200 | 2 | 1103 | 249 |
| 200 | 3 | 1335 | 1033 |
| 200 | 4 | 1769 | 1284 |
| 200 | 5 | 2124 | 1747 |
| 200 | 6 | 2792 | 2106 |
| 200 | 7 | 3073 | 2708 |
| 200 | 8 | 3510 | 3061 |
| 200 | 9 | 4126 | 3467 |
| 200 | 10 | 4350 | 4042 |
| 200 | 11 | 4847 | 4368 |
| 200 | 14 | 6487 | 6182 |
| 200 | 15 | 6681 | 6499 |
| 200 | 18 | 10749 | 9307 |
| 200 | 20 | 11787 | 11464 |
| 200 | 22 | 12859 | 12410 |
| 201 | 1 | 509 | 105 |
| 201 | 3 | 3704 | 3237 |
| 202 | 7 | 5296 | 4817 |
| 205 | 2 | 117 | 323 |
| 205 | 5 | 1669 | 2148 |
| 206 | 2 | 546 | 196 |
| 206 | 3 | 841 | 632 |
| 206 | 4 | 1622 | 777 |
| 206 | 9 | 5466 | 5035 |
| 209 | 1 | 472 | 86 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 209 | 3 | 1510 | 1280 |
| 210 | 3 | 3175 | 2363 |
| 210 | 6 | 5281 | 4868 |
| 210 | 8 | 5619 | 6002 |
| 211 | 4 | 1708 | 3756 |
| 212 | 1 | 919 | 2 |
| 213 | 2 | 1107 | 1826 |
| 214 | 2 | 2106 | 1237 |
| 214 | 4 | 3677 | 3132 |
| 217 | 6 | 3548 | 3162 |
| 218 | 1 | 1 | 1218 |
| 218 | 3 | 2731 | 3378 |
| 218 | 5 | 4188 | 4667 |
| 219 | 3 | 1386 | 910 |
| 219 | 4 | 1595 | 1344 |
| 220 | 2 | 794 | 1144 |
| 221 | 1 | 110 | 295 |
| 221 | 2 | 326 | 880 |
| 221 | 4 | 1496 | 1825 |
| 221 | 5 | 1907 | 2200 |
| 221 | 6 | 2169 | 2555 |
| 221 | 8 | 3425 | 4246 |
| 221 | 9 | 4233 | 5111 |
| 221 | 12 | 6419 | 6757 |
| 221 | 13 | 6751 | 6987 |
| 221 | 14 | 6911 | 7120 |
| 221 | 16 | 7400 | 7909 |
| 221 | 17 | 7963 | 8199 |
| 221 | 19 | 8597 | 9079 |
| 222 | 17 | 11376 | 11597 |
| 223 | 6 | 5328 | 5008 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 223 | 12 | 12189 | 13307 |
| 223 | 13 | 13291 | 13716 |
| 223 | 14 | 13601 | 13434 |
| 223 | 17 | 15331 | 15068 |
| 223 | 19 | 15940 | 17160 |
| 223 | 21 | 17710 | 19089 |
| 223 | 23 | 19800 | 20708 |
| 223 | 25 | 22857 | 22027 |
| 223 | 26 | 22757 | 23365 |
| 225 | 1 | 756 | 394 |
| 225 | 5 | 3793 | 2945 |
| 226 | 1 | 141 | 536 |
| 226 | 2 | 521 | 871 |
| 228 | 8 | 5473 | 4835 |
| 229 | 7 | 6749 | 6057 |
| 232 | 2 | 1461 | 910 |
| 233 | 5 | 3359 | 3063 |
| 233 | 11 | 7226 | 7456 |
| 236 | 1 | 3 | 482 |
| 237 | 1 | 1 | 219 |
| 237 | 3 | 1197 | 991 |
| 237 | 5 | 2009 | 2329 |
| 237 | 6 | 2319 | 3056 |
| 237 | 8 | 3261 | 3701 |
| 237 | 10 | 3900 | 4763 |
| 237 | 11 | 4730 | 4963 |
| 238 | 11 | 9966 | 9238 |
| 238 | 19 | 16613 | 17728 |
| 238 | 29 | 26812 | 27663 |
| 239 | 2 | 1576 | 4245 |
| 239 | 5 | 6393 | 6956 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 239 | 6 | 6902 | 7237 |
| 240 | 5 | 1537 | 1809 |
| 241 | 1 | 228 | 1040 |
| 242 | 9 | 6581 | 7015 |
| 242 | 10 | 6988 | 7368 |
| 242 | 12 | 7488 | 7928 |
| 245 | 2 | 1670 | 1251 |
| 247 | 2 | 1558 | 1812 |
| 250 | 4 | 3210 | 2998 |
| 251 | 1 | 622 | 2 |
| 252 | 3 | 2598 | 2383 |
| 252 | 4 | 2911 | 2564 |
| 253 | 1 | 1 | 345 |
| 253 | 2 | 359 | 898 |
| 254 | 1 | 2 | 307 |
| 254 | 3 | 318 | 4 |
| 256 | 5 | 3768 | 4040 |
| 256 | 7 | 7292 | 6639 |
| 256 | 9 | 9589 | 8465 |
| 257 | 2 | 992 | 294 |
| 257 | 4 | 4528 | 3596 |
| 257 | 7 | 6894 | 6718 |
| 257 | 8 | 7252 | 6884 |
| 257 | 9 | 7986 | 7231 |
| 258 | 2 | 544 | 804 |
| 258 | 3 | 1224 | 2921 |
| 258 | 4 | 2964 | 2728 |
| 258 | 5 | 2919 | 3752 |
| 258 | 6 | 4120 | 5298 |
| 261 | 1 | 3 | 362 |
| 264 | 1 | 582 | 361 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 264 | 2 | 881 | 561 |
| 264 | 3 | 1367 | 879 |
| 264 | 4 | 1966 | 1361 |
| 264 | 5 | 2316 | 1945 |
| 264 | 6 | 2636 | 2295 |
| 264 | 7 | 3194 | 2634 |
| 264 | 8 | 3531 | 3055 |
| 265 | 2 | 398 | 817 |
| 265 | 4 | 1583 | 1071 |
| 265 | 6 | 3293 | 3009 |
| 265 | 7 | 3186 | 3046 |
| 266 | 1 | 451 | 2 |
| 266 | 4 | 1983 | 2225 |
| 266 | 7 | 2540 | 2325 |
| 268 | 1 | 798 | 1223 |
| 268 | 2 | 1912 | 1265 |
| 270 | 4 | 3977 | 4186 |
| 270 | 6 | 4397 | 4573 |
| 271 | 5 | 2719 | 3066 |
| 271 | 6 | 3041 | 3352 |
| 271 | 9 | 6278 | 5862 |
| 271 | 10 | 6550 | 5993 |
| 271 | 14 | 10291 | 10004 |
| 272 | 3 | 1870 | 1199 |
| 272 | 4 | 3378 | 1831 |
| 276 | 5 | 2350 | 1994 |
| 276 | 8 | 3702 | 3103 |
| 276 | 9 | 4441 | 3692 |
| 276 | 10 | 4595 | 4416 |
| 276 | 12 | 8173 | 7382 |
| 276 | 14 | 10001 | 9762 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 276 | 15 | 11065 | 9890 |
| 276 | 17 | 11642 | 11250 |
| 276 | 19 | 12892 | 12503 |
| 276 | 21 | 13302 | 13099 |
| 276 | 22 | 13663 | 13271 |
| 276 | 23 | 13995 | 13642 |
| 276 | 25 | 15065 | 14211 |
| 276 | 27 | 16293 | 15955 |
| 276 | 29 | 18482 | 16563 |
| 276 | 31 | 19951 | 19016 |
| 279 | 3 | 1469 | 1675 |
| 279 | 4 | 1600 | 1923 |
| 279 | 5 | 2269 | 2105 |
| 279 | 10 | 7698 | 7279 |
| 280 | 3 | 3138 | 2968 |
| 281 | 4 | 2055 | 2552 |
| 282 | 1 | 316 | 2 |
| 282 | 2 | 456 | 1232 |
| 282 | 3 | 1957 | 1346 |
| 283 | 1 | 1 | 450 |
| 283 | 3 | 1098 | 1556 |
| 283 | 5 | 2062 | 2238 |
| 283 | 7 | 3127 | 3312 |
| 286 | 3 | 2883 | 2698 |
| 287 | 4 | 2359 | 2180 |
| 290 | 10 | 8820 | 9074 |
| 290 | 11 | 9008 | 9172 |
| 291 | 2 | 1103 | 855 |
| 291 | 3 | 2622 | 1123 |
| 292 | 1 | 2 | 283 |
| 292 | 2 | 701 | 330 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 292 | 5 | 2459 | 2866 |
| 292 | 7 | 4252 | 4995 |
| 292 | 9 | 6704 | 7096 |
| 292 | 10 | 7066 | 7827 |
| 292 | 12 | 8377 | 8622 |
| 292 | 15 | 11502 | 12674 |
| 292 | 17 | 13326 | 13727 |
| 292 | 18 | 13738 | 14778 |
| 294 | 1 | 117 | 623 |
| 294 | 2 | 905 | 723 |
| 294 | 6 | 2496 | 2272 |
| 295 | 7 | 4274 | 4510 |
| 300 | 4 | 3525 | 3337 |
| 301 | 6 | 6714 | 4852 |
| 301 | 13 | 10150 | 9914 |
| 301 | 16 | 11316 | 11657 |
| 301 | 18 | 13199 | 14398 |
| 301 | 19 | 15724 | 14657 |
| 306 | 3 | 1135 | 2727 |
| 306 | 4 | 2742 | 4025 |
| 306 | 5 | 4004 | 4552 |
| 306 | 6 | 4527 | 5117 |
| 306 | 7 | 5131 | 5466 |
| 306 | 9 | 5642 | 5968 |
| 306 | 11 | 7000 | 8013 |
| 306 | 12 | 7926 | 8138 |
| 306 | 13 | 8180 | 8908 |
| 306 | 14 | 8899 | 9120 |
| 306 | 15 | 9118 | 9510 |
| 306 | 16 | 9508 | 9963 |
| 306 | 17 | 9964 | 11313 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 306 | 18 | 11319 | 11570 |
| 306 | 19 | 11540 | 11707 |
| 306 | 20 | 11626 | 11856 |
| 310 | 2 | 1126 | 176 |
| 310 | 5 | 4215 | 3556 |
| 311 | 4 | 5671 | 6006 |
| 311 | 5 | 6173 | 6778 |
| 311 | 6 | 6833 | 7225 |
| 311 | 7 | 7236 | 7520 |
| 311 | 8 | 7492 | 7926 |
| 312 | 2 | 859 | 1506 |
| 312 | 3 | 1449 | 1808 |
| 312 | 4 | 2043 | 2306 |
| 313 | 4 | 3568 | 3122 |
| 319 | 1 | 3 | 881 |
| 319 | 2 | 832 | 1185 |
| 321 | 1 | 638 | 898 |
| 321 | 4 | 1862 | 2131 |
| 321 | 5 | 2168 | 2548 |
| 321 | 6 | 2470 | 3159 |
| 321 | 7 | 3069 | 3395 |
| 321 | 8 | 3461 | 3733 |
| 324 | 1 | 3 | 692 |
| 324 | 2 | 867 | 1592 |
| 324 | 4 | 2392 | 3021 |
| 327 | 6 | 5052 | 5213 |
| 330 | 5 | 3745 | 3464 |
| 333 | 2 | 998 | 717 |
| 333 | 3 | 947 | 1534 |
| 335 | 2 | 1024 | 521 |
| 338 | 11 | 8869 | 8591 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 340 | 5 | 3931 | 3608 |
| 341 | 6 | 3484 | 3155 |
| 341 | 7 | 4348 | 3482 |
| 341 | 8 | 6419 | 4332 |
| 341 | 10 | 9264 | 7672 |
| 341 | 11 | 10777 | 9245 |
| 341 | 12 | 12026 | 10779 |
| 343 | 1 | 459 | 262 |
| 343 | 4 | 3905 | 2661 |
| 345 | 4 | 3467 | 3201 |
| 345 | 14 | 15320 | 16447 |
| 345 | 16 | 18409 | 18927 |
| 345 | 18 | 19974 | 20465 |
| 347 | 1 | 763 | 1155 |
| 350 | 5 | 3273 | 2980 |
| 351 | 1 | 693 | 280 |
| 351 | 2 | 1268 | 654 |
| 351 | 3 | 1716 | 1222 |
| 353 | 4 | 2749 | 2546 |
| 354 | 1 | 2 | 298 |
| 355 | 16 | 8911 | 9399 |
| 355 | 19 | 12476 | 12904 |
| 355 | 22 | 15766 | 15608 |
| 355 | 23 | 17165 | 17461 |
| 355 | 25 | 18313 | 19104 |
| 355 | 26 | 19092 | 19598 |
| 355 | 27 | 19692 | 19495 |
| 355 | 28 | 19734 | 20198 |
| 355 | 29 | 20196 | 20471 |
| 356 | 2 | 2204 | 1536 |
| 356 | 4 | 2887 | 2537 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 356 | 5 | 3167 | 2859 |
| 357 | 1 | 381 | 4 |
| 360 | 3 | 3167 | 2877 |
| 361 | 1 | 7 | 909 |
| 363 | 1 | 1405 | 167 |
| 363 | 6 | 7178 | 8404 |
| 364 | 1 | 41 | 331 |
| 366 | 2 | 1386 | 1598 |
| 367 | 19 | 8690 | 8941 |
| 368 | 4 | 1786 | 1947 |
| 369 | 4 | 1652 | 1428 |
| 372 | 6 | 5262 | 4534 |
| 376 | 2 | 625 | 293 |
| 377 | 1 | 331 | 2 |
| 379 | 4 | 2975 | 3142 |
| 382 | 3 | 2951 | 3277 |
| 382 | 4 | 4183 | 3320 |
| 383 | 6 | 6158 | 5637 |
| 386 | 9 | 5725 | 6027 |
| 387 | 2 | 486 | 980 |
| 390 | 2 | 1668 | 2057 |
| 390 | 3 | 3499 | 2867 |
| 391 | 1 | 2 | 154 |
| 392 | 5 | 5163 | 5387 |
| 394 | 1 | 1 | 375 |
| 394 | 8 | 6437 | 7585 |
| 394 | 9 | 7542 | 7967 |
| 394 | 11 | 10354 | 10713 |
| 395 | 5 | 1957 | 2229 |
| 395 | 9 | 3869 | 4216 |
| 395 | 11 | 4571 | 4960 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 398 | 1 | 395 | 1180 |
| 399 | 7 | 5691 | 6134 |
| 399 | 10 | 7662 | 7820 |
| 399 | 14 | 10111 | 9845 |
| 399 | 22 | 16699 | 16481 |
| 399 | 29 | 28519 | 28244 |
| 401 | 1 | 189 | 4 |
| 401 | 2 | 178 | 1044 |
| 401 | 3 | 1038 | 2141 |
| 401 | 5 | 3517 | 3939 |
| 402 | 3 | 919 | 1269 |
| 404 | 1 | 578 | 12 |
| 405 | 1 | 293 | 643 |
| 405 | 3 | 1926 | 1501 |
| 407 | 1 | 80 | 406 |
| 407 | 4 | 3188 | 3670 |
| 408 | 5 | 3037 | 2681 |
| 408 | 6 | 3786 | 3475 |
| 410 | 2 | 811 | 1092 |
| 413 | 2 | 742 | 1314 |
| 413 | 3 | 1275 | 1532 |
| 414 | 2 | 908 | 678 |
| 414 | 3 | 1137 | 1889 |
| 414 | 4 | 2738 | 1959 |
| 416 | 3 | 1945 | 1709 |
| 418 | 1 | 3 | 350 |
| 418 | 2 | 331 | 930 |
| 419 | 2 | 619 | 296 |
| 419 | 4 | 937 | 773 |
| 419 | 5 | 1305 | 910 |
| 419 | 6 | 1183 | 1521 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 419 | 7 | 1859 | 1299 |
| 419 | 8 | 2170 | 1850 |
| 419 | 9 | 2483 | 2160 |
| 419 | 10 | 3399 | 2470 |
| 419 | 11 | 3708 | 3397 |
| 420 | 3 | 1649 | 1452 |
| 421 | 6 | 3983 | 3510 |
| 424 | 1 | 797 | 3 |
| 424 | 2 | 513 | 851 |
| 424 | 3 | 1029 | 733 |
| 424 | 6 | 1859 | 1551 |
| 424 | 7 | 3076 | 2780 |
| 425 | 1 | 52 | 384 |
| 425 | 2 | 1031 | 777 |
| 425 | 3 | 1127 | 1936 |
| 427 | 2 | 1488 | 1114 |
| 427 | 3 | 2114 | 1464 |
| 430 | 2 | 1334 | 1489 |
| 431 | 1 | 420 | 196 |
| 431 | 2 | 634 | 269 |
| 432 | 2 | 1133 | 1372 |
| 432 | 3 | 2014 | 1439 |
| 432 | 6 | 3869 | 3378 |
| 433 | 1 | 292 | 2007 |
| 435 | 1 | 706 | 131 |
| 435 | 2 | 1730 | 1047 |
| 439 | 1 | 1 | 627 |
| 441 | 1 | 1 | 513 |
| 441 | 7 | 10592 | 7974 |
| 443 | 1 | 31 | 744 |
| 447 | 2 | 744 | 322 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 449 | 1 | 3 | 212 |
| 449 | 2 | 471 | 286 |
| 449 | 3 | 551 | 393 |
| 451 | 1 | 823 | 314 |
| 452 | 2 | 322 | 714 |
| 452 | 6 | 2806 | 3342 |
| 452 | 7 | 3358 | 3792 |
| 454 | 1 | 1033 | 2 |
| 455 | 3 | 3214 | 3837 |
| 455 | 5 | 4078 | 4488 |
| 455 | 6 | 4965 | 4117 |
| 455 | 8 | 5123 | 5473 |
| 457 | 1 | 940 | 35 |
| 461 | 2 | 476 | 691 |
| 461 | 4 | 1548 | 1991 |
| 461 | 5 | 2322 | 1948 |
| 461 | 6 | 2664 | 2449 |
| 462 | 5 | 2810 | 2064 |
| 464 | 2 | 2162 | 1530 |
| 465 | 1 | 1762 | 38 |
| 465 | 3 | 2373 | 2050 |
| 467 | 2 | 652 | 1260 |
| 467 | 3 | 1149 | 1442 |
| 469 | 2 | 922 | 1101 |
| 470 | 2 | 971 | 1768 |
| 473 | 2 | 450 | 220 |
| 475 | 1 | 1 | 969 |
| 477 | 2 | 1064 | 843 |
| 482 | 1 | 1 | 534 |
| 484 | 1 | 130 | 543 |
| 484 | 2 | 1320 | 1159 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 487 | 2 | 1258 | 1929 |
| 488 | 2 | 509 | 162 |
| 488 | 4 | 2247 | 1945 |
| 489 | 1 | 1 | 396 |
| 489 | 2 | 560 | 255 |
| 490 | 2 | 1096 | 458 |
| 491 | 5 | 5167 | 4433 |
| 491 | 6 | 5975 | 5247 |
| 491 | 7 | 6811 | 6041 |
| 494 | 1 | 650 | 3 |
| 497 | 5 | 3351 | 3536 |
| 497 | 8 | 4757 | 4308 |
| 497 | 10 | 5229 | 5086 |
| 497 | 11 | 5967 | 5671 |
| 499 | 1 | 663 | 247 |
| 502 | 2 | 1324 | 851 |
| 504 | 1 | 3 | 650 |
| 507 | 2 | 727 | 906 |
| 507 | 3 | 840 | 1010 |
| 510 | 3 | 2056 | 2574 |
| 512 | 2 | 854 | 300 |
| 514 | 2 | 1067 | 669 |
| 518 | 5 | 3119 | 2970 |
| 520 | 1 | 3 | 467 |
| 520 | 2 | 452 | 231 |
| 520 | 4 | 2218 | 1859 |
| 521 | 2 | 988 | 821 |
| 522 | 1 | 409 | 885 |
| 524 | 1 | 579 | 4 |
| 525 | 1 | 1 | 144 |
| 525 | 2 | 86 | 352 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 529 | 2 | 5731 | 6147 |
| 533 | 1 | 1044 | 157 |
| 536 | 3 | 587 | 1462 |
| 539 | 7 | 6180 | 6662 |
| 540 | 1 | 198 | 476 |
| 543 | 3 | 2179 | 1835 |
| 543 | 4 | 2404 | 2177 |
| 543 | 7 | 3924 | 3700 |
| 544 | 2 | 1004 | 870 |
| 546 | 2 | 497 | 324 |
| 547 | 3 | 717 | 965 |
| 549 | 2 | 371 | 135 |
| 550 | 1 | 527 | 3 |
| 550 | 2 | 864 | 709 |
| 550 | 3 | 1540 | 1277 |
| 550 | 4 | 2039 | 1509 |
| 552 | 5 | 4681 | 5073 |
| 552 | 8 | 8390 | 8223 |
| 555 | 1 | 470 | 267 |
| 560 | 1 | 635 | 210 |
| 560 | 2 | 834 | 514 |
| 563 | 2 | 1215 | 1469 |
| 564 | 1 | 8 | 511 |
| 564 | 2 | 1019 | 555 |
| 564 | 3 | 577 | 744 |
| 565 | 1 | 321 | 4 |
| 565 | 5 | 1266 | 1619 |
| 567 | 2 | 1055 | 531 |
| 571 | 3 | 1149 | 886 |
| 573 | 1 | 208 | 666 |
| 573 | 2 | 651 | 1148 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 573 | 5 | 2558 | 2809 |
| 575 | 1 | 262 | 2 |
| 584 | 1 | 268 | 110 |
| 584 | 4 | 1310 | 795 |
| 584 | 5 | 1329 | 1574 |
| 586 | 1 | 771 | 4 |
| 588 | 1 | 346 | 56 |
| 588 | 2 | 1078 | 434 |
| 589 | 1 | 1 | 555 |
| 591 | 1 | 217 | 2 |
| 592 | 2 | 674 | 868 |
| 593 | 1 | 190 | 2 |
| 593 | 3 | 1035 | 1268 |
| 601 | 1 | 77 | 274 |
| 601 | 2 | 172 | 576 |
| 602 | 2 | 759 | 415 |
| 604 | 6 | 2868 | 2416 |
| 606 | 1 | 271 | 798 |
| 607 | 2 | 633 | 797 |
| 613 | 1 | 420 | 82 |
| 616 | 2 | 593 | 435 |
| 616 | 4 | 975 | 730 |
| 619 | 3 | 641 | 817 |
| 620 | 1 | 863 | 3 |
| 621 | 2 | 1493 | 2014 |
| 627 | 1 | 113 | 763 |
| 628 | 1 | 2 | 163 |
| 631 | 1 | 1 | 516 |
| 631 | 3 | 1715 | 1521 |
| 633 | 1 | 280 | 2 |
| 634 | 3 | 1139 | 1387 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 637 | 2 | 1613 | 738 |
| 637 | 3 | 1597 | 2208 |
| 637 | 4 | 2242 | 2694 |
| 637 | 7 | 3550 | 4545 |
| 637 | 9 | 4767 | 5171 |
| 639 | 1 | 175 | 2 |
| 640 | 2 | 468 | 689 |
| 643 | 1 | 496 | 320 |
| 645 | 1 | 1 | 537 |
| 645 | 2 | 539 | 1024 |
| 647 | 1 | 64 | 855 |
| 647 | 2 | 1419 | 895 |
| 649 | 1 | 2 | 364 |
| 651 | 1 | 539 | 3 |
| 653 | 2 | 738 | 550 |
| 656 | 8 | 7784 | 8587 |
| 657 | 2 | 1356 | 967 |
| 657 | 3 | 1708 | 1376 |
| 661 | 1 | 2 | 244 |
| 664 | 3 | 1149 | 820 |
| 672 | 1 | 546 | 10 |
| 673 | 2 | 1207 | 1827 |
| 676 | 1 | 443 | 790 |
| 679 | 1 | 998 | 219 |
| 682 | 3 | 749 | 1171 |
| 685 | 1 | 176 | 511 |
| 685 | 2 | 498 | 199 |
| 685 | 3 | 480 | 947 |
| 685 | 4 | 1000 | 1443 |
| 686 | 4 | 1567 | 2001 |
| 686 | 5 | 3238 | 1712 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 686 | 7 | 2965 | 3435 |
| 686 | 8 | 3441 | 3067 |
| 686 | 9 | 3752 | 3339 |
| 686 | 10 | 3530 | 3826 |
| 688 | 2 | 628 | 894 |
| 689 | 2 | 582 | 331 |
| 690 | 1 | 275 | 90 |
| 690 | 2 | 487 | 248 |
| 696 | 1 | 239 | 9 |
| 696 | 2 | 1237 | 233 |
| 696 | 3 | 1424 | 1200 |
| 697 | 1 | 20 | 520 |
| 698 | 1 | 29 | 313 |
| 698 | 2 | 217 | 483 |
| 701 | 5 | 1061 | 1534 |
| 707 | 2 | 855 | 538 |
| 709 | 1 | 1 | 675 |
| 710 | 1 | 3 | 416 |
| 712 | 1 | 674 | 96 |
| 713 | 1 | 933 | 139 |
| 713 | 2 | 1125 | 1436 |
| 716 | 2 | 1226 | 765 |
| 721 | 1 | 3 | 371 |
| 726 | 1 | 543 | 94 |
| 729 | 1 | 19 | 210 |
| 731 | 1 | 532 | 2 |
| 736 | 2 | 309 | 644 |
| 738 | 1 | 561 | 4 |
| 740 | 1 | 488 | 3 |
| 749 | 2 | 20 | 475 |
| 751 | 1 | 1 | 456 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 751 | 2 | 454 | 774 |
| 753 | 1 | 76 | 729 |
| 754 | 1 | 761 | 21 |
| 755 | 2 | 345 | 539 |
| 756 | 1 | 1 | 375 |
| 764 | 2 | 528 | 1088 |
| 772 | 1 | 1 | 558 |
| 772 | 2 | 432 | 866 |
| 775 | 1 | 706 | 2 |
| 778 | 2 | 992 | 834 |
| 780 | 1 | 52 | 351 |
| 782 | 1 | 3 | 557 |
| 783 | 1 | 28 | 609 |
| 791 | 1 | 1 | 582 |
| 791 | 2 | 859 | 641 |
| 791 | 3 | 1235 | 711 |
| 797 | 1 | 2 | 289 |
| 797 | 2 | 287 | 3 |
| 801 | 2 | 598 | 191 |
| 805 | 1 | 1 | 414 |
| 806 | 1 | 392 | 3 |
| 810 | 1 | 3 | 317 |
| 810 | 2 | 407 | 3 |
| 815 | 2 | 443 | 282 |
| 819 | 1 | 39 | 668 |
| 830 | 1 | 291 | 4 |
| 830 | 2 | 476 | 162 |
| 834 | 1 | 561 | 46 |
| 834 | 2 | 953 | 453 |
| 837 | 1 | 3 | 317 |
| 837 | 2 | 320 | 589 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 839 | 1 | 1 | 753 |
| 841 | 1 | 1 | 489 |
| 855 | 1 | 308 | 3 |
| 861 | 1 | 1 | 330 |
| 863 | 1 | 451 | 221 |
| 870 | 1 | 21 | 503 |
| 890 | 2 | 1548 | 1255 |
| 895 | 1 | 3 | 140 |
| 896 | 1 | 2 | 400 |
| 897 | 2 | 244 | 498 |
| 902 | 1 | 1 | 300 |
| 904 | 1 | 294 | 4 |
| 910 | 1 | 143 | 3 |
| 917 | 1 | 36 | 518 |
| 918 | 1 | 3 | 167 |
| 918 | 2 | 116 | 373 |
| 920 | 2 | 243 | 515 |
| 922 | 1 | 669 | 259 |
| 926 | 1 | 2 | 394 |
| 927 | 1 | 119 | 556 |
| 928 | 1 | 493 | 179 |
| 930 | 1 | 526 | 344 |
| 933 | 2 | 257 | 418 |
| 936 | 2 | 243 | 683 |
| 937 | 1 | 341 | 3 |
| 942 | 1 | 58 | 228 |
| 945 | 1 | 318 | 4 |
| 953 | 1 | 254 | 48 |
| 959 | 1 | 1198 | 164 |
| 959 | 2 | 1740 | 1123 |
| 963 | 2 | 462 | 232 |

Table 3: *E. faecalis* - Putative coding regions of novel proteins not similar to known proteins

| Contig ID | ORF ID | Start (nt) | Stop (nt) |
|---|---|---|---|
| 965 | 1 | 403 | 2 |
| 969 | 1 | 360 | 4 |
| 970 | 3 | 673 | 314 |
| 972 | 1 | 3 | 470 |
| 973 | 1 | 2 | 700 |
| 974 | 1 | 2 | 235 |
| 974 | 3 | 270 | 467 |
| 981 | 2 | 154 | 405 |
| 984 | 3 | 164 | 337 |

# SEQUENCE LISTING PLACE INDICATOR

PAGES 280 TO 2076, WHICH ARE THE COMPLETE SEQUENCE LISTINGS FOR THIS APPLICATION ARE LOCATED AFTER THE DESCRIPTION, CLAIMS, ABSTRACT & DRAWINGS.

(1) GENERAL INFORMATION:

(i) APPLICANT: Charles Kunsch
Patrick J. Dillon
Steven C. Barash

(ii) TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

(iii) NUMBER OF SEQUENCES: 982

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.

(B) STREET: 9410 Key West Avenue

(C) CITY: Rockville

(D) STATE: Maryland

(E) COUNTRY: USA

(F) ZIP: 20850

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

(B) COMPUTER: HP Vectra 486/33

(C) OPERATING SYSTEM: MSDOS version 6.2

(D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE: herewith

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: A. Anders Brookes

(B) REGISTRATION NUMBER: 36,373

(C) REFERENCE/DOCKET NUMBER: PB369PCT

(vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (301) 309-8504

(B) TELEFAX: (301) 309-8512

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4315 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

AAGTCAATCA CTTGCAAGTC GTTTTCTGTC ATATGGCCGA CTAGCTCAGT GACACCAGGA 60

ATATAGCCAC CAGGGAAAAT ATAACGATTA ATCCAAGCAT TTTTAGCCCC ACCTTGTTGG 120

CGACTGATCC CATGAATCAA CGCCGTACCT TTAGGCGCTA AATTCCGCTG AACGACATCA 180

AAATATTCAT GTAGATTTTC CGCACCGACA TGTTCAAACA TCCCAACACT CGTAATATGG 240

TCAAAAGACT CTCCTTTTAA ATCACGATAA TCCATCAATT TGACAGTCAT TCGATCTTGT 300

AGATCTTCTT TTTCTATAAT ATGGCGAATA TGATGAAATT GCTCTTCACT TAATGTAATC 360

CCAGTTGCTT TGGCTCCATA TTCTTTCACC GCAGTTAAAA TTAACGTGCC CCAGCCGCAG 420

CCAATATCCA GTAAAGTGTC GCCCTCTTTG ATAAACAATT TATCTAAAAT ATGATGAACT 480

TTATTCACTT GCGCTTGTTC TAATGTATCT TCAGGCGTTT TAAAATAAGC ACATGAATAC 540

GTCATTGTTT GGTCAAGCCA TTTTTTGTAA AAATCATTTC CTAGATCGTA ATGGCTGTGA 600

ATATCCTCTT GCGAACGTTT TTTTGAATGA CTTTCTTTAG GAAGCCATTT AATAAATTTA 660

GCATTGTGTA AAAAGCTATC CTTTTGGTTA TACACATCAT AAATCAGsGC TTGGATATCG 720

CCTTCGATTT CAATTTTGCG ATCCATGTAG GCTTCCCCTA AAGTTAACGA AGCGTTATTC 780

AGTAAATCCT TCACAGGAAT TTTTTCATTG AATACAATTT TAAAAACCGG ATCCCCCGAC 840

CCTTGCCCAT ACTCTTTGAC GGTACCATCC CAGTATGTGA CTTGTGTCTT TTTTGAAAAA 900

GACCATTTAA ACAGTTGACT GTACGTTTCT TTTTCTAACA TTGCATTCCC TCCATTAAAT 960

ACCATTTGAA GCCAAAACAA AAAGAAGTCG CTTTCCGGTA GTTCGTCAAA ACAAACACCA 1020

CAGTCCGTTC TAAACTGAAG CACAGAAAAG TTATCACCCC TTCTATGTTC CGCTTCTTTT 1080

TTTGCAATTA CAGTTCTATT CTACTCCTCT TTTAAAAATT TGAACATTCT TTTAACGTAA 1140

TACCTACTAT TGTTATTCTT TATCACAAAA AAACTAGAGC CAGTCCTTGA CAGACTCCTC 1200

TAGTTCTAAA TATTATGCTT TCTTACGCAT CCGTTGTTCC GCATGAGTGT AAGCGCCATG 1260

```
CCATACGTGC CCCACATAAG GATTAACTTG AATACCGTGT TTAATCGCCG CTGCTACAAA    1320
TTTTTTTGCT AAAGTTACTG CTTCTAACAC CGAATAACCT TTCGCCAAGC CAGCTGTGAT    1380
TGCCGCTGAA AAAGTACAAC CTGCACCATG ATTATAATCA GTAGGATATA ATTCATTTTC    1440
CAAAAGATGC GCGGTGTGAC CATCGTAAAA CAAGTCCAGT GCTTTTTCAC CAGCTAAGCG    1500
ATGTCCCCCT TTAACCACGA CATGCTTGGC TCCCATTTGT ACAATTCGTT TTGCCGCTTC    1560
TTCCATCTCC GCCACGGAAG AAATTTCGCC TAAACCAGAT AAGATGCCCG CTTCAATTAA    1620
ATTAGGCGTG GCCACTAATG CTAATGGCAG TAAATCGTTT TTTAGGCCTT CCACACTTTT    1680
GGGTTGCAGA ATTTGTGCCG TTCCCTTACA AGCAATGACT GGGTCAATCA CGACTTTTTG    1740
AATTTTCTCT TGTTTAATGT ACTTACTAGC CATTTTAATA TTTTGTTCAT TACCCATCAT    1800
CCCTGTTTTC AAAGCCGCTA CTGGACCGCC TGCAAAAACC GAAATCAATT GTTTTTCTAA    1860
GAGCGTTTCT GGCAATTCAG TTACTTCATG TGACCAACCT GTCGTAGGAT CCATCGTCAC    1920
AATCGAGGTT AAACTTGAAA ATCCAAAAAC TCCATACTCT TCAAATGTTT TTAAATCTGC    1980
TTGAATCCCT GCCCCTCCAG TTGAATCGGA GCCTGCAATC GTCAATACTT TTTCCATTAA    2040
ATCACCTAAC CTTTTTCTCC AAGTATACTG AAGAAACAAG TCTGCTAAAA CAGCCAATTG    2100
GCTTATTTTT TAGCCAGCCA ATTTCTAAAC AAAAAAAAGA CCAGAGAATA AATTCTCTGG    2160
TCTTACGTCC GAATACCCCA GTTTTTTACG CTGGTTAAAA CTATAGTTAA AAAGTTAATT    2220
ATTTAACGAT TTCAGTAACA ACGCCTGAAC CTACAGTACG TCCGCCTTCA CGAATAGAGA    2280
AACGAGTTCC GTCTTCGATA GCGATTGGGT GAATTAATTC AACGTCCATA GCAACGTTAT    2340
CACCAGGCAT TACCATTTCA GTACCTTCTG GCAATTCTAC AACACCAGTA ACGTCTGTTG    2400
TACGGAAGTA GAATTGAGGA CGGTAGTTAG TGAAGAATGG AGTGTGACGT CCGCCTTCTT    2460
CTTTTGATAA TACGTATACT TCAGCTTkGA ATTTgGTGTG TGGAGTGATT GTAGCTGGTT    2520
TAGCTAATAC TyGyCCACGT TCGATATCTT CACGTGCTAC ACCACGTAtA AAGCACCGAT    2580
GTTGTCGCCT GCTTCAGCGT AGTCTAATAA TTTACGGAAC ATTTCAACAC CTGTAACAGT    2640
TGTTTTAGAT GTTTCGTCTT TAATACCAAC GATTTCAACT TCGTCACCAA CGCGAACTTC    2700
ACCACGTTCA ACACGGCCTG TAGCAACAGT ACCACGTCCA GTGATTGAGA ATACgTCTTC    2760
GACTGGCATC ATGAATGGTT TGTCAGTATC ACGTTCTGGA GTTGGGATAT ATTCGTCAAC    2820
TGCAGCCATT AATTCTAAGA TTTTTTCTTC ATAAGACTCG TCGCCTTCTA AAGCTTTCAA    2880
AGCAGAACCT GCGATAACTG GAACATCATC GCCTGGGAAA TCGTATTCTG ATAATAAGTC    2940
ACGAACTTCC ATTTCTACTA ATTCTAATAA TTCTTCGTCA TCAACCATAT CCATTTTGTT    3000
TAAGAATACA ACGATGTATG GTACACCAAC GTTACGTGAT AATAAGATAT GTTCACGTGT    3060
TTGAGGCATA GGACCATCAG CAGCAGAAAC TACTAAGATA GCTCCGTCCA TTTGAGCAGC    3120
ACCAGTGATC ATGTTTTTAA CGTAGTCCGC ATGTCCTGGG CAGTCaACGT GTGCaTAGTG    3180
ACGAGTTwCA GTTTCATATT CGATATGAGA AGTGTTGATT GTGATTCCAC GTTCTTTTTC    3240
```

TTCTGGAGCG TTATCGATAG AATCGTAGCT TTGTGCTTCC CCGCCACCGT GTTTTGATAA 3300

TACAGTAGCA ATTGCAGCTG TTAATGTAGT TTTACCATGG TCAACGTGTC CGATAGTACC 3360

AATGTTAACA TGGGATTTAG AACGGTCAAA TTTTTCTTTT GCCATTTTGA ATGTTCCTCC 3420

TAAAAATATA AATTTTATTG TTTGTTTTAT ATGTGATAGA CGTTACTAAA AATCGTCTGA 3480

TCATCATAAG CTATTTTACA CGAAATCTTG AAAAAAGGAT AGTTCTAAAA ACAAACCTTT 3540

TATTCTCAAG GAAAATTATG CGTtGCCGCC GTTTTTCTTG ATGATTTCTT CTTGAACAGA 3600

CTTAGGTACG TCTTCATAGT GGTCAAATAC CATCATAAAT GTACCACGTC CTTGTGTTGC 3660

TGAACGTAAT GTTGTAGCAT AACCAAACAT TTCAGCTAAA GGTACCATCG CATTAACGAT 3720

TTGAGAGTTA CCGTGTGCTT CCATACCTTC AACACGTCCA CGACGACTTG TTACGTGTCC 3780

CATGATATCA CCTAAGTAAT CTTCAGGAAC AGTAATTGTA ACTTTCATCA TTGGTTCTAA 3840

AATAACTGGG TTTGCATTTT TAGCCGCAGC TTTCAATGCC ATAGAAGCAG CTACACGGAA 3900

GGCCGTTTCA TTTGAATCGA CATCGTGGTA TGAACCATCG TAAAGTTTTG CTTTGATATC 3960

AACTAATGGA TAACCAGCAA GAACACCGTT GTTCATAGAG TCTTCTAACC CTTTTTCAAC 4020

TGCTGGGaTG TATTCACGAG GAACCACACC ACCGACAATT GCGTTTTCAA ACTCGAAGCC 4080

TTTACCTTCT TCATTTGGTG TAAATTCAAC CCaTAcGTGA CCGTATTGTC CTTTACCACC 4140

AGACTGACGA ACGAATTTAC CTTCTGCTTT AGTCGCAGCA CGGAATGTTT CGCGGTAAGA 4200

TACTTGAGGA GCACCTACGT TAGCTTCAAC TTTGAATTCA CGTTTCATAC GGTCTACTAA 4260

AACGTCTAAG TGCAATTCAC CCATaCCAGA GATAACTGTT TCACCAGTTT CAACC 4315

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2716 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

ATTGTGCCAT AGATTTGACC ATCTGCTGCT GTTGTAGTTT CCCAGTTTGT AATTTGGTAA 60

TGTTGTAAGA nCTTAGTTGG GACATTAATG GCAAAAGTTT TTTGGCGATA GGTATTAAAA 120

TTGGTTTTAC CCACCCAACC TTCATAGGCT TCAAATTCAG CTCCACCTAA CAAAGGGTTG 180

TCGGAAATAA TCCAAGTGTC TTTTTGATGA TTCCGAATAA AACGAACGAT TGTACTGTTA 240

AAGCCTTTCG TTGTTGTACC GCCGTAATCT TTTTCTGCAG ACCAATGGTT CCAAACCGAG 300

TCATACTCTA AAGCATTCGG AAATTCATTG GTTGTAAACC AACCGAGTGA ATTAATTTCT 360

TTACTTAGTT GACGACTTGC CCAACCTGAT TCGCCTTGGT TGCCCCAAAC ATCAACATAA 420

ATATAATCTA GATTCGGTGc TTTTtGTTTT AATcnTTAAA GCGCTCATAG CGACGACCAC 480

TCAATGTAtC GGGtCTTkGT TTAATAAAAT AAGACGGATC TAACCAATCC CAGCCACGCT 540

TCGTTGGATC AACTAATTCC TCATTAAATC CTTTTGCTTC TGGGTAAGAC TCGGTGTCAT 600

TAATATGCAC ACCGAAAACG GCATTTAATT TATGTCCTTC ATCAATTAAT TGATTCAACG 660

CTTGTTCCCC ACCAGGTCGC TGACCAATAG CACCGTAATC TGGATGCGCA GAATCATGTC 720

CTTCATTTTG ATACCCTTTT AGTAAATTCA TTTGTCCTAA TCCATCTGTT AAATTGTAAA 780

TACGTTTTGA TTCGTCTAAC GTCACTAAGA ATGGGTTTGT CGCCTGACTA GCAAAGTTAA 840

AAGGAATCCG TTGGTTGACT AATTCAGGGA CTTTTTCCGC ACCCATTGGG TTATTCATAA 900

TTGAACGGAA CCCAATTGCC GCATCTTGCC AATCCACCCG ATTGTCGTCG TTGCTATCTT 960

TTGAGAATTT AACTTTCACT TCTGGCGTTT CTCCAGTTGT GTAATCTTCC GGTGCATCAA 1020

ATGGTCGATA GGTCCATGCC CCACTTGACA AGGTTGTTAC AAAACCAGTC GCCGCTTCTT 1080

TTGTTTGTTT ATGGATTCGA TTGTTGTCAC TACCATCGAC AGAGCCGTCT CCATAGGCGT 1140

TCGTCCAAAA ACTTGCAGCA TAATTCGCCG TATTTAAGAA ACCATACATA TATTTTTTCG 1200

TTTCTTCTTG GCTTGTTGTA GTTAAATCTT GGATAGTATC ACCATTTTTA TTTCCGTTAT 1260

TTGCATTTGT CCCAGTATTC ATCACAACGC CAGCAAAAGT CGCCCCTTCA TCTGTACTAT 1320

TGACAGAAAT CAAGCCTTGA TTTGGAATTG AAATTGTGTG TATTTTGGTT CCTTCTTCAC 1380

GAATATCTGA GAGTTTCATA GCTAATTCTT GGCCTTCGGA AACAGTCAAT GTCGTTGTGA 1440

AACGCAAACC AATTTCTGGG ACAGCTACGT TATAAACGGC TTGaTTTGCT TCaCTCTTCT 1500

GTTTTTcAGC AGTTGCTTTA tATGCTTTGC CAyCAAyCAT TACTTGGTCa ATTTTCTCgT 1560

GTTGGCCATT AAAAATCACT TGATTACTCT GAGGGTCTTC GTAGCGAATA ATTCGTGGAA 1620

ATTGTGGATC AACCACGGCT GTTAAATCTG CGGAAGAGAT TTTGTCGCCT TTTTCAGCTT 1680

CATTGTCTTT AGCAACAACA ATGCTGGCTT TTGCTTTGAT GTTTGTATTT TCCAAAGTAC 1740

CGTCGACTTC AAAAGTTCCT GGTTCTTTAT ATGCATCAGG ATCGATTTCA TTCCAACGGA 1800

CTGCTTCATT GGCTTCTTTA CCAGTATTAT AGGTCACTTT TACTTGTTTC GGtAATTTTG 1860

GTGCGACACC AATTTTTGTA AAAACTTTGA CTGGCGCAAT TTGTGTCACT TCTGGCACAA 1920

TTTCTGGAAT GCTATCTACG GGGCCATTTT TAAAATAGTC ATAATTGACA ATTTTTTTGT 1980

CGTACCACAA ACGGACACCC ACATGTCCGG CTTCTGTTGG AATTTTGTCT CCATTGGCCA 2040

AAACAGGTTC TCCTTCGTAA ATCAACGTGG TGTTAAGCCA GATAGTAATT TTTTCACCAA 2100

CATAACGTAC TTTCAACAAA TAATTCGTAT CTTCATTTAA CGTTGGCCCA GAAATTGAAT 2160

CATTCCACGA ATTAGGGCTT TCAACTAACC ATTTGCCATT CGCGTTATAG CCTACAAAAA 2220

CCCAGCTATC TTTGGTATTA CCACGTATAA TCACACCCGT TCGACCGCTT CCAGCAGTGT 2280

ATTTAAACTT TGTTTCTACT TCACCATCAG CTAATTTAGG GGAATCTAAA TTTAAAGATA 2340

CGGCATTATT ACCTGCTGCT GCATCTCGTG AAATCGCCAA TGACTCTCCT GCTAATTCTC 2400

GATTTGTTTT ACCAATCACA TCTTGCCACT CACCAGTTTT CCCTCCTGGA AAATCAGATT 2460

GCCACAATTT TTCTTCTGTT GCGGTAGGTT CCACCATTGA TGAGCTTGTT TCAGTGCTGC 2520

TGTTTGTTGT TTCTTCTGCA GACGTACTTA ATGGTACTAA AATCGTTGCG CTTGCCAATA 2580

GTGTCAATGT ACTAAATCGT TTTATTTTTC CATGTTTCAT CTGaTAAATC AGCCCCCTGA 2640

ACATTTTTGG AACACGCTTT CATTTATTTA ACAATTAAAG GGGGGTTnCC AAAAATTAAA 2700

TTGTnTTCAT TTTTGG 2716

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1311 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TTGAAATTGC TAGCGTTTTC CGCCATTGGC TATGTTTAGT CGCAACGGAG GGACCsCtGA 60

CTTTTTAGCG AAGCATcgTT GCCTGTTAAG aAAGTGACGA AAATTTCTAA TCCTGAAGGG 120

GAAACCGTGC TGGATGTCAT TCGCAATGGG AATGCGCAAG TGGTAATCaG TACCATGGAT 180

AAAAATCGTT CCAGTGcCAA CCAAGATGGA TTTTCAATCC GTCGGGAAGC AGTCGAACAT 240

GGGATTCCAT TGTTTACCTC ACTTGACACA GCAAATGCGA TTTTAAAAGT TTTAGAATCA 300

AGAGCCTTTA CGACAGAAGC CATTTAAGCA AAAATCTGGA ACAATCGTCT GAGGAGGACT 360

GTTCCAGATT CTTTTACGAA GAAATAGAAA ATCATTTATA GGGGAATCAG AGAATAAAGG 420

AGTGAAGGAT ACTGATGCAG AGAAAGCAAG AAATGATGAC CATTGTAGCT CAAAAACAGT 480

TAGCGCCACG GATTTATCAG TTAGACTTAC AAGGCGAGTT AGTGAAAGAG ATGACACGAC 540

CTGGACAGTT TGTGCATATC AAGGTTCCGC GTGCAGATTT ACTTTTGCGG CGACCGATTA 600

GCATCAACCA AATTGATCAT TCGAACGAAA CCTGTCGTTT GATTTATCGA GTCGAAGGCG 660

CTGGAACGGA AGTGTTTGCC ACCATGAAGG CTGGTGAACA ATTAGATATT TTAGGTCCTT 720

TAGGAAACGG CTTTGATATA ACTACCGTAG CAGCTGGTCA AACTGCTTTT ATCGTTGGCG 780

GAGGAATCGG GATTCCCCCA TTGTATGAAT TATCAAAACA ACTCAATGAA AAAGGCGTGA 840

AAGTGATTCA TTTTCTTGGT TATGCATCAA AAGAGGTTGC TTATTACCAA CAAGAATTTA 900

TGGCATTAGG AGAAACACAT TTTGCCACAG ATGACGGCTC GTTTGGCGCT CATGGCAACG 960

TGGGCCGTTT ATTGTCAGAA GCGTTAGCAA AAGGACGGAT CCCTGATGCA GTGTATGCTT 1020

GTGGTGCGAA TGGTATGTTA AAAGCAATTG ATTCTTTATT TCCAACACAT CCACACGTTT 1080

ACCTTTCTTT AGAAGAACGG ATGGCCTGTG GAATTGGGGC TTGCTATGCC TGCGTTTGCC 1140

ATAAAAAAGG AGACACTACT GGAGCAAAAA GTGTCAAGGT CTGTGATGAA GGTCCGATTT 1200

TTAAAGCAAG TGAGGTTATC TTATGATGAA AAAyCCTTTA GCCGTTTCAA TTCCAGGTTT 1260

AACATTAAAA AATCCAATTA TTCCCGCCAG TGGTTGCTTT GGGTTCGGGG A 1311

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6948 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TCCTCTGATG AACCGATTGA AAGAATTAAG GAAAGAAnAA AATATTACTT TAGTTGAGTT 60

AAGTGAGGAG TTAGGTATTC CACGTTCCAC TCTTAATAGG TATGAAAACG AAGATAGCGA 120

ACCAAAACAA GAAaCTTGGG AAAAATTAGC TGATTATTAT GGTGTTTCTA CGGCTTATTT 180

AATGGGGATA TCCAACCAAA AGGTTAGCGA AGAAAAAGCT TTGACGGCCG CAAAGAAAGT 240

TTATCAAGTC TATCTTTCCG ACGACGATTT AGGAAAAGAA ATTCGAAAAG CTCTAATGTA 300

TTTTAATAAA AATGATTTAG ATAGTGTCTT AAAACAAGCA ATGCAGCAGT ATTTTACTAT 360

CCCAGCCGTT GAATGGAATA CAGAATTTCA ATCTTTAAAA AATACTGGTT TTCTATACAG 420

TTGGTTAGAA GGATATTTGG TAGACAGGTA TCGAAAAGAA GTGAAAACTA ATCATAATTT 480

AATTACAAAT ACGTATTATA GTATCCCTTC TATTGATGAT GTGAATGAGT ATGGAATTTT 540

CTCTGATGAA ATTCAAATAC CGTTGACTGA AGATTTTAAA CAAATTTTCA AAGTTGAGGA 600

TCTATCCACG AAAACACATG AAGAAATAAA AAAACTTTTT AAATCCGATT CTACGCTCAT 660

TGcTTACAGC TATGAAAGCT CCGTAGACGA TAAATTAAAA GATGATATTA ATATGATTTT 720

AGATAATGCT CGTTCCGAAA TTCTTAATCT AAAAGAAAAA TATCCTGATA AACCTAGTAG 780

AATAGaACAA ACAACCGTAC TCATATCTAA AGACGAAGAT ACTTCCTTTT GGTCAAGGAA 840

TGGCTCAAAT GATTATATAG ATGAATTGAA CTTATCTGAG GAAACTAAGA AATTATTTGT 900

TCATTTAGGT TCAGTATTTA TTGAGGAAAA AAAACGTAAA TCAAAAGCTA AAGAATGAGT 960

TTTTTTCTTT GTATAATCCA GTTATAGAAA CATCTCTCCC CTAACATCTA CATACTATTG 1020

GCGTGGTGGT AGATAGGAGA AAAACAATGA CACAAATTAA CAAGTATACA AAAAAAGACG 1080

GTACAACCGC CTATATGTTT AATGCTTATG TAGGACGTCA TCCCAAAACT GGTAAAAACG 1140

TTTATCGTAA GCGGCAAGGA TTTAAAACAA AAAAACAAGC ACAAATTGCA CTTGCTGAAA 1200

TTCTGAAGGA TATTGAAGAA AATGGACTAG ATAATAAACC CTCTGTTTTG ACGTTTAAAC 1260

AGCTTTACGA GAAATGGTTA GCACAGCAAC GGTTAACGAT TAAACCATCT TCAATCGCTG 1320

TGAATAAGCG TTTTGCTGAA AAACACATAT TACCTTATCT AGGTGACTGC AAACTTGACG 1380

AAATAACCGT TATACAGTGT CAGGACTTAG TTAATAAATG GTTTAATCAA GGACACAAGC 1440

AATACTCGTT CTATAGAAAA CTCACTGCTC AGATTATGCG ATATGGCGAA TCTATGGAAC 1500

TAATGAATAC TAACCCAATG AGAAAAACAA TCCTTCCTAA GTGGAAAGAG GAAGAAACAA 1560

AACTGGAGTA TTACACTAAA CAAGAATTGA ATCATTTTTT TGATTGTCTA AAACAATATG 1620

GTAATTTCAA ACAGTTAGCA TTTTTTAGGT TGCTAGCCTT TACAGGTTGT CGTAAGAGTG 1680

AAGTTCTAGC CTTGCAGTGG AAAGACCTAG ACATAACAAA CAAGAGCGTA TATATAGGCA 1740

AGACCGTAGC ACAAGACGAA TTCTATAATA TACTCACTCA AACGCCCAAA ACAGCTTCTA 1800

GCACACGTTC TATCAGCTTA GACAATGAAA CACTTCAAAT TATGGCTAAA TGGCGTACAA 1860

TCCAACGTAG TGAcTATTTC CAAATGGGCT TTAATACATC TAGTGAAGAA CAATATATTT 1920

TCACGAACAA CCATAATAAA CTGCTCTCAC CAAATGTAGT TAATATATGG CTCAATTGCT 1980

TAATAAAAAA ATATGATTTA CCTAGTATTA CACCACATCA TTTTAGGCAT ACACATGCAA 2040

GTTTACTTCT TCAATCAGGA GTGCCTATTA AAGAAGTTGC TGAAAGATTA GGACATACTA 2100

GCACTGCTAT TACTGATCGT ATCTATTCTC ATGTTATGCC TGAAGAAAAA GAAAAAACAG 2160

CAGATAAGTT CGCTCAATTC GTCGGTTTTT AATAAATGCA CAGACAAAAG CACAGATCTT 2220

TACTATTTTT ATGTAAAAAA AAGAGCTAGA ATCCTTATAA ATAAAGGTTT CTAACTGGCA 2280

TTAGTAACGC TATTAATGAA ACAACGTGTT TAATCATATA ATAAACCAAC AAATAAGTAA 2340

ATATCAAATT CTAAATTCTG TAAAAAACGA ACAATCACGA TTAAAAAAAG TATTCCTTTC 2400

TAGGAATACT TTTTAAATTA AATCGTTTGT GTGGTCGGGC GAACAACCAT TTCACTGATT 2460

GACATACGAT CTGGCGTGTC TATGGCAAAG ACCACTGCTT GCGCAATGTC TTCCGCCTTT 2520

AAGCCCCAAC TTAACTGTTC TAGATGTAAC GTTTCAGCTA CTACTCGATT AGAAATTGTT 2580

TGATAAAGTT CTGTTTGTAC AGCACCTGGT GAAATAATCG TTGATTTAAT GTTATTCTCT 2640

CTTTGTTCTT GTCGCAATCC TTCCATAATT GCTCGAACAG CAAACTTTGT TCCACAGTAG 2700

ACAGCCGAAT CTGGATAAAC GACATGTCCT GCCACAGAAT CAGTAGCAAT AATCTGCCCT 2760

GATTTTTGTT CAACCATAAT TGGAAGAACT GCCGCAATGC CATTTAAAAC ACCCATAATA 2820

TTAATATCTA GCATTTGACG CCATTCTCCT TTTGGTGCTT CAATAAGAGG CGCCGTTGGC 2880

ATAACTCCTG CATTGTTAAA AAGAACATCA ATTCGTCCAT ATTTTTCCAT TGTAAGCTTG 2940

ATTACACGCT GAACTTCCTC TTCTTTCGTT ACATCTGCTT GTTGCACGAG AATCGTTGCT 3000

TCAGGAAGTT CTTTTTTTAT AGCAATTAAA CGTTCTTGTC GACGTGCTGC AATGACTAAC 3060

TTCGCTCCTT TTCTGGCAAG TAAACGGGCT GTTGCCTCAC CGATGCCACT GGAAGCGCCC 3120

ATGATAACAA TAACTTTTTC TGATAATGAT TTCATTATAT TTCCTCCTTA TCTTCCTGAA 3180

TCCAACGAAT AATTTTAGCT GGTGTTCCTG CAACAATAGC GTTAgcTGGc ACATCTTTAG 3240

TTACAGTCGC ATTAGCTGCT ACGATAGCAT TTTCACCAAT CGTAATACCT GGTAAAACAG 3300

TGACACCCGC GCCTAACCAC GCAAATTTTT TAATGGTAAT TGGTGCTAGC AATACGCCTC 3360

GCCGTTTCAT TGGCAGTTCC GGATGATTGA CACTTAAGAG ATTAACGCGC GGACCAATTA 3420

AAACATTGTC TTCAATTATA ATGCCCCCTA AATCAACAAA CATAGCTGCC CGATTAATAA 3480

AAATATTTTT ACCAAAATGT ATATTGCGTC CAAAATCTGT ATAAAAAGGT AATAAAATCG 3540

TAACGGTTTC GTCAATTTTG TCTTGGATAA TCGTGGCAAG AAGATTCCTT ACTTCATTAT 3600

GTGTATAGGC TTGTTGATTT AATTCTATTA AATAGCGACT ATTTTCCTCT TGAATCTGAT 3660

GAATAAGTTG AAATAGTGGT TCATCTTTAT CAATGCTACC TGTTGCAATT AGTTCTTGGA 3720

GATTTTTTGT AAGCATTCCT AATCGCCTCC TTTTGCTAAA TTTATCGTAA TACGTGACAA 3780

CTCCTCTGTC TAATACTTAT AATAAATAGA TAAGTATGCG TTTTAGTTAT AGGAGGGAAA 3840

GTAATGGAAT TACGAGTGAT TCACTATTTC TTAGCAGTGG TCCAAGAAAA AACGATTAGT 3900

GGCGCAgcCA AACAATTGCA TGTATCACAA CCAACATTAT CTAAGCAATT AAAAGAATTA 3960

GAGGAAGAAT TAGGCGTGAC ATTATTTATA CGAGGAAATC GACAAATACA ATTAACACCT 4020

GAAGGAGAAT ATTTAGCTAA ACAAGGGCAA GATATTTTAA GCTTAGCGAA TAAAACAGTC 4080

ACCAACTTGT CGCAAAATGA ATTCATTAAT GGCGAAATCA CGATTGGTGG CGGCGAAACA 4140

AAAGCTATGT CTTTTTTAGC GAATGCACTA CAACAAATAA CAAGCCAGCA CTCAGCTGAT 4200

ATTCACCTTC ATTTATACAG TGGGAATGCA GATGATGTAA TAGAACGGCT AGACAAAGGA 4260

TTATTAGATT TTGGTTTAAT TATTGAGCCT GCACCTAAAC AAAAATATAG CTATTTAACA 4320

TTACCAGTTG TAGATACATG GGGCTTAATT ACTGTAAAGG ACCATCCCTT AGCCACTAAA 4380

AATGTTATTA CTGCAGCCGA TTTAAAAGAA GAACCTTTAT TTATTTCTCG ACAAGCACAA 4440

GTTCCGAGCC AACTCTCTGA TTGGCTCGAA GCAAGTCTAG ATCAGTTCCG AATCGTTGGG 4500

ACCTACAACT TACTTTACAA CGCTTCACTG ATGGTAGAAG CTGGGCTAGG TAGCGCCCTA 4560

AGCATTGATG GTATTCTAGA AACAAAGCAA ACAAATTTAC GTTTTATCCC GCTATATCCT 4620

GCTTTAACAG CCAAAATTAG TTTAATTTGG CGCAAAAATA CGGTTCTTTC CACAGCTGCT 4680

GCATTATTTT TAGAACAAAT AAAAAAAAGT ATTCAACGCC CTGAATAATA GGTCGTTGAA 4740

TACTTTTTTA AATCATCACT GTGATCGTAG CTGCTGCCAA AATCAAAACT AAGCCGATAA 4800

TCGTTACGGT CATTTCCTTG GCTGTTTTCT TTTGACCTAA GAACCAAATA CCTGTCAACG 4860

TTGCAAGCAC AACTGATGTT TGGGATAGAA TAAAGCCAGT TGCCAAGCCG TTCATATCTG 4920

GcTGCGCTGA AATTAAATAA GTTAACGCTG CAAAGGCGAA GAAAAAGCCT GAAAAAATTT 4980

GTTTGTAAGA AACGGCTTCT ATAAATGGGG ATATTTCCCC ACCCTTAATA GTGACAACCG 5040

CTGAGTAAAT AACTGCGACA ATTACCATGC CTATTGCTTG TGGTAAAAAG GCATGCATGC 5100

CATCAATTGC CGTTGCTTGA GGTGCTGCAG AGTACGCCCA ATAGCCGATT TCACCAACGG 5160

CTAATAGAAG AACTGCTTTT TTCATAATCC CAGCACTCTC AGCGGACTCC GTTTCCGACC 5220

AGACCGTCAT TTTGGCACCA ATCATGATTA ACACTAAGGC AAATGCGCCT AACAATTTGG 5280

CTGTCGCACC TGGCCAGTTT CCTAAGAAAA AGACACCCCA GAGAGAAGCA CCTAAAAGTT 5340

GGAACGCGGT CGTTACTGGC ATGGCACGTG ATGAACCAAT CATCGTGAAG CATTTAAACG 5400

TAATGATTTG CGCACAGGCC CAACCGACAC CAGATAAAAT TGAAAAGAAT AAATCCATGC 5460

CCGCTGGAAA AGCTAAACCA TTAATCATTG AAAAAATAGC GGCAAAAATC AATGTTCCTA 5520

AACTGGTACC GAGAATTTGA TTCACTGGAC GACCACCAAT TTTGGAAGCG ATTGTTGGGA 5580

AAAGCCCCCA ACCTAGAAGC GGTCCTAAAC CGATTAATAA TGCAGTCGCA TTCATCTCTA 5640

TTTCCCCTCT CTCGTTTAAA AGACAACGTT ACTTTCTAAT AAAATATTTG CGTAAGGAGT 5700

CATTTCCCCT GTCCGAATAA ATGCATGGCA GTTGTTTAAT TCTTGTTTCA TTTCTGTATG 5760

AGGAATAAAA CTGATGGGAG TTTCTGGTAA ACGAGTTTCA ATTGCTGCCA ACATGTCAGG 5820

ATTTTCTGTT TTAATTTCTT CTGCTAAATA AATTCGTTGA ACTGCCAATT CCTCAAGGAC 5880

ATTGTTTAAA ACTTCCATAA AACTAGGGAC GCCATTGGTA ACTGCTAAAT CAATTTTTTC 5940

TGTAGTCCGA GGCACTGGCA TGCCTGCATC ACCAATACTT AATTTATCAA AATGACCCAT 6000

TTGTGCGATG ACGCGGGAAA TATCTGAATT AATAACTTTT GTTTTTTTCA TGTCGCTCGC 6060

TCCTTTTTAT TTGTCTGCCA ATTCGTGTTG ATAAGGAATC GAAGGTTGGG CTCCAAAACG 6120

TTGAACAGTC AACGAAGACG CCTTGTTTCC ATAACGAATA GCTTCTTCCA AATTGCTAAA 6180

ATCTTTTTCT AATATACTAC TTAATGCGCC AATAAAAGTA TCGCCAGCAG CCGTTGTATC 6240

AACCGCTTTC ACTTTAAAAG CAGGCACAAT ACCACTTCGT CCATTGACGT CATAAAAGGC 6300

GCCTTTACTA CCTACTGTAA TAATTACTGC TTCAAyCCCT AATTGATGAA GTGCTTCTGC 6360

GGCTTTACGC ATACTCGCTT CaTCTGTGAT TTTAATGCCT GTTAAAATTT CGGTTTCTGT 6420

TTCATTTGGT ACAATCATAT CTGTCACGTT TAGTAATTCT TCAGGAACTT GTTCTAACGC 6480

AGGTGCGGGA TTTAAAATCG TTTTGACGCC TGCTTTTTTA GCAATTTTAA ACGCTGCAAT 6540

CGTACTATCA ATCGCACTTT CAAATTGGGC AATCACAAAA TCACTTTTTT CAATAATTTC 6600

TTGATGTTCC TGGACTTGTT TTGGCGTAAA GGCGTTATTG GCTCCTGCGT AAATCATAAT 6660

GCTATTTTCA CCAGCGTTAT CTACCAyAAT AAAGGCTTGG CCAGTAGCTG TTTTCTCTAA 6720

AGTGGTGACC CCAGTTAAAT TAATTTCATC TTGACTCATT AAATCAGTCA TCATAGCTCC 6780

AGCGCCATCA TTTCCCACAG CACCAATAAA ATATGTTTCT GCGCCTGAAC GTTTCGCTGC 6840

AACTGCTTGG TTTGCTCCTT TACCACCACC TGCTGTAAAA TGTTCAATGG CGTGAATTGT 6900

TTCGCCTGGT TTTGGCATTT CTTTGACTCT TAATGTTGTG TCTAAATT 6948

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8486 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

AATTTGTTGA ACCAGCGATT GCGCnGCnAA TACTGGTTTT TCCAGTGCCA GGCGGCCCAT 60

AAAGAATCAT CGATGAAAGC ATTTTGGCTT CGACCATCCG ACGAATAATT TTACCAGGGC 120

CGACTAAATG TTGTTGACCA ACTACTTCGT CTAAGTGACG AGGTCGCATT CGATAGGCTA 180

AAGGTTGTTG CATAGAAGGA CCTCCTTTTT CATAATCATC TTATTATAAC ATAAAGAACG 240

TACATyCGCT TTTTGGTGGk TTTTTGGGkT TTAGATGGAT ATAGCGGAAG GTTcGTGtAT 300

GATAGGGAGT GGTGAAAGAA ATGAAAGAAA CGATAAATAT CTTTAACACA AAAACAACTG 360

AAGAAGTGGC ACAATACTTG TTAGGGATGT ATTTAGmGCA CGAAACAGCA ACAGGCGTTT 420

TAGGTGGGTA CATTGTGGAT GCAGAAGCTT ATTTAGGTCC TGATGATGAG GCGGCCCATA 480

GCTTTGGCTT GAGAAAAACG CCACGCTTAC AAGCCATGTA TGACAAACCA GGTACGATTT 540

ATTTATATAC TATGCACACG CATTTAATTT TAAATATGGT GACACAAGAA CAAGGCAAAC 600

CACAAGGGGT GATGATCCGC GCGATTGAAC CTGTTGAAGG TGTCGACAAA ATGATTGAAA 660

ATCGCCAAGG ACGTCAAGGC GTGGAATTGA CCAATGGACC AGGTAAATTA GTTGCAGCCT 720

TAGGAATTGA TAAGCAATTA TATGGGCAGT CGATTTTTTC TAGTTCGTTA CGGCTCGTGC 780

CAGAAAAACG AAAATTTCCG AAAAAAATTG AGGCACTACC ACGAATTGGG ATTCCCAATA 840

AAGGTCGTTG GACAGAGTTG CCTTTAAGAT ACGTCGTGGC TGGCAATCCA TATATTTCTA 900

AACAAAAAAG AACAGCAGTA GACCAAATAG ATTTTGGCTG GAAGGATGAA GAAAATgAAA 960

AAAGCAACAA TGCTCACATA CTTAGAGGAA CAACTTGAAA AACACTTAGG AGATTACGAA 1020

gTAGGCCTTG ATTGGGATCG CAAAAACCAT ACCATCGAAG TCATTGTTcG TTTATATGCA 1080

GAAAaCAATG AGCmAGTGGC GATTGATGAT GTTGATGGTA CGCTTTCAGA AGwAGAATTC 1140

ATTGAATTTG AAGATGGTTT GTTATTTTAC AATCCGCAAA AGTCTGTCGT TGATGACGAA 1200

GAGTATTTAG TCACAATTCC TTATGAAGGG AAAAAAGGgC TACGCAAAGC AGTTTTAGAC 1260

GGATTCATTC ACTATTTAAA AGTGGTTTTA GATGAAGGGC AAAGTGATTT GTTAGACTTT 1320

TTATCAGATG AAACAGCGGA AGTTTTTGAA TTGCATTGGG AGCCAGCAGA TTTTGAAGCG 1380

ATGATTAAAA AAGTGGCAGA AACAGAAAAA GAACAATGGA TTGCGTATCC AAGTTATTAA 1440

TTTAGAGGTG AGAACATGAA GTGGACAGAA GTAAAAGTCG AAACGGCTAG CGAAgCAGTT 1500

GAAGCAATTT CAAACATTAT GATGGnArCT GGCGCAAGTG GCGTGGCCAT TGAAGATGCG 1560

TTGGACATTG AAAATTTTGA AAGTGATCTG TATGGGGAAA TTTTAGATAA AGAGCAATTC 1620

ACCCACATTA AAGAAGGGGC GATTGTGATG GCTTATTTTC CTGAAACAAC CTTCTTACCA 1680

GAAATTTTAC CATTTATGAA AGAAAATATT TTACGCTTGC CTGAGTACGG CTTATCCATT 1740

GGTAAAAACG AAATGACAAT TAGTGAAGTA GCAGAAAGCG ACTGGGCAAC TGCTTGGAAA 1800

AAATATTATC ATCCTGTCCG TGTCACACGA TTCTTAACGA TTGTGCCAAG TTGGGAAGCC 1860

TATCACGCAC AAGATGAAGC AGAAAAAATT ATCACATTGG ATCCGGGCAT GGCTTTCGGT 1920

ACAGGCACGC ATCCAACGAC TCGTTTAACC TTGCAAGCTT TAGAAaCTGT TTTACGTGGT 1980

GGCGAAACAG TTCTAGATGT AGGAACGGGT TCTGGTGTTT tGAGTATTGC AAGTCGCTAT 2040

TTAGGGGCTA AAGACGTCTA CGCATATGAT TTAGATGAAG TAGCAGTTGC AGCGGCAAAA 2100

GAAAaTATGG ATTTAAATCC CATTGCGGCC GATGTTCATG TGTCAGCCAA TGATTTACTA 2160

AAAGGGATTG ACCATTCTGC TGATGTAATC GTCGCAAATA TTTTGGCAGA TATTATTGTG 2220

TTAATGATTG AAGATGCTTG GCGTTTGCTA AAACAAGACG GTACCyTCmt TAtTTCTGGA 2280

ATTATTGAAG wTAmAAAAGC aATGGTTTTA GAAGCACTAA CGAAAGTCGG TTTTGTGGTG 2340

GACCAACTCT TTAATCAAGG CGATTGGTAT GCGATTaTCT TGAAGAAACC aGAGGAAGAG 2400

TAAAaTGCAA CGGTATTTCA TGAAAGAaGA TTATCCAgAA AAgATTTGTA TAAgTCGCAG 2460

ATGAAAACTA TCATCATATT GTGCGAGTTA TGCGAATGAC GCCGAATGAT CGTTGTTATT 2520

TAGTATTTCA AAATAAAACT GCCATTTTAG CTGAGATTGT GGAAATTGAT TCAACATCTG 2580

TTTACTTTAA AGAAATTAGT AAAGAAGAAA TGGACAAAGA ATTGCCAATT GAAGTGACGA 2640

TTGCGTGTGG GTTACCAAAA GGGGACAAGT TAGAATGGAT TGTTCAAAAA GGCACGGAGC 2700

TAGGTGGCAA TCAGTTTATC GGCTTTCCTG CGAAAAtTCC GTTGTCAAAT GGGACCACAA 2760

AAAAAGAGCA GCCAAAGAAA AAAGATTACA AAAAATTGCC ACAGAAGCGG CGGAGCAATC 2820

GCATCGACAG CAGACACCAA GTGTTTCTTT GGTGGAAAAA ACACAAGAAA TCATTGCGCA 2880

ATTTGACTCA TATGACACAG TCTTGGTTGC GTATGAAGAA TCAGCAAAAC AAGGAGAAAA 2940

AAGTCAATTA GCACAAGTAT TATCCACTTG CCAGCCAGGC GCACGTTTGT GTGTGCTCTT 3000

TGGACCTGAA GGTGGCTTTG CACCGCAAGA AATTGAACAA TTTCTGCAGG CTGGGGCGAA 3060

ACTTTGCGGG TTGGGTCCGA GAATTTTACG AGCAGAAACA GCACCGTTGT ATCTTTTAAG 3120

TGTGGTTAGC TACCAAATGG AACTTTTAAA TTAAACGAAA GCCCGCCTGT AAAATTCTTG 3180

TAGGCATTCT CTTTAAATAA ACAGAGAGTA AAAAAGCGAC AGTGATTGAA ATTTAAGGGA 3240

AATTTAGTAT AATAGAGTTG TTGAAAGTGC TACTGTTTCA AGTGGTACTT TTCTTTTTTA 3300

TTAAAATAGA TAAAACGTGA ACGAATgnTC TTTATCCAAA GGTCACCTTC ATAGAGGAGT 3360

GAATACCATG CCAAAAGAAG AAATACTGAC AGGCCCAGCG GTCATCAAAA TTGTAAGTAC 3420

CTATATGGGA CCAGAACATG TTGAGTTGGT TCAAAAAGCA TTAACTTATG CTGAAAAAGC 3480

ACACGAAGGT CAAGTCCGAC AATCAGGTGA ACCTTACATC ATCCATCCGA TTCAAGTGGC 3540

AGGCATTCTG GCTGAATTAC ATATGGATCC CCATACTGTA GCAACAGGAT TTTTACATGA 3600

TGTTGTCGAA GATACGGACG TAACACTGGA CGATTTAAAA GAAGAATTTG GCGAAGACAT 3660

CGCAATGTTA GTCGATGGCG TGACGAAACT AGGAAAAATT AAATATAAAT CTCATGAAGA 3720

ACAATTAGCT GAAAATCATC GGAAAATGTT ATTGGCAATG GCGCAAGATT TACGGGTCAT 3780

TATGGTGAAA CTAGCCGATC GTTTGCACAA TATGCGCACA TTAAAACATT TACGTGAAGA 3840

TAAGCAACGA CGGATTGCAC AAGAAACTTT AGAAATTTAT GCGCCTTTAG CACATCGCTT 3900

AGGGATTAGT CGAATTAAAT GGGAATTAGA AGATACTGCC TTACGTTATA TTAACCCCAA 3960

CCAATACTAT CGAATTGTTA ATTTGATGCA AAGCAAACGC GATGAACGTG AAGCTTATGT 4020

GGCAGAAGCA GTCGAAGATA TTCGTTTAGC yACGGAAGAT TTAGAAATTT ACGCrGAAAT 4080

ATATGGTCGA CCAAAACATA TTTATTCAAT CTATCGTAAA ATGAAGGACC AAAAGAAACA 4140

GTTTAATGAA ATCTATGACT TATTAGCCAT nCGTGTTATt GTCGATTCCA TTAAAsaTTG 4200

CTATGCCGTG TTAGGGcGAk cATACACGTT GGACACCAAT GCCAGGTCGC TTTAAAGATT 4260

ATATTGCTAT GCCAAAAGCG AATATGTATC AATCTATCCA TACGACAGTT ATTGGACCAA 4320

AAGGGAATCC AGTAGAGGTC CAAATTCGGA CCCATGAGAT GCACCAAATC GCCGAATTTG 4380

GGGTGGCGGC TCACTGGGCC TATAAAGAAG GCAAAGCGGA AAAAATTGAA ACAGATGAAC 4440

TGACGAAGCA AGTCGATTGG TTCCATGAAA TTATTGAATT ACAAGATGAA AGCTATGATG 4500

CTTCTGAATT TATGCAAGGT GTGAAAGAAG ACATTTTTAG CGACAAAGTC TATGTCTTTA 4560

CGCCAAGTGG AGATGTAACC GAACTGCCAA AAGGCTCCGG ACCACTAGAT TTTGCTTACA 4620

GTATCCATAC TGAAATTGGG AATAAAACAA CCGGTGCGAA GGTCAACGGC AAAATGGTAC 4680

AGTTAGACTA TGTCCTAAAA AATGGGGATA TTATTGAAGT CCTAACATCA CCGAATTCTT 4740

TCGGACCTAG TCGCGACTGG CTGAAAATGG TTAAAACAAG TAAAGCCAAA AATAAAATCA 4800

AACGTTTCTT TAAAGAACAA GACCGTGAAG ATAACATTAT TAAAGGGCAT GATGCTGTCA 4860

TTAAATACAT GACAGAAATC GGCTTTACAC CGAAAGAATT TTTAACGAAG AATAAAATGG 4920

CGGAAGTGTT AGATAAATTT AAATTTCAAA CGGAAGATGA TTTATTTGCG GCAGTAGGTT 4980

ATGGCGAAGT AAGTGCGCAA GTAGTGTGCA ATCGTTTAAC AGAAAAAGAA CGCCGCGAAC 5040

AAGAGTTGGA ACGCCAACGT CAAGAAGCAG AAGAGTTGTT AAATCAGCCA GCTAAAAAAG 5100

AATCAGAAAA AATGAAAGTT CGTCATGAAG GCGGCATCGT GATTCAAGGG GTGGACAACC 5160

TATTGATTCG TTTAAGCCGC TGTTGTAATC CAGTGCCTGG TGATGAAATC GTTGGTTACA 5220

TTACCAAAGG ACGTGGCGTT TCAATTCATC GTGCGGACTG TCCTAACGTA CAACATCAAG 5280

AAGAACTAGC ACAACGTCTA ATTGAAGTTG AGTGGGAAGA TACCGAGCAC AGTCGCAAAG 5340

AATACGCTGC CGATTTGGAG ATTTACGGCT ATGATCGTAG TGGCTTACTG AGTGATGTTT 5400

TACAAGTCAT TAGTTCGATG ACAAAAAATT TAGTCGGCGT TGAAGCTCGT CCATCCAAAG 5460

ATAAGATGGC CCAGATTCAT GTTACTGTGA AAATTCAAAA CCTTTCTCAT TTAAAAACCA 5520

TCGTTGACAA AATTAAAAGT GTACCAGATG TTTATAGTGT TCGTCGAACC AATGGGTAAG 5580

TGTTAAAAGT GAGGCAAAAA AGATGAAAGT AGTTATTCAA AGAGTCAGTC AAGCACAAGT 5640

TGCGATAGAA GAGCAGATAG TGGGACAAAT CAAACAAGGC TTCATGGTGC TGGTGGGGAT 5700

TCATCAAGAA GACACGCCAG AAGATGTGGC TTATGTTGTT GGCAAAATCA GTAAATTGCG 5760

TGTTTTTGAA GATGACGAAG GAAAAATGAA TCGCAGTATT CAAGAGATTG AAGGATCAAT 5820

TTTGAGTATC TCTCAATTTA CCCTCTATGC AAAAACTAAA AAAGGCAATC GTCCCAGTTT 5880

TATTGAAGCG GCTCGTCCAG ATGTGGCGAT TCCTTTGTAT GAATTATTTA ATCAACAATT 5940

AGAAGCAGAA GGAATAGCGG TAGCCACAGG AGAATTTGGC GCGGATATGC AAGTGTCGTT 6000

GACCAATGAT GGTCCAGTGA CCATTGTGAT TGATACACGA GAAAAATAAT TGATTTTCCT 6060

GTCGGAACCC CTTGACTTTC TAAGAGAACT TTAGTAGTTT ATAGATAATA TTAAAAATCT 6120

ACGAAGGAAC GAGTAAGTTT CATCTGTCTG CGAAGAGAGA GTTGCCGTGG CTGAGAGCAA 6180

CTTCAACAGT GAAAGTGAAA GACATTCTGG AGATGGACGG AAGAAAAGCC GTCCCGTTAC 6240

GAGCGTTAAT CGTTGAGGAA AGAGTAAAAT CTTTCAAAAC TAGGTGGTAC CGTGTAAAAG 6300

AGACTTTTAC GCCCTTGTAT GTTCAACATA CAAGGGCTTT TTTCATTTTT AGAAAAAAGA 6360

AAATGGCGCA CACATTTTTG TGGAATGATA GAAGGAGAGA AAGACAATGA GTTATCAAAA 6420

ACCAAAAGGA ACAAACGATA TTTTGCCAGG AACTTCTGAA AAATGGCAAT TTGTGGAAGA 6480

AACAGCTCGT TTGATTTTTA AAGATTATCA ATACCAAGAA ATCAGAACCC CGATTTTTGA 6540

ACATTATGAA GTAATATCTC GCAGTGTTGG CGATACCACA GATATTGTTT CAAAAGAAAT 6600

GTATGATTTT TATGATAAAG GAGACCGTCA CGTGACGCTA CGTCCTGAGG GGACAGCGCC 6660

aATTGTTCGG GCCTTCGTTG AAAATAAATT ATATGGTCCG GAATATACGA AACCATATAA 6720

AACCTATTAC ATGGGGCCGA TGTTCCGCTA TGAACGCCCA CAAGCTGGTC GTTTGCGTCA 6780

ATTCCATCAA ATTGGTGTGG AAGCATTTGG TAGTGAAAAC CCAGCATTGG ATGTTGAAAT 6840

CATGGCTATG GCTTTGGACT TCTTCAAACA ATTAGGCATC CAACAAATCA AATTAGTTAT 6900

TAATTCCTTG GGGGATAAAG AAACACGTGC TACGTACCGT CAAGCATTAA TCGATTATTT 6960

AGAGCCCCAT ATGGCAGAAT TAAGCGAGGA TTCACAACGT CGCTTACACG AAAACCCATT 7020

GCGGGTGTTA GACAGCAAAG ATAAAAAAGA CAAGGTGATT GTCGCAGAAG CGCCCTCCAT 7080

TTTGGATTAT TTAAATGAAC CATCTAAAGC ACATTTTGAA GCAGTAACTG ATATGTTAGA 7140

TTTACTAGAA ATTCCTTATG AAATTGATAG TAATATGGTT CGTGGCCTGG ATTATTATAC 7200

ACACACAATT TTTGAAATTA TGAGTGAAGC GCCTAAAATG GGTGCGCAAT CAACTATTTG 7260

TGCAGGAGGC CGATACAATG GTTTAGTTGA AGAATTAGGC GGCCCAGACA CACCAGGTTT 7320

TGGTTTTGGT ATGGGCATTG AGCGAGTGTT GTTAACAATG GAAGCTGAAG AAGTTGTGAT 7380

TCCAGCGTTA TCTGAATTAG ACGCATATGT GGTTGGGATT GGTTCAGACA CCAACGTCGC 7440

AGCTTTGCAA CTTGTTCAAA GCATTCGTAA CTTTGGTTTC TCAGCTGATC GTGATTACAT 7500

GAATCGCAAA CCAAAAGCGC AATTTAAAAC GGCCGATAAA TTACAAGCAA AATTAGTTTT 7560

AACAATCGGT GAAAATGAAT TGAATGAAGG CATTGTCAAC GTAAAATCAA TGGCAACACG 7620

CGAAGAAAAA GCCTTCCCGT TAAGTGCTAT TCATGATTCA TTTGATGAAG TGTATGACGA 7680

AATGATGACA AAAATGATTG AAGAATAAAA AGGAGAGATT GAGAAATGGA AAAACGCACA 7740

ACCTATTGCG GCAATGTATC TGCCGAATTT ATTGAAAAAG AAGTTGTATT AAAAGGATGG 7800

GTTCAAAAAC GTCGTGACTT AGGAGGCGTT ATCTTTATTG ATTTACGTGA CCGTGAAGGA 7860

ATCGTCCAAG TGGTCTTTAA CCCAGAAAAG TCAAAAGAAG CATGGGAAAT TGCTGATAAA 7920

TGTCGTAGTG AATACGTCAT CGAAGTTAAA GGTCAAGTCG TTTATCGTGA TAAAGAAGCG 7980

ATTAATCCCA AAATGAAAAC TGGCGAGTTT GAAGTGATGG CGACTGACAT TACGATTTTA 8040

AATACTGCGA AAACAACGCC ATTTACCATT GAAGATGACA ATAACGTTAA TGACGAATTG 8100

CGCATGAAAT ATCGTTATTT AGATTTACGT CGTCCATCAA TGACAAACAA CATTAAATTA 8160

CGTCACCAAG TAACGAAAAC AATCCGTCAC TATTTAGATA ATCATGACTT TTTAGACATT 8220

GAAACTCCTT ATTTAGGCAA ATCAACACCA GAAGGCGCTC GCGATTATTT GGTTCCTTCT 8280

CGTGTCCATG CCGGTCATTT TTATGCTTTA CCACAATCAC CACAACTCTT TAAACAATTA 8340

TTAATGGGGG CAGGCTTCGA TCGCTATTAC CAAATCGTTC GTTGTTTCCG TGATGAAGAT 8400

TTACGAGGCG ACCGTCAGCC AGAATTTACA CAAATCGATA TTGAAACGAC CTTTTTAACA 8460

CCAGAAGAAA TTCAAACATA TACAGA 8486

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 789 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

AATGTCAAAA TTATTAGTTG TTAAAGCACA TCCATTAACA AAAGAAGAAT CACGCTCAGT 60

TCGTGCGTTA GAAACATTTT TAGCTTCATA CCGTGAAACA AATCCATCAG ACGAAATCGA 120

AATTTTAGAT GTTTATGCCC CAGAAACAAA CATGCCGGAA ATCGATGAAG AATTATTATC 180

TGCTTGGGGT GCACTTCGCG CTGGCGCAgc ATTTGAAACA TTAAGCGAAA ACCAACAACA 240

AAAAGTGGCT CGTTTTAATG AATTAACTGA TCAATTTTTA TCTGCAGACA AAGTAGTAAT 300

TGCTAATCCA ATGTGGAACT TAAACGTACC GACACGCTTA AAAGCTTGGG TAGATACAAT 360

CAACGTTGCT GGAAAAACAT TCCAATATAC TGCAGAAGGA CCAAAACCTC TAACAAGTGG 420

TAAAAAAGCC TTACACATCC AATCAAATGG CGGCTTCTAC GAAGGAAAAG ATTTTGCTTC 480

TCAATACATT AAAGCGATTC TAAACTTTAT TGGCGTTGAT CAAGTTGACG GATTATTCAT 540

CGAAGGAATC GATCATTTCC CTGATCGCGC GGAAGAACTT TTAAATACTG CCATGACCAA 600

AGCAACTGAA TACGGTAAAA CATTCTAAAT TTTCCGAGTG ATTTACATCA AAAaGAGACA 660

AGCAGTTCGT TCACTGCTTG TCTCTTTTGT TACGTAGATT CTTGTTGaTC TGGcTTATCt 720

TCATcATACA AACCGGaAAC GaTTTGGtAC CACTCaAAAA TGTGTGTAAT ATCACnTTAG 780

CTGCGGCAT 789

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4159 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CGTCGTTTAT TACGTCGTGA CGGTAATGCA CGGGAAAAAA ATTAGGCATT AATGAAGCCT 60

TTTTATACAA ATTAGTGCCT GTTGTTGGCG AAATCATGGT AAGCTACTAT CCAGAAGTAT 120

TACAACAAAA AGACTTCATC GAAAAAGTGG TTCGGACAGA GGAAGAAnGT TTCCATGAAA 180

CCATCAATGA AGGTTTAAGC ATGTTAAATG AGGTTATTAA AGAAGTTAAA GATGCAAAAG 240

GCGATACATT AGATGGAAAA ATTATCTTCA AACTTTATGA TACTTTTGGc TtCCCCGTAG 300

AATTAACGGA AGAAGTTGCA GAAGATGAAG GCTtGAAAGT CGACCAtGCT GGGTTTGAAA 360

CAGAAATGGA AGCACAGCgT GAACGTGCTC GTTCTGCCCG CAGTAAAGAA ACTTCAATGG 420

GTGTTCAATC TGCTTTATTA ACAGATATTA AAGTAGAAAG TAAATTTGTT GGCTATACAG 480

AATTAACACA TGATAGTGAA TTATTTGTTA TTATTCAAGG TGACGCACTA GTAAATGAAG 540

CATCTGCAGG AACAGCCGAA TTAATTTTTG CTGAAACACC ATTTTATGCT GAAATGGGTG 600

GACAAATTGC TGACCGCGGC TATGTAAAAA ATACCGCAGG GgAAGTCGTT GCCAACGTGG 660

TGGATGTGAA AAAAGCACCA AACGGTCAAT TTTTACACAA AGTAGAAGTT TTGGCGCCAT 720

TAGCAGAAGG TCAAATTTAT CAATTGCAAG TGGACGAACG GATGCGGACA CGTaTTTTGA 780

AAAATCATAC GGCAACCCAT TTATTGCATC GTGCTTTAAA AGATGTGTTA GGGGaGCATG 840

CCAACCAAGC AGGCTCATTA GTTGCACCAG GACATCTACG TTTTGACTTT ACTCATTTTG 900

GGCAAGTTAC ATCAGAAGAA CTAGCGCGGA TGGAAGCCAT CGTTAACGAA AAAATCTGGG 960

AAGCTATTCC TGTTGTCACA ATTGAAACAG ATATTGATAC AGCGAAAAAC ATGGGCGCAA 1020

TGGCGTTATT TGGCGAAAAA TATGGCAAAG AAGTCCGTGT AGTTAATATT GGGGATTACT 1080

CTATCGAATT ATGTGGTGGA ACACACGTTG CTAATACAGA AGATATCGGG ATTTTCAAAA 1140

TTGTTTCTGA ATCTGGAATT GGTGCAGGAG TACGCCGTAT TGAAGCAGTG ACGAGTAAAG 1200

AAGCGTATCA ACTTTTACAA GAAGAAGAAC GTCAATTGAA AGAGATTGCT ACATTAGTTA 1260

AATCACCTCA ATTAAAAGAA GTGGTTACGA AAACAGAGCA ATTGCAACAA CAATTACGTG 1320

ATCTACAAAA AGAAAACGAA CAATTGGCAG GCAAATTAGC GAACCAACAA GCTGGTGACA 1380

TTTTTAAAGA TGTCAAAGAC ATCAATGGTG TTCGTTATAT TGCCGCTCAA GTCAATGTTA 1440

AAGACATGAA TCAATTACGC CAATTGGCTG ACCAATGGAA ACAAAAAGAA TTGTCTGATG 1500

TACTAGTTCT AGCAACTGCA CAAGATGAAA AAGTAAGCTT GTTAGCGGCT ATGACAAAAG 1560

ACATGAACGG AAAAGGCTTA AAAGCCGGTG ACTTAATCAA AGCAATTGCG CCAAAAGTTG 1620

GCGGTGGCGG CGGTGGTCGT CCTGATATGG CTCAAGCTGG TGGGAAAAAT CCAGCAGGGA 1680

TTGCGGATGC CTTAGCAGAA GTTGAAAACT GGTTGGCAAA TGCCTAATTA AAAATGTTTT 1740

CCCTATTATT AATATAAGGA GAATGCTCTT TATAGCaATA ATGACAGCCC TCTTCAAAGT 1800

TATGAAATTG ATTTTGAAGG GGGTTCTTTT TTTtAAATGA TTAAGGAGAG GAAACAGTAG 1860

ATTGCACGAA TGTGTAAGTT TTCATCTTCA ATTGACAATG GTATAATAAG CAACGGAGGA 1920

ATCAAAAATA TGAATGAAAA TCAATTaTCa AAACGTCTCG CAACGGTTGG TGATTTAATC 1980

CCTAGAGGAA GTCGCTTAGC AGACATTGGT TCAGACCATG CTTATTTACC AGTTGCTTTA 2040

ATGTTAGAAA ATAAACTGTC TTTTGCTGTG GCTGGGGAAG TGGTGGAAGG CCCCTATCAA 2100

TCTGCCAAAA CCCAAGTCAG TAAATCCAAT TTGACGGATA AAATTATCGT ACGTTTAGCG 2160

AACGGGCTAG ATGCAATTGA ACCTGAAGAT CAmATAGACG TCATTAGTAT TTGTGGAATG 2220

GGCGGCACAT TAATCCGAGA TATTTTAGAA GCTGGACGCA AAaAAAATCG TTTAACAGGG 2280

AAGGAACGCC TTGTTTTACA ACCGAaTATT GGTGAACCAA CTTTACGCCG CTGGTTAATG 2340

GCAAATGACT ACAGTATTAT TGATGAAACA ATCGTGGAAG AAAACCGTAA ATTGTATGAA 2400

ATTATCGTAG CCGAAAAAAC AGAGCAATCC GTTTCTTATA CGGACCAAGA ATTACTATTT 2460

GGGCCAGTTT TAATTAAAAA ACAAGGACCG GTTTTTACTA AAAAATGGCA ACGAGAATTA 2520

AAACAGCGCA AAACTGTTCT TGCACAACTA GCAAAAGCAA GTGGAGAACA TATTGAAAAG 2580

CAAGCGAAAC TACAGCAAGA CCAACAATTA ATTGAGGAGG TACTGGCCAA TGGCTGTGAA 2640

CGGTAAGACA TTGATTCGAC GCTTCAATGA CTATTGTCCC GAATGGTTAG CAGAAACGGG 2700

AGATCCTGTG GGCTTGCACA TTGGTACATT AGATAAACCA ATTGAAAATG TGATGGTTAC 2760

CTTAGATGTT CGTCCTGAGG TCGTTGCAGA AGCAATTGAG AAGCAGGTTG ATTTAATTAT 2820

TGCGAAACAC CCACCTATTT TCCGCCCAGT TAAGCGATTA ACCACAGATA ATTTTCAAGA 2880

AAAAATGTAC GCTGATTTAT TAAAACACGA TATTGCTGTT TATGCAGCGC ATACTAACAT 2940

GGATATTATT GACAATGGGC TAAATGATTG GTTTTGTGAA TTATTGGGAA TTAAACAGAC 3000

CACCTTTTTA ACTAAAACAC ATACTGTTCC TTATAAAAAA TTAGCCGTTT TTGTGCCAAT 3060

TGATGAAgCT CCGCAAATGA GAGAAGCTTT AGGTCTAGCT GGCGCCGGGT CTCAAGGTGA 3120

TTATTCTAAA ACAAGTTACT CTTTAATTGG AACAGGACGT TTTACACCAA CACAAGGAGC 3180

AAATCCAACG ATTGGCGAAA TTGGTCAAGA AAGTGTCGTC CAAGAAGCAA AAATCGAAGT 3240

GATTTTCCCT GAAACTAAGC AAGAACAAGT TCTTGCAGCT ATGTTACAGG CGCATCCTTA 3300

TGAAGAACCG GcGTATGATG TTTATACCAT TGAAAATCAA TCAAAAGAAT TTGGTTTAGG 3360

TCGGGTTGGA GTATTAGACA AGCCTGTGAG ATTCTCTTAT TTTGTACaGC AAGTCAAGGa 3420

AGCTTTCCaA TTGGATGGCT TGCGGkTCAT TGCAAAAGAT GATACTAAAA TGATTCAACG 3480

GGTAGCCATT TGTGGTGGAA GTGGTGAAAA ATTCTATCAT GATGCATTAC GTAAACAAGC 3540

CGATGTCTAT ATTACTGGTG ATGTGTATTA TCATACGGCA CATGACATGA TTGCTGAAGA 3600

TTTACCAGTT ATCGATCCTG GTCATTACAT TGAAGCGCTT TGTAAGCCAr AACTTGTGGA 3660

ATTAATGAAT CAATGGAAAC mAGAAAaTGA GTGGgCTGTA TCCATTTTkG AaTCAGAAGC 3720

GAATACCAAC CCATTTCGTT TTAAATAAAA AAATGAAGAT TAAAGGAGTG TTCTTTTATG 3780

TATGAAAATC TATTACCTCG TTTTTTACGG TATGTGAAaA CAGAAACACG CTCGGATGCA 3840

ACTAGCACAA CGACACCATC AACACAAACA CAAGTAGCCT TTGCACAAAC CTTAAAAAAA 3900

GAATTAGAAG AATTAGGGAT GAGCGATGTT ATTTATAACG AAACAAATGG TTTTGTGATn 3960

GCTACGTTAC CTAGTAACGT GGAGAAAGAC GTTCGTTCAA TCGGCTTTAT TGCCCATATG 4020

GATACCGCTG ATTTTAATGC AGTGAATGTT TCTCCGCAAA TTGTTGAGAA CTATGACGGA 4080

GAATCAACGA TTCCTTTAGA TAAAGAAGGC AAATTCACTT TAAACACGAA AGACTTTCCT 4140

AACTTAAAAA ATTATCGTG 4159

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8377 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

ATCGATACAA TTTTTCTTTT GTCACAACAG AAGAAGAATT ATTAGAGATT TTAAAAGATG 60

CCTTAAAAGA TAAAGCCTTA GTTGTCAGTA CATTAGTCAG TAAACAACTA ATCACAGCGG 120

CGAAAGAATT TAGTGAACGA ACAGGGTTGT TATATTTAGA TTTAATGGCG CCATTTTTTG 180

AATTAATTCA AGCGAAAGCC GGAGTAGATC CTATTGAAGA GCCTGGACGA CGCCACCAAC 240

TAGATCGTGC CTATTTTGAT AAAATCTCAG CGATTGAATT TGCTGTAAAA TATGaTGATG 300

GCAAAAATCC TCAAGGGTTT CTTGATTCTG ATATATTGTT GTTAGGCGTT TCGCGGACCT 360

CAAAGACGCC AGTCAGTATG TATTTAGCGA ATCAAGGCTA CCGCGTTTCT AACTTACCAT 420

TAATTCCAGA AGTTCCATTG CCGCCAATTT TGGAAGAAAT GGATCCACAA AAAATGATTG 480

GTTTAGTTTG TTCGCCAGAA ACATTAGGAC AGATTCGTAG CAGTCGGTTG GCTTCCTTAG 540

GTTTAGGTAA TGAGACCAGT TATACCAATG TTGAACGGAT TGAACAAGAA TTAGCTTATG 600

CCGAAGAGAT TTTTGCGAAG TATGGCATCC CAGTGATTGA TGTAACCGCA AAATCTGTCG 660

AAGAAACAGC CTTTTTAATT AAAGAAAAAC TAGATGAAAG AAATTAAAAA GCTTGGCAAC 720

CTCACTAAAT TTCAGTGAAG TTGCCAAGCT TTTTATACTT AATTGAGCTT CTTTTTGTTA 780

GTCAATATTT TTTTGATCGA CAATGAGTAA AGCGATCATC ACGTAGACAA TCCAACTACT 840

TTTAGAATAA AGTGTATAGC GGGGTTGCCA TTTGCTAGCA AATTTTTCCT TGTATTCGCG 900

TAATCCTTGG AAGGAGTAAA TTTCTGAACC AAATTCATAA ATTAACGCAG CAATCCGTTC 960

TTGAATAAAA CTTTTTCTTG ATTGACCAAC GTTGGCTAAC GGAGCCATCC CTAAATTGAA 1020

ATAATGAATA TTTTCTGTTT TCATATATTC AAATAAATGA ATAAATAAGA AATCCATACT 1080

TCCAGAAGGG GCTTTTTCTT TATGGTGACG CATCAAGTCG ATGGTACCAA CTTCATTTGT 1140

GTAACTAGGA ATGATATTAG CGAAAGAGAC CATTTCTCCT TCAGCATTTC GGACAACAGC 1200

AATTGGGTTT CTTTGAAGAT ACGCTTCATC GAAAAATCCT AGTGAAAAGC CTTTTTCTTT 1260

TCGATTGTCT AACCAACTAT CCGAAATAGC TTTTAATTCG TGCATTTGTT CGGTACTAAA 1320

AGGAGGTTGG AGTACGTCAA ATGAATAGCC GGCTTTAGTA ATTTTATTAA ACACTGCGCG 1380

GGTACCTTTG AATTTTTTCC CAGACATCGT AAATGTTTCT AGATTAACTA AGGCTTCTTC 1440

GCCCATTTTG ATAAATTCGT AGCCATATTC GTGTAGTAAC ATCACGCTGT CCTCATTGGT 1500

TTCGTAAAAG ACGGGTACAT AGCCGAAACG ATCTGTTTCT TTAAGGAATG CATCAATCGC 1560

CGCAGGAAAA TCTTCTTTGT TCCCAGAAGG GTTGCCCATC ACAACGCATT TATTATTAAT 1620

GGTGTTGAAT TGTAAAAAGA CAGTCGGTTC TTCTTTGGGA TATAAGAAAA CCTGTTTATC 1680

ATGAAGAAAG ACAAGTTCAC TGTCAATATT CCCGCCATAA GTTGTTAAGA TATTATGCAA 1740

GATACTATCT TCTGGAAACT CACCAACTTG GTGTTTTTTC CCTTGGAGGT AGCGAACAAA 1800

TAGAAAGTTA AAACTCATAA CTAAGAGGAT AGCGATAAAT CCTGACAACC AGATCCGTTC 1860

GGAAGGGAAC AAGAAAAATT CCACAAAATG ATGACGATGG TGTGGGAAGT TCGGTAAATT 1920

ATAAACACCA ATTACCAGGT ATAACAAAGT CAACAAGCCA AAGATAAAGC CATCAATCGT 1980

CATCCACTCC CAAGAGTAAA CCAGTTGTTC ACGGTACAAT TCTGATTTTG AAAAAATAAC 2040

GATAAATAAT AAAATAGCTA AATAGAAAAT GGAAAACATA CTAAAGTCAA CAACAAATGA 2100

ATACAAGATA GCACCAATTA ATAAAATAAT CGTTGGATAA TAGGCTCTTT tAaCmCGATC 2160

AGCATCCCcG TCCCATAATT mATAATGCAA AGCCTAAAAC AATACTCGGA ATTTGnATAA 2220

TAATATGTGA ACGGAAGGGG TTAATATCTC TTAACCAATG CACTTGAGTG AAGGCTTCTG 2280

GAATAGTGGC TAGGAGCACC ATCATAATTC CTGAAAAATA AAGTAAAACC ACTTCTAGCT 2340

TGTGGGCAAT TTCCAAAGTA ATTTGTTTTA GCAGTTGACG ATAATGAGTA TCCAACTTTT 2400

GGCTTAAGTG TTTTGTAAAG AACAAACAGC CAATTAAGAA CGGAATAATA TAATAGAATA 2460

AACGATAGAG CAATAACCAC AAAACAATAA TTTCACGATC GACACCTAAG TTGCTCAGTC 2520

CTAAAATCAT CATGACATCA AAACTACCTA AAGCCCCTGG AATCATCGAA GCAATCCCAA 2580

TAATCGAAGC TGCTACATAT AAAGGAACAA TATCAATTAA AGGAATCGGT ACATCTAATA 2640

AGTAACCAAT ACTGATAAAG GTAATTAAAA CACCTGTCCA TTCTAAAAAC GAaGAAACAA 2700

CTAGAGATAA GCGATTTTTT GCATCGAGGT TACCAAAGGA ACTATTTTTT TGGAAATGTA 2760

AAATAAGATA AAGTAACGGG AAATAAAGAC CGCCACCTAA TAACCAAATC CAATATTGCT 2820

GTAAATAAGC CATCTTTGGA TCAAACTGAA CAAGGAAAAA GCAAAGAATA CTATAAATAG 2880

ATAGACCGGA TAACACAAAA AGTAAAATAT GGGTTAACGC AGGTAAAATT TTCTTTTCCT 2940

CTGTTTTATT TCCATAAAAT TCAGAACGAA GTCCCATGCT GACCAAACCA CCGAATCCAC 3000

CAaTATTATT GATTGTATTG ACAATCCAGC TAGATTCAAG GAGTTCCCGT TTTGAGACTT 3060

TCAAGTTTAA AATGCGATTT AATGTTAAGT CATAATTTAG CATCGGAGTA ACGGAAACTA 3120

AACCGATAAC CATCATCAAT AACAATTTCC ATAGCGGAAT CTCATCAAAA ATCTGTTTTA 3180

ACTGTTCAAA TGAGATCGTT TTACCAATTG ATAAAAGTTG GGCAACAACA ATAATTGAGA 3240

CCGCAATAAC AAACAGTGTT TTTAGTAATC CTGTATGCGT TTTGATCCAA TGAATAATTT 3300

GTTTCATTTA GACACCTTCA TTTCCTTTAA GAATTTTTCG ATTGTTTTAT TAAAAATATC 3360

TGGTGAGTTT TTGGCAAATA AATGACCCTG TTTATTTACC AAGACCAATT TGCCGGCAGG 3420

AATGGcATCA GCAATCGCCC GTGAATGGGA AAGCGCAATT AAATCTCGAC TGCCAGTAAT 3480

AACAAGAGTA GGACAGGGGA TGTGTTCTAA ATCTTGTTTT GTCAGTCCAG TATCTTCAAA 3540

TAACAGAGCC ACGATAGGGA GCCGTGCGCG CAATTTTTTT GAAAAAAAGG CGCCAATACG 3600

ACAAAACCAA TATTGAAAAT AAGTGGCAAT CGTTACATAT TTTTTCAGAC CAGACATGGT 3660

GACATTTCCA GAATTTAAGA CGAGTTTGTT GACGAAATCA GGATATAAAT GGGTAAAAAC 3720

CATGGCAATA TTTGCTCCGT CACTAAAGCC CAATAAATTA ATTTTTGTTA AGTGTTCTTG 3780

TTGAATAATA CCAGCTAAAT CTTCTGCCAA AAGCGGAAAA GTTAATTGAC TTTGTGTATT 3840

GGTGGAACGT CCATGTCCGC GGGTATCCAC CGCAATAACT TGATACTCCT TGGAGAAATA 3900

AGAAATTTGA TAATTAAAAA AAGTATGATC ATTGCCATTA CCATGAAGTA ACACAAGGGG 3960

GTAATTTGTT kCTTTGCCAA AAATTTGAyA AAAAATTTGA CTGCCATCTA TTGTTTGAAA 4020

ATATTTTTCT ATTGCCATTC TCCGTCACCT ATCTCATCTA TAATCAATCG AGTCCTCTCT 4080

CATTATAGAG TCTTTTGTCT TTTTTTTGTA ATAAAAAACA GAAAATTTTC CGAACAGTTT 4140

TGTAAGGTGT TGTAATATAG TGTGATATAG TAGAAGATAG CCACCTAAAA ATTAGACTAG 4200

AGGAGGGAAA GTTGATGTCT TTAATCGTTT TAATTGGTGC CCAAGCTGTC GGGAAAATGA 4260

CCGTTGGCAA GGCTTTAGAA AAGCAGTTAG ATGCTAAATT ACTATTTAAT CATCAAACCA 4320

TTGACTTATT TGCAAACTAT CTTGGGTATA CGGAACGTGc ATTTCAATTG TCTGATTCCG 4380

TTAGAAAGGA ACTTTTTCAT GCCTTTGTCG AAAATCCAGC CACAAATACA ACAAAAAACA 4440

TTATTTTTAC AGTGGTGATT GCTTTTGATC AAGAACAGGA TTTTCAATTT TTAGAAGAAA 4500

TATCAGAAAT TTTTCTTAGT AAACATGAGT CCGTTTATTT CGTCGAACTC GTCTCTTCAA 4560

CATCAATCCG CCTAGAACGA AATGTTCATC CAGAACGGCT AGAGGCAAAA CCTTCAAAAC 4620

GCGATGTCGA CTTTTCTCGA AAAGAATTAC TCAATACCTT TAAAGAACAT CGATTAGTAA 4680

GCCACGAaGG TGAACTAGAA AGTATATTTC CAAATGTAAA AATAGCGAAA ATAAATAATT 4740

CAAACATGAC ACCAGAAGAA GTAGCCACGA AGATAATTGC TACTTTTGGT TTGACATAAT 4800

GTGAAAGGAA CATCGGAATC AATGAAAAAG TTTTTTAAAT ATCTCGGTAG CTTGATGGTC 4860

CTTCTTTTAA GCATAGCTTT AATTTTGATA GGATATTTAA CCTTCCGAGA GTTTCGACCA 4920

CGAGCGGTTG AACCAGTTGC TATAAATAAA CAATTTTCAG AGCAAACCAT CGAAAAAAAG 4980

CAATCCATTT CTTTAGTCAC GTATAACATA GGCTATGCCG GCCTAGGACA AACAGAAGAT 5040

TTTTTTATGG ACGGTGGCAA AACTGTCCAA CCAGTTAATA AGGCAATGGT CCAACAGAAT 5100

TTAACAGGCA TTGAAGAAAC CTTAAAAGAA TTACCAGCGA TGATTTACTT ATTCCAAGAA 5160

GTCGATCGTC GCTCGCAACG TTCTTATGAA GTCAATCAAG AGGAAGAATT ACAAAAACAG 5220

CTACAGCTGA ACAGCGCATT TGCTTATAAT TTTAAGGTAG ATTATGTGCC CATTCCTTGG 5280

CCGCCGATTG GACGAGTAGA AAGTGGTTTA TTAACATTAT CAAATGAAAA AATAACAGAG 5340

GCCAAACGCA TTGCTTTACC CAATCCCTTT CGTTGGCCCG TTAGCATTAG TAACTTAAAA 5400

CGAGCACTTC TAGAAaCGCG CTTTCCAATT AAAGGCACAG ACAAAGAATT AGTGGTTTTT 5460

AACTTACATT TGGAAGCCTA TGACAATGGG GAAGGGAAGA TTGCTCAAAG CAAAAAGTTA 5520

GCTGACGTAT TATCCCAAGA ATACGCTAAA GGAAACTATG TAATTGCTGG AGGTGACTTT 5580

AATCaAGTTT TCaAAGGCAG TCATCGATAT CCTGACTTAG GGGAAGCTGG TTGGGTACCA 5640

GGAGAAATTG ATCCAGCAGA CTTACCTAAA CATTTTTCAC TTGCTTATGA TGATCAACAA 5700

CCGACTGTGC GGGTGCTTAA TAAGCCTTAC ACTGGCTCTT ATGAAACATC CCAAGTGTAT 5760

GTAATTGATG GTTTTATTGT TTCTGATAAT GTGGCTGTTC ATAGCGTGAA AACAAAGGAC 5820

GAACAATTTA AATATACAGA TCACCAACCT GTGAAGATGG AAGTTGGGTT GAAATAGAAA 5880

AAGGAGGAGA TAGATGAAAT TAGGGGTTAT TGCCGCAATG GATTTAGAGC TTCAAAAATT 5940

ATTGGAATAT TTTCCACCAC AACGAAAGAT ACAGCTAGCA AAAAATACAT TTTATATTTA 6000

TGAACAAAAA ACATCGCAAG TCATTATGGT TTGTGCAGGA CAAGGAAAAA CGAATGCTAC 6060

GCTATATAGT CAAATACTTA TTGATTCATT TCAAATAGAA CAACTAATTA ATATCGGTAT 6120

TTGTGGCTGT TTAAATGAAA AACTGCAACT CTTTGAGATG GTGCTAGGTG AGGAATACTG 6180

CCATTACGAT ATTCGTGAAA GACAGTCAAT CAATAAATTT CCTTATCAGC TGTACTATAA 6240

AGGAGATAAA GGAATGCTGG CCGAACTTCA ACATCTAGAT GAAAAGATAA AGTGTGTTCG 6300

TTTTGGAACT GGAGAAGGCT TCGTTTGTGA CACAACTGAA AAAAGAAACT TAATTGAACA 6360

ATTCTCTATT GACTGTGTGG ATATGGAATC CGCAGCTATT GCACAATGCT GCTTTTTAAA 6420

TGATTGCGCT TTTGTTTCTA TAAGAATTAT TTGCGATAGA GCGGATGCTA ACGCAGTCAA 6480

AGTAACAGAA GATACACAAA CACAGGCAAT GGAAAAAGTA TTTGAATTGG TCTGTGCATA 6540

TATAAATTAA AAAAGAAAGA GCTTGGAACA AAAATCCAGA AAGATTTTTG TTCCAAGCtT 6600

AAAAAGTAAT ATAGGTTGGG GATAAAAGTA ACATTGTTAT TGTTCCAATT ATTTTACAGA 6660

TAGCCTTTAG TAATAGAATG AGTGGAATTT AGCATATTTG CTGGGGTTCC TGTAAAAAGA 6720

AGCTGTCCGC CATTTTTTCC ACCTTCGGGA CCTAGTTCAA TCAACCAGTC GGCACTTTTA 6780

ATCACATCAA TGTGATGTTC TAATAAAATC AAGCTGTTTC CTTCTTCTAC CATTCGATTG 6840

AAAAGTTGAA GACTTTGTTG GATATCAATT AAGTGTAAAC CATCTGTAGG TTCATCCATT 6900

AAATAGATAG CTTTTTTTTG GTGAAGTGTG TCTGCTAATT TTACCCTTTG CAATTCGCCA 6960

CCAGATAATG TTGAGAGCGA TTGATTGAGT TTAAGATAGC TTAAGCCAAC TTCCAGTAAA 7020

TTTTTTAATG AAAGAGCAAA AGGCTGGTCT TTGAAAAAAT CATAGCCATC TTTAACGCTT 7080

AAAGCTAGTA CTTCAACGAT ATTTTTTCCG TTATACAGAT AATGAAGCAC CTCTTCTTTG 7140

TAACGCGTTC CGTGACAGGT TTCGCAAATA CTAGTAACAT CCTCCATAAA AGACATATCG 7200

GAGACAATTA TGCCTTTCCC CTTGCAAGTA GGACAGGCGC CTTTGGAATT ATAGCTAAAT 7260

AAAGCTGGAC TAACATGATT TTCTTCCGCA AACAGTTTGC GGACCTTATC AAAAATATTA 7320

AGATAGGTCA TAGGTGTGGA TCGTAAATTT GCGGTAATGC TTTTTTGTGA AAGATGAATA 7380

ATTTCTTGGT TATCAGCCTG GGCTTTTTGA TAAATTTCTT CCGCTAATGA TGATTTACCT 7440

GAACCAGCGA CCCCACAAAT AACAGTCAAA ACTCCTAACG GAACTTCGAC AGATAGATTG 7500

TTTAAATTAT GCAGTGTCGC ATGTTCGATT GATAAAGACT TCCTTGCTTT TCTTGGTTGG 7560

TCGTTTAAAG GGAGCGGCTC TTGAAGCGCT TGACTAGTCA AGGTTTTGGA GACTAAAAAT 7620

GCGTCATACG TTCCTGAAAA CTGGACATGG CCACCATTTT CGCCTGCGAA AGGTCCGATA 7680

TCGATGATAA AATCAGCTTC TCTAATTAAT TGTGGATTAT GTTCCACGAG AACCACGGTG 7740

TTTCCTTTAT TTTTTAAATT GAGCAATGCA CGACTGATCC GTTCGATGTC CTTTGGATGT 7800

AAGCCAGCAC TTGGTTCATC TAAAATATAC ATAATATCAT TTAAGGCGCT GTTTACATAC 7860

TTAGCAATTT TAATCCGCTG TGCTTCACCA CCAGAAAGTG TTTCAGTTGC TCGTCCTAAT 7920

GTGAGGTAGG AGAGACCGAT ATTAATTnAT GCTTCTAGAC GTACAAGTAG CTCTTCTTGA 7980

ATAGTTTTTA TTAAGGATAG GTCCATTGAA CGAATAAAAG AATGTAATTC AGTGAGTGGC 8040

ATGTCAACAG CATCAGCAAT ACTTTTCTGA TTAATTTTGC AGCTACGAAC ACGTTCATTG 8100

ACTCTACTTC CTAAACAATC AGGACATCTT TTTACGGTAA CAAAGTGATT AAGGTATTTT 8160

TGATGACGTT TGCCTTCGTC TGTATGCAAT ATGCTACGTT GCATACGCGG GACGATTCCT 8220

TCATACAAAG CTGTATGAGG CCACTCTTTG GGTGGATTAG CTAGTTTTTG TTGTGGAGCa 8280

TATAArmrTA ATGCTAACTC TTCAGGAGmw TAGTCCTTGA TTTTwkTATC CAAATCAAAT 8340

AGACCACTAT GGGCATAACG CTTCCACCGC CAGTTGC 8377

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1102 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GTTTTTAAAC ACGAACGTTA AACGGCCGAA TGTTCGGTTT TTACGTTATT TTTGTAAATA 60

ACTAGGAGGC AGTAAGTATG AAAGTTTTTG ATTATGAAGA TGTCCAATTA ATTCCCAATA 120

AATGTATCGT GAATAGTCGT TCTGAATGTG ATACTACAGT AACTTTAGGA AAACATTCGT 180

TtAAAATGCC TGTCkTcCmG CGAACmTGCm AACatTATgr TGrAACgATT GCGGAAcCCT 240

TGCTGAAAtG GTTATTTTTA TATCATGCAT CGGTTTGGAT GAAGAGGCGC GAGTGCCaTT 300

TATTAAAAAA ATGcAACAAA AGGGCTTGAT CACGTCAATT AGTGTGGGTG TGAAAGAAGG 360

GGAATACGCT TTTGTTGAAA CGTTAGCCCG AGAAGGTTTA GTTCCTGACT ATGTGACAAT 420

TGATATCGCC CATGGTCATT CAAATGCGGT GATTAATATG ATTCAACACT TGAAAAAATC 480

TTTACCAGAA ACGTTCGTTA TTGCTGGCAA TGTTGGAACA CCAGAAGCCG TTCGCGAATT 540

AGAAAACGCT GGAGCCGATG CCACCAAAGT GGGGATTGGA CCAGGAAAAG TTTGTATTAC 600

TAAGATTAAG ACTGGCTTTG GGACAGGTGG CTGGCAGTTA GCGGCTTTAC GTTGGTGTGC 660

CAAAGCTGCT CGGAAACCAA TTATTGCAGA TGGTGGAATT CGGACCCATG GAGACATTGC 720

AAAATCTGTG CGTTTTGGGG CaACTATGGT CATGATTGGT TCaTTATTTG CTGGACATGA 780

AGAATCACCa GgTGrAACaA AAGTGGaAGA TGGTGTTGTT TACAAAGAAT ACTTTGGTAG 840

CGCCTCTGAA TTCCAAAAAG GCGAAAAGAn AAACGTTGAA GGCAAAAAAA TCTGGTTACG 900

CCATAAAGGG AAGTTGGCAG ATACTTTAGT CGAAATGCAA CAAGATTTAC AATCTTCCAT 960

TTCATACGCG GGAGGTCGCG ATTTAGAAGC GATTCGTAAA GTTGATTATG TGATTGTAAA 1020

AnATTCTATT TTTAATGGTG ATACCATTTA ACATTCAACT ATAGACTGAA AAGAAGACCG 1080

CACGTAGnAA GGAAATTAAA AT 1102

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1009 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

nAAGAATACG GTCTGTCTTT AACAACTTCA ATTGATGATA TTTTAGAAGA TAAAGAGATT 60

CAAATCGTTG TGGAATTAAT GGGAACCATT GAACCAGCCA AAACGTATAT CATGAAAGCT 120

TTAGAAAAGG GCAAACATAT TGTGACTGCC AATAAAGATT TACTGGCGCA ACATGGGAGT 180

GAATTAGTAG CGTTAcCCAA AAACATCATT GTGATTTATA TTACGAAGCA AGTGTTGCTG 240

GCGGGATTCC AATTTTAAGA ACAATTGCTA ATAGCTTAGC GGCAGACAAT ATTCAACAAG 300

TTTTAGGGAT TGTGAACGGT ACGACGAATT ATATGCTTAC GCAAATGGTT TCTGCTGATA 360

AAAGTTATGA AGAAGCTTTA GCAGAAGCGC AAgcGTTGGG GTTTGCTGAA GCGGATCCGA 420

CAAATGATGT CGATGGAATT GACGCAGCCT ATAAAATGGT TATTTTAAGT CAATTTGCTT 480

TTGGGATGAA TGTTTCATTA CCACAAGTTG ATATTCGTGG AATTCGCGGA TTGTCTTTAG 540

ATGATGTCGC TATGGCTAAA CAACTGGGGT ATGAAATCAA ATTGATTGGC TCTGCTGAAC 600

AAAATGAGAA TAGTATCTCG GTGGAAGTGG CGCCTATGCT AGTCAATCAA AAGCACCCTA 660

TTGCATCTGT TCGGAATGAG TACAATGCAG TATTTATTAA AAGTGCAGGC GTGGGGGAAT 720

CAATGTATTA CGGTCCAGGA GCTGGAGCTA AACCGACAGC AACCAGTGTG GTCAGTGATT 780

TAATTACGAT TGCTAAAAAC ATTCGCTTGG CAACGACTGG CCATATGTTC AATTCGTATC 840

AACATAAAAC ACAATTGACG AGTTCCGAAA ATGTATTTGG GCAGTACTAT TTTyCATTAG 900

aTGTTCCAGA TACACCAGGA CaATTTTTGC AGTTGACGCA ATTAATGACT AAAGCAGAAG 960

TCAGCTTTGA TCAtTAGTTC AACAGaAATC AGATGGACAG CGAGCAAGA 1009

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5558 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
nCGGnTAAAA TTTTTTCAGC ATCCTTATAT TACCAGTTTG GTCTTTCAAg AcTACTTGcG      60
ATcAGCaCtT TgTAcAgTTA aTTGTGCGCA ATATGCCAaC AaTCCAGaGC TTTtAACTaG     120
TCAaTTATTT GGACATCTTA AAGgCGCTTT TACGGGAGCA GaTTCCGATA AAATTGGGsC     180
CTTTCAAAGT GCAGAAGGGG GCGTTTTATT TTTAGATGAA GTTCATCGTT TGAGTCCTGA     240
AGGGCAGGAA AAATTGTTTA CATTTTTAGA TCAGGGTATT ATTTATCGAA TGGGTGAAAC     300
CAATCGGCCC ATTCCTGTTA CGTGTCGGTT GTGCTTTGCC ACAACTGAAG AAATTTCTAG     360
CACCTTTTTA ACGACATTTT TAAGACGAAT TCCTATTCAA ATTAAAATCC CaAGTCTTGC     420
TGAACGCACT CAAGCGGAAA GAAAACAATT AATTATGCGA GCATTTTATG AGGAACAGCA     480
AGCTATTCAA AAAGCAGTTA CTATCACGCC ACAAGTTATC CAACTTTTAG AAAACCACCA     540
TTATGTAGGC AATGTAGGTG AATTAAGAAA TAATATAAAA ATTATTACTG CACGCAGCTT     600
TGCCGCCAAT CTTGATAAAT GTGTTATTCC AATCACTTTG CACGATCTAC CGAAAGAATT     660
TCTGGATCAA TCAATCAAAC TAGTTCCTGA TGAAAATGAA TTGCCTATTC GTTTAGATGA     720
GCAAACAAAT CTGGTTAGTT TGTTAGAGGA AACAGAGCTT GCGCAArGAA GAATTATCCA     780
GAGTTATGAA CGAATCCTTC GTCTTTATGT CTCTCACCAT CATCATCTTT CCACTGCTAA     840
TAACGATATC TCGAAAGAAA TTGAACGCTT ATTCGATGAT TTATTATTTG AAAAGAAGCG     900
AGAGAAAAAT CACGAAATGC TATTGTTTAT TACGCAAAAT ATCCGTCAAT TATTGGAAAC     960
GATTGAATCG TCTTATCAAA TTCGGTTTAA TGGCAGTTTA GTCTATGCAT TAAGCACCTA    1020
TCTGTTTCAA CGAAGATGCA TTGATTGGTT TCCAGAACAA GAACCTACAA CAGTCATTGA    1080
CGAACTTTTA ACAGAAGTTC AAACCAAATT AGCAACTAGT TATGGtACGC TGAACAACTG    1140
TTAACTTTAG TGAAACGTTC ATTAGATATT GAACTATCCC AGATGGAwCG AATTATTGTC    1200
ACCATTTATC TTCACTATTC GGGTTCAGTA AAGAAAAGTC ACTATCCGAA AGCGGTGATT    1260
GTTGCCCACG GATATGCCAC AGCAAGTAGT ATTGCGAATG TTGCAAATCG GCTGTTAAAT    1320
GTGCCAATTT TTCAATCGTT TGACATGCCG TTAGATGTTA CTCCCArAAA AATTTCCGAA    1380
CACTTAATTC ACTATATGGA ACGTCAAGAA ACGAGGAATG GTTTAGTAAT TCTTTTTGAT    1440
ATGGGTTCAT TGAAAGAwAT TTATCAGTAT TTtCCGGCAG AAGrAGAAGG kCCtTTTTTA    1500
tTAATGraTa ATGtGACmAC mAGcTTGGCT CTGkCTaTwG GAGrAGCCmT TwAAGmCGAA    1560
kTTCTTTTGA AGAATTgCCC CAAAAAGcGT TAmCGGTyCA TCCmAATGaA TGGGAAaTCa    1620
TATTACCTGA AAATAAAACG GAACgTGTCA TTTTAACAAC GTGTAgTACC GGAATTGGCA    1680
CAGCCGTTAA AATTAGAGAT TTACTTGAAA AAAGTTTACC TGCGGAAGCA CAATTAAAGA    1740
TTATTCCTTG TGAATATAAT CAATTACGTA ATGCAGAGTC GATCAAAGAA AGTTTTCCAG    1800
AATATGAGAT TGTCGGCATC ATCGGTACAA ACAACCCGTC TTCAAATGAC CTCCCGTATA    1860
TCTCTTTAGA AGAGTTGATT GCAGGAAAAG GAATTACGAC CCTTTTAGAA TGGACAAAGA    1920
GAGAATTAAC TAAAGATATG CTGTCCTATG TAAATCATGA GTTGATTCGT AaCTTTTCAT    1980
```

TAGATCGGGT CATTCAATCA GTGACGATTT TAGATACAGA AAAAATTATT CGACAAGTAG 2040

AGGTCTTTTT AATTCAATTA GAAGAGCGGT GGCAACAAAC AATTCAAAAT GATCGTAAAT 2100

TAGCGGTATA TGTCCATGTC AGTTGTTTAA TTGAACGATT AATCCGGAAT GAACCTATTG 2160

AAAATTATAA TGGCGCTGAA CAACTAAAGC AGTGTCAAAG AACGGTCTTG CAAGAGTTAA 2220

AAGAGGCGTT TAGTGTCATT GAAAAAGTTT ATAGTGTCAA TATACCTGAA TCGGAGCTTT 2280

TTTATGTCTA TGATGTGTTG TTTGGAAAAA CGGAGTTTAA CAACGCAGAA AGCGATTTTT 2340

AAAATAAAAA GGAGGGAGTG CCACTTTTTT GGCACTCTTT TTGCTATATA AAGAAGTAAG 2400

TGAGGTGAAG AAAATGGAAC GTAAAATTAT TCTGGCAACA CATGGTAATT TCGCTGCAGG 2460

GATTCAGACA TCATTAGCTT TAATTTGTGG TGAAACAAAG AATGTGGAAT GTCTTTGCGC 2520

TTATATGGAA CAACCTTATG ATTTAGCTAA GACGGTCAAA GACATTTTAA AGAACAATCA 2580

ACAAAATGAA TTAATTGTGA TTACAGACCT ATTTGGCGGA AGCGTGAATA ACGAATTCTT 2640

TAATTATGTG GGGAAACAAG CAATGCATTT AATTTCAGGA TTAAATTTAG CAACATTAAT 2700

TGAAATATAT ACCCAAATAA ATACTGTTGA TTCATTAGTT GATTTAGTAA AACGTGCCGT 2760

TGAAAACGGA CAAGAAAGTC TTTGTTATTG TAATGAGTTA TCTAGTAAAG AAATTTTAGA 2820

AGAAGaATTT TAGGAgGAat gAAnTGATTA AATTGTTAAG AGTCGATCAT CGCTTGTTAC 2880

ATGGTCAAGT GGCTTTTTCA TGGACCCAAG GCTTAGGAGC CGATTGTATT TTAATCGCAA 2940

ATGATGACGT ACCTAAAAAT GATATTCGAA AAACAACCAT TAAGCTAGCT AAACCAGCAG 3000

GAGTAAAACT GGTCATTAAA ACGATTGAAG ATGCGATTGC CGCACTAAAA AGTGGTGTTA 3060

CAGAAAAATA CAAACTTTTT ATTGTAGTTG AAAGTGTGGA AGATGCTTAT CGATTGGCTA 3120

CTGCTTATCC AGAAATAAAA GAAATTAACT TAGGAGGAAT GAAAGTAAGG GAAAATACTC 3180

GAAACATTTC TAAAGCAATT AATATCACAC CGGATGAAGA AAAAATGGTT AAAGAATTGG 3240

TCGCAAATGG TTGTGAAGTT GAAATTCGAC AAGTCCCAAA TGACAAAAAA GTAGCAGCAT 3300

TAAATGTCTT GTAGAAAGGA GAAATAAAAA TGATAGGACA AGCGATTCTT TTAGGTTGCA 3360

TCGCTTTTAT TGCGCAGTCG GAATATGCTT TAGGAACCTC GTTGCTTTCA AGACCAATTG 3420

TAACCGGTTT ATTTGTCGGA ATTGTATTAG GTGATGTCAA AACGGGCGTA ATCATGGGCG 3480

CGACCTTAGA GTTAGCTTTC ATCGGTTCTT TTTCAGTGGG CGCATCAATC CCACCTGATG 3540

TAGTGACTGG TGGGATTCTT GGTACAGCGT TTGCTATTAC TGCTGGCGCA GGAACAGAAA 3600

CAGCCTTAGT CTTAGGCTTG CCAATCGCAA CGTTAACGTT GaTTCTTAAA AATATTTATT 3660

TGGGTTTGCT CATCCCAATC ATGAGTCAAA AAGCTGATCA GTATGCTGAA GAAGGAGCAT 3720

ATAAAGGGGT GGAGCGCATG CATTTACTTG CAGGCTTTGG CTTATCGTTC aTGctAGCGC 3780

TCATTGTAAC TATTTCATTT GCCGTGGGCA GCAACACCAT TAAAGGCTTA CTTGATCTGA 3840

TCCCTGAATT TATTCAACAT GGATTGTCTG TTGCAACGGG GATTATTCCA GCGCTAGGAT 3900

TTGCGATGCT TGCACGCTTA CTGATTAACA AACAAGTGGC GCCATATTTC TTTTTAGGTT 3960

TTGTCTTAAT GGCCTACTTT AAGTTACCTG TAACAGGAGT AGCAATTTTG GGTGCCATTA 4020

CAGCAGTAAT TGTTGTGAAT ATCATGAATT ATGCCAAAGG AAAGAATGAA ACAATACAAA 4080

CAACAAGTGG GGAGGTAATT GATGATGACG AAGACTTCTA AAACTAAGAC AGAAAACTTA 4140

ATTACAAAAA AAGAACTTAA TCAAGTCTTT TGGCGTTCAT TTCAAATGGA ATTTTCATGG 4200

AATTATGAAC GTCAGATGAA TATGGCTTAT GTCTATGCCA TGATCCCTAT TTTAAAGAAA 4260

TTATATCCTC AAAAAGAAGA GATGGCGGCA GCTTTAAAAC GGCATCTTGC TTTCTTTAAT 4320

ACTACGCCGC ATATCGTGAC GTTTATTTTA GGAATTAATG CTGCAATGGA AGAAGAAAAT 4380

GTTCTGGATG AAAATTTTGA AGTTGGAACA ATCGATAGTA TTAAAACCTC GTTAATGGGA 4440

CCTTTAGCTG GTTTAGGTGA TTCCTTTTTC TGGGGAACCT TACGCTTGAT TGCAACAGGT 4500

GTTGGAACTT CGTTAGCTTT ACAAGGGAAT GTGTTAGGGC CAATTCTATT TTTACTCATT 4560

TTTAACGTCC CCCATGTGAT TGTTCGTTAT TTACTGAATC GTTGGGGCTA TAAATTAGGA 4620

ACTGGTTTTC TCAAAAAGAT TCAAGCAAAT GGCATGATGG AAAGTCyCAC ATTAGGCGCT 4680

TCmATAATTG GCyTGATrGt GGkArGTGCy ATGaCCGCCT CCATGATTGA TTTAACTATC 4740

CCgATTACAA TCAGCgGAAG TGGGaAAAAT GCaGTCACGG TyCAAAATAT CyTTGACGAT 4800

ATCATGCCTA AATTATTACC GCTTGCATCT TTTGGTTTTG TTTTTTATTT ATTGnAAAAA 4860

GAAGTAAAAC CTTTAGCGAT TTTAGGTGGC ATGGCAATTG TCGGAATTCT TGGCTCATTA 4920

ATTGGTCTAT TTTAGGAGGC AAAATATGAA AACGATGCTG GATTATrTAA ATGAGGAGCA 4980

aGCAGCTCTA rCTCAGaTTC TTAACCAATT TCAaTTAACT GATACAGACG TTTCAAAGGT 5040

CACGCACTGt CTAATTTTAG CAACTGGCTC GTCACATAAC GyTTkTCAGG CGGCTAAArT 5100

TTACCTTGAA GAAGTGGCGC CTATTTATGT GGAAATCCAA GAGCCGTTCA ATTTTGCTCA 5160

TTATGGCAAA GTGGATGATC GGTTTGATCT GGTCGTAGCG GTTTCGCAAA GTGGTAAAAG 5220

TGCTTCAACA ATTGGATGCC ATTGCAAAAA TTAAAAACCA AACGTCTGTA AAAACAGTGG 5280

CCTTGgACAA GTGaTGTCAC CAGCCCGATT ACGGaGGTTG TGGaTGaAGT CCTGGaTTTa 5340

AAAATAGGGc TTGAAACAGT TGGtTTTGTC ACAAAAGGGt ATACGGcAAC ACTTTTaAAT 5400

TTAtTTCtAt TTGGtTTgAA ACTGGGCTAT CAAAAGCATC AATTAACAAA AGTAGAGGTT 5460

GAAGAAGAAT TAGCAAAACT TnACCAAAGC AATTCAGGAA ATCGATTCGG TGATTTATAA 5520

AACCGAACAA TTTTGGGCAA AAAACGGAAA AAAGGAAn 5558

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9047 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

AnATnTGAAT CGTATAGCGA AGCAATTCTC TGATCAGAAT nATTTAACTT CGTTTGGGAA 60

```
AAATTTCTGG CGCAAAACGT ATAACGGAAA ATCACCCACT CAGGAACTAC TGAAGGTTGC    120
AACGGAGGAT TTAGTACAAC ATAAAATGAT TAAACAGCTG GCAACGGAGT TAGAAATAGC    180
CCACTCACAA ACATTTGGGc AAGAAAAAAA AGAATGGCAA GACCAAAAGA GCAGTTTAAC    240
GCTCTGGCAA TTTTTAGATG CAAAAGATCA ACAATTACAG GACCAAATAA AAGAGAAATT    300
AATGGAGAAA GAAAAGCCAA CGCAAAAAGA ATTGCGGcAG GCCTTTGAGC AATTGGATGA    360
TAAATwTAAA AAGACCGATT ACTTTGTTGA AGCAATTGAA ATTCCAAATT TTTCAGGTAA    420
GCAAGCAGAA TTAGAAAAAA TAGCCGAAGC AATCTCTCCG AATCTATCTT ATGAGGAGAC    480
GTTATTGGAG TGGCAAAATA AATTACCTAA TCTTGTTATT GAATCGTATC AATTGAAGAG    540
TGCAGAAATC CAAAAAGAGG ATATTTATTC GTTGTCTGTT GGGGAGATTC TTAGTGAAAA    600
AGCTGTTGGT ACCGTGGTGA AAGGATACCA CGAAAACCAA TACTTTTATA TTTTTAATAA    660
GGAAGGAGGG CAACTTTTAC AGTTTGAAGA AGCACCACAA TTTGGCAAAA ATGCCTATAT    720
CAACACACGC TACAAAGAGA AaTTAGCTAC GTATCmACmA GCAACCAAAG TAGACTTGTT    780
GGAGAAAGAC AGAGAAAAGT TTTTTCAAAA TTATCAAAAT AAAAACAGTT AAGGAGCAGA    840
AAATGAGAAA AAAAGCGTAT TTTAAAAGTG TAACCATCTT CCTTTCTATT TTGCTTGTTT    900
TTTCCCAGCT TCAATTATTT AGTCTTCCTG CTTATGGGGA AACAGTGAGT GATCAACCGT    960
TACTTTTTCG TTCAGTAGGT GTCGCTCAAA GTGGGACGAC CTACTATGTG GATGGGGAAG   1020
GAGGAAATAA TGCCAATGAT GGGCAGTCGC CAGCCAGTGC TTGGCGTGAT TTTGAAAAAG   1080
TGAATCAGAC GGAGTTTCAG CCAGGCGACC ATGTTCTGCT GAATGCGCAA AGTACCTGGA   1140
ATAACCAATT GTTGCATCCA AAAGGGAATG GGACAGCGGC ACAAAAGATT GTCATTGATT   1200
TTTATGATAC GAATGACAAA GGAGAAACCA TTTTTGAAAC AACACGGCGT CCGATTATTA   1260
ACGGTGGTGG CACCTACAGC ACAGGGACTT TCAAACGTGC AATCTCAGGG GCCGTGCAGT   1320
TAGTGAACCA AGAATATTGG GATATTTCAA ATTTAGAAGT GACAAATACA CCAGAGCTAG   1380
ATAATCTCGA AGGCTATAAG AAACCAGGAG ATGCTCAAAG AGCGGGGATA TTAGTACTAG   1440
GATATGAACA AAATCGAACG TTCAATAGTG TCACGATAAG GAATAACTAT GTTCATGATG   1500
TTCAAACAGA ATATTATTTA AATCTAAGTG GGAATACTGC AACCAAACGA TTAAAAGCGG   1560
TAGGTGGAAT TATTGTTTTA GGAAGTTGGT TTGATGAGAA TGGCAATGTA GTAACGGCAG   1620
CTAATGATCA TCGAACAACG ACTGGATTTA ATGATATTTT GATTGAAAAT AATGTGATTC   1680
AACGAGTGGG ACTTGAAGGA ATTCGAACAA AAGCTGATTC AGATACCtCA AGAGGCAACA   1740
CTTTCTACAA AACGTTTTCT AATATTACGA TTCGTAATAA CTATTTAGAA GATATTGCTG   1800
GCGACGGGAT TGTCTTGTCA GAAGCCAAAA GTGGCGGTGT CGTCGAAGGA AATGTTGCTG   1860
TCCGAATGTG CAATGCCGAT TATGGTACAC AAAATTACGC GGGCGTTTGG GCAATGTCTG   1920
TAGATGACGG ACTGTTCCAA TATAACGAAg TATACGGAAT CAAATATGGA TTTAACGATG   1980
CTGAAGCTTA TGATGTGGAC ATGCAGTCAA ACAATGTTAT TTATCAATAT AACTATTCAC   2040
```

```
ATCATAATAC TGGCGGCTTT TTATTGCTTA TGTCTGATCA AAAAAATTCG GTTATTCGGT 2100
ACAATATCTC AGCTAATGAC GGCGGTGGTA ACCGAGGAAC AGGTAAAGAC AATCCAGGCG 2160
GAGCTGGTGG CTACAATTAT AAAGAACAAA GCATTTTCCA TTATTGGGTG AAAAATGATG 2220
GAGCGGCAAT GCCAACTATC CACAATAATA CGATTTATGT TGGCGACGGT ATTTCAACAT 2280
CATTATTTGG TGAAGGAAaT AGTTCTGACA ACAGTGGTAC AGTAGCCAAT TTTTACAATA 2340
ATATTTTATA CAAAGAAGGT ACCGGACAAT TAAAATTCTT ATCTAATTAT CCAACAAATG 2400
GGACACAACC AATTGAACGT AAAATGGTCG ATAATCCAGA AAAATATTTC AAAAATAATG 2460
TTATTTGGCC AAAAGAAATT GCTACAGAAA AAAGTGGTGC AACTGTTGAA AAATTAGTAT 2520
CAAGTGGAAA CATTTTTGAA AAACCACAAT TAGAAATCAC AGATAACCCT GAAAAAGTGA 2580
AAGAGTTGGC AGAGCAAGAA TTTACGACAT TGAAACCAAC AAAAGATAAT GTGGTTGAAT 2640
TTACTTCTAA AGAACGCTTA CGTCAGCGAG CACAAATGTT CCGATTAAAA GAAAACTCTC 2700
CAGCGATTGG TAAAGGCTTG TCGGAAGTGA ACAGTCCGGC CGAAGACTTC TTTGGAAATA 2760
GCTTGAAAAA TAAAGTTTTA GATATTGGTG CGCAACAAGC ATCCACAATT GAAAAAAGTA 2820
TTCGTTATCA AAATCAAGTG TTAGAAATTA GTTCGGCAAC AGGTGTGTAT CCTAATTTGC 2880
CAGAGCAAGT TGAATTAACG TACGAAGAAG TCGTTAATGA AGAAGTGGTT GCGACAGGCA 2940
AAAAAGAATT TAAGTTACAA TGGGAAGCAA TCCCTCAAGA AAAAATTAAT ACAGCAGGCA 3000
CAGTTGAAGT TGCAGCAACA GTTATCGGCT TACCGATAGA TGCTGTCAAA GTAACAGCCA 3060
AGGTCTCTTT TGAGGGTGAA CTTGGTGAAG GAAAAGACAC CGTTAAACTA AAAACAGCTC 3120
AAACTGCCTA TGTTCAAAAG AGTGATGGTA ACCGAGCATA CTCTGCTATA GCAGGCGGTA 3180
CCGCTGCCAT TAGTTCaGGA GACGCCTATA AATATCCTTA TGGAGTAAAC TATACGGGCA 3240
ATTACGCATT -AAAATTAAAA AATGCTAGTT CTGCGGGGTA TAATCGCCGG ATTTACGTTG 3300
AAATTGATAC ACAAGAATTG AAGAACTATC AATCACTAAA AAGTGCCAAT TTAGAACTAA 3360
ATGTGATGCG TTATGATGCG TGGAATGGTG CTGGAAATAC AAATGATGAA CGTTTGAAAA 3420
ATACGCAGTT CCAAGTAGAT GTTTATGGAa CAGATACTAA TTGGATGAGT AACACCATTA 3480
CTTGGAACAA TGGTCCAAAT AATTTAAATG TACCAAATGA AGAATTTATT GCTCGACAAT 3540
CTTTTACAAA TAGCTCAATT ATGAATAATC AAAATACAAT CTCTATTGAT ATTAGTAACT 3600
ATCTACGGAA GCTAATTCAA AGTGGTGAGA AAATTCCAGC CAAGTTAAGT TTCTTATTAG 3660
CGATTACAGA CAGTCGTTTG CCTGGTTATG ATTCTGATAA TGCTGGGTTT GATGCTTTTT 3720
CAAAAGAAGG CGCACAAAAG GCCTATCAAG ACTTTTTAAC AGGCAAATTG ACCTTGCCAA 3780
CTGGTCAACA GCTAGCAGAG GATAGTTTAG CTCCTAAAAT TGTTTTATCG AATGTTTTCG 3840
AAGTCAAACA TGAGTCGATT GAGGTCACGA CAGAAGCGGG TCAAGCCCCT AAACTACCAG 3900
AAAAAACAAC TATTTTTTAT TCGGATGGTT CACAACGAGA AGTGACTGTC AATTGGTCAG 3960
AAGTTCCAGC ATCTTCTTAT CAAAAAGAAG GAATATTCAC TGTTGTAGGA CGAGCAGCAG 4020
```

GTGTCTCAAT GCCAATTATT GCCAATGTAA AAGTCACCGC TAAACATATT GTTGGATTCA 4080

AGGAATTGCC TGCGCTGGAT CGACTAACAG GCACCTCGCG AGGCGAACTG AATTTGCCAA 4140

CTGAAGTAAT TGCAAAATTA GATGATGGTA GCGAAACGAA ATTGAAGGTT ATTTCTTGGG 4200

ATGACGATGT TAGTAATTAT TCGCCAAGTA GCCCGCCAGG TACTTACCAG TTCCCTGCTG 4260

CAGTAGAAGA AAAAATAGGA ATTGCGAATC CTGATGAACG AAAAATCTTC CAAGTCGTGC 4320

AAACTCATGC GATTCCTGAA AGAATTCAAT TTGCTACAGA AACAGCCACT ATAAAATCAG 4380

GCGAAAATTA TCAAATTCAA TCTAAAGTGA TTGGTCAAGC GCCGCACACA GAAACGGATG 4440

CTTGGAGCTC GCAAGTAACG TATGAGTTGG TTACACCAGA TGCTGGCAAT ACAGTAAGTG 4500

TAGATGAAAA CGGTCTGATT AGGACAGAAG CAACGACTGC TGCAGGAAAT TATCAAATCA 4560

AAGTAACAAG CAAAGTTTTG CCGATAGTGA CTGCACAATT TTCAATTAAA GTCACAAAAT 4620

AAACGAACTG TTTGAAGAGA CTGGAACAAG AATTGCTGTA CCTTGCAATC TTTCTAGTCT 4680

CTTCGATTAT CTAGAATCAA TAAATATTCT ATTGCCTACT ATTCATTCTT TTTAAAACAA 4740

GGTAAAATAA TAAATGAAAG TAAACGATAA AGGTGTGTAG ACAATGAATA CTATCTTAAC 4800

AGGTGATCGT CCAACTGGTA AATTGCACCT AGGACATTAC GTAGGTTCCT TAAAAAAACG 4860

AGTAGAAATG CAAGCAGACC CTACAAATCA ATTATTTGTC ATGATTGCAG ATTTACAAGC 4920

ATTGACAGAT AATGCAAGAA ATCCCGAAAA AGTTTCAGCC AATGTTTTAG AAGTTGCTTT 4980

GGACTATTTA GCAGTTGGTT TAGACCCTAC GAAAACAACA ATTTTTATCC AATCGCAAAT 5040

TCCACAGTTA GCTGAATTAA CAATGTACTA TCTAAATTTA GTGACCACGT CACGTGTTCG 5100

TCGGAATCCA ACGGTCAAAG CAGAAATTGA GCAAAAAAAA TTTGGGGAAG GTGTTCCGAC 5160

AGGATTCTTT ATATATCCCG TTTCACAAGC AGCTGATATT ACTGCCTTTC AAGCGAATTT 5220

AGTCCCGGTT GGGGAAGACC AAAAACCAAT GTTGGAACAA GCGCAAGAAA TTGTGCATAG 5280

TTTTAATCAG ACGTACGGTG AAGTTTtAGT TAGACCGGAA GCTGTGTTAC CACCAAAAGG 5340

CATGGGACGC TTGCCAGGAA TTGATGGCAA CGGAAAAATG AGTAAGTCTT TAGGCAATGG 5400

TATTTATCTT TCAGATCCAG CTGAAGTGGT ACAGAAAAAA GTAATGAGTA TGTATACCGA 5460

TCCAAATCAT ATTCGTATAG AGGACCCAGG ACAAGTTGAA GGAAATATGG TCTTTACTTA 5520

CTTGGATGTT TTTGGGAAAA ATAAAGAATA TATTGAAGAA TTAAAAGAGC ACTATCGCCA 5580

TGGCGGTTTA GGAGATGTGA AAATTAAACG TTACTTAATT GATGTCCTAG AAGAAGAACT 5640

CGCACCGATT CGTCGACGTC GTGAAGAGCT GGCTAAAAAT CCAGAAGCGA TTATGGAAAT 5700

GTTGCATAAA GGAAGTCTTG CTGCTGAAAA AGTGGCAGCA CAAACCTTAA CAGAAGTCAA 5760

AAAAGCAATG GGAATTCAAT ATTTCTAAGC AATAAAAGAA CTGGAACGAG TCAAAAGAAG 5820

AATGACTTGT TCCAGTTCTT TTTAAGTGGT TTGAAATTGA CTATTATATA AGCTGGCGTA 5880

AAGAGTCGGA GCTTGTAATA GTTCCTGATG CGTTCCTTTT TCAATAATAT CGCCATTTTT 5940

CATCACTAAA ATTAAATCAG CATTTTCAAT TGTGGATAAT CGATGAGCGA TAACGAAGCT 6000

TGTTCGATTT TCTGTGACAG TTTCCATCGC TTTTTGAATA TGGGCTTCTG TTCGTGTGTC 6060

CACACTCGAA GTTGCTTCAT CGAGAATAAC AACGGGCGGA TTTGCTAAAA TGATTCGGGC 6120

GATGGTTAAT AACTGTTGTT GCCCTTGTGA TAATGCACCA TTTTCGCTAG AAATAATTGT 6180

GTCATATCCT TGAGGAAGGG TCCGAATAAA ATGATCACAT TGAGCAATTT TAGCTGCTTC 6240

AATTATTTCT TCACGAGAAG CATCTTTTTT TCCATAGGCA ATATTATCTG CTACGGTTCC 6300

TTCAAATAGC CAAGTGTTTT GTAAAACCAT GCCAAATAAA TTTCTTAGAT TTTGTCGAGA 6360

AAGTTTCGTA ATATCAATCC CATCAAAAGT AATGGCACCT TGATTTATTT CATAGAAACG 6420

CATTAATAAG TTGACTAATG TTGTTTTACC AGCGCCGGTG GGTCCCACAA TGGCCACTGT 6480

TTTTTTCGGT TGAACAGAAA AATCAACATT CTTCATTAAA ATTTTTTCTG GTGTGTAGCC 6540

AAATTGAACA TTTTTAAATT CAATGGCTCC TTTAGGAGAA GAAATAGTTT CTAAATGAGT 6600

TGCTTCTGGC TGTTCATCAG CTTCATCTAA GATAACAAAA ATCCGATCAA TGGAAGCCAT 6660

CGCTGCTTGA ATTGAGTTAA TGACATAAGA AGCGGTAGAA ATCGGCTCAG AAATTTGGTT 6720

GATATATTGC AAATACGCTT GCAAGAAACC AATCGTAATA CCACCAGATA AAACGAGCAT 6780

TGCGCCTAAG ATAGCACTAA TAATAAAAGC CAACTGATTA ATAAAACGAA TAGCTGGATA 6840

AATCGCAAAA TTTAGAAACT GTGCTTTTTT AAAGGcACGA TAATGTTGTT GATTAACAGC 6900

ATCAATTGTT TTTTCAGCAT TTTGCTGTTG ATTAAAGGTT TTTGTGACCA AATTTCCTGC 6960

TAAATATTCT TCCATCTTAT TATTTAATTG ACCTAATTCA GCTTGACTTT GATCAGCAAA 7020

CACCTTGTTT TTATTCGCGA TTTTTGTCGT CATGAAAGTA CTGCCGCCAA TTAGGAGCAA 7080

CACTAAAATG GTTAATTTTG CATCAATATA GAACAACATG ATTCCTGCAA AAAGGATAGT 7140

CACGACGGAA GTGAAAAATT GGTTGATGCC TGTTAAAAGA ACTTGTGACA ATTGGTTTAA 7200

GCCAGTTGTT GAACGACTAA TGATATCGCC CACTTGGTGA TTGTCAAAGA AGGCCATTGG 7260

TAAAGTTTTA AACTTTTTTG TCACTTCTTT TCTAATTCTT AAAGTAACCC GTTCGCTTAA 7320

AGAAGCCATT GCTCGTTCCT GAATAAAAGA AGTGATACTA CTAATGATTG AAAAGAGAAT 7380

TAAAAGTAAG ACGGGACCTA AGAGCGCTTC TTCAACTAAA GGAAGTGTCA TACCTTTGAG 7440

CCCGACACGC TTGATAGCTT CTAAGAGGTT ATCAATCCCG ATTGCCATAA TATAGGTCAT 7500

GGCAACAATT AATAAGTTTC CAATTAAACT GCAGATCATT AGACCATAAA AGATTGGATG 7560

CTCAGGTTTA ATCATTTTCC AGAAACGTTT AAAGCTTTGA AGATTCATTG TTTTTTTCAT 7620

TGTATGGCTT GAAAAGAAAA CTTTCTTTCA AGCAGTCACA TCCTTTCTTT AATTACTTGG 7680

GTAAGATACC TTGCGATTTA GCAAAGTCTT GGTAATAGCT ATTAGTGGTA AGTAAATCAG 7740

CGTGGGTGCC TTGACCAACA ATCCTTCCTT CATCTAGGAC GATAATGTTG TCAGCGTTCA 7800

TGATTGTACT TAACCGTTGA GCAACAATGA GTAAAGTTTT GTCCGACATT TGTGCATGTA 7860

AAGCGGCACG TAGAGCGGCA TCAGTTTTGT AATCTAATGC GGAAAAGCTA TCATCGAAAA 7920

TATAAACGTC TGCCGGTTTG ATTAAGGCTC GTGCAATACA CATTCTTTGT TTTTGACCAC 7980

CAGAATAGTT GGACCCGCCT TGTGCTACGA AACTTTCAAT CCCTTGTGGT AAGGAATCGA 8040

TAAATTCAGA AGATTGTGAA ATTTCTAGTG CCGTTCTTAT TTCTTCTGTA GTTGCTTTGG 8100

CATTTCCCAT TAATAAGTTT GATAAGATTG TCCCACTGAA AAGAAAGGCT TTTTGTGGCA 8160

CATAACTGAT GACTTGGCGA ATCGTTTGCT GAGATAATGA GCGGATATCT GTGCCAGAAT 8220

AGCTAATCGT GCCGGCTGTG ACCTCATTTA TTCGTAAAAG TAACTTAACT AAAGTACTTT 8280

TACCAGCGCC AGTTGCACCA ACAATCGCCG TTGTTTTTCC TTTAGGAATG ACAAAACTAA 8340

CATTTTCTAA AACAGGATCA GCTGTTTCTG TGTATTGAAA CGTGACGTGA TCGAAGCAAA 8400

TCAAGGCTTT CTCAGGATGA TAAGCAATGG TTGCCTCAGT ATGAGGATCG CTAATTTCAA 8460

TTTCTTCTGA CAAAACTTCT TGCAAGCGAT TCAAGGAAGC AACGGACCGT GGAATGACAA 8520

CTAATACCAT AGCCGCGATC ATTAAGTAGG CCAAAGTTAG AAGCGTATAT TCAATCACAG 8580

CTGTAACGGT GCCAATTTGT AGGGTGCCTA CTGCAACTAA ATTACCGCCA AACCACAAAA 8640

TAGAGGCGTA CACAACGCCC ATTAATAAAA AGGCAACGGG AGTAATCCAg CAAAGATTTG 8700

ATTCACTTTA CTCATGCGTT CAGCATAATT TTTAAATGTT TGATTGGTTC GCTGCTCTTC 8760

AAAATCTTGA TTATTGAACG CACGAATCAT ATTAATTCCA GTAAAGAACT CTCGTAAAGT 8820

CGTAGTAATT CGATCCATTT TTGGTTGAAT TTTTAAAGAT AAAGGGGTAC CTTTTTTCAT 8880

TAAAAAATAA ACAACGACTA AGTAAAATAC GATGGAAGCC AATGTTACCC AAGCGAGTGA 8940

AACTGAATAA GAAAATGTCA TCACTAAGGC AAATAAACTA ATAATAGGCG CTGGAAAGAT 9000

TAATTGGCAA AACAGCACAA TCATTGTTTG AACATTATCT ACATCGT 9047

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1751 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TTGTAATTTT ATTGGGTTAA CGTTATTTTT TGGTATCGGG ATCTTAGTTG CGAATGGCGC 60

CAAAATTGTA GCGAATAAAG GGAAAAAGAG TAAGGGgACC ATTCTATTAA TGGTGATGGA 120

ATTGCTCTTG tTAGGTATCT TTTTATTTGT GaCTTTGCCa AGTAAAAATC TTTCaACCGT 180

GTTATGGATT AATCTGATTG GCGCAGTGAT TGCTAGTGGG GTGTTATTAA GTACACGCTT 240

AGCTAGTAAA GTCGAAGTTC CTGGTATCCA AAAGGGCCGC AAAGAAGTAG GGAAGAATAC 300

CAACAAAACG GCTAACAAAG ACTGGTCTGG CAAGATTTTG GGTGTTGCCG TTTTGGCAAT 360

GATTTTGTTG ACTATTGTCA GTAGTATTAC CAGAATCTCT TCTATTGATG AAATTTATCA 420

AACGATTCCT TTAAAAACGG AAGAAAAAGC TGAAGTTTTA ACTTCGGCGA AAGAAACACC 480

CaTTGCGATT GCACCACAAA CGGCGAAACG AAAAATGTTA CAGAAATTTT CTGtCATTCC 540

CAATTcCAAT ATGTTTACTT TGGaTGGtAT AACGGCACAA GTTGTCAATG GCGAGTATGT 600

TTATGTAGCm ACAGTCGAAT TTAATGGCTy CTTTAAATGG CTAAAACtGG GGGAAgTACC 660

TGGCTATTTC ATCATTAGTG CGACAGATAT CAACGCACAG CCAGAATTTG TAGAAAAACC 720

AATTGTTTAT ACTCCTTCAG CTTACTTTGG CAAAGATGCC GCACGAAAAA TTTACTCCGC 780

TTATCCAGGT TATGCAGCAA CGGGCACAAT CAATTTAGAG CTGGACGACC AAGGGAATCC 840

TTACTATATT CAAACGCTTT ACAAAGAGTA TGGGGTTTCT GGTCGCATGC ACTATAATGA 900

ATTTAAAACG GCAGTATTGA ATGCTACAAC AGGAGAAGTT AACGTTTATG ACAGTCAAAA 960

AGCCCCAAGT TTCGTTGATG CGCCGATTAC TAGCTCAGCG GCGAATAGCT TAAATGAATT 1020

CTTTGGTCGT TACAGTCAAG GCTGGTGGAA CCAGACaATG TTTGGTGCCA AAAAAGATGT 1080

CAAAATTCCA ACTGAAAATG GTATTTATGC ATCAGGACAA ATTACGCCGA TGATGAACAA 1140

GGAAGGGAAC CAGTTGCTTT ATTTTACGGA CTTTACAAGT AGTGAAAAAG ACCAAGATTC 1200

CGCTTTAGGT TATTCCTTAA TCAATGCCCG AACAGGCCAA GTTACGTATT ATCGTGATAC 1260

TAAAGTCGGC ATTATGGATA GTGATGGCGC CATTTCTATC GCTTCAAAAA TTTATCCAGA 1320

AAAGAAATGG AAAGCTAGTA TGCCAGTCCT TTATAACATT GATGGCGTGC CAACGTGGAT 1380

TGTTTCGTTA ATGGACAGTA AAGGGATTTT CAAAAAATAT GTCTATATAA ACGCCGTAGA 1440

TAATGATATT GTAATTGATG CAGACACGGC ACAAGGCGCA TTAGACGCTT ATCGAATTGA 1500

GTTAGTAACA AAAGGAAGCA ACAATGAAAA TACAGACAAA GCCAACTTGA AAACGATGCA 1560

AGGTAAAGTA GCTCGTGTCT CAGTAGTAGC TGGGGAAGCG CAAACGGTAG TTTCATTTTT 1620

ACTAGAAGAT GAAAAAaCCA TTTTTAGCGT AACGaCTAaT AACAGTcCcT GGCGCTATTC 1680

TTAAAaGAGG GCGaCCAAGT GACaTTTAAA GCCGTAGtGA CTGAAAATGC TAAAGCAGCG 1740

AATGTTGAAG A 1751

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3928 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CGTTTATTTT ATCAAACCAA AATGGATTTG TAAAACATTT CACATGTTAT TTAATAAACG 60

ACTATTACTA GAAAAAAGAC AAGCCAAGTT TCCTTTATTT AGGAAACTTG GCTTGTCTTT 120

TTATTCTTTT TACTTCGTTA TTTTTTGATA GTTTGCGAAT AATTGATTCG CATCTCCTTC 180

TCCATCAATT TTCTTGAGAT GGCCGTTATC TATTTTAAAC ATAGATGGCA CATGTAAAAC 240

TTTATTTTCT CGGATAAATG TTGACCATTC CTTCATGGTC TCtTTGTTGG TTAATTCTAC 300

ATAGTAAACC TGTTTAGGTG CTTGTAATAG TTTCTCGAAA TCCTGACAAT CTGAACAGTT 360

TtTCTTTTTT AAAACAACAT AAAATGTTTC GCCAGAGTGT TTTTTTTCTT GATACTCATT 420

```
GAAACTTAAT TCCTTTTTAT GCGGAATAAA TGTAAAACCT ACCAACAAAC TAAGTAACAA      480
AATTAACAGA CTGAACGAAA TTTTTTTATT CATGTTTTTT CTTCCTTTTT AGTTTGTCTA      540
ATTTTTTCTT TTTTTATTTT TTGTTATGCC AAAGCTAATA GCCACTATAA GGACAATTAA      600
ACCTGCGATC ATGTAAAGTG GATTTACTGT TGTATTAGTC ATTGGTAAAT GAATTCCTTT      660
ACTTGCTGaT TTTCCAAAAT TGaCTGGGTT CGTTGGTGCT TGATGATTTG TAGGCGTCAA      720
CGGTTTTTTC GGTTTTAGCG GTTGTTTAGG TTCAGGTTTT TTTGGGTCTT CTGGCGTTGG      780
TGCAGGCGGT GTTACGGTTG GCTTGTTGGA ATGCAACACG TCACCTTCGT TGCCAAAGTT      840
TAAGTCGGCT TGGTTGGGAA TCCCGCCTTG TTCAATATAA GGCGCTAATT CTTCATCCGT      900
TGCGTCAGTT TTAATTTTAG TGGTAATAGT CATTGTGTAT GTATGACCAG CTAAGTAAGA      960
GTAGCTGTCA TCTTTTTTGT TCATAGTAAA AGTTACTTTG TTATTTTCTT GTGTTACTAT     1020
GCCATTATCT GTAACATCTT TGCCATTTTC mTCAGwAACT TTCACGTCTG TGATGTCTAA     1080
CACTTTATTA ATGTCATCCA CCATGCTGGC TtGGGTCCAT GTGCTTGTTT CGTTACCGAA     1140
AGCTGTTTtG ACGTTCCATT TAAATTCTTG GTCACGGTTG GTTAAATCTA AATGTTCTTG     1200
GCCTTCGATA TCTTTTGTAA TCGTTGGATC TTCTGGcGTT GGTGCAGGTG GTGTTACGGT     1260
TGGTTTGTTG GAATGCAACA CGTCACCTTC GTTGCCAAAG TTCAAGTCGG CTTGGTTGGG     1320
AATGCCACCT TGTTCAATAT AAGGTGCTAA TTCTTCGTCC GTTGCGCTAg CTTTGATTTT     1380
AGTAGTAATG GTCATTGTGT ACGTATGACC ACTTAAATAG TCATAGCTGT CkGCtTGkTT     1440
GtTCATTTCA AAAGTTACTT TGTTATTTTC TTGTGTTACT TTGCCGTTAG CTGTAACATC     1500
TTTACCATTT TCATCTGTGA CTTTTACATC AGTGATGTCT AACACTTTAT TAATGTCATC     1560
TACCATGCTG GCTTGGGTCC AAGTGCTTGT TTCGTTACCG AAAGCTGTTT TGACGTTCCA     1620
TTTAAATTCT TGATCGCGGT TGGTTAAATC TAAATGTTCT TGTCCTTCTA CGTCTTTAGC     1680
AATATTTGGA TCAACTGGCG GTGGTGTTAC GGTTGGTTTG TTGGAATGTA ACACGTCACC     1740
TTCATTGCCA AAGTTTAAGT CGGCTTGGTT GGGAATCCCG CCTTGTTCAA TGTAAGGCGC     1800
TAATTCTTCG TCCGTTGCGT CAGTTTTAAT TTTAGTGGTG ATAGTCATTG TATACGTATG     1860
ACCACTTAAA TAGTCATAGC TGTCTGCTTG TTTGTTCATT TCAAAAGTTA CTTTGTTATT     1920
TTCTTGTGTT ACTGTGCCGT TAGCTGTAAC ATCTTTACCA TTTTCGTCGG TGACTTTCAC     1980
ATCAATGATA TCTAGCACTT TATTAATGTC ATCCACCATG CTGGCTTGGG TCCAAGTACT     2040
GGTTTCGTTG CCAAAGGCTG TTTTGACATG CCAATCGAAA CTATCTTCAC GATTGGTTAA     2100
ATCTAAGTGT TCTTGATTTT CGATATCTTT TGTAATCGTT GGATCTTCTG GCGTCACGTC     2160
ACACGCATGC GCATTGTCAT TTTTAGGAAT CGTCGCTTCT GGAAATTCCT TAGCAATTGT     2220
TCCTTGATTC ATTGTCCCAC TACTTACAAG GGTTGCTGCA TCAATCGCTG TATTTTCTTT     2280
GACTTCGTAG TGAATCGTGA TGTTTCCTAC GTAACCTTCT GTACTAGTTG CTGAAATGGT     2340
TTGTCCGTTA ATTTGAACAG GCACATCATT TCCAGCATCG TCCGTAGCGG TCGCAGATTG     2400
```

AATATCAAAT TGATTGGCAA TCGTTAATGA AGCTGTTGCT GGTGTCGATT GCGCCAGACG 2460

ATCTTTGACA ATTTGTTTTA ATTGGGTTGT AAAATCATCA ATAGATTGGC TTGTAATAAA 2520

ATCTTCTGGT GTGCTAGACC CTTGTTGCAA CTCTTGTTTT ACAAAAGGAC CCACTTCTGT 2580

TTTATATTTA TCAAAGTATG AATTCACTGA ACTTAAAGAT TCAACACTTT CCCAATACGC 2640

ATTAATCATT TCATAGCCTT GGTTAGTAAT TTCTTGGTTT AACGCTAAAA CTTCTGCTGC 2700

TGCACCTTGG TAGTCATTAC TATATTCCAC TGAGACTTGA AGAGGATGTC TTGGATCTGG 2760

ATATTCATTG ATTGAATCAT TGGTATTGGT CTTATGCAAG TAACCATCTA AACGTGTATT 2820

AGCGACCCCA TCTGTCACTA ATAGGAAATA CGTTTTTCGA TTCGTTAAAT CTCCGTGTGT 2880

TTGATTGTAC GTATCTAAAG CGAGTTTCAA TCCTGGGGCG GTTGGCGTAC CACCATACGT 2940

CCGAACGTCT CCAAAACCAG AGACAAATTG GCTTTTATCA TAAGTCAATT GCGTATTGAC 3000

GCGCACATTC ATATCATAAT CAGCTGAATT AATTTTTGTC TTTCCATCAG GAAACATAAA 3060

TTGTTTTCCG CCGCGATATG AAGCCAGCAT CACGCGGTCT TGATCAGATA AGCCTTGAAC 3120

CACTTCATCA ATCGCTTGTC TTACATGTGG AAAATTATCT GAAAAACTAC CACTAGCATC 3180

TTCAACAACT ACTAAATCGA CTGGTTCCGT TTCACCCGCT TGAACTGGGA AAGTAGCTTG 3240

TCCTTCCAAT ATCCGTTTAC AATCAGCCAG ACTGTCGTCT TTCACTTTCA CCGTATTGGT 3300

CATTCCTTCT GTATCTCCTG CTTTAACCGC TTCTTCAGCT AAAGCCAATG TGGGACTGAA 3360

ATTTCCCACA ACTGTTCCAG CGACAATCAC CATTGTCGAT AATAATTTCG TTGCTTTCTT 3420

AAACATTTTA ATCCTCCCTA TTTTCTTTCG CTAGTGTTTG GGACACTTGC TTCTCTTTAC 3480

TTCAACTACA GCTCCTTAAA TCTTTCATCA AGTCGATGTG TCTCCTTGTT TTTGTGACAC 3540

ATCTCTTACT CAATTTCTGA TTCTATCACC TTTAAAGGCT AAAACTTGGC AGTATTCGGC 3600

TCTTGAAAAC GTAAAATTTG TCTTTTAAAA ATTAATGAAA AGACTATCCA CTGACAAAGT 3660

TGGTTTTCAC TTACTTTTTG CTACTTGATT CCTTTTAAGA CTCTTTTTTT ACACAAAAAA 3720

AGTAAAAGAA CCTTTACAGT CCTTTTACTT AATTGTTAGA TGCTATTCAT AATTTTCTTA 3780

GTTTTGTTCT ATAAGCTGAA TAGTTTAGCC TGTTTATTTC AACAATATTT GACGTTAGTC 3840

AGCCGTTTCG TTATAGTTAA AGAAAAGGCT TTTTGGTAGG AGTGAATACG ATGAATGAGT 3900

GGCCCCATCA GTTAATGACG CTTTTAAC 3928

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5259 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CGTTAGGTAC ATTGCTGGTT AATGGAACAA GTGAACAGTT ATATGAAACA AaCGAmaCCA 60

TTCACTTGGC TTTTCCAGCT GAGAAATTAG TTATTTTAGA AAAATAAAAA CCATTTTGGA 120

GGAAGATTTA AAAATGAAAA AACGCGTAAT TTTAGGGACA TTAGTCGCTG CAACGTTATT 180

AATGACTGCT TGTGGAAACA GCGAAgcAAC TACGAAAAGC GAGAGCAAAG GTGGAAGTAA 240

TGCTTTAGTC GTTTCAACTT TCGGATTAAG TGAAGATATT GTCAAAAAAG ACATTATCGC 300

TCCATTTGAA AAAGAGAATG AAGCGAAAGT TACCTTAGAA GTAGGCAATA GCGCAGACCG 360

CTTTACGAAA TTAAAAAATA ATCCCAATGC GGGAATTGAT GTCATTGAAT TAGCACAAGC 420

AAATGCAGCA CAAGGTGGAA AAGATGGGTT ATTTGAAAAA ATTACAGAAA AAGAAGTACC 480

TAATTTAAGT CAGTTAACGC CGGGAGCAAA AGAGGTTTTT GAAAGTGGTG CTGGCGTACC 540

AATCGCTGTA AACAGTATCG GGATTGTTTA CAACAAAGAA AAATTAGGCA AAGAAATTAA 600

AAACTGGGAT GACTTATGGT CAGCTGATTT GAAAGGTAAA ATTTCTGTTC CAGACGTTGC 660

CACGACGGCA GGTCCTTTAA TGTTATACGT TGCTAGTGaA CATGCTGGTC AAGATATTAC 720

AAAAGATAAC GGGAAGGCCG CTTTTGAAGC GATGAAAGAA TTAAAACCAA ACGTTGTTAA 780

AACGTATTCA AAATCGTCAG ACTTAGCyAA TATGTTCCAA TCTGGTGAAA TTGAAGCAGC 840

TGTGGTTGCT GATTTTGCGG TTGATATTAT TCAAGGCGCA CAGAAAACGT GACGTATATC 900

GTTCCAGAAA GCGGCACGTA TGCCAACTAT AATACAGTCA ATATTCCTAA AAAAGCACAA 960

AACAAAGAAA CCGCCTTTAA ATTTGTTAAC GCACGGATTA GTGAAGAATC ACAAAAGGCC 1020

AAAGCGATTT CGTTAAATGA AGGACCTACA AATCAACAAG TAACGTTAAG CGAAAAAGAA 1080

GCGAAAAATA AAACATATGG TGCCATTGCC GAACGTGCCA AAACAGTCGA TTTTAATTTT 1140

ATTAATAGTC AACTAGCGGA TTGGATTGAT CAATGGAATC GGACGATGAA TCAATAGTTT 1200

TAGAAAAGAG GGACAAAATG AACGTCGATG TATTACTAAA AAATGTGTGG CTGTATCAAA 1260

CAGTCACTCA AACATTCGTT CAACGAAATG TTGCGATTAA AAATGATAAA TTTTATTACA 1320

TTTATGAAGA AGAGAACGTC AATTTGCAGC CCCAAAAAAC AATTAATGCC GAGAATCAAT 1380

GGATGATTCC CGGCTTAATT GATGCACACA TGCATATTGA AAGTTCCATG ACGACACCTA 1440

CTATTTTTTC CAAAGCAGTA GTGCGTTAcG GTGTGACAAC GGTGATTGCC GATGCCCATG 1500

AAATGGCTAA CGTATTTGGC TTAGAAGGCT TAAAAGCTTT TATGGCCGCT GAAACAGAAC 1560

TAGATATCTT TCACGCAATT CCTTCCTCGG TTCCTTCGAC CACTCCCGAA TTAGAAACAA 1620

CTGGTGGGAT TATCGGCTTA GCAGAAGTAG CAGAATTACT CAAAGAGCCA AAAGTGATTT 1680

GTTTAGGGGA AGCCATGAAT TTTAAAGGGA TTTCCTATGA GCCAGATTCT TTAATTCGTC 1740

AAATTATCGA TTTGTGTCAA AAACAACGAC CAACCATGCC TTTAGAAGGC CACTGTCCTA 1800

AAATCGAAGA TCAAGAGTTA GCAGATTTTC TATATAGTGG TATCACTTCA GATCATACCC 1860

ATCAATTTCC TAAAACGTTA AAAGAAAAAA TTGAAGCGGG CGTCTTTATT CAATTTCAAA 1920

ATAAATCAAT CACTCCTGAA AATATTCAGG TGATAGTAGA CAATGACTTC TACAATTATG 1980

CAAGTATTAT CACGGATGAC GTGATGGCTG ATGATTTGCT AAAAGGTCAT CTAAATGAAA 2040

ATGTCAAAAA AGCTGTTCAT GCAGGGTTGC CTATCGAAAA AGCCATTTAT ATGGCAACTT 2100

ATACGCCAGC CAAACGAATG GGTCTCCATG ATCGCGGCGA AATTGCACCT GGTAAAAAAG 2160

CTGACTTTTT GTTATTAAAT GACTTAGAAA GTTTTGATAT TAATACGGTT TATAAATCTG 2220

GAAAGGTCGT TTTTGAAAAA GGCGAGCCCT TCCATTATCC AGAAAAAATC GAAGAGTTTC 2280

CAGCCACTTA TCAACAAACC ATCCAATGTA AAAAGTTAAC AGAAGAAGAT TTACTTTTAA 2340

AGGTGGCTAC AACAAAAGAA ACAGTTCGTT GTAATGTGAT TCAAAAACAA GAAATTGGCA 2400

CATTTACAGA GCGAATTACA AAAGAAATTC CTGTTGAAAA TGGATTGTTG CAATGGCAAA 2460

AAGCCAACTG CGCCTTATTA ATTGTGATGG AACGTTACGG AAAAAATGGT AATATTTCTT 2520

TTTCTTTAAT GGATCAACCT CTTTCCGAGA AAGGAGCCAT TGCGACAACC TGGGCCCATG 2580

ATCATCATAA TTTAATGGTG ATGGGTAATA CGATTGAGGA TATTCTGTTA GCGCAGAATG 2640

AATTATTAGC GATGCAAGGC GGCTATTTGG TTGCTTCGGA TCAGCAAGTG ATGGCAACTT 2700

GTCCTTTACC AATTGGTGGG ATTTTATCGC AAGCACCGAT TGAACAATTA GGCGCTTCTT 2760

TACAAAAAGT CCGACAAGCT ATGCAAGCGT TAGGCTATCA AAATATGAAT GAAATCATGT 2820

CCTTTTCTAC ACTATCATTA CCTGTCTCGC CAGCCATTAA AGTGACAGAT TTTGGGATGA 2880

TGGATACGAA GTCACAACGT TTCTATCCAT TAGTTTTTCC AGAGGATGGA GTTCTTTTGC 2940

ATGAAAACAC TCATTAAAAA TGTGCATATA CTAACAATGG ATGAGCAATT TTCAGAAATC 3000

AAAGCCGGCT ATTTAGTGAT TGAAGAAGAC ACAATTGTGG AATTAGCACC AATGACCACA 3060

CTTGATGAAA AGCGAATGGC TGCAAATCAA GTAATCGATG GTCAAAATGG AATTTTAATG 3120

CCTGGGATGA TTAACACCCA TACCCATGTA GGCATGATTC CGTTTCGTTC GTTAGGAGAC 3180

GATGTGCCAG ATCGACTCCG GCGTTTTCTT TTTCCATTAG AACAATTCAT GACAAAAGAA 3240

TTAGTAGGAT GCAGTAGTGA TTATGCAATT GCCGAAATGT TACTGAGTGG TATTACGAGC 3300

TTTTGTGATA TGTATTATTT TGAAGATGAA ATTGCTAAAA GTTGTGAAAA AATGAGTGTT 3360

CGTGCTTTGC TCGGAGAGAC GATCATTGAT ATGCCCACTT GTGATAGTCC GGAGCCTTCA 3420

GGCGGtCTTT TTTACGCGGA AACCTTTATT CGCAAGTGGC AAGGCCATCC GTTGATTACG 3480

CCTATGCTTG CGCCACATGC ACCGAATACC AATTCACCAG AAGTGTTGGC GAAAATTATT 3540

GAACTTAGTC GGCAATACCA AGTTCCTGTG ACCATGCACG TTGCTGAAAT GACTTATGAA 3600

ATGGCTGAGT TTGAAAAAGC CTATCAAAAA ACACCAATTG CTTTCTTAGA AGAACTGGGT 3660

TATTTGAGCG AGCCGTTTAT TTTAGCGCAT TGTATTTTGG CAACAGATGA AGATCTTGCG 3720

AGTTTAGCTG CTACTAATGG AAAAGCGCGT GTCGCTCATT GTATCGGTGC GAATACTAAA 3780

TCAGCCAAAG GCGTAgcGCC GATTAAGCAA ATGCTTGATC AAGGGATTAT TGTCGGTTTA 3840

GGCACGGATG GACCTAGTAG TGGGAATACA TTAGATTTAT TCACCCAAAT GCGCATGGTT 3900

GCGAATTTTC ATAAGACAGC ACACCAAGAT CGCTCCTTGT TTCCTGCTAA AGAAATTGTT 3960

TATCTGGCAA CGATGGGGGG CGCTAAAACG TTAGGCTTGG CGGAGCAAGT CGGCTCATTG 4020

GAAGTGGACA AAAAAGCGGA TATAACATTA ATTGAAACGC AATCAGTTAA TATGTTTCCG 4080

ATTTTTGATG CCTATTCAGC GTTGGTTTAT TCAGCAAATG CTAGCAATGT TGAAGCCGTT 4140

TGGGTGAACG GTCAACAGTT AGTTGCCAAT AAAGAATTAC AACAAGCTAA TCTCAAAGAA 4200

ATCAAGGaAA AATTATATCA GGCCATGAAT ACGTTTGTGA AAGAAGCTAA AAAAAGAGCT 4260

GCTCTCTAAA CAGAACAAGT AGGACGTATC GGATTGATTA AATAACGCTG GATTGACGTT 4320

TAGTGATTCA GGTTCTAATA GAGCGCAGAC AAACTTTTAA TGAAAGAGTT GTCTGTGTTC 4380

TTTTTTGTTA TAAACATTTT TCTTTTTCTA TATATATTTT TAAGGTAAGT TGTATTATAG 4440

TAAGTTTGAG AAGATGAGAA TGAGAGGAAA AGCAATGAAT ATACAACGGT TTATAGAAAA 4500

ACGCAAAGCA CGTGGTTTGT CGCAAAGTGA ATTAGCTAAA GGGATTTGTA CGCAAGTTAC 4560

GGTTAGTCGC TTTGAAAAGA ATGGTCAAGT CCCAACTTTA AAAATATTGA TCCaACTATG 4620

TAATCGTTTA GAGCTACCTT TGGGGGAACT CTTCCCGCGC GTAGGGATTA AACAACCTGA 4680

AATTCTTGAA AAAATGGAAG AAGCAGAATT CTTCTTGATT ACCAGCGAAC ATGACCAATT 4740

GCAAACAATT TTAAAAAATA TTCCTTTTGA TGAAATCAAA GATTCCCAAC TGTTATTAGA 4800

ATATTATTAT TTACAAGGCT TTGTCATGAT TTTTCAAAAT GCTTCGTTGA TGGATTGCTT 4860

ATTTACCTTT GAAAAACTTC TGTTTGAAGA GCAAAAATAC ACGAGCGATA TTTACCGTTT 4920

GTTAGCTTTT ACTGGAATTG GCATGGCCTA TGCCAAAGAA GGCGAAATTG AGAAAGCCGA 4980

ATTTTATTTC AATAAGGTCT TCAAGGAAAT TTATCTCTAT ACTATTCAAT CAATGGAAGA 5040

TACTTGGCGA GTATTAAATG TTGTCTTTCA CTGTGGTGTT TTTTATGCAG AAAAAGGCGA 5100

TTTAGAAACA AGTGATGCTT TGTTAGAGTA TGCGATTTCC ATCTGTTCTG ATAATCACGT 5160

CACGTATTAC TTGGCAAGAG CAGCTTTCCA ACTTGCGAAA AATGCCTTAG CTGAAGAAAA 5220

GCCGCAAGAG CAAATTTTAG AACTACTACA AGATGCCCG 5259

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7684 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AGGCCGTnTG AAGTGGaCCc TnGAgGAaTT wGtTCaAAAr GCmAAgGnTT TAATTGGaGA 60

AGGCcAmCGT GGAACnTGCG CaACmATTcA TTGrAGAACA TAAAGaTGAA TTAGGCGAAC 120

ACTATCATAA AATTACAGCG ATGTTAGCTG ATACAACgAC CgATGgTTTG TTTGATAAAA 180

TTAAAGGATT GTTTCAATAG TTGAAAAAAA TCTGAGATAA AAGTGTTTAG CTCCAAAAAC 240

CAAGTAGGTA CGATGATTTC ATTTGTTCAT GCTTCACGGT GAAACATTGC AACTCTCGCC 300

GTTTAACAGT TTTTGAGCGT GGAGCAAAAA TCCAAAGTGA TTTTTGTCCC ACGCTCTTCT 360

TTTTTTATTT AACTTTTAAC CATTGTTTTG GCAACACATA CGCAATTGGA TTAAAAAGAT 420

AATCCGCTCC CCGTTTCACC ACAGGCATAT TCGTAATAGC TAAAATGAAG AGGGCTAAGA 480

AGAAAAGTGC CAGGATTCGG AAAGTAAACC ATTCTTTAAA GAGTGTGTCT AATTGCAAAC 540

CACTTTCCAA CAGAATAATT AAGGGCATAT GCAAGAGATA AGGGTAGAGG GTATCTTTTC 600

CGAGGGCGCT TAAAGGTAAC TTTCGTTTCG GCAATAATGC AAACCAAGAA AGAGACACCA 660

ATGCTGCAGC AACGTAAAAG AACAGCTGAA TGAAACTACT GTATTCAAAG GCCGGACTAA 720

TTyCAGCGTA GGAAGCCCAA CCAGTTAAGA aGCCCCAAAA TTTCGTTGCC AATGACCATT 780

GACTATTTAC AAATAAAAAG GAGCTTAAAA TTACGCCTAT CGCTGCCCAT TTGACTTTGC 840

GATGATCTAA AAAGGTTTGT AGCTTTTGAA TATCCATTAA GTAGCCACAA TAAAAGAAAG 900

GGAAGAAGCC CAGAGTCCGT CCTAAAGACA AGAAACCTTC CACAGGTAAA TAGCCTGAGG 960

CGCAgCcGAT AACAATGATG ATTGGCAACA TGATTTTCGG CTGAACGCGC TGAAGAAAGA 1020

TGGTAATTAA GTAGAAACTT GCTAAGGAAA GTAAGTACCA AGGAACGCCT GTTTCGTTCA 1080

GTAAATCAAA ATTCATTTTA TTTGCAAAGG CGAAATTAGC AAGCGTAATC AACAACTTCG 1140

CTAAGAGATA AAGCCAAACA AAGGAAAGTA AGCGTAGATA TTTTCGTTGT TTAATGGCAT 1200

GTTTACTAAA GAGTCCAGAG ACAAAAATAA AGGCTGGCAT ATGGAATGAA TAGATCCAAA 1260

AGTTGATCGT TTCCATGAGT CGTGAAGGAG AAGCTTCACG ATCTAAAATA TGCCCAAAAA 1320

TCACTAAGAA AATTAAAAAA AGTTTCACGT TGTCCCATTT ATAAATTCGC TTGTCGTTTT 1380

CCATAACGAT CACTCCTTTT AGTAGAAAAC AGTTCACTTT TGTTTAAGGC AGAAGAGGAC 1440

AATTTTTCCA TTCATTGAGT AGATAATTGT AGATACGAAC CGGCAGCCGT TAAATGAGCC 1500

CCCACATACT TGCGCTTAAA ACAAAAAATT AGGCAGGATG TTAGTCTTAA ATTATACTGA 1560

AAGAGTGACG ATCCTCCAAA GATTTTGTTT ATTAAGAAGT GAAAGATGCG GTGAAAATAA 1620

ATGATAAAAT TTCAAAATGT TTCAAAAGTA TATCAAGAAA ATGCCGTTTT GTCCGATGTG 1680

AATTTAACAA TCCAACAAGG CGAATTTTTT GTATTAGTAG GGCCGAGTGG TAGTGGAAAA 1740

ACGACGACTT TGAAAATGAT TAATCGGCTC ATTGAGCCAA CGACAGGTGA AGTGGCTCTC 1800

AATGAACAAG CGGTAACCAA TTATCCTTTA CGTGAATTAC GACTAAAAAT TGGCTATGTT 1860

CTTCAGCAAA TTGCCTTATT TCCTAACTTA ACCGTGGCGG AAAATATTGA GTTGATTCCT 1920

GAAATGAAAA AATGGCCCAA AGAGCAACGC AGAGAGCGAA CCATTGAACT ATTGAAAAAA 1980

GTTCAGCTGG ATCCTGAGGA ATATCTGCAT CGTAAGCCAG CAGCGCTTTC AGGAGGCGAG 2040

CAACAACGTA TCGGGATTTT ACGCGCCATT GCAGCCGAAC CAGAAGTGAT TTTAATGGAT 2100

GAACCTTTTA GCGCGCTAGA TCCGATTTCA AGACATCATC TTCAACGTTT GGTGAAAGAA 2160

CTTCATCAGG AATTAGCGAG TACCATTGTT TTTGTGACCC ATGATATGAA TGAAGCTTTA 2220

CTGTTAGGAG AGCGCATCTG TATTATGGAC ACTGGCAAAA TTATTCAAGT CGATACGCCA 2280

GAAGCCATTC AAAAACATCC GAAAAATAAT TTTGTCGGTC AATTTTTCAA ACAATCTTCA 2340

CCAGAACTAG CGAACTATTG TGGCGCTGAC TTGACTGCAT TTTTAGAAGC CGCAGATGAG 2400

ACGATTGAGC CAGTGCCGAC TGTTTTCATG GAAACCAATT TAAAAGAGAT TGTCCAACTC 2460

ATTAATGCGC AAGAGAAAGC AAATTTAGTA GAAAAACAAC AAGTTTTAGG GTGTATCACC 2520

TCAAAAACGA TTACTCGTTT TATGGAACAG CTGTTGGAAA GCGAGGGACA GTAGATGCAA 2580

GCACTGATTG AAACTTTTTT CGCTCGAAAA GAAGAATATC TGCAAGCGTT AATCGAGCAT 2640

ATGCAATTGT CCTTTATCTC GTTATTATTC GCTGCGTTGA TTGCAATTCC TTTAGCCATT 2700

TATTTAACAA GTCATCGTAA GTTAGCGGAA AGTCTGTTAC AAGTGACTGG GATTTTCCAA 2760

ACCATCCCTT CTTTAGCCTT GCTAGGCTTA CTGATTCCGT TAATTGGTAT TGGTCGACCA 2820

CCGGCAATTG TTGCGCTAGT GATTTATGCC TTATTCCCAA TTTTACAAAA CACCTATACT 2880

GGTTTGACGG AAATTGATCC TTCGTTAGAA GAAGCTGCAG AAGCATTTGG GATGAGTAAA 2940

AAAGAAAAAC TGTTTAAGTT TGAGTTGCAA ATGGCGTTAC CATTTATTAT TTCTGGGATT 3000

CGAACGGCTA CCGTTTTAAT TATTGGTACG GCCACGTTGG CGGCGTTAAT TGGAGCAGGT 3060

GGTTTAGGGA CCTTCATTTT ATTAGGAATT GATCGAAATA ATTTATCGTT GATTTTTATT 3120

GGAGCCCTTT CATCGGCTGC ACTAGCAGTC TTATTTAATT ATGGAATTCA TTGGTTGGAA 3180

AAAGCGAATG GGCGGCGACT AATCATTGGC GGTACTATTT TAGGCCTTTT GTTAGGTGGT 3240

TCATTTTTCT GGAATCAACA AACTTCTTCA AAGGAAAAGC AGCTGACAAT TGCTGGCAAA 3300

TTAGGAGCGG AACCAGATAT TATTATTAAT ATGTATAAAG CCTTAATTGA AGAAAATAGC 3360

GACATTCAGG TGACGTTAAA GCCAAACTTT GGCAAGACTA CTTTTTTATA TAATGCCTTG 3420

AAATCAGATG AAATCGATCT TTATCCGGAA TTTACAGGAA CAGTGATCGA AACCTTTTTA 3480

AAAAATCCTC CTCAGCTAGA CAATCAACCT CAAGTAGTCT ATGAAGCGGC ACGCGCTGGC 3540

TTGAAAAAAC AAGAAAACCT TACTTTGTTG AAACCGATGC GCTATGAAAA TACGTATGCC 3600

GTCGCAGTTA AGCGCTCATT TGCAAAAGCG CATCAATTAA AAACGATTAG TGATTTACAG 3660

AAAATTAGCA ATCAGTTGAA AGCAGGTTTT ACGTTAGAAT TTATTGATCG GCAAGATGGT 3720

TATAAGGGAT TACAAGAAAA ATATCATTTA AATCTGAATG TCCAATCGAT GGAACCGGCG 3780

TTACGCTATC AGGCGATTAA TAATGGTGAG GTAAATGTGA TTGATGCCTA TTCAACGGAT 3840

AGTGAATTGA GACAATACGA TTTAGTAACG TTAGAAGATG ATCAAGCGCT ATTTCCGCCT 3900

TATCAAGGAG CACCACTTAT TAAAACAGCA ACGTTGGAAA AATATCCAGA ATTAGCAGAA 3960

ATATTGAATA AATTGGCAGG AAAAATTAGC GAAGAAGAAA TGAGTGAAAT GAATTATCAA 4020

GTCAATGTGG AAGGGCAAGA TCCTAGCATA GTGGCGAAAG ACTATTTAAA AGAAAAGAAT 4080

CTTCTTAAAT AAGTATTTTT TACTGAAAAA GACTGGTATA TAAACATTTA TCCAGTCTTT 4140

TTCAGATAAA TAAAGAGTTT AATGTGAAAA GTACCTGTAT TTTTACTCAT TTTTGGATAT 4200

AATAAAAGAA ATGAATGGAG AGGAGTGGTC TATATGAAAA AAGCAGAACG TCAACGTTTG 4260

ATTAAACAAT TAATTATGCA ACAAGAAATT GAAACACAAG ATGAATTAAT CACTCGTTTG 4320

GAAGAAATTG GCGTCCGAGC CACTCAAGCG ACTGTTTCAA GAGATATCCG TGAAATGAGT 4380

ATTGTTAAAA CACATGGTGC AGATGGTCGT GTCAAATATG CAATTTTTTC TCAAGCACAA 4440

GGCACAAGCA GTGAAGAGAA ATTACGAGAA TCAGTAAAAG ATTCAGTTGT TCGAATGGAG 4500

CGAGTACAAT TTATTGTCAT CTTGCATACA GAGATGGGAA ATGCCGATGT GGTGAGTAAC 4560

TTTTTAGATG AAGTAGCTTA TCCAGAAGTG GCTGGTACAG TGGCTGGTGC GGATACAATT 4620

ATTGTGATTA CACGTTCAGA AGAGGATGCG GAACACTTTA TCGAACGCAT CGAAAATATG 4680

ATTTTTTAAA ATTGAAAGCA TCAATAGAAG GGAGTTTTTT TCTTGGCAGA ATTGAAAGCG 4740

CTACAAAATG GTTCGGATAT TCGGGGAATT GCTTTGGATA CGGAAGAGCA AACTGCAACA 4800

TTAACCGCAA CAGCAGTAGC CGAGATTGCG GTGGGCGTGG TTCGTTGGTT ACAGGATAAA 4860

AAACAGCTGC CACGAAAGGC ACAACAGCGC CTGACAATTG CAATCGGACA TGATAGTCGT 4920

CTAACGGCGG AGcatTAAGC AAGCATTAGT AGATACGTTT CTTTCATTAG GTATTCAGGT 4980

GATTGACGTT GGTTTAGCTA CGACGCCTGC TATGTTTATG GCTACTCAAT TTCCAACGCT 5040

TCAATGTGAT GCAGCGATTA TGATTACAGC TAGCCATTTG CCTTATTACT TTAATGGCTT 5100

GAAGTTCTTC ACAGCAGAAG GTGGAGCAGA GAAAGAAGAT ATTCGGTATA TTTTATCTCA 5160

TACTGATCCG CTGACGGCCA ATGAAAATGG CACACTGATG AAGCAAGAGT TACTGCCAAT 5220

TTATGCGGAG CATTTGGTGG AAAAAATCCG CCAAGGAATT CATTCGCCAG AAGAGAAACC 5280

TTTACAGGGC TTTCGAATTA TCGTTGATGC GGGTAATGGG GCTGGGGGCT TCTTTGCGGA 5340

GCAGGTTCTT CAAGTTTTGG GCGCTGATAC GACGGGATCA CAGTTTTTAG AACCAGATGG 5400

GCATTTTCCT AATCATGTAC CTAATCCAGA TAATTCTGAA GCAATGAAAA GTATTCAAAC 5460

GGCTGTTTTA GCTAATCAAG CAGATTTGGG CATTATCTTT GATACAGATG TAGATCGTTC 5520

GGCCGTGGTG GACCAATCAG GTGAGGTACT AAACCGAAAT AATTTAATTG CTGTTTTAGC 5580

GGCAATTGTT TTAAAAGAAG CGCCTGGTAG TTATATTGTG ACAAATTCGC CAACCTCAAG 5640

TCATTTAAAA ACATTTATTG AAGAAAAAGG CGGCCAACAG ATtCGTTATA TTTCTGGTTA 5700

CCGAAaTGTC ATTAATAAaA TGATTGAATT GAaTCATGGG GGCTTTCAGA CACCGTTAGC 5760

CATTGAAACC AGTGGTCATG CAGCGTTTCA AGAAAACTAC AATCTGGATG ATGGTGCATA 5820

TGTCGTGGCG AAAATTTTAA TGTTGTTACC AGAATTGAAA CAAAACAATC AAACGTTAGG 5880

AGATCTGATT GCTACATTGA AGCAACCGGC AGAAACCAAC GAATTTCGCT TCAAAATTAC 5940

TGCGGAGGAT GTTACTTGTT ATGGGCAACA AGTCTTGCGA GACTTTGAAT TGTTTGTAGA 6000

AAATCAAGCA GACTTTGCTG TGGACCGTGA AAACCAAGAA GGTGTTCGGG GAAACGTGTC 6060

TGGTCAATAT GGCAGTGGTT GGTTTtTGTA CGCTTAAGTC TGCATGAACC GTTGCTTGTT 6120

CTTCAGGTAG AGAATGACCA AAGCGATAAA AATGCGTGTG TCATCGAAAA AATTGCCACT 6180

TTTTTACAAA AATACGAGGA AATAGATAGT CAGCAAATAG AAAAATAATC TAGGAAATTA 6240

AAGGTGTTCT TATGGAAAAT GTTTTAAAAG GAAAACACGT TGGCTTTTCT CATTTTAGAG 6300

AAGAATATAT CGAACATATT GCAAAACAGC AATGGGATAA TGAATTGCTT CGGCATTTGA 6360

GCTGGGATGC TTTACATCCG TGGGGGCTTG AAGAATGGAA AGATTTCACA ATTAATAAAG 6420

GAGAAGATGA TCGATTTCTA TTTGCTATTT TAGAAAATAC GACAGAGGAA TTTATTGGTT 6480

GGGTTTCGTT GTCTGATGTA CAGTTGAAAA ATAGAGGGGC CAATTTAGGG ATTGCTATTT 6540

TACAAAAAGA GCAACGAGGT CAAGGATGTG GCTTTGAGGC GGTGTCTTTA ATTTGTAAAT 6600

TTGCTTTTTA TGAATTGGGT TTGCACAAAA TCAGATTAGC TGTCAATAGT AACAATCAAA 6660

AAGCGATTCA TGTTTATGAA AAAGTAGGCT TCAAAAAAGA AGGAATTGAT AGAGAAGCGT 6720

TATTTCAAGA TGGCCAATGG CTGGATATCT ACAATTACGG AATTTTACAA AAAGAGTGGC 6780

TACAAATGAT TAAGGCTGAG AGTTAGCTCA GGAACTGTTT AACACCAAGA CACTGAGCAA 6840

ACGCAGATAT GCCGTTCTCA GCAGTGAATG TGCGGTAAAA TAAAAAAGCA GAAGAAATCG 6900

GCGGTTACCG TCCGATTTCT TCTGCTTTTT TTTACCATAG GCCAGGTGTT GCAGTCACAA 6960

CGATTTGTTT TTTTGTAAAC GGATGGAGTA AATGAAGTTG TGAAGCATGG AGCATTAAAC 7020

GAGCTGCCTG CGATTTTGGG TTGTATAACG GATCACCAAT AATTGGATGC CCTTCGCTAG 7080

CTAAATGGAC CCGAATTTGG TGCGTCCGAC CTGTTTCTAA AACACAGGTA ACTGCGGTTT 7140

GATTATTAGA AAGAGCTTGG ACCACTTGGA CATGTGTAAT CGCTGAATCC CCTTTTTGAG 7200

GATCAATCCG TCGTTTCCGG CGATCATGTC GATCACGACC GATCGGCTGA TTGATTGTCT 7260

GCTCTTTTTG GCTAAAGTGA CCAGCTACAA TCGCTTGATA TTGGCGGTAA ATTTGTTTTG 7320

TTTCTAATAA ACGACCTAAA ATAGGTAAGA CAAAAGGATT TTTAGCAAAG ACAATCGCAC 7380

CACTGGTTTC TTTATCTAAT CGATGAACGA CATAAGGGAC CTGATTTTTC TCGGCTAAAT 7440

AGGCAGCAAC ATGATTTAAC AAGGTGTCTG TTTCCTGTGG CTGATTTGGA TGGGTTTTAA 7500

TGCCATAAGG TTTATTTACA ATCATTAAAT GTTCGTCTTC ATATAAAACC TTCACATGCG 7560

TTGCTTCCCC CAAATGAATA GTGGGACGTA CATAATCGGT TTCTTCAAAC GTTAAGGTGA 7620

CACAATCGCC AGCTTGGACT GGAAAATGAA AAGCAGCTGT TTCTCCGTTG ATTTTGACAT 7680

TTTT 7684

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32768 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

CGGAACACCT GnCGCCTATA AAAAGGTTGG CCAACAAACC CTGATGGCTG GGGCGATTGA 60

TAAACCGAAC ACTTACACGG TCAAAwTTga TGTAGAAAcC GAGGGTTCTT TAAAAATCAA 120

AAAAGTTGAT AAAGAATCAG GTGCTATTGT ACCAGGAACG GTTTTCCATT TAGACTTTGG 180

AAAAACTTTA CCTGCAAAAG ACGTGACTAC TGACAAAGAA GGCATTGCTA CATTGGATGG 240

GATTCCTCAT GGTACAAAGG TAACCATTAC CGAAAAATCG GTGCCGGCGC CTTATACGAT 300

TGATACCATA CCCATGACTG CTACGATTAA AGCAGGCGAA ACCATTTTTA TGACTTCAAA 360

AAATACGCGG GAAAAAGGTC AAATCATTCT GGATAAAACC GGGATTGAAA CAGGAAGTGA 420

TCTTTGGAAC GACAATTACT CTTTAGCTGG AAACACGTTT GCTATTCGCA AGGACAGTCC 480

TACTGGTGAA ATTGTTCAAG AAATGACCAC GGATGAAAAA GGTCATGCAG AAACACCAAA 540

AGAGATTGCT AATGCATTGG AATTGGGAAC TTACTACGTG ACCGAAACCA AAGCATCTAA 600

TGGCTTCGTG AATACCTTTA AAGCAGTAAA AGTCGAATTG AAGTATGCCA ATCAAACCGT 660

GGCCCTTGTG ACCAGCAATA TCAAAGGGCA AAACCAAGAA ATTACTGGGG AAACAACTCT 720

GATAAAAGAA GACAAGGATA CTGGAGATAA AACACAAGGC AGGGCAGTCT TTGAGGGGAC 780

TGAATATACC TTGTTTACTG CTAAGGATGG AAAAGCGGTT AAATGGAGTG AAGCTTTTAA 840

ACCAGAATTG GTAAAGGGAA CGAAAGCTTC TGATGAAACA GTGACTCTGG CTTTAGATGA 900

AAAGAaTCAA GCTGCCGTTa AACACtTAGC CATCaACGAG TACTACTGGC AAGAAACCAA 960

AGCACCTGAA GGATATACCT TGGaTGaAAC GAAGTATCCT GTwTCyATCA AAArGTwGA 1020

TgATAACGAA AAAAaTGCCG TrATTACyCG AGATGTTACG GCAAAAGAAC aGaTTATTCG 1080

TTTTGGGTTT GATTTCTTCA AATTTGCyGG ATCrGCyGmT GGCACTGCCG AAACTGGGTT 1140

TAATGACTTA ACCTTTAAAG TGTCGCCATT rGAAGGGACC AATGAGATCA CAGGTGCTGA 1200

AGATrAAGCA ATCACAGCTT ATAATGAGCA ATTAGGCTTT GATGGCTACG GAAAGTTTGA 1260

AAATCTTCCT TATGGGGATT ATTTACTTGA AGAAGTAGAA GCGCCAGAAG GATTTCAAAA 1320

AATTACGCCA TTAGAAATCC GTTCTACTTT TAAGGAAAAC AAAGAAGACT TTGTGAAGAG 1380

TGAGTATGTC TTTACCATTA CCGAACAAGG CCAAAAGCAA CCAATTAAAA CGGTGACCGT 1440

TCCTTACGAA AAGCTGACGA ACAAAGCATT TTCTGTTAGT TTGAACCGTT TGATGCTCTA 1500

TGATTTGCCT GAGGAAGAAG ATAGTTTGAC TTCTCTTGCA ACTTGGAAGG ACGGAAATAA 1560

AAAATTGAAT ACCCTTGATT TTACCGAGCT AGTTGATAAA TTGAGATATA ACTTGCATGA 1620

AATCAAAGAA GACTGGTATG TCGTAGCTCA AGCCATTGAT GTGGAAGCCA CAAAAGCTGC 1680

CCAAGAAAAA GACGAAAAAG CCAAACCAGT GGTAATTGCC GAAACAACGA CCACTTTGGC 1740

GAACAAAGAG AAAACAGGTA CTTGGAAAAT TCGGCATAAA TTAACGACCG AACAAGTTTT 1800

GGATAAAACT ATCGTCTTGT TTAACTATGT GTATGAAAAC AAGGAAGCTT TTGAAGCAGG 1860

CGATGAGCCA GTAGCGAAGG ATGCTAGCTT GAATAATCAA GCCCAAACTG TTAATTGTAC 1920

GGTGGAACGT CATGTTTCCA TCCAAACAAA AGCCCACCTA GAAGATGGTT CTCAGACCTT 1980

TACTCATGGT GATGTGGTGG ATATGTTTGA TGATGTATCG ATTACCCATG ATGTGCTGGA 2040

TGGATCAAAA GAGGCTTTCG AAACAATTCT GTATGCACTA CTACCGGATG GTACAAACAA 2100

AGAAATTTGG AAATCAGGCA AGATTGATTA TGAAGTAAAT GATAAAGAAT TTACCAAAAC 2160

TATGCTTGCG GAAAAAGTAG ATACCGGAAA GTTTCCAGAA AGTACTAAGT TTACTTTTGC 2220

AGAAATCAAT TATGATAAAG ATGAGAATAT CAATGGGAAA CACAATGAGG ATCTAAAAGA 2280

AAAAAATCAA ACCTTAACGC CCAAAAAAGT TCCAACCACA CCACGTACGC CGAAACAACC 2340

GGAAACACCA ACTGTTCCAA GTGGCTATCA AGAATCTAGT CCCACAGTGA AGACATTCCC 2400

TCAAACTGGG GAGAAAAATT CCAATGTTTT ACTGTTCATT GGTTTTACTT TGATTTTTGC 2460

GACGGCGGGC TATTATTTCT GGAATCGCCG TAACTAAGGT GATGAGATGA TACAAGTAAA 2520

TCCTAAGAAC CAACGTGCGC GACCGCCTCC TAACAAAGTG GTCGCGCTAT TTTTGTATCC 2580

AAAACTAACA ATGAAAAGAG GAAAAGAAGA TGGAATTGAA ATTTGTTGTG CCAGATATGG 2640

CCGAAACGTT TGGAAAGTTA AGTTATGCCG GTGAAGGAGA AATTTTAACA GAAGGATATG 2700

GACGGAATAC TACGGTGATT GGTCGCAGTT ACCATCTATA CTCCAGCAAA CAACGGGCCG 2760

ATGATATTGA GGTGGTAGTA GCTGCGGAAG CCGGTGAAAA GGATTTTGAT CAAGACCAAC 2820

CGCTAAAAGC CGTGAATCCC CATTTGGTTG CCAAAGGCTA TGAGATTGAA AATCGTGGGT 2880

TTACTGATTA CGTGTTGTAT GTTGATGATT TAGTGAAAGA ATAGGAGGAA AaGAAGATGA 2940

GATTAACAGA AGGTATTGTC GTAGATTCAG TATTAACATT TGGAAAGTTG CGTTTTTCCG 3000

CATTACGCCG AGAAGTACGG AAGCAAAATG AGGATGGAAC GGTTAGTAAT GAGGTAAAAG 3060

AACGAACATA TAATTTGAAA TCCTCTGCAC AAGGGCGAAT GATTCAAgTT AGTwTCCCGG 3120

CTAATGTGCC CTTACGCGAG TTTGCTTATG ATGCAGAGGT AGAGTTGGTC AATCCAATCG 3180

TGGATACAGT CGCCAATTAT GTTTTTCGAG AAGGAACTAC TGTCAATTGG TTTATTAAGG 3240

CAGACGATTT AGTcTTGAAA CGACAACCAA ATCAAGGAAA TTCTACGAAC CAAATTGAAG 3300

GAAAGAAATA GGTGAAGATT ATGGAAATTA ATGTTGAACA TATTTTAGAT TGTCTCGATC 3360

AATATGGAAA AGGTGAGTTA ACTGAGGAAC AACTCGTTAA AGCATTAACC TATGATGAAA 3420

AAATGTTTCT GATTATGCAT CAAGGTTTAC TTGGAGAAAA TAGCGACAGT GAAGACGACT 3480

TTGATGTACA AGACTGGTTA GAAGAAGAAG AGTCCTTCTT TATGGTTGTC GAAGTAAATG 3540

AAAACTTATG TCGGCAGGCA GAATCCGTCT TGGAAGAAAT CGGTGTGGAG ATGCCAGATG 3600

CCATTGAAGG CTTTTTAAAT CAATTGGTTG AAAAGAAACA ATTCCCCGTA GTAATAAATA 3660

GCTAGCATAA CTATATGAAA ATGCCCAATG AAATCCTATC GGGCGTTTTT TGATCGGAGG 3720

TGGAAACCAT CAAAATGTAC AAAGAACATC GCATACGTGC AAGAGATCAG CATTTGGTTT 3780

ATCACTTTAT TCTTGGCTGG CTAATTGCTT TGTTAATCAG TTGGATGGGT GTCTTTTATT 3840

TTCAAGAGTT TAGACAATTT GATATTTCTA GGGTATCACT TTCTACTATT GAAACTGTTT 3900

GGTCCATGAA GGAGTTAATC TGCTTACTTG GAAGTCTCGG TTTTTCAGGA GCTATGCTGT 3960

TGCTTTATAT TCACTTTTTT CCTGATCATT GGCGAAGCTT GTGGCATCGG CAAAAACTAG 4020

CTCGGATGAT TCTAGAAAAC CATTGGTATG AAGTGAAACA AACTCAAAGT GAGGGCTTTT 4080

TCAAAGATTT AAATAGTAGT CGAACCAGAG AGACTATCAG TTACTTCCCC AAAATCTATT 4140

ATCGAATGAA GGAGGGTTTA CTTTCTATTC GCGTTCAAAT TTCACTAGGG AAATATCAGG 4200

```
AGCAGCTATT GAAGTTGGAA AAGAAGCTAG AAAGTGGATT GTATTGTGAG TTAGTAGAAA   4260
AGGAACTCAA AGACTCCTAT GTGGAGTATA CTTTACTGTA TGATATGATC GCCAATCGAA   4320
TTGGGATAGA CGAAGTAGTG GCAGAAAATG GGACTCTGCG ATTGATGAAA AATCAAGTAT   4380
GGGCTTATGA TTCTTTACCC CATATGTTAA TCGCTGGTGG CACAGGTGGG GGGAAGACCT   4440
ATTTCCTCCT CACCATTATC GAAGCACTAT TGAAGTCAGA TGCAGAATTG TTTATCCTCG   4500
ATCCCAAAAA TGCAGATTTA GCTGATTTAG GTACGGTGAT GCCTCACGTA TATTCTCAAA   4560
AGGAAGAAAT TTCTGCTTGT GTGGAAGATT TTTATGAACG CATGATAGCT CGTAGTAAGG   4620
CAATGAAAGA AATGCCCAAC TACAAACCAG GAGAAAATTA TGCCTATCTT GGACTTCCAC   4680
CAAACTTTTT AATCTTTGAT GAATACGTGG CTTATATGGG TGCGAATCGT TTTCCTACAA   4740
GCATCGAATA GATGTGAAAT GGAAGCCTAT TGAGATTATC AATAGTGACA ACTAATGTTT   4800
TGAGAGATGG AATGAAAGCC TTCAAGTTGA GCTGAAACAT CTCCGCTAAT ACTCCTACAT   4860
GCTCTAATCA ATGAAGTAGA TTGGTTTTGG TATGAAGCTA GGTGAAGTCG TCTGAAATCT   4920
GCCTAAGTGT TTGAGCATAA GGCTGACTAT AGGGCGGGCG GCTGAAAAAG ATAACTATGG   4980
TAAGAATGTC TGAAAATGAC TGACGAATCT GCGAATGTAC GGGTCTAAAA GGCTGATCTT   5040
GTCGAAAGAT AGGAACATAC GGACGTATGG TGAAATACCG ACGAGtaagG TTCCATGATA   5100
TGAAAATCGG GATAGCTAAC GAAGAAGCTA TAATTTCACA GGCTCATAGT AGATACCTAA   5160
GGTTATACAT GATAGCTAAA AACATTGGAA CGATGTAAGC AGAAAACTAT CAAATAAATG   5220
ACTACTTCTA GATAGGTGCA TATAAGTAAT AACGAAATTG CTTTGTTTCT GTGAAAGTAA   5280
GGGCACAGTA CCGATGAAGC GTCTAACAAA CGTGGAGGGA TAGCCCTAAG TCTATTTAAT   5340
AACAATGAAT GTTGAAACAG ATGAAGTTCA AGGATCTGGT AAGAACGTGG GGAAATCCGT   5400
TATGGAGGTG CCACGCTTGA ACACAAAAAT ACGTTACTGG GAATATTACG ATCTACAAGA   5460
AACATTCGAT CAGCTATTTA CTCAAAGTAA AAACGGGAAA AAGTTTTATC AATTATACGA   5520
GTTAATCATT TCTGAAAATA ATATTCTTTT AGCATACCGC ACTATCAAGG CAAACAAGGG   5580
GTCTTCGACA CCAGGAACGG ACTCCTTCAC AATTGATAAC TATAAGGAAA TGAaCCAAGC   5640
AGAGTTTATA CATCTTATTC TCAGTCAGTT AGAAAATTAC AAACCAAAAT CCATAAAGCG   5700
AGTAATGATA CCTAAACCCA ACGGCGAGAA ACGCCCATTA GGAATTCCaT GCATGATTGA   5760
TCGAaTTATC CAACAGATGT TCAAGCaGGT ACTGGAGCCA ATATGTGAAG CAAAGTTTTA   5820
TGAACATAGT TACGGGTTCA GACCATTGAG AAGTGCAAAA CATGCTCTAG GTCGCATTAT   5880
GTATCTCATA AACATCAGTA AGATGCACTA TGCAGTAGAC ATTGATATCA AGGGTTTCTT   5940
TGATAATGTA AATCACCGTT TACTGATAAA ACAATTATGG AATATAGGAA TTTGTGACAA   6000
ACGAGTCTTA GCTATTCTTT CAAAATCTCT GAAATCACCC ATTCAGGGAG AAGGTATTTC   6060
GAGCAAAGGA ACGATCCAAG GAGGTATCAT TTCTCCACTA TTATCTAATG TTGTTCTAAA   6120
CGACTTAGAC CATTGGGTCT CAAAGCAGTG GCATACCTTT GAAACAAAGT ATCCCTATAC   6180
```

GAAAGGTTAC AATAAATTTC GTGCTTTAAG AGATACAAAT CTAAAGCAAG GCTATATTGT 6240

CAGGTACGCA GATGATTTTA AAATCATGAC TAATGATTAT CCGTCCGCAT TAAAATGGTT 6300

CCACGCTGTC AAGCTCTACC TAAAAGACAG GCTTAAATTA GATATCTCCA ATGAGAAGTC 6360

TAAAATCGTC AACTTGCGTA AACGCAAATC GGAATTTTTA GGATTCACCA TTTGCGTGAA 6420

ACAAAAGGGG AAAAATGGGT TTGTAATTCG CATATTTCTG ACAAAAAGAA AAATCAAATT 6480

AAAAAAGAGA TTAAACAAAG AATTAAAGAT ATTCAAAAGA GCCCTACAGC ACAAAATGCT 6540

CTCCTATTTA ATGCTCTAGT GCTTGGTGTT CATAATTATT TCAGGTATGC CACTCATATT 6600

AACTTAGATT TGAATAGGAT GAACTACGTT CTTTCATTCA CTACGTATAA CCGATTATCA 6660

AATTGTAGTA CAAGATGTTA TCCCAACAAA GCTTCCCCTA CTTACAAAAA GTTTTACTCG 6720

ACAACGATGA AAACGTATAA AGTAGCCGGA GTCTATTTGT ACCCTCTTTG CGACGTATCT 6780

ACaAAAACAA TTTACGGATT TAATACGGAA GATACACCGT TTACCCCGTT TGGtCGTCaA 6840

aGGTTGGAGC ACAAAAGCTT ACAGTCTCTT GTATACCAAG AATTAAGAAA ATTGATGGAA 6900

TCGAAAATAC TGAATAGAAT GGTTGAGTAT TTAGATAATA GAATATCTCG TTACTCAATG 6960

AAATCAGGCA AGTGTGAGAT TACAAAACAG TTTTTGCCAG CAAAAGCCGT TCACTGTCAT 7020

CACTATTTAC CAAAAAGTCT TGGTGGAGAC GACAAATTTG ATAACTTGAG AATTATTCAT 7080

AAGGATATCC ATCTTTTGAT CCATACCACA AATAAAATGA TAATCGATCA CTACGTCAAT 7140

GAACTGAAAC TGCTGCCAGA GCAAATAGCA AAAATCAACC TTTATCGAAA AATGTGTAAT 7200

CTGCAAAACA TTCAATAGAA AATAAATAGA TGGAACGCCG TATGAGGTGA AAGGCTCACG 7260

TACGGTGTGG AGCGGGGGAA AAGATGGAGA TTATATCAAA TTCTTACCTA TCGTTATAAA 7320

TGTTAACGAC GAAAGAAAGT GCGGTGATTT TAAATAAGCT CAAACAAATC GTGATGTTGG 7380

GTCGTCAGTC TGGTTTCTTT CTAATTCTGG CTTGTCAAAG ACCGGATGCA AAATATTTAG 7440

GAGACGGGAT TCGCGATCAA TTTAACTTCC GAGTGGCTTT GGGTCGTATG AGTGAACTCG 7500

GGTATTCCAT GATGTTTGGA GAAGTTGATA AAAATTTCTT TATGAAACGC ATCAAAGGTC 7560

GAGGCTATGT GGATACTGGA GGGAGTGTGA TTAGTGAGTT TTATACGCCA CTTGTGCCAA 7620

AAGGGTATGA TTTCTTAGAA TCCATCAAAC AAGTAGCACA ATCAAAAGAA AAATGACAGA 7680

TACAGAAAAA TTTAATCTAC TAGTTCATCG GGTTAAAAGT ATGATAAACT AGCTTTGAAG 7740

GTGTCAGGGA GGTGCAATAA AAATGACAGA GGATGATTCT TTACAATATG CCaAGGAGAA 7800

CAAGAAGGAA ATTATTTCTT CTGTTATTGA TGGAAAAGAA AAAGAAGAGG AAAAAACTGC 7860

CATTTTTATG GCTGGAAGTC CTGGGGCTGG CAAGACTGAG GCTGCCCAAA CTTTAACGGT 7920

ATTAAATAGT AATCTTTGTG TGATTGATGC TGATAAGTTT CGTGTATTAT TTCCAGGTTA 7980

TGTTGGGAAT AATTCAGATG AATTTCAACG AGGATCATCA CTCTTGGTAG ATGCAGCTCT 8040

CGATTTAGTT CTTAAAAAAG GCTATAGTTT TATTCTTGAT GCTACTTTTG CGACCTCTAA 8100

GGTAAAACAA AATATTGAAC GGGCACTGAA AAAGAATTAC AATGTGCTTG TCTATTATGT 8160

TTATCAAGAT CCATTTATTG CTTGGGATTT TACCAAAAAA TGAGAAGTGG TTGAAGGGCG 8220

TTTTGTTTCT AAAGAGCGTT TTATCAATGC TTTTTTCCAA TCACGAAAAA ATTTGATGCG 8280

TGTGAAAGAG CAAATTTCAA GATCAAGTGA CAATCAATAT CTTAGTGAAG GACTTTCAAA 8340

ATACGATTTC TGATTTTCTG ATGGATATTG ATAATGTTGA GTTGACATTG CCAATCAGTT 8400

ATACAAAAGA AATATTGGAG GAAGAGCTAC ATGACTGAGA AAGAAAAAGT TGAGGAAATT 8460

ATGGAAAAGT ATAACCGTAA CTTTTCCACA TTACAAAAGA ATGCTTCTGC GAAAGAGCTG 8520

AAGACAGTCT TTAAATTTGT AGCAGATGAA TCCAACCGAA AACAACGTGA ATTGATTGGT 8580

TTGGATAAAG AAAAATAAAT ATGAAATGAG AATAGTTGTA GAAGCCAAGA GCAAAAAATGT 8640

TCTTGGTTTT TTTCTATGCA ATGAAAGTAG ATATGGAAGG AAAGGAAATC ACAATCAGCA 8700

ACTCAACTGA TAGGCACTGA CCGCGGAGGA TTTGCGAAgc AGAAATCGTC CGTGGTCCTG 8760

TGCCTGTTCA TGGACGGCGC TAGcgGCCAT GATGTCTCAC CCCCCGCATC TAACAGGGGG 8820

GTACAAAAAC ACAAGAAATA AGCAAGCAAC GATTAAGAAT TGCTGGTGTA GCAAGGATAT 8880

TAAGTAATTT GCATAGGTGT TAACTCATTT GAAAAAGTTA ACACCTTGAA TTTGGAGGTG 8940

AGAACTTGGT ACAACGGAAT TTAGATTATC GGTTATTAAA AGATCGTCGC AATGAATATG 9000

GTATTTCGCA AAATAAATTA GCTACCGCTT GTGGGCTTAG TCGACCCTAT TTAAATCAGA 9060

TAGAAAACGG TGGCGTAACT GCCTCTACAA AAACCATGAG AAAAATTTTT AACCAACTGG 9120

AAAGTTTTAA TCCCGATTTA CCTTTATCGT TACTGTTTGA CTATGTAAGG ATTCGTTTCC 9180

CCACAACGGA TGCACGGAAA ATCATTCAAG AGATTCTCCA CCTGAAATTT GATTATATGC 9240

TCCACGAAGA TTACGCCTTT TACTCCTATC AAGAACAATA TGTCATGGGC GATATTGTGG 9300

TGATGTTGTC TCATGAAGAA GATAAAGGAG TTCTTTTGGA ATTAAAGGGT CGTGGTTGTC 9360

GGCAGTTTGA AACTTTTTTA CTCGCCCAAA AGCGTAgcTG GtACGACTTT TTCGAAGATT 9420

GTCTAAAAGC CGGTGGCGTA ATGAAACGCT TAGATTTGGC AATCAATGAT CTAGTAGGGC 9480

TATTGGATAT TCCTGATTTA ACGAAAAAAT GTCAGAAGGA AGAATGTATT TCCTTGTTTC 9540

GTACCTTTAA AAGTTACCGT TCGGGGGAGC TGTTGAAAGC TGATGAAAAG GATGGTATGG 9600

GGAATACCTT ATACATTGGC AGTCTCAAAA GTGAAGTTTA TTTTTGTTTG TACGAAAAAG 9660

ATTATGAACA ATATATCAAG CTAGGGATTC CATTAGACAA AACTGAAACG AAAAATCGCT 9720

TTGAGATTCG ACTAAAAAAT GACCGTGCCT ATCATGCAAT TCAAGATTTG TTAAAGGGTC 9780

GTAGCATTGA AAGTACCACG TTTTCCATTA TCAATCGTTA CTTGCGATTT GCGGATAAAG 9840

TAGAAGGCAA AAGGCGAACT AACTGGCCAT TAAATGAACA ATGGGGGCGT TTTATTGGAC 9900

GGAATCGGAA AGAGATACAA CTGACTTCTG AGCCGAAACC GTATACCATT GAACGAACCT 9960

TAAATTGGCT AGGTAGACAA GTCGCCCCTA CGTGGAAGAT GGCCAAGGAG TTGGATCGGT 10020

TGAATCAGAC AACCTATATC CAGGATATGG TACAGAATGC GCGCTTATCA GACCGGCATA 10080

AAAAGATTTT GGAACAACAA AGTATGGCAA TCGAAAACTT GATTATATGA AAAGAGGAAT 10140

TGTGGTAGGA AGTGTAAAAG ATATGTTATT GGCATTTGGT GGCAGCATGC TTAGTGTTAC 10200

CGTGATGTGC TTGATGCATG CAAGTGTAGC GGCGAACCAA GCAATGGAAA AGAAAGTCTA 10260

GTGGTAATAA GTCAAGATAA TTTTTCAAAA GATTTTTAAA GGTAGTTAAT TTTTGAAGCG 10320

CACTGAACTA AACCCGACCA AAATTAAAAA TTTTTTCGGG TACTTCCAAG GATTAGCGTC 10380

AATAACATAC TGTTATTTAT CTATCACGCT CCTTGGTGGC ATATAGAGTT GATGGTTACG 10440

TAGTAGCGTA TCCACCAGAC GCGTAAATTT TCTTGCGGTT AGTACGATGG CTCGTTTGTG 10500

TTTATGCTTG GGTGTTTCAG CGTATTTTTT CTTATAATAC GCATGGTACT CAGGCTCGTA 10560

TCTCCTTACG GAGTTGGTGG CTTCAACTAG ATAATAACGG AGATAACGAT TGCCTCGTTT 10620

GGCCATAGGA GTATTTTCAT ACTCAGAGTT ACCTGATTGG TTCTGTCGCC AATATAAACC 10680

GGCATATTTA GCCAATTTGC TTTCATCTGA AAAGCGATTG ATAGGGCCAA TTTCAGCGAG 10740

TATACCCGCA GCATAAACAG GACCGATTCC TGGAATACTG GTTAAACATT GATATTCAGG 10800

TAGCACAACG ACTAAATCGT CAATGGCTTT GTCTAAGTCT TTAATTAACT TTTCAAGACT 10860

ACGAATTTCT TTGGCTAGAA CACCCAAAAC AATATCTACA GAGTCCTGAT GGACTTTTCC 10920

TAACCGGTAG GAACTACGTA TGGCTTTAGA AATTGCCTTT GCAATTCCtT CGGGaTTTTT 10980

AAAACGATAC TTCCCTAGTT TTTGAATTAG GTTCGCAAAC TCTTCTAATG ACAGTTCGGC 11040

AAGTTGGTCC ATTGTGTAAT CTTCCGTCAT TAGAGTTACT AATGTGGcAC TTAAAACGGA 11100

TCCACCTTCT GCTTTTAGTT CTTTAGAAAG CGTGTTGCAT TTATAATAAA TGTTTTCGAT 11160

AAAGTGTTGC TTTGTACGAA CCAGCTGTTT AATTAGCTGA TAACGCGTTC TAGTTAAGTG 11220

TTGGAGGGCA AGGTATTCTT CCTCTTTAAT GATGGAGTTG ACTTGCCTTT GGATGCGAAA 11280

ATAGTCAGCG ATATAAAAAG CATCAATTTG ATCGTTTTTG TTTTCTTCAA AAATATCACG 11340

ATACTTTTTA ATTTTATTTG GTTGTTCAAC TGAAACCATA AGGTTCAATT GATTTAACTC 11400

AAGATCTTCT TTAAAAAACA TAGCTGGATG AAAGCTATAT AAAGATGTGG CCTCCATTCC 11460

AATCACTAAG CTCTCTATCT CAAATTTTTG AGAAAATTCG AGAATGTATT CTTTGATCTG 11520

GCTTGCGCCA AGTTGACTAT TTCCATAAGA TGCTTCTTTT AAAACAGAGC AAGTTGAATC 11580

GTCTGTCATA AAGCAAGCAT CTAATTTTTC AGAACTAACG TCTAAACCGA CAAATAGTTT 11640

CATGTGGAAG GACCTCCTTT CATTGATTAT TCGGAAATTG CTTGGGTACT GTGGGATATC 11700

CCAGAATCAC GTTGTCTAAC GACAGCCTCG CATAATAGAG TTTCATTCAT TAAACCTCTA 11760

AGCCACCCAA TCTcTACTAG GaTTGGTAGA GGTCAATGGa CTCAGCTAGT GTGTTGGaGG 11820

TTATACCAAC GGTTCGGGTC ACAGACTTTC TGTCGCAGTA AAACTGCAGA AGGTGATAGC 11880

GTCTTCCCGT AAGTCCTTTC CAAGACTTAT TATTCAGGAA TATCCTACAG TATCCAAGTA 11940

GTAATCGTTA TTTTATTTAA ACATATTAGG AATAGCCCCT AAGGGmGwTT GAAGGAATGA 12000

AAGTTTCTTC AAAATATATA TTACGAGGAA AGTGACAAGA TGAATTTCGG ACAAAATTTG 12060

TATCAATGGT TTCTAACCAA TGCCCAATCC TTAGTCTTAC TGGCAATTGT GGTCATTGGT 12120

TTGTTTTTAG GATTTAAGCG GGAGTTTTCT AAATTGaTTG GCTcCTGGTG ATTGCCTTGA 12180

TTGCGGTAGG TCTGGTCTTT AATGCTTCtG GTGTAAAAGA TGTCyTACTC aATTTATTCa 12240

ATCgAATCGT CGGTGCGTAA GGAGGAGTCA CGATGGAACa GATGCGTGTT TATATTGCGa 12300

ACTTAGGTCG ATATAACGAA GGCGaATTAG TTGGTGcTTG gTTTACACCA CCTGTTGATT 12360

TTGATGmGaT AAAAGAACGA ATTGGATTAA ATAATGAATA TGAGGAATAT GCGATTCATG 12420

ATTATGAACT TCCTTTTGAG ATTGATGAGT ATACACCAAT TGAAGAGATT AATCGGTTGT 12480

GTGGTTTAGC GGAAGAGCTT GAAGGAACGC CGATTGGTGA GGTTGCTTCA GAGATTCAAC 12540

ACGCTTTCTT CAATTCATTT GAAGAAATGG TGGAGCATCT GGATAATATT GTTTATTATC 12600

CAGATTGTAA TGACATGAGT GATTTAGCGT ATTACTTGAT TGAGGAAGCA GGCGCTTTAG 12660

GTGAAGTTCC TACTCATTTA CAAAATTATA TTGACTATGA GGCTTATGGT CGAGATTTAG 12720

AAATTGAAGG GAATTATTTG GTTACGAGTC GTGGGGTCTT TGAGCTATGT GAGTGAGAGT 12780

ATATTGAATA CTGCCCTTAA AGATACTACA ATAGAAAATA TCGAGTAGTA TTTCAAAGGA 12840

GTGGAAGTGG TGAATATTGA TGAGGCAAAA CAGCTTGCAA ATTATTTTTC AGACGGATCT 12900

CCTGGATATA TATTGGAATT CAATAATCTT GAATGGGACC TATATACTAA GAAAATAGTT 12960

GGATTTGGAA TTCAAGAAAG ATTTGGTGGC TCTAAAGGAA ATTCCTTCAA AGAATTTCTT 13020

GTCTCAGAGG ATATTAGTGA TCACCTAAAG TTCAAATTAA TATACTCATa AATGaAACTA 13080

TGAAGCAAAT gCTTCTATTC GCCAATTTGa AAATCCAAGT TCaTATAAGT CaCCTAAAAk 13140

GAATCCACTT GtGACTCyAw TGaTTgAAGA GGCaTATCaT AAGaAAGATA GTGCaAGATT 13200

GCtTAATGGT ACTAAGTTTT CGAAAGAAGT GAATTTAGAG TATATTCAAC AACTTCCTAC 13260

TCGTATAAGT AAGGATATTC AAAATAAAGA TTTCGATTCT GTTTTAACAA AATCAAACAC 13320

TATGATTGAT GAAGTCTTGA AATATATCAT TGAGAACACA AAAGGAAAAA GCATTGATAA 13380

TAATACTGTA AGGTCTAAAG ACTTAAGAAA TCAAGCCTTT GAAAGACTAA ATATAAAAAT 13440

TGACAAAGAT ATGGATAmAa GAATAAAaGC TTTAGTTGGT GCTCTAAACA CTTTGyCAGA 13500

CAAAATTCTy GAGATGAGGA ATAGTCAAGG TGATGCCCAT GCTCAGGGCT CCAGGAGGAT 13560

TGTAATTAAT GAAGAAGAAG CTATCCTTAC TGCGAACTCA GCAATGATTT TATGTGAGTA 13620

TCTATTTCGA AAGTATGAAA AGAGAAAAAG GTTCAGTGAT GCAAAAGAGA GTTGATTtTT 13680

ATCAGCTCTC ATTTTCTATC TTTyCAATTT TTGCGAATAG CTTCATAAGA ATACTAGTTt 13740

GTTCTATAAC AAATTCTACT TCTAGGTCAG TGAATGTTTG ACTATGTTTA ACATTTTCAT 13800

TTTGATAGCT TATAAGTCCt TTAATAATTA AACGGACCAA ATTGATTATT TCTTTTGGAA 13860

TGCCCtTTTT TCTCAATAAG GGCAGGAGTT CTCCTTCGAT AGGCTGGTTT TCTAAAGACT 13920

TTGAGTTTTT TAAAATATCC TGCATAAAAA ACTCAAAGGC CAGCCTGATA CTGTCTAAAG 13980

TATAACGGTT ATAGATGCTG TTGTCGTAAG ACTGTATGGC TTTTTCAAAT TCTTCTTTTG 14040

CTCTTGAGGA AGCAGTCAAC CAAGTGTTAG TTTCATTGAT GATATTTCTA TCTATTTTAT 14100

CATTAGAATA AACTGAACCA TAATTTTTTA GTAACAGTTT TAAAATCTTA GCAACGTCTT 14160

TATTTGCTGG ATTTTGAGCA AGCACACTTT CACAAAGCTC TTTTATAATT TTAAATTGCT 14220

GTTCTTCTGT GAAAGATTGT AAATTATTTT GAATCGCGGT AGCCTTAGAA GTAAAGTTCG 14280

gAAAGGGAGT CGAAATAAAG GGAACAGAAg AGTTAAATTC TGCTCCATAT TCAGCAAATA 14340

AAGTTGCAAT CTCAGTTCCC GAGTAAACTT CGTTAAGTGT TCCACTTGCC ATTGTTATTA 14400

AGTTTTGCAT ATATACCTCC ATAAAAAATT AAGCATGATG TTTATATTAT ATTAGAnAGT 14460

GAGCCGATTG AATATGAAAA AAATAAAAAG CTACACCAGT ATCTGGGCAG TTGAAAAGGT 14520

CATTTATGCA ATCAATGACT TTCAGCTACC TTTTCCTGTG ACCTTTAATC AGATGGCTTG 14580

GTTTGTTTTG TCCTTGCTGT TTGTCATCGT TTTTGCCCAT GTGCCTCCGC TATCTATGAT 14640

AGAGGGGGCC TTTTTAAAAT ACTTAGGCAT ACCGGTGGTG GTCACTTGGT TCATGTCGCA 14700

AAAAACCTTT GATGGTAAGA AACCACTGGG ATTTTTACGA TCCTTTCTCA GTTATCACTT 14760

ACGGTTTAAG GTCACTTTTG CAGGAAAGAA GGTCAAAGAA CAGAAGAAAC GATGGGATGA 14820

ACCCATCACT CTAGTAAGGA GTGTGAACTA TGTACCCGAT TAAGTATGTT GAAAATAATC 14880

TTGTCTTTAA tCAAGAGGGa GAATGTTTTG CCTACTAwGA ACTTGTGCCT TAwAACTATT 14940

CCTTCTTATC GCCTGAACAG AAGTATCAAG TGCACGATAA CTTTCGTCAA TTGATTGCTC 15000

AGAATCGTGA AGGCAAGATC CATGCCTTAC AAATTGCCAC AGAAAGTAGT ATTCGTGCCA 15060

CGCAAGAACG ATCAAAAAAG GAAATTACAG GCAGACTCAA AGAAGTCGCG AAACAACGAA 15120

TTGACCTTCA AACAGATGCA TTAGTATCTA TGATTGGCGA TAGTCAGATT GATTATCGCT 15180

TTTTTATTGG CTTTAAACTA ATTGCAACAG ATGAAGAGGT CAATTTAAAG AGTCTCAAGA 15240

AATCTTTCCT TTCTGGGCTT CAAGAATTTG TCTATGGGGT GAACCATCAT TTGATGGGTG 15300

ATTTTGTTTC CTTATCCAAT GAAGAAATTC GCCGATATAC CAAGCTTGAA AAACTAATGG 15360

AAAGTAAATT GGCTAGAAGA TTCAAAGTCA GACGGGTCAC ACCGAGTGAT CTGACGTATT 15420

TAATTGAACA TATTTATGGC GAAAAAGGCA CACCGTTTGA AGAATATGAA TTTCAGCTAC 15480

CAAAGAAAAA GCTAAAATCA GAAACGCTGG TGAAACGTTA TGATTTATTA CATCCTAGTC 15540

GTTGCTTGAT TGAAGAAAAT CCTCGGTACT TACGCATGGA ACATGAAAAT CmTGAATCGT 15600

ATGTAGCCTA TTTGACAATC AATACGATTG TAGGGGAGAT GGAGTTTCCT TCTTCAGAAC 15660

TTTTTTATTA CCAGCAACAA CAATTCACTT TTCCCATTGA TACTTCAATG AATGTGGAGA 15720

TTGTGACCAA TAAAAAAGCT CTTGCCACTG TTCGCAATAA GAAAAAAGAA CTGAAAGATT 15780

TAGACAACCA TGCCTACCAA TCCGATAATG AGACAAATTC CAATGTCTTA GATGCTTTGG 15840

ATTCAGTAAA TGAGTTGGAA ACTACCTTGG ATCAATCCAA GGAGTCGATG TACAAATTAA 15900

GTTATGTAGT TCGAGTGAGT GCGGAATCCG TGGATGAATT AAAACGCCGT TGTGATGAAG 15960

TGTTCGATTT TTATGATGAT ACTAATGTAA AACTAGTCCG ACCCTTTGGC GATATGATGG 16020

GGCTACATGA AGAATTCTTG CCCTCAAGCA AACGCTATAT GAACGATTAT ATCCAATATG 16080

TTACTTCTGA CTTTTTAGCG GGACTTGGTT TTGGGGCTAC TCAAAAGCTA GGGGAATTAG 16140

AAGGCATCTA CTTTGGCTAC AATGTAGATA CGGGGCGGAA TGTTTACCTG AAACCGGCAT 16200

TAGCTTCACA AGGGGTCAAG GGTTCAGTTA CCAATGCTTT GGCTGCTGCG TTCCTTGGTT 16260

CACTCGGTGG CGGAAAATCC TTTAGTAACA ATCTCTTGGT CTACTATGCT GTTCTTTTTG 16320

GTGGCCAAGC TGTGATTGTC GATCCAAAAG GAGAACGAGG TGGTTGGAAA GAAAACTTGC 16380

CAGAAATTGC GGATGAAATC AATATTCTCA ACTTGACTAG TAAACCAGAA AATCAAGGCT 16440

TACTTGATCC TTATGTGATT ATGAAGAAAA AGAAAGATTC AGAAAGTTTA GCAATTGATA 16500

TCCTGACTTT TTTAACAGGA ATTTCCAGTC GGGATGGGGA GAAGTTTCCA GTATTGCGCA 16560

AAGCCATTCG GACGGTTACC CAAAGTGATC AACGAGGATT GCTGCAAGTC ATTAATGAGC 16620

TTCGAAAGGA GAACTCTCCA GTCGCGGAAA ATATCGCCGA TCATATCGAG TCCTTTACCG 16680

ATTATGATTT TGCCCATTTG CTTTTCTCAG ATGGTTCGAT TACCCAATCA ATTAGTTTAG 16740

ATAAGCAGTT GAATATCATT CAAGTAGCGG ATTTGGTTTT ACCAGATGCA GAAACCACTT 16800

TTGAAGAATA TACCACGATG GAGTTGCTAA GTGTCTCTAT GCTAATCGTC ATTTCCACGT 16860

TTGCCTTGGA TTTTATCCAT AGTAATCGAG AAATCTTCAA GATTGTGGAC CTGGATGAGG 16920

CTTGGAGTTT TCTTCAAGTG GCTCAAGGAA AGACCCTTTC AAATAAGCTC GTTCGTGCTG 16980

GTCGGGCCAT GAATGCCGGT GTTTACTTTG TCACGCAAAA TAGTGACGAT TTACTGGATG 17040

AACGCTTGAA GAATAATATC GGATTGAAAT TTGCTTTTCG TAGTACCGAT ATTCATGAAA 17100

TCAAGAAGAC TTTGGAATTT TTTGGGTTGG ATCAAGAAGA TGAAAGTAAT CAAAAACGCC 17160

TACGAGATTT GGAAAACGGC CAATGCTTGA TGCAGGATTT GTATGGTCGG GTAGGCGTGG 17220

TTCAGGTACA TCCAGTCTTT GAAGAACTCT TCCATGCCTT TGATACCCGT CCACCTGTTC 17280

AAGAGAGAAT GGAGTGAGGG GATGAAAAAG AAACTTTTTC GGATTGTGGG AATCGTTTTT 17340

ATTAGCATAC TCACTCTGAT AGTTCTATTG AGTTTAGTTG GTACGGTAGC CGAGGCGACA 17400

GGGCTAGTTG ATGATACAGT TGAATCAGGA AATCTATATT CCAAATATTC ACTGAACAAC 17460

TATCAACTAG ATTTCTTTGT GGATAGCTCT TGGGATTGGT TGCCTTGGAA TTGGGGTGAT 17520

GGCTTAGGGA AAAGCGTGAT GTATGGTCTT TATGCCATTA CTAATTTCAT TTGGACGGTG 17580

AGTTTGTATT TGTCGAACGC CACGGGTTAT GTGGTTCAAG AAGCCTATAA ATTAGATTTC 17640

ATTTCGGATA CGGCTGAAAG TATTGGAAAG AATATCCAAA CATTGGCTGA TATTACTGAA 17700

AATGGACTCC AGACCTCCGG ATTTTATTTT GGGTTTCTAT TATTGATGAT TTTAGCTCTG 17760

GGGGTGTATG TGGCGTATAC CGGATTATTA AAACGAGAAA CCACGAAAGC AGTTCGAGCC 17820

GTGATTAATT TTGTGATGAT TTTTCTACTA TCGGGTTCGT TCATTGCTTA TGCGCCTACC 17880

TATATCACAA AAATCAATGA CTTTAGCTCA GATGTTAGTG AAGCGGCCCT TACCCTTGGA 17940

ACCGAGATCG TTGTACCCAA TTCAGAAAGC CAAGGAAAAG ATAGTGTGGA TTTGATTCGG 18000

GATAGTTTGT TTTCGATTCA AGTCCAACAA CCTTGGTTAC TACTGCAATT TGATGATTCC 18060

AACATAGAAG AAATTGGTGA AGACCGGGTC AACAAGATTC TATCTGTGAG TCCAGATGAA 18120

AACAAAGGCA AAGATCGGGA AGAAGCGGTC AAGGCAGAGA TTGAAGACAA TGACAATGCG 18180

AATCTCAGTA TTACCAAGAC CATGAACCGT TTAGGGATCG TGGTCTTTTT GGTATTGTTC 18240

AATATTGGCA TTTCCTTCTT TGTTTTTCTT CTAACAGGGA TTATGTTGTT CTCGCAAATT 18300

CTCTTCATTA TCTTTGCGAT GTTTCTGCCA ATTAGTTTCT TATTGTCCAT GTTGCCGACT 18360

TATGAGAGTT TGGGCAAGAT GGCGATTATT CGTTTATTCA ATACGATTAT GATGCGAGCG 18420

GGTGTCACCT TGGTGATTAC GACTGCATTC AGTATTTCTA CGATGTTTTT CAATATTTCG 18480

GCTACCTATC CCTTCTTTAT GGTAGCTTTT CTACAAATTG TGACCTTTGC CGGTATCTAT 18540

TTCAAATTGG GTGACATTAT GAGTATGTTC AACCTTCAGA GTAATGACAG TCAGTCAATG 18600

GGCAGACGGG TCATGCGAAG ACCGCAAATG TTGATGAACC GTAAATTGCG GCAACTCAAT 18660

CGAAATGTGG GCCGAACTCT GGCTTTTGGC GGTGCTGCAG CAGTygGaaT GGCAAAATGC 18720

ATTACtCTGA TTTATGGTAC AATTTTCCCA TTAATCCGCC GTATGTACCA ACGGAAAACC 18780

CAACAGGTAT TTACAAACGG ACATTTGCTA TAGATGAAAC ATTTCACGAT AAAAAAATCA 18840

TTTTACGTTT TTGTGGCGTA GATTCAGCTT ATCATGTTTG GGTGAACGGC CATGAAGTTG 18900

GCTACAGTAA AGGTGCACGA AATGAAGCAG AATTTGACAT TACGCCTTAC GCTAAAATTG 18960

GTGAAACAAA TGATTTAACG GTGCGAGTCT ATCAATGGTC AGATGGTACG TATTTAGAGG 19020

ATCAAGATAT GTGGTGGTTA AGTGGTATTT TTCGTGATGT TGAATTACTC GGTGTACCAG 19080

AAAATGGGCT GGAAGATTTC TTTATTATTT CCGATTTAGA TGATTCGTAT CAAAATGGGC 19140

ACTTGGCAAT TACGGGTAAA TTTTGGCAAG ACAAAGGACA ACAAGTTCAA TTGGAGTTGA 19200

TGGATCAGCA GGGAAAAACG GTTCTTAAAG AGACTGTCGC AGGAAATCAA GGTACGGTTG 19260

AATTTTCAGC AAGTTTGCCG TCTGTCACTG CTTGGTCAGC TGAAAAACCA TACTTGTATC 19320

AACTATTTAT CACTGTTTTT TCTGAAGGGG AAGTCGTCGA AGTTATTCCG CAAAAAGTTG 19380

GTTTCAGAAA TATTCATGTA TCTGGTGAGA CTTTCTTAGT CAATGGTGTG GCCATTAAAC 19440

TAAAAGGAAT GAATCGTCAT GATTATAATC CGAAAAATGG TCGAGTGGTT AGTCGTGAAG 19500

AAATTGAAAA AGATATTCGT TTAATGAAAC AATTTAATAT TAATGCCATT CGAACCAGTC 19560

ATTATCCAGC ATCTGCTTAT TTCTATGATT TATGCGATGA ATATGGAATG TATGTGATTG 19620

ATGAAACTGA TTTGGAATGT CACGGCTTTG AATTAACCGG AGAATATGAT TGGATTAGCA 19680

ACGATCCAGA ATGGGAAACG GCTTATGTTT CTCGGATGGT ACGGATGATT CAACGGGATA 19740

AAAATCATCC ATCAATTCTC TTTTGGTCGT TAGGGAATGA ATCTGCGTTT GGTCATAACT 19800

TTATTGAAAT GGCCCGTATT GCAAAAGAAA TGGATCCAAC ACGTTTGGTT CATTATGAAG 19860

GTGATTTTGA GGCAGAGGTC ACGGATGTCT ATTCTACAAT GTATACGTGG TTGGAACACC 19920

CAACTAGAGA ATTACTGATG AATACGATTA TTGAAAATTC AAAAAAACCG CATATTTTAT 19980

GTGAATATTG TCATGCGATG GGTAATGGCC CTGGGAATTT AAAAGAATAC CAAGAGCTAT 20040

```
TTTATGCTCA TGATAAATTA CAAGGTGGCT TTATTTGGGA ATGGTTTGAC CATGGCATTG    20100
AATCTGTAAC AGATAACGGT GAAGTTTATT ATCGTTATGG CGGCGATTTC GGCGATGATC    20160
CAAGTAACAA AGATTTTTGT ATTGATGGGA TGTTGATGCC TGACCGTACG CCATCTCCTA    20220
GCTTGTATGA ATACAAAAAA GTCATTGAAC CCATTACAAC TAGCGCTATT GACGTTCTGT    20280
CAGGTGAATT CTCATTATTG AGTCGGTTTG ACTTTGAAAA CTTGGCTATC TTTAAGCTGG    20340
TTTATACAAT TACTGAAGAT CAAACAGTGA TTCAATCAGG GACAGTAGCA GTGCCAGCAA    20400
TTGCTGCTAG AGCAGAAGGA AGACTGCACT TACCTTATCA TTTAGATTTT CCTAAAAAGG    20460
CAGGAGCCGC CTATTATTTA ACTCTTTCAT ATCAATTGAA AGAAACGACT GCCTATGCAT    20520
CTGCTGGTCA TGAATTAGCT ACCGCACAGT TTGAATTGCC AATTGCAACA CCAGGGATTG    20580
AGATCACACC AGTTGGCTCG CTAATGGCAA AGGAAATTGG CCCTCATTTA TACATTGAAG    20640
GACCAAACTT TTCAATTAAT TTTGATAAAG TGAAGGGCGC CTTAACGAAT GTGACACGAG    20700
ATGGTAAAAA ATTATTGCAT AAAGGGCCGA AATTTACATT TTGGCGGGCA CCAATCAGTA    20760
ATGATATGGA AATTATTGAT GAAATGAAGA AGAAATATTT CTTACATTTA GAGCATGAAA    20820
TTGTCCGTTC TTTTGAGTGG AAAAAAGTTG ATGATTTTAT CCAAGTCATC GTAAAAACAA    20880
TTAACGGAAC GACTAACAGT GCTTGGCATT ATCAATGTAC CTATCAGTAT TTAATTGCAC    20940
CAAATGGCGA AATTTTCTTC GACTTAAAAG GTAGTCCTGC TGGTAAAATT GAAAATGCGC    21000
CAGATATGTT ACCACGTTTG GGTGTGACTC TCCATTTAGA CAAGTCTCTA TCAGAAGTAA    21060
AATATTTCGG TAAAGGGCCG AGAGAAAACT ATGTAGATTC ACAAGAAGCT GGTTTACTTG    21120
GCGTTTATGA TGCTACAGTG GCGGAAATGT TCACTAATTA TGTTGTTCCA CAAGCCAATG    21180
GTAACCACAT GGCGACTAAA TGGTCCGCAT TCACGGATGA TCGCGGCCAA GGGGTAGTAG    21240
CAACAGCGGC TGATTCATAT AATTTTAGTG TTTCTTATTT TGAAGAACAA GCGCTAGATG    21300
TTGCTAAGCA TACCAATGAA TTACAAGAAA GTGAGTATGT TGTTTTAAAT ATCGATTACA    21360
AACAAAACGC CCTAGGTAGC TACTCTTGCG GCCAGTGGCA ATTAGAAAAA TACCGAACTA    21420
CGTTTGAAGA ATTCCAATTG GCATTCCGCT TAACACCTTT TAATAATAAA GAGATTCAAG    21480
CAGCGGATGT CGCACATGAA CGAGTAAAAC GACCAACAAT TTCTTAAATA AAAGCTCAGA    21540
ACAGTGCACT TCGGTCTCCT GTTCTGAGTT TTTTTAATGA AATTATCAGA AATTATTGCT    21600
TTTTTTTCAT GGAAATAAGC AAACTATTTT GTTATCTAAT AAAAATAGTT CATAATAGTA    21660
ACATAGCGAA TTATTCGCTA GATAAAGCAT TTAGGAGGAA TTATTTTATG CATACTTTAT    21720
GGGTTTTAAT CGTTGGTGCT GTTATCGGTG CTAtCgcTGG CGCGCTTACA AGCAAAGGTC    21780
AATCAATGGG TTGGATTGCG AATATCGTTG CCGGTTTAGT TGGTTCATTT ATCGGAGAAA    21840
AAATCTTAGG TAGCTGGGGT CCTCACGTAG CAGGTATGGC AATTGTGCCA TCAATTATTG    21900
GGGCAGTAAT TTTAGTTGCT GTGGTTTCAT TCTTCTTCGG ACGAAAAAAT AGCTAATTAA    21960
GTAAATAATT GTTTAATTAG TTAAAGGAGG TCGGGACAGA ACCGTTTAGC TCCGAGAAAT    22020
```

AAGAAGAAAT TTCCGAAAAT TGCTTTTTAA TTTTTGGATA ATTTCGGCTT ATTTCCGAAG 22080

GAGTTGGTTC TGTTCCCGCC GTTTATCAGT TTTTGAGCGT GGAGCAAAAA TCCAAAGTGA 22140

TTTTTGTCCC ACGCTCCTTT TTTTATTTTC TCGTTAATTT TACGACAACT TCACAACGAG 22200

CGGTTTGTGG AAACATATCA ACAGATTGTA AATAGTCTAC TTGATATACT TTGGCTAATT 22260

GAACTAAATC ACGTGCTAAC GTGGATACAT TACAAGAAAT ATACACCATT TTCTTAGGTG 22320

GTTGTTTCAA AAGGGCCGTT AATAATTTAC GATCCAAACC AGTCCGTGGG GGATCAACTA 22380

CAATGGCATC AGGTTTAAAG CCTTCTTTGA ACCATTTAGG TAATAAGTCT TCAGCTGTTC 22440

CAACAGCATA ATGTGTGTTC GTAACACCAA GTCGTTTTGC ATTGAATCTA GCATCGTCAA 22500

TTGCGGCGGG AATGGTATCC ATACCACGAA CTTGATGCGC TTGTTTCGCT AAACTTAAGC 22560

CAATGGTCCC GACGCCACAA TAAGCATCAA CAACAGTTTC GTTTGGTTGT AAATCTAAGG 22620

CCTTGATTCC TTCATTATAT AAAACTTCCG TTTGTTCTGG ATTCAATTGG AAAAATGCTC 22680

GCGGCGATAA ATCAAAGACA ACGTCGTTAA TGTGTTCTTC AATGCTTTCT TTGCCCCATA 22740

AGTGAAGGGT ATCATCACCC ATTACTAGAG AGGTTTTTTT ATTTTGGACA TTTTGCATAA 22800

TAGAAACAAC TTCTGGTAAT TGTTCATTAA TTGCGCGAAC CATTTTTTCT TTTTGTGGAA 22860

ATTTCTGCGA TTGTGTAATG AACACTACTT GGACTTCGCC AGTTTGGATG CCCACTCGAA 22920

CCATTAATGT TCGAAAGATA CCACTATTTT TACGTTCATC ATAAATTGGT AATTGAAAGT 22980

CATTCAATAA ATCAACTAAT GTATTCATCA CCTTGGTTGT CGCAGGCTGT TGGACCAAAC 23040

AGTTATCAAT CGGTACTAAT TGATGTGAAT CAGCTTGATA GAGACCAGCT TCTACTTGGC 23100

CATCGATTTG GCGTAATTGG AATTGGGCTT TATTTCGATA ATTCCAAGGA TTATCCATCC 23160

CAATTGTTTT TCGTAGTTCA TAGTTTTGAT AATTGGCTGG TTTAAATTTG TTAAGCGCCT 23220

GCTTTAATAA ATCTTTTTTA AAGTCTAATT GTGCTTTGTA GGCTAAATGT TGTAGTTGGC 23280

AGCCGCCACA TGTTTCATAG ACAGGACAAG GAGGGACCAC GCGATCTTTA GCAGCTTTTT 23340

TGACAGATTG AAGGGTTCCT TCAGCAAATT TTGGTGTAAC ATTTGTGATC GTGGCAGTCA 23400

CTTCTTCTTT CGGTAGCGCC CCAGGAACGA AAATAATTAA TTTTTTATAA TAGCCGATAC 23460

CTTCGCCATT AATTCCTAAT CGTTTAATTT TTAAAGAAAT TTGTTGGCCT TTTTTTACTA 23520

GTTGTTGGGT CATATTCTTT CATCCTTTAG TCATCTCACA TTTTTCACTT CCTCCCATTT 23580

TAGCACATTT TTATTTATTT GGCTTCCAAG GATTCTATGG TATAGTAGAA TATATGTTCT 23640

GATTTAGAGG GGCAGGAGAA TAAGATGACA ACAATCACAC GAATTAGTAA AGATAAAGGT 23700

GAATTTTATC TACTTTGGCT ATCTTCAGGC GAAAAACTGC GTGTCTCGGA AGATATTTTA 23760

GTACGGCAGC GTCTGCTAAA AGGCCAAGAG TTATCAGATA CTTTAATAGA AGAAATAAAA 23820

AAAGCCAGTT CTTATGATGT CGGGTTACAA ATGGCGATGA ATTATTTAAG CTATCAATTA 23880

CGTTCAAAGA AAGAGATTTT TACTTATTTA AAAGAGAAAG AAATTGTACC TGAAGATCGA 23940

GTAAAAAATCG TTCAACGTTT AGAAGAGCTG CGGTTGTTGG ATGATGCAAT TTTTAGTGAA 24000

```
AGTTACGTAC GGACAGCTAT GCGAACCAGT GATAAAGGAC CACGAAATGT GGCACAACAA     24060
TTAAAGCAAA AAGGTATTAG TGAAGAAGAT ATTCAGCATG GTTTGACCTT TTATACGTTA     24120
GATGAACAAC TTAATGTTGC CACAGCAaCT GCAGAAAAAG CCATGAAACG CTATCGAACT     24180
AAAAGTTTTa AAGATGCACT TCaGAAAATG CGTCTGCATC TGATGCAAAA GGGCTTTACA     24240
AATGAAATCA TTGATCTAGC TTTGGAATCT TTAGCTTTTG AAAAGGACGA AGAACAAGAA     24300
CAACAAGCGT TAGACAAGGA AGGCGAACGT TTATGGCGTG CCAATCAGCG ATTTGATTTT     24360
TCTAAAAAAG TCCAAAAAGT CAAACAAAGC TTATTCCAAA AAGGGTTTGA CTATGATTTG     24420
ATTCAGCAAT TTATTAGCGA AAAGGAAGTA GAACATGACG AATAAAGAGC CTTGGGAGGC     24480
GTGGAGTCCC GCAAAAGTTA GTTCATTCCA GGAAGATTTT TTAGCTTGGT ATGAACGAGA     24540
AAAACGCAAT TTACCTTGGC GAGCGAATAC AGATGCATAT CGTATTTGGA TTTCTGAAAT     24600
TATGCTACAA CAAACTCGCG TAGATACAGT CATTGATTAT TTTTATCGAT TTATGGAATG     24660
GTTTCCGACG ATTCAAGATT TAGCGGAAGC GCCAGATGAT AAGTTGTTGA AAGCTTGGGA     24720
AGGGTTAGGT TaCTATTCAC GAGCGCGTAA TTTAAAAGTG GCAGCGCAAC AGATTGTTTC     24780
AGAATTTGGT GGGAAAATGC CTGACACAAT CGAAGATATT CGGAGTTTAA AAGGAATCGG     24840
CCCCTATACG GCTGGTGCGA TTGGCAGTAT TGCCTTTAAT CTCCGCGAAC CAGCAATTGA     24900
CGGTAACGTC ATGCGGGTAG TGAGCCGTTT ATTTGAAATT GATGCAGATA TTGCAAAAGC     24960
GAGTAGTCGT AAAGTATTTG AAGCGGCGAT GCTTAAAATC ATCGATCGTG AGCGTCCAGG     25020
GGATTTTAAT CAAGCTTTAA TGGATCTAGG CTCAGCCGTC TGCACCCCAA CTTCCCCTAA     25080
GTGTGAAAGT TGTCCGCTTC AGCAATATTG TGCGGCCTAT CAGGCAGATA AAATGACCGC     25140
TTATCCAGTC AAATCAAAAA AAGTTAAGCC GAAAGATGTG TATTATGTCG GCACTATTAT     25200
TGAAAATAAA AAGCAAGAAT TTTTATTAGA ACAACGCCCT GAAACCGGTT TGTTAGCCAA     25260
TATGTGGCTG TTTCCGATTG AAGAAATTAG CAAAAAACAA TTTCAACAAT TGCAAAAACT     25320
GGCGCAACCA GCAGAAACGG AAAAACAATT AACTTTAGAG TTAGAGCCTG TAACAGAGCC     25380
TTTGGTCGCT GAAGAACCTG TCAACTTTTT TACGGACTAT GAAACAGTTG TCTGGCAAAA     25440
ACGTACCTTA GGCGAAGTCG TTCATATTTT CAGTCATTTA AAATGGCATA TCTTAGTCTT     25500
TTATGGCCGA AATACAGGCG AATTAGCCAC ACTAGAGTCG CAGCGATGGG TGGCAGCACA     25560
ACAATTTTCA GACTATGTTT TTCCAAAACC ACAGCAAAAA ATGGTGGAAT TGTTCAAAAA     25620
AGAACACGAA AAAAAATAAC AAAAAAGAGA GGCTTTTTAA CAAACAAGCC CCTCTTTTTT     25680
GTTTCCTCTA GGTTATACCC CTATATAAAT TCTGCTATGT TTAGTTTTGA AAGGGACTGG     25740
TAAAATAGAC TTGTCTCTTG TCTGCTTAGA GGAGTAGTAA TGATGGTGGA GACTTTTTAA     25800
GAGAGAGGAA GTACAGCCAA TGAGTAGGAA GCGAAAAATC AGCTTAATTA GTTTAGTCAT     25860
CATTTTGGTT TTTGTCACAG TCGGCTCAGC ATACTTTGCT GTAGCGGGTA GCTATTTAAA     25920
GAAAACAATT GATAAAGGCT ATGTTCCCAT AAAAAATGAT TATAATGAAG CGCAAAATAA     25980
```

AGATAGTCAA TCGTTTTTGA TTATGGGGCT AGACAATACA ATTGAACGGA AATTAGGCAC 26040

AACTAGGACT GATGCTATGA TGGTGATTAC CGTGAATAAC AAGACGAAGA AAATAACCTA 26100

TTTAAGTTTG CCACGGGATA GTTTTGTTCA AATTGATGCG AAAAATTACC AAGGGATGCA 26160

GCGAATTGAA GCCGCCTATA CCTACGATGG ACCAACAGCT TCTGTTAACA CAGTTGAGAA 26220

ATTATTGAAT ATTCCAATCA ATCATTACGT TGTGTTTAAC TTTTTATCTT TTATTAAGTT 26280

AATTGATGCG GTTGGCGGCA TAGATGTCAA TGTCAAGCAG GCGTTTGATG GTGTCACCAA 26340

AGACGGGCCA GGATCCATTC ATTTTGATGC AGGGAAACAG CATTTAGATG GTACGAAAGC 26400

TTTATCTTAT GCCCGTGAAA GACATAGCGA TAACGATATT ATGCGTGGAT TCCGACAACA 26460

AGAAATTATT CAAGCAGTTG AAGACAAGTT GAAATCTGGT CAATCAATCA TGAAAATAAT 26520

GGACATTATT GATTCGTTAA ATGGAAACAT TCAAACTGAT GTGGATTCCA ATGAATTGAC 26580

TCATTTAGTC AAAGAAGGTT TGACTTGGAC CAATTATGAT AAACAACAGC TTTCTTTTGA 26640

CTGGCGCACT TTTAGTAATG AAGGGCGCAG TATGGTTGAA CTATACCCAG ATAGTATTGA 26700

AAATGTCCGT CATCAATTAC GTGTGTCTTT AAATTTAGAA AAGCCAGATG AACGAGATCA 26760

AGACGGCTAT GTCTTCCATA CGAACGGTGA ATTTTTATAT CAAAGTGATT ATACCGTTCA 26820

AGATGAAGCA GCTGAGGAAA ACGAAATGAC TTCCATCAAC GGCAATACGT ATATTGGTGT 26880

TCCTGGTAAT ACACAGACCG GCCCGTTGCC ATCAGTTAAA ACGGAAAATG GCTTTATAAA 26940

ATAATTTATT TTCCAAGTGT AGAAAATCCG AACATAACAG ATTAAATAAT CCGCTTACTT 27000

ATTATTCTTC TAAGGAAAAA ACGAAGGATT TTAGCAAGGA AACAGCGTAT CATTTGATAT 27060

AAACCCAACA TTCCCTTTAT AATGAAAGTA ACAAGCAAAT TTTTGTTACT TAGAAGGGGA 27120

AGGGAGCAAG GAGTTATGTT TAGTTATACA GATACCATTT TGTCAATTAT GCAAAGAATT 27180

GAAGTATACA ACGAAATTTT CAAAGCCATC TCTAAAGAAG TTCAAGAATA CGGTTGTAAT 27240

CAGGCTGCAC ATCGGAGAGG CCGTGACACT TATATGTTTT GCCGAAATAA TGTGAATCGT 27300

TTTTTTACAG AAGAAATGAT GTTTCGTAAA AATTTAGAGC CGTATTATGA TAAAGAAGCA 27360

ACGAAAATCT TGCTAGAAGG ATTAGATACG TATAAAGAAG GAGTATTTTT CTGGTTAGAG 27420

GCATTAAGTG AATCTTGCCA AGTGACAGAT GAATTGAAAT ATCAATTAGG AATCCGCCAA 27480

AAAGAATCCT CTTATTTATT AATTAATCAA GCTTGTAAAG AAGCCTGTGA GGGCATTCAA 27540

AGCGCTCATA GTGTTCATAA GATGTAAAAA GTTAGGGATA GCTGTCAATC TGGCAGTTAT 27600

CCCTAACTTT TTTATTTCAA TATCATCTGA TAAGCATAAC GTTCGCCATC GTTAACATCT 27660

AAAATCACGT CACCTTGATA AACAAAGCCA GCTTTTTTAA TCAGGTGCTG CATAGCGAGA 27720

TTTTGTGGAT GTGTATCGAT TCGAATATCA GTAGCACCTT GCAAACGAGC GGCTGTAACC 27780

AGTGAGCGCA AAAGCACCAA AGCCAGTCCT TGACCTTGAA AGGAACTTTT TAGAGCAACA 27840

CAATGAATTG CCGTATACGT AGTTTGAGAA TGAGGTTGCC AAGAACCATT AAAAATCGCT 27900

TCATAAGCAG GATCAATCTC CGAACATAAA ACACCAACGC CTGCCAAGGC ATCCTCAACC 27960

ATTAAAATAT AGCAAGTTTG CTGGTCGATA TCTTTTGCTA AAATTTCTTG ATTCGGGCCA 28020

TCGCCGTTTT GCCATTGGGG rATGCCACTT TTTTGTAAGG TGCGGCGTCC ATCTTCAATG 28080

ATAGCAAGGA TGTCAGGCAA ATCAGAAGGT TCTGCTTTTC TTAAGTAAAC AGTCAAAAAA 28140

TATTCCCTCC TCTTAGAAAC CTGTTGTTGT TTTCAATTGG TAAATTTTTG AAGGGCGGCC 28200

AATGCCAACG GGACGTTCAC CTAACTCTTC AAAATATTGT GTCATGGCTT TTTTAAAATT 28260

TGAATGGTCA ATGCTGCGAT AATCAACGCC GAGGAACTTC GCAAAAACTT TACGGGCTTC 28320

CGTAATGGTG AAATCTTTTC CTAAGACTTG AAGCACTTGC GGTTCATGTT CCATTTTATC 28380

AACAACACGA TTAAACGCCT TAATTATAAT TTCACTATGG TCGAAGGCTA ATGTATCTTT 28440

ACCAAGAGAG GCCGCTGTTT TCAAGTCTAA AGTAATTTCC ACATCTTCAT GAGATAACGT 28500

GATGTGTTGA CCATGCCGTT CAAGATTGAA CCAATGGACT TCTTTGGCAT CGTCGCCAGC 28560

AATTAAAGGT TCTTCACCAA TAAAGGCTAA ATAGCTGACT GTCACGACCC AGCCACGAGG 28620

ATCTCGATCT GGGCGACTAA AGCTGTGTAA TTGTTCAATA TTTTCTTGAG AAATCACCAC 28680

ACCAGTTTCC TCTTTTGTTT CTCTTAAAAC ACTATCTTCA GTAGATTCGT TGCGATTAAC 28740

AAAACCGCCA GGAAGCGCCC ATGAGTTTCT AAAGGGGTGA CCTTTACGTT GAATCAGTAA 28800

GACTTTTAAT TGATCAGCTT CCTTGTTGTA ACAAAGAAGA ACCATATCAA CGGTTAAAGA 28860

GGGTTTTTCA TATTCAGGTA ATTCTTGTTG ATGATACCAA GTTAAAAACT CTGCTAAGCT 28920

AGCTTGGCGT TCATAATAAT TTTTTTCTTC GGCTTTAGAA GCAAATTGTG GCACAATTGC 28980

TCCTCCTTTC AAAAAGAAAC ATGGTTTTCT TTTATTGTAT CATTAAGTGG CTGAAATACA 29040

ATATTCAAGC GTAAAAAGAA AAGCAAGCCT GGGACAAAAA TCACTTTGGA TTTTTGTCCC 29100

AGGCTTGTAC TATAAGAAGT TTGTTCTAGA CGTTATTTAT TGTGTTAATT CTAATAATTT 29160

ATTTTTCAAA TCTTCTTCCA TATGGCCAAG TTCGATTTCG GCAGCTTGGC GTTTTTCTTT 29220

TCCTTCTTGT TGAATACGTA ATGTTTCTTG GATAGTTTCA ATTAAATCAT TTTGTGTTTT 29280

TTGtAATGTT TCAATATCGA CTAGGCCACG TTCGTTTTCT TTTGCTGTTT CAATTGCAGA 29340

AATTTTTAGC ATTTCTGAAT TTTTCTTCAG TAAGTCATTG GTGGTTTCAG AGACTTGACG 29400

TTGAGCAGTC ACAGCATCTT TTTGGCGGAG TAAGGTTAAG GCGATGACTA CTTGGTTTTT 29460

CCATAACGGA ATCGCCGTAT TCACCGATGA TTGAATTTTC TCAGCCAAAG CTTGGTTAGT 29520

ATTTTGGATT AAACGAATTT GTGGGGCTTG TTGAATCGTA ATTTGACGAG CCAAGCGTAA 29580

ATCATGGGTT CTTTTTTCTA AACGGTCTAG GAATTGGTTA TAATCATTAA CAATTTGGAC 29640

ATCCATTTGA TCTCCAGATT CTTCCGCTTT TTTCATGGCT TCAGGGATAA TTTGCGTCCG 29700

TAGTTCTTCT GCTTTTAATT CCCCAGCGGC AATATAGATG TTTAAGGCAT CAAAATAATC 29760

TTTATTTTTT TGATACAATT GATCTAACAT GGCATTATCT TTTAAAAGAC CATCTTTTTC 29820

ATGGGTTAAT TTGACAGCAA TTTTGTCAAT TTGCGCACCA ATTTTTTGAT ATTTGGCGGT 29880

AATTTCGTAA ACGGATTGTT TGACTTTACC TAACATTTTT TGAATAAAGT TGCCTTCGCC 29940

AGCACGTAAT TCATTGGGAT TGGCTTCATT TAAACGAAAC ATTAGATCGG TCAGTGAATC 30000

ACCAACAGGA CCAATGTCTT GTGCTTGAAC GTGGTTTAAC ATCGCTTGCG AAAATTCGCC 30060

GAGTTTGATT TGGGCATTTG AACCATAGGT AATTACTGCT TGTGAATCTT GGACATCAAT 30120

TTTTGAAGCT AATTCTTTTG CTTGGGCTTG ACGTTCTTGT GGTAATTTAT CAATTAAACG 30180

AGGAGCTGCT TGCTGTTCTT TTAACTCAGA TAGTTCATTT TGTTGCGTTG GTGTTAGTTC 30240

ATCCATTGAT GAAAATGGAT TACTTAATAA ATCGTCCAAT GTATTGGTTA CAGATGTTGT 30300

TTCTTTGGGT TGTTGCTCAG TCATTGTTTT TCCTCCCTAA TGATTATTGC TGGTCTTCGT 30360

TTACTTGAGG TGATTCCTCA GTTGCAGCTT TTTGCGACAA GCTGCTTTTA GCGATCGACA 30420

TTTCGACATC TAAATCGTCT AAGTCATCGG AAACGATTTC CTCATAATCA TTTTTAACTA 30480

ATTTTGACAA TTGGTCAATG ATTTGTGCGC TTTCTTCTAA TTTTTCATAC GTTTGTTTGT 30540

TTTTTACTTC GTGTTGTTCG ATTTCTAAAT GTTTACTTGT TAAGTCAACG ATATTTGGTA 30600

AATGTGTATA GAGAAAGTGA TTGGCTAAGT GTAACTTTTT AGGTTCTTTC ACTAACTCTT 30660

TAAACAGAGC TTTAGAAACC TTCGTAGTAT CATTACGTAA GTCAATCGCC CGTAATTTAG 30720

TTGAACGATT CATGTTTTCT TGCAATTGTA TGATTTGTTT TTTGGCTGTG CTCATTGTAC 30780

TTCTGAAAAA ATCAATTTCT TGTGGAGATA ACCCCAACTC ACGATAATGG GCTTCTTTAT 30840

CTTTCGTTAA GTAAGGCATT TGATCGCTTT TGTCGGCGGC ATTTTTCTTT TTTCGATAGA 30900

GAACGAATAC TAAAAAGCCT ACCAGAATCA CAAAAGCGAA GAAGCCATAT TCTCTAGCGA 30960

CAGcAATGaT TAAGGCTCCT AAAAGAATAT AGCCAATCAG TAATTTTGTA TTTAATTTGT 31020

TCATTTATAT AGTTCCTCCT ATCTATCTAA TCCTCGAAAG AGTCAGCTCT CGTTACAATA 31080

GTTATTTTAG CATACTCTGT GAGAGAAAAG ACATACAGCG GAAGATGTAT TTTATGATTT 31140

AGTGAAATCC TAAGAAAAAA AGTATAAAAC GCAAAAAAAG ACCCGTCCTA ATTGGGGGTA 31200

GGACGGGAAA GGAGTTAAAA ATGAAAAAGT GTTTTGTTAG GGTTGTTTGT TGGTATGCTT 31260

AAATAATACA AGGAAGTTGT GAATTTTTTA TGTCTATTTT TTTTCGAAAA TGTAATTTTT 31320

TTCTAAAGCA ATAAATTTTT TTCTTTGaAA AGGGTTTAAC TAGAGAAGAA GGGGTGCGAT 31380

TTTTATCATT TTAGGGTATG ATAAACAAGA AAGGTTGGTG AAAGtATGTT AAATTATGAG 31440

GACTTCGAAG rAAAAACACT TCmACgTCGA GrAATTTTTA AAGGrAAGrT TATTGATGTT 31500

TTTTTGGATG ACGTGGCTTT ACCGACAgCG GCACAGCTAA ACGCGAATTG GTTTTTCATT 31560

CAGGCGCAGT AGCTATGATT CCGTTAACCG CAGAAGGAAA AATCGTCTTA GTTAAACAAT 31620

TTAGAAAACC CTTGGAACAA GTAATTTTAG AAATTCCGGC GGGAAAAATT GATCCTGGAG 31680

AAGAAAATCA ACTGGAAACG ACGGCTATGC GGGAATTAGA AGAAGAAACT GGGTATCGTG 31740

CAGGTCAGCT TACCTACATA AATTCAATGT ATCTGTCACC AGGATTTGCT AATGAAAAAT 31800

TGGCTCTCTA TTTGGCAACG GATTTGCAGA AAGTCGAAAA TCCGCGTCCT CAAGATGAAG 31860

ATGAAATCTT GGAACTTTAC GAGCTAACAA TAGCTGAAGC CAAAGCAGAG GTGGCAAAAG 31920

GCACGATTTG TGATGCCAAA ACACTTTTTG CTATACAATA CTGGGAATTG TATTTACTGC 31980

AGAGACAGTT TAAGGAGGAC ACGCAATGAG CAAAGGTCCG TTAGTCACTC GGACAGAGCT 32040

TCGCAAACGC AGAGAAGCAG AAGAAAAAGA AGCGGAACGT CGTCAGCAAG AAGAGCAGAA 32100

GCTGGCGGAA AAAGCGTATA AGCGAAAAGA AAAAGAAATT TCGACGTTTT ATCGTAAAGA 32160

AAAGAAAAAA CaAAAACCGA TCAACAAGTC ACGAGTAGGA GAATACTCGA AGCGTCGAGA 32220

ACGGAGTACT TGGTTAAACA AGGCAATTAT TATTGTAGCG ATTTTATTAG CCGTTGTGGC 32280

ATATATCGTT TTGAATTTAT AGAAAAGAGG AATCACTATG AAAATTGGAA TTATTGGAGC 32340

AATGGATCAA GAGGTCAAAA TTCTAAAAGA AAAATTGACA GACACGATGT CATGGGAACG 32400

AGCAGGCGCT TTATTTGTTT CTGGTTCGTT AGGAAGACAT GAGGTGATTG TGGTTCGTTC 32460

AGGAATTGGT AAAGTGGCCT CAGCTGTGAC CACGACTTTA TTGATTCATC AATATGGCGT 32520

AAATATGGTG ATTAATACAG GTTCTGCTGG TGGTATCGGT GAAGGACTAA AAGTTGGCGA 32580

TTTAGTCATT GCAGACAAAT TAGCTTATTT TGATGTTGAT GTCACTGGTT TTGGTTATGC 32640

CTATGGACAA TTGCCTGGTG GTGTGCCTTT GTATTTTGAA ACAAGTGAAT ACTTGCGGAT 32700

GGAAATTGCG AAAGCCGCTG AAAAAACTGG TTTAGCTATT CGTAAAGGAT TGATTGTTAC 32760

TGGAGATA 32768

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6930 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

AAGTAGAACG TAAATGGTAC GTAGTAGATG CAACAGATGT TCCTTTAGGA CGTCTATCAG 60

CAGTTGTTGC TTCTGTACTA CGTGGTAAAA ATAAACCAAC ATTCACACCA CATGTGGATA 120

CTGGAGATTT CGTAATCGTA ATTAACGCTG ATAAAGTAAA ATTAACAGGT AAAAAAGCTA 180

CAGACAAAAT TTACTACCGT CACTCAATGT ATCCTGGAGG ATTAAAATCA GTGACTGCTG 240

GTGAATTACG CGACAAAAAC TCTCGCCGTC TTATCGAAAC TTCTGTAAAA GGTATGCTAC 300

CTAAAAACAC TTTAGGACGT AAACAATTTA CTAAATTGAA CGTATACGGT GGAGCTGAGC 360

ATCCACATGC TGCACAACAA CCAGAAGTAT TAGATATCAC GAACTTAATT TAAGGAGGGA 420

AATTCATTGG CACAAGTACA ATATAGCGGC ACAGGCCGTC GTAAAAATGC AGTTGCCCGC 480

GTACGCTTAG TACCAGGTAC TGGTAAAATT ACTGTAAACA AAAAAGACGT AGAAGAATAT 540

ATTCCACATG CTGACTTGCG TGAAGTTATC AACCAACCAT TTGGCGTAAC TGAAACAAAA 600

GGCGCTTACG ATGTAATCGT AAACGTAAAC GGTGGTGGCT ATGCTGGACA ATCAGGAGCT 660

ATCCGTCACG GAATCGCTCG TGCATTGTTA CAAGTAGATC CTGACTTCCG TTCTGCTTTA 720

AAACGCGCTG GGTTACTTAC TCGTGACGCA CGTATGGTTG AACGTAAAAA ACCAGGTCTT 780

AAGAAAGCCC GTAAAGCTTC ACAGTTCTCA AAACGTTAAT ATATCTTATA TATCAACATC 840

TCAAGACACT TTCCAAATTT TGGAGGGTGT CTTTTTTTGT GTTTTTAGGT ACGACAATGG 900

GTACGAGTTT TTTAAACCAG CTATTCATTT TATATTTTTT GCTTCTGCTA ATCGTTGTTC 960

AAATTTTTCA TAAGTGGATA GATCAACGAT GTTTGGTGTA TTCACATAGA CTTCTGTTGT 1020

TTTTGTATCA GAATGGGTGA GCGCTCTAGA AATTTGTTCC ATGGTTGCGC CACCTTCATA 1080

TGCTAGCGTA CTAAAAGTAT GTCTTAACTT ATGAGAAGAA GTATTTATTA AGTGTTTGTG 1140

TCGACGTTTA ACTGAGTTGA TGCGGTAGTT AAGATAGTCT ATGTGTACCG GGACATTCAC 1200

ATTTCCCTTT CTATCAATAT AAGTGAAAAG AAATTGCTCT GCATTTTGnT tAATTCCAAG 1260

TTGTAAAAGT TCTTGTGCTT GTTCTTTTTT CCATTTCATT AACAACTCTA TAAGGGAGGA 1320

AGGCAATTGG aATTTAGTGT TTTTTCTTCC TTTGGTAGGA GTTAAATTTC CTAGCTTGwC 1380

TTTGCTTTGG ACAAGTAGTA TTGTTCCATT CTGTAAATCG ACATGTTTCC ATTGCAGAGC 1440

ATAACTTTCA CTTTTACGAT CGCCTAAATT TAATGTTAAG GTAAAAAGGA CAAAATCTTG 1500

AAAGGTTAGA TGTTTGTTAT TGTAGTCTTG TTGAATGGCT TGAATCCACT CAAGAAGTTG 1560

TTCTGCGGTT AGTGCTTCTC CTTTTAGCGT TCACTCCATT TTTCTACGTT GTTTTAAAGG 1620

ATCTCCTACG TAACGAGTAA TCTTAGCAAC ATGATTGTAT TCGATGTAAT CAAGAAGTTC 1680

TGCAATATCA AACAATTGAT TTACATAGCT TTTAATCGTT TTGATATTTG CATATTTTTG 1740

CGCTTTCGCA TTTAATTTTC GCAAAACTAA ATCTTTATTT GTATTCAATT CAAAGAGAGA 1800

GTAACTACCG AACATAGGTA GAATATGAAG ACGAAACAGG TCTTTAGTGT TTAAAATCGT 1860

TTGTTCTGAT GGAATCATTC GGTTTCTCCC ACTAGAACCA TTGCAATACA TGTCTAACCA 1920

GACCTGTTCA TAAAATTCTT TGAAAGAAAT ATTTGCTTTT AATTCAAAAG CTCGTTCATG 1980

TTTTTCAAGT TGTACTTTTT CAATTTTTTT AAGAATATCT TTTTCAGCTA CCAGAGCTTC 2040

ATTTTTGGAA GAGAACGTTT TTCTGTATCG CTGTGAAGTA ACACCCAGAA TTTCTCTTAC 2100

TTCTTTAGGA TAAAATACTT CGACTTTATA ACTTCCGCTT TTCATTTTTT TGATTGCCAT 2160

TAATATTTCC TACTTTCTTG CGTGAATCAG AACGTAGAAA GACTATCGTT TGATTATAGA 2220

ATAGATACTT GAATGTTTAT AGGTTGGTTT AAGCATATTT TTTTAACCAA TTATCAATCT 2280

CCAACTTATC ATAAAGTACT ATTCCGTTAA TACTGATTCG TGGAAGTCCA ACTGTAATTA 2340

ATTTATCAAT AGTATTGTTG GAGAGATTAA GATATTGGCA TAGCTGATGC TTTTTAAGAT 2400

AACGATTATT TAATCCTAAA TCAATTTGAA ATTGTGTGAT TTCATTTTCT AGTAGTTTGT 2460

GAATGGCTTC CTGTAATCTT TGTTTTTGAG TGTCAGACAA TAAAATAGTG AGTTCGTTCA 2520

TTTATGCGTT CCTCCATTTT TCTTGATTTT TGTGTAACGA ACATTTTGTT CTTCAAAACC 2580

TATTTCTTTG TTTGGGATAT TTTGTTCATT CCAGTCTTGA AAGAAGACAG GTCTTTTACT 2640

TACATAATTG TTGGTAAATA AGACTTTTTG ATCTTTTACT TCATAAGGTC GATCTGTCAG 2700

TCTTTTTATT TCGAGCGGAT GCTTCAAGTT TTGAGATTTA AAGAAGGCTT TTTTCTTATG 2760

CTCTTTTAGT TCTAAATCTT TATCAAAATA CTTGCTGACA TACCGTCCTC TATTTTCAGC 2820

GCTATCTATG TCAATTTTGT TGATTTTGAT TAGTCCATGT CCCCAAATAG CTGTTAGTTC 2880

TGCATGAGGT AAATAGGGGA AGCTGAATAA AATAACGTGA TAATGAATAG CGCCATGTTT 2940

TTGTTTCTCC CACGTAGCTA GGTACTTAAT TTGGGCTTTC TTTGTCTTGT ATAAATGGTA 3000

GTTTAGTCGG TTCATAAATT TTTTGAACTC ACTATTTGTA TAGGCTATAT CTGTTATATT 3060

CTTTTTAAAA GTTAAAGTTA AAAATTTTGT CTGGTTATCC AAATTCATAT CAATGAGTCG 3120

AGCAACGACA AAGCGTTGAT TTTTGTAATG TTTTTGTTTT CTCTTTAAAC TGTCATATTG 3180

TCCTTGTGCG GTTAAATCAT CAAATGTCTG TTTTGTAGTT TTTAGCCAAT CAGGTTTTTC 3240

TTTCTTTTCA TCATTATCTG TTTTTCGTTT TGTTAAGATT AATTGTTCAT ATTCCCAAAT 3300

TTCAATGTAC GTAGGGGTTT CAATGATTTT TTGATTGTAT CCATGTAGCA ATCGCAATCA 3360

CTCCTACTAC AGAAATTGAG TGTTCTTACT GGTTATTCAG TTGTCAAAGA ACGAGTGTAT 3420

AACTTTTTGA TTTTTCGCCA GATAATGTTC TATAAATCAA GTTAAGTAGG AAGAAAGCCT 3480

ACGCCTTTCT TCCAGAGCAA TCATCCATCG TGTGGTTGAT TTTCGAGCGG AAAGCGGCTT 3540

GAGCTTTGAT GGCTCTACCG CTTTCGTCGA AACCACCCGC TGGAAGTCTG CTCTTGGGTG 3600

GATAGTGATT CTTCAAGTTG AGCCAAGCGT TGTTCTAATG TTTGATAGCT TCTTTCCAAT 3660

TGTTGGTACT TTTTTGTTAA TGCTTGAAAC AACTCCTCTT TATTTGCCAA CTTAACTTCA 3720

ATTTGATTGA TGATTTCAGT TAGTAATTCA CTATGATTCA TGGTATCTAA ATGATGTTTC 3780

TTATCAAAAA GTTTTCGTAA AGAAGTATTT CTGCCCAATA TAGGGATCAG TTTTATTACT 3840

TTTTGAAATT TTTGGATTTC TTCCAGTGAA AACATTCTGG TTTTGCGGTG ATTGATCGTT 3900

CTTTTAGTTC TTGTGTGACT GGCAAAACGA ATATAATCAT AGTGAAATTG GTAATCGCAC 3960

TCTTTTTGAA TAAAATTTGC CCATTTGTTT AAAGTAGAGA GGGAATTCAG CCCCTCTACT 4020

TTTTTTAGCA CGTCATGGGC GGTGAAGTAA TCAGTCATTT GTTTGGAACT AGTGTGGCTT 4080

TATAGGCTGT AGCAGTTAGT CCATAGCTAA CAAATTGTCC ATTTACACGA CCCCAAGGAG 4140

TAATCATAAG ATTATCAAAT CGAACGATTG CCATACCTTT TTCATCTAAT GTTTCTTTCG 4200

TGATTACAGG AGCTTCACTT TTAACTTTGA CACTGATTTT TTCAAATCGG AGTTTTGGAA 4260

GAACACATTC ATACGTCCAG CCCAACGTTC GCCGTCTTTT TGCCAAAGGG AGACTTGGAC 4320

TACTTGGTAT TCATTTTCAA GCTTTTCTTT AGGTAAAGTT AAATCGTTAA TTTTTACTGC 4380

CATTTCATAT CCTCACTTTC TTTTAGTACA TTTTTTAGTA CAAACATATT CTAATGGAGC 4440

TAGAAAATTT TGTTAATATT TTTTAGCACA AAAATTAGTA CAAAGCTGTT TTACTCTTAG 4500

TGCATTATTT GTTATGATTG TTTTGATAAA AATTGGCTTT TAACCATTGA GGAGATTGAG 4560

CTTTATGAAA CCAAACAAAA AAAAGTAGGC GAAAGAATTC GAGAAATTAG GACAAATCTA 4620

GGCTACAGTA TGGACGAATT TGGAAGTTTA TTAGGTGATT CTCCACGAAt TCTGTGAACA 4680

ATTGGGAAAA AGGGGTCAGT ATTCCTAAAA GAGATAAACT AGAAAAAATT GCGATATTAG 4740

GGAATATGTT ACCAGACCAG ATACTATATG GTCGAGCAGA TGAATATTTA TATGATTTAA 4800

TTGATAGCAG TTTCAATGTT AAATTCAGCG ATAATATTCT TTATGAAATT TTTGAAAGTG 4860

TACCACCTGA AAAGCGTTCT TATGATGATT TGATGTGGTT AGCCGTTGCG AAATATTTTA 4920

TTGATAATGG TACATATGGA AAAGTTGTTG GAAAATTCAA TTATATTTCA ATGCTAGGTA 4980

TACCGAACCT TTATACAGGA AGCTATCAAA ATGAATTTAT TGAACAACGT GAAGATTATA 5040

AACAACAAGA AGTTAAATAC TATGTATATG CAGATCCGGA AAAAAACATA TTACACATTA 5100

TGCCATTTGT TCCTAATGAG AAGAATATAG AACTGCTTTA TGAATTTCCT GAATTATTAG 5160

AACAACAAGC AGATCACGAA GTATTTACTG CAAATTTTGA ACAAATTGGG CTTACCTTAA 5220

AAGGATCAAC CATTGTGTAT TATGGAATTG ATAAACAAAA ATTAGAAGCT AAAACACAGG 5280

CATATAAATA TGATGAAGCG GTTGATTTAT ATAGAATAGC TCAACTACCA GAAACTTTAT 5340

CAATGTTTGA TGATGAAATC CAAAAGCAAA TTTCGTTTTT AAAAAGTAAA TAATTGCATA 5400

TAAAAAGAAG TCAACAGTGT TTCATAACTG TTGGCTTCTT TGTTTTGCTA TTTAGTTTAT 5460

AAGTTAGGCC AGAATCAAGT ATATCGTAAG TCTAAATTCT AACCTTTTCT CTTCTTTTTA 5520

AATTTTCAAG CTGTTTTTCT AAACTCTCAG TACGTTGTAG AAAGTCATTG AGAAAGATTT 5580

TGAATATTTG GGCATTTGCA TGGTATTTTC GAATGATCTT TTTTTTATTA ATTGTTTCTT 5640

GTTCAAAATC AAAAAATTTG TTTAATATCT CTAATGTTGA TGTCCATGAT TGAAGTTGTT 5700

CAAGTGTTAT TTCTAGCGCA TTGATATCTT GTTGAGTAAT GATTTTAATC GGCTGATTCT 5760

TCTTTGTTAG TAATTCTTTA TTTTCTTTTT CCGATGTCAT AAGGTTTCCT CCTGCTATTT 5820

TACAAAATCC ATGTCTATAC GAACAACAAT CTAAAATTAT TATACTATAC CATGTTTATA 5880

TAGACAACAC GGGGAGTGAA TATGAATGAA CGAAAATGAA GGTGGTAGAA AAGTGGAAGT 5940

GCTGATCAAC GATTTATTAG CAGAACGTCA AATGTCTTTA AGAGAACTCG CAAGATTGTC 6000

GGGAATTGAG CCTTCAAATT TAAGCAATTT GGCAAATGGa AAAAGACAGA AAATTTATTT 6060

AGAACACATT GAAAGAATCG CAGATGCGTT GGAGATTGAT GATATATCGA GGATATTAAA 6120

GTTGAATAGG AAGTAGTATT GGGGATATAA AGAATAAATA TAAAAATAGT TAAAGATATT 6180

GTTAAATACT TTAGATTAAA TGAAAAAATT ACAGAAATAA ATTTGATAGA AGAAAAGTAT 6240

AAAGCGAACT ATTTGATTTT TAATATTTAC TGATGTTTAC GATGAAGGTG GAATATAAAT 6300

GAAGAATGAT AACACTCTTT CTGAAGAAGG AATTGATACA GATAAAGGAA TAGTTGGTTC 6360

AATTTTTGAA AGTATGCTAT CGGATACTTT ATCAGAAATT AAAAATTTAG ATAATGTTAA 6420

AATACAGAAG ATAGTAGAAA ATATAGGGAA ATATGATAAA AAAGAGTTTT TGCAAAGACT 6480

TGCTGCACTA AGAATTCCCT TTGAAAATAG AGATAAGGCT GTATTATTGG ATGCAACAAC 6540

TACAGCAACA CTGAATTGGT TATCGGAAAA CAACTGGAAT TTTAATGGGC TTTCAATGAG 6600

CTATGGTAAA TTCAAAAAAG TTATTCAACA GATTAACCAA TTAGACTCTA AAATGGCTAT 6660

TGATCCATTG GATAATCCCT ATATAGATAA TATCCaaTTT TATGGTAATC ATAAGGTAAT 6720

GCCTGGAATA AATTTTGCAA GCTCGTATAA TTTGCAAATG ATGATCCAGT CTATATTTTT 6780

ATCAAATCGT ACAAAGTTAT CTGATGAGAA AAATAGATAT ATATCCATTT CATTAAATGA 6840

TAACCTTATG GTTTCTACCG ATGTTTGTAA AAATGTCAGA TTCCAAACGA ATTnCCATCA 6900

TTCACTCGCG AAATATnTAT nCCTAATAAA 6930

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25580 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TGTACTTTCG TCCAATATTT ATTTCTTCTT CAAGCTGATA AATAGACACA TCTCTATCTT 60

TTGCTAGTTT TTTTATACGT TCCAAAAGGT TCATATGAAA ACATTCTCCT TCTAGTTACg 120

mGTTATTTmC AAAAAAAGTT AGTAAAAATG TTTGACATTT ACAATATTTG TTAGTAATGT 180

AATCACGTAA GCTAATTTAT TAGCTAATAA GTTCTCAAAT AAAACCTATA AACAAAACTT 240

GAAAATCGTT GGGGAACGGT AAAAGTATTG TTTTAGAAGG CTTTTAAAGT CTTATTTAGC 300

TATGGGTTCA TTTTACAATA TGTGTTAGAA AGTGTCAACG AATTTTATTA AATTAGCTAA 360

TTTTTTAGCT TACAAATTAA AAAGAAGGAA TGGAAGAAAT GAAAATTGAC GAaGCTGTTT 420

CAAAAGCAAT GAAAAAAGGa AAATATATTT ACAGGGAGTC TGAAAATGAC TGTTCAGCTC 480

ATGTAAAATAT ACTTCCTACA AATACCTATG ATTGTTGTCT ATTACTACAA GAAAATAGTA 540

ATTCTGTTGG CAAACGGTGG AACCCTACAG CAAATGATTT AATGGCGAGT GACTGGGAAA 600

TACGTTAGTC AAGGCCCAGT GTTTTTGTAA TAACTTTAGC TGCGATTTCA CTCATTATGT 660

CAATTGAGAC ACTTGcTAGT TTGGATGATA TTnTTTTTGA TTCCTTCCAA tTTTCGGATC 720

TCGAATATTG TCCAAATATT TATGACCATC GAATGTAAGG GAAGAGATAC ATGTCATATA 780

AGGTTTGTTA GATGCAAACT TAACATATCC TTTTATAAGT TGTGCATCGT CATCACCTAA 840

TTTGGAGACG GCATAATTAA TTTCCTCTGA ACTATATTTT TTTAGTAGAT CTGATTTAAA 900

TATTTTATCA CCAAAAGCCG GTTCATTGTA CGGAAGGCTT TCGCTAAATA ATAAAATATC 960

ACGAATACAA TCATGAGACA GTTTCATATC ATTCACCACC TTATCAGTTA TTTCAGCAGA 1020

CCACTTGCTG ATAAGGAAAT TATATCAAAA AAAGAAAGTG AGGTAATTAA ATGTCACAAG 1080

ATTTAGCTAT TGAAGTAAGG GCAGCGCTAA TTCGTGCAAA GAAAACTCAA TCTTGGTTAG 1140

CGAAACAATT AGGGATTTCA AGTCCGTACT TATCTGATAT TCTTCATGGT CGTAGACGTT 1200

CAGAAGAGCA AGTTCGGAAT ATCAmAAAAA TATTAGATAT TAGATAGGAG GTGTAGGTAA 1260

AAGTGGAAGA AGCGATTATC AAAGTTGACT TACAAAACTT AAAGAAGTTG ATCAAACAAG 1320

CAAAGGAGCA AGCTGACCAA CTTCAAAAAA CTTTAGATGA GATAAATAAA ACTAAAATCC 1380

TAATTTCTTA GCTACCCATT CTGATCCAGC AGCAGACTTC ATATCTTCCC ACGAATCGAA 1440

GTTTGTGTTT GAGCTAATGA AATTATCTAA TTTATTGTCA TCTATTGATT CCATATCAGA 1500

GAAATCAAAT CCGGATTTTT CAATAAACTC ATCAATATTA GAGAATTTAG TATTTTCAAT 1560

CATAAATTTT TTAGTGAATA GTTTATCGAA TGGTACAGAA TGTTCGTCAT CTAAAGACTC 1620

GCCATTTTTA GCAAATTGAT TTAGCTCATG TTGTAATTCG TCAAATCCAT TTAATTCGAA 1680

TTTCATAATA TTTCACCTCG CTTTCAAATT CATTTTACCA AGAGGTGAAT TGTAAAACAA 1740

TCAGTATAGG AGGTGATAGC AATGGAAGTG ATTTTAACTC CAGAAAATGA AGCTTCTCTA 1800

AGAGATTTTG TACACGGAAT TATTGTTGAT GAAATAGAAA AAGCaCGaAG aGaTACCGcA 1860

GTTGaTAAgC gaGtCtTAAA tCaAACaGaG rTTGcaAAAT AtTTcaATGt TTCCACAACA 1920

ACAATAAGGG AATGGGAGAA GCTAGGTCTT CCACATGGAT CAGTAAGTAA ACAAGGGAAG 1980

TTCTACGACA AAGAAGGGTG TCGGAGATGG CTTCTATCAC AAAAAAGATA AATCTTGGGC 2040

AAGCGAAATC AAGGGAGGAA ATAAATATGA AAAAAATATA TCACTTAAGA CGTATGGCGG 2100

CACTGTTGGT TGTATTCGGA CTAGGTCTAT TGGTAGGTGG CAATATTGGC CCATTAATCC 2160

AAAACATATA TATAGCAGCT TTTATCATTT GGTTGCTCTA CTACGATTTA GCGTTGGAAG 2220

ATCGAGAAGA AAAAAAACAA AAATAAAGAC CCACTTCGAC GGCCATCAAA GTAGGTCAGT 2280

TACAAATATC AAATTCAAGG AGAGTGTACC ACATGAATAG AAAAATTGAA AGAATGATTA 2340

TTGAACTTGA AAAAGAATGT AAGGCACAGA ATGTTGAACT TCTTTTATGT GCTACAAATT 2400

TTGAAACAGG CCAAGGAAGT ACTGCGTTTT GTGGTTCAGT TATCGGTTTA GCTATACTCT 2460

TGCAAAAACT TGTAGGTGAT CTAAAAGAGC AATTAAGTAT AAGCGAATCT TGTGATTGTC 2520

CAGAATGTGT AGCAGAAAAA GCCGAAGATG CTGCAAATGA AAAATCTATG GATGAACTAC 2580

TAACTGCATT TTTACGAGGT GAACTGCAAT GATTGAAGTA AGAGGTTTAA GTGATGATGT 2640

TTACGAATTA ATGTTAGCGA ATGCTCAAAA TAGGATTATT CAATCAATTC GAACTGCAGC 2700

AGCAAATGGT AATACAAGTT GCGTAGTGAA TAGTAAAGGT CTTACATCAA CGTTTTTATC 2760

TCAATTAGAA ACAGAAGGAT TTGATCACGT TGAACTTGAA GAAAACAAAA CGAAAATATT 2820

TTGGGAGTGG TGAAAATGCC TGAATTTGAT TCATTAGGGG CTAGACAAGA ACCTCCAGAA 2880

GAAAAAGAAG CATTAGAGCC AACATGGGAA TATGACGAAG AAGAGGAGAA TGACAATGAG 2940

TAACGATTTA ACACAAATAA CACAACGATC TTTAGATGAA CAAGTCATCG GAAATTTGAA 3000

TAGATTGCAA GAGCAGGGAT TAGAAATGCC ACCAGGTTAT AGTCCACAGA ATGCATTGAA 3060

AAGTGCTTTC TTTGAACTAA CCAACAATTC AGGAGGAAAC CTTCTTCAGT TGGCAGCTAA 3120

CAATCCAGAA ACTAAAACAT CTATTTCTAA CGCCTTGCTT GATATGGTCA TCCAAGGATT 3180

ATCGCCAGCT AAAAAACAAT GTTATTTCAT CAAATATGGA AATAAAGTTC AGCTTATGCG 3240

CTCATATTTC GGAACCATGG CTGTATTAGA TCGAGTAACT GGAGGGGCAG ATATCACGCC 3300

TGTTGTAGTA AGAGAAGGCG ATGTATTTGA AATTGCTATG GATGGTCCCG ACTTAGTTGT 3360

TGCTAAACAT GAAACGGCCT TCGAAAACCT AGACAACGAC ATTAAGGCTG CTTATGTGGT 3420

TATTAAGCTA GCAAATGGTA AAGAAGTAAC AACCGTCATG ACAAAGAAAC AAATTGATAA 3480

GTCATGGAGC AAAGCAAAAA CAAAAAATGT TCAGAATGAT TTTCCAGAAG AAATGGCAAA 3540

AAGAACTGTC ATCAATCGAG CTGCTAAATA TTTAATCAAT ACTAGTAACG ATAATGATTT 3600

ATTTGTGCAA GCCGCTAAAG ACACACTCGA AAATGAATTC GAACGAAAAG ATGTAACACC 3660

AGAGCGAGAA GAGCAAGCTG CGGTACTTGA AGAAAAACTA TTTTCCAACA ATATAAAAGC 3720

TGTTGATCAA GAAAACGAGA ATGAACGAAT TACACGTGTA GCTGACGTAC CAGAGCAACC 3780

CGATATTGAA CAAGCCAAAC CAATTGAAAA AGATAATTTA ACGAAAGTAG CTGACCAAAT 3840

TTTAGAAGAA CCAGTTCAAG AAACTTTAGA TGTGATGGCT GGTTATGAAA CCAATCAGAA 3900

AGAGAGTGAA GCTGATGTCT CAACGATTGA AGAAGACGAT TATCCTTTCT GATGAAAATT 3960

ATTATTCACA AGAAGCGGAC CTAGCTTATA TGTCTGTCTC TCAATATAAA AAATTTCTTG 4020

AATGTGAAGC TGCAGCTCTT GCCAAGTTAA AAGGCGAATG GACACCAGAA AACGATCCTA 4080

AAGCATTGCT AGTTGGTAAT TATGTTCATT CTTACTTTGA ATCACCAGAA ATTCATGAAG 4140

CATTTAAAGA AGAAAATAAA AGCAAGATGT TTTCTTCAAG AAAACCGTTT GGCCTACTGA 4200

AAGATTTCCA AATTGCGGAG CAGATGATTG AAAGATTAAA ACAAGAAGAA GCCTTTTTAA 4260

ATATTTATCA AGGTGAAAAA GAAGTGATCG TCACAGGTGA AATTGGTGGT GCAATGTGGA 4320

AAGGGAAAAT TGATTGTTTG AATTTAGAAG AAAAGTATTT TGTAGACATC AAAACAACCA 4380

AAGATATGCA TGAGAAGAAA TGGGATGAAC GTTTAAACAG AAAAGCAAAC TTCATTGAAC 4440

GCTTCGGTTA CGTGTTACAA ATGGCTGTTT ATTGCGAACT GCTTCGGCAA CAATATGACA 4500

AAATTTTTCT TCCTCTCATT GCAGCCGTTT CGAAACAAAC ACCTAGTGAA GCAAAACTAA 4560

TCACTCTTAG CGAAGAAAAA ATGATTTACG AATTAGAAGA ATTAAAAGAA AACATCGAGC 4620

ATGTTGTGCG AGTGAAAAAC GGCGAAGAAG AACCAGTTAG TTGTGGGATT TGTGAATATT 4680

GTAGAGGACA CAACAAAATT ACAAATTTTA CCAGTATGGA CGATTTATAA AAGGAGGCGA 4740

GCAGGTTGGC CATTGGAGGT TGGATAAAAC TTTATAGGAC CATTCAAGAT CACTGGATTT 4800

GGGAAAATCC TCAATACCTT AAATGGTGGC TTGATTTAAT ATTCATGGCC AtCACCAgAC 4860

AGAaAGATTC TTTTTGATGG AGAGTTAAAA cGGTGAAAGT CGGAGAAAGA wTTACATCCG 4920

AAAAAAAaCT TtCTGAAaGG TGGGAaGTAG CCGAAACACC GTCAGAAAAt TTTTAGATTT 4980

GCTTGTTAAA GATGACATGA TTGAATTAAG CAGAAGTAGA CAAAACGGGA CATGGTACAA 5040

AGTCAGTAAC TACGCAGAAT ATCAAAACTT TTCTGAAATA AAAAAACAAC GAACTGAACA 5100

ACGAAGGGAA CAACAAGCTG CACATCAAAC TGAACAACGA ACTGAACATA AACAAGAACC 5160

TAAAGAACCT AAGAACCTAA GAAATAATAA TAATAACAAT AAAGGGGCGT TCATTCGTTC 5220

AATTTGGGAA AATAACGGAT TTGGATTGAT GTCGTCTAAA ACCATGACCG ATTTTGATTA 5280

TTGGATTTCT GATTTTGAAA AAATCGGAGC TAGTCAAAAA GAAGCTGAAC AATTAATTGT 5340

```
TAAAGCTATT GAAATTGCTA TTGATGCAAA CGCAAGAAAC TATAACTATA TCAATGCCAT     5400
ATTGAAAGAT TGGGAACAAA GAGGGTTCAA ATCTGTTGAG GAACGAGAAG CGGCAAGGAA     5460
GCAAAAGAAA ACAACCAAAA AACAGAAATC AAACACAGGT CATTCTGATT ACGATGATCT     5520
TGGATTTTAG GAAGTGAAAG AATGCAGTCA GCATCAGATG GATTTTCAAA AATGATTAAA     5580
ACGTTGCTTT ATATCACGCC TGATCCATGT CCAGAGTGCG GAGGAAATCT TTATGCTTGG     5640
CGTGCAAAAA ACAAAGATGG GTCCGATAGG TGTCCGCCAA CTTGCATGGA ATGTGGCTAT     5700
AAAGCACGCA AAAAAGCAGA AGACCTTGAA ACAGAGAAAA TGTTTAACGA TAGTTTGAAA     5760
GCCAGAGCGA TTAATTATCT GAAATATAGT TCGCTTTACA CCGACAAAAA TTTAATTAAT     5820
TGTCGTTTTA AAACTTACAA AACAGTAGAC ACAGAAACTA AGCTTGCTTT TGAAATTGCA     5880
AATCGAGCCA CAACTGAAAT TCTTTTGAAT AAACCAATTC ATATGATTCT TTCAGGCAAA     5940
AGTGGTGTTG GAAAAAGTCA TTTAGCTATG TCAACGGCTT GGGAAGTGTT GGAGAAATCA     6000
AACTATGATA AACGCTGCTT ATTTATTAGC TATGCGGAAC TCTTAGAACA ACTAAAATTT     6060
GCGATGAATG ATGAGCAAGC CAGAAAAGAG ATAACAGGTA GTTTGATGGC AGAAATCAAA     6120
AGCGCAGATT TAGTAGTTCT GGACGATTTA GGAGCCGAGT TGGGAGTTAA ACAAATTGAT     6180
GATAGGAATA AAAGTACTAA CTTCAATAAC GACACCTTGA ATCGCATTGT AGAAGCTCGG     6240
CAGAATAAAG CAACTATTTT TACTACGAAC TTAACTGGTA AAGAAATGAG TCAAGCTTAT     6300
GGGGAGAGAA TTCTTTCTCG TATCATGAGT AATTCACAAG GTTTTGTGAT GAAAATTGAA     6360
GGGACATCAG ACAAACGAGT AGCAGGCATC TAAAATGTTA TTTTTAGCGA ATATATTCAG     6420
CGTAGAGCGG TTTTACAATC AAGTGAATAT AAATAGATAC AAAGAAAGAA AAACGGATTA     6480
AAACGCATTT TAAAGCCTTA AAAACAAATC GATAGAAAGG GGAATCATTC AATGCCGTAT     6540
GTAGTGAAAA TTTCAGCCTA TCTTGGCAAA GATGGTCGAC CTGTAGCCAA TTTAAAAGAT     6600
GCTGTGCTAT TTGAGCAAAA AGAGACAGCA GCTATTGCAA CAATCGTATC TGGCGGAACC     6660
GTTTCAGAAG TAAAGGAAGC CATTATAATA CCAGAAAAAC CAAAGAAACA TATAGGGAAA     6720
TCTATTAAAC GGATTGATAA GAAGGAACCG ACCGAAAAAG CTACCAAAAG TAATCAAGCC     6780
TGGATGAAAG GGGCTAAATA AGAATGAAGT GTGTTAGATG TCAAGATCAG CGCGTGATTT     6840
GGGGCAAAGA CAGATTTAAT TATGCAACAC CTATTCCATG TCCTGAATGC AACAAAGATG     6900
GAAAAGCAGT TCGAGCGGAA ACTGCGACCA AGGAAAGGGA GTTAAAACAA TGCAATCACC     6960
AACAGCCCTG AATAAGCGAG GAAACAAAGT CACAATTGAT GGTTACACAT TTGATAGCCA     7020
GAAGGAAGCT AACTTTTATA CAAAGTTTGT CAAAAAATTGT GGGTTACCTT TTGAAGTTCA    7080
TCCGCGTTTT AGACTAACCG AACTTACACC AACTGCGGAT GGTATAGGCA AAATTTCGGC     7140
GATAGCTTAT TCACCTGACT TCATCATAAA AAACTTAGAT GGGAGTTGGA GACATGTCAT     7200
TGATATTAAA AACTCTTTTG GCGTGTATGG TATTGACCAA TCCGTTAAGC TTCGTTTTCG     7260
TCTATTTGCC CTTCGATATG GTCATCCAGT TGAAGCGATT GTTGTTCGTG CTAGAGATTT     7320
```

TAAAGTGATC ACTCAAGGTG TAACTAAGCC TTTAAACGAA AAAAGACCAT TCATAACCGA 7380

TAATTTCGAT TACGAATGGA AAGATGCAAC TAATTATTAA ACGAAAGTAG GAAAATAAAA 7440

TGACAAAACA AGTAAATTTC AGACCAGAAG TGAAAAAAGT GACATCTAAA TCAAACGGAA 7500

ATATCGAAGT GCTATTAGTG GTTAGCAACG CTTCATTAAA AGGGAAATAT GAAAGTTTAA 7560

ACGAATTTTT AGGTAAAACA GTATCAACGA CCATTGAGCC AGAAACAGTA GAATACAaGG 7620

TACCAGTTAA CAAGCrGACC AwTAAACCGn ATGTCGAATA TATTGyGmnT AACGACGGAA 7680

CAGTTGAAGT TCTAAAAGAA GAACAAACTT CTTTAGAAAT GGGCGATGAT GTGCAAGAAG 7740

TTGAAGAAGT TGCTGTGCAA GTATCGAAAG AAACCATTGA CGAATTCATC AAGAAGGCAA 7800

CAACAATCGA ATGGCCAGAA TCAGTAACAA TCAACGTTCG TGGCGTGTTG CATCGAATCG 7860

ATGAAGGGGA AGCCTTAGAA GAAATTGCAG CTGATCATGA TGTTTCAGTT GAAAATCTAA 7920

TCAACCAAGT TGAACTTGCA CGCCAACATT TTGCACCGTT TGCAGATTCT TGGAGCAAAA 7980

ATAAAGAGAA CATCATTTTC CCTGAAAAGA CAGTTGAAGA TGATGAAGAA GAAATCGAAT 8040

AATAATCTCG TAGAAAGTGA GTGTTCATTT TGCTTGAGAT TTATTATACG CCAACATCCG 8100

CAATTATTGC GGATGCACTG GCTAAAACAT ATGAAGTCGT TTCTTTAGAA ACAGCTAGAA 8160

ATATTGCCAA GAAATTTAAG GCTAGCTTGA AGCAGAAAAC GGATCTTTAT GTAATTGAGG 8220

GAATTTTGAT TGATGCTGGT TATAAAAAAG AGCCAGTGAA TTTATAGAAG GGAGTGGAGG 8280

TTTGCGGCCG CGATTAAAAA GCTTTTTACT CCGTTGAGAT TATGAATATT TACCAAAAGA 8340

AAATGGGTAA AGTTATGAAA GTCCTTTGCG GTAAAAGAAG TGAAAACATT GGGTAAATAA 8400

AAAGTGAAGA AGAGTAGCTT TTTACTACTC TTCTTCGCTA GAAAAAATTT TTGAACAGAA 8460

TAAAAAGCCA AGGTCAGTTA CATGTATGGA GCCACGAATA CTTATAATGC TAAACCCTGT 8520

ATCTTTAAAA AAAGAACTGT TTGCAAAAAA ACTATTAAGA TTATATTTCA AAGTTTTAAA 8580

ATAAGTTGAA TTTTCAAAAT CATCATAAAT TTTTGTATGT TTTTTTTGTG TTAATTCAAT 8640

AGAAGGTCTT ATATCAATTA ATCCATAAGA GCGTAAAGTT TGTAAAGAGG GACTTAAATC 8700

AGAAAAATTA TTTTCAACAA TTCCTTTCTG AGTAAGGATT AAATTTTCTA TATAATCTAT 8760

ACCGTTACCA TTAGGAGCCT GTACTCTAAC CTTTATAGTT GGGGTAATAC TAAGTTTATA 8820

CAATTTTTTT AATAGATTTG CATCTATAGT CGTTAGTTCA GATAAAATAA TTGAAAACCT 8880

TGGCGAAGAA TCTTGATTTT TTCTGTTATC AACTAATCCA GCAaGTAAAT TAGAAAAGTA 8940

TTCACGCATA TCTTCATCTT CTAACTGATA TCTAGCGTCT TCCATTGTTT TTAAAGCAAA 9000

CCCTAATTTC GAATCATCTC TGTTATCTAC AGGGATACTA TCGGTTTTAG AATTAATCTT 9060

TTCTACAAAA TCAATATAAG ATTTTTCACT AACAACATTT AGTTTTCTAA GAGGACCTAC 9120

AACATAATTT ACAATCCCGC CAAGGCCTTC TCCAATTGAT TTTGTAGCAG GTCCAAAGAG 9180

ATTAGTTATC AATTCCTTTG GTAATAGTTC AATGTTAATT AAAGATTTTT TATCGTCAGA 9240

CATTTAAACA TCTCCTTTTT CAAAATATTA TACCAATAAA GAAAGGAATT TTCTATGCAA 9300

```
CTTAGACCAT ATCAAGAAGA ATCTCGTTCC GCTGTTCAAA ATGAATGGGA GAATAAGAAA     9360
AAGAAAACAT CATTAGTTTT GCCTGCAGGA TGAATCTGAA TGAAAATATA TTGATATGAA     9420
AGTTCAAAAA ATACTGGGTG TTGGACAGAA GGCCTAGATC ACTCAAATAA AAAATTTGAA     9480
AAGGAAAAAG GCTACTTGAT GGTTGACGAT CCTGAGGAAG ATGAAACAGT AAAACTAGTC     9540
ACAATTGAGG TGATAAACGT TTTAAAATCC TTAAGAATGA CCGTAACTAT CAAAAAGGCG     9600
ATATCTTACG CTTAAACGAA TATCAAGACG GACAGTATAC AGGTGATGTT CATGTCTCAG     9660
AAATAACATA CATTACAGAT TATACTCAAC AAGATGGTTA TCTAGTGTTA GGAATTAAGT     9720
GAGGAGGATT AATAAATGGA ACAATTCTTA TTAACAAAAA CTGGTAAAAA ACAAATTGAT     9780
ATAAAGGGTA CGGGAATGGA TCATAACGAA ATTGTCTTTA CGTTAGCTGC TACTTTAGTT     9840
GGATATAGCA AAGAATTAGG CCTAACAAAA GCAATACTAA ACGAAAGTAT GTACGTGCTG     9900
TGGAAAGATG GTGAATAAAT GAAACGTAAC TGTAAAAGTG TAATAAATAA AGTTAGTGGC     9960
ATTGCAATAA TGATTCTTGT AGCAAAAGCA ACCGTGAGCC ATTTCGTGTA TGGAAATGAC    10020
ATAACAAGCA GTGACCTTGT TTATTTCCTT TCATGCTCGT TTATTTTGGG ATTAGGGCTA    10080
TATTTAGGAG GTTCCAGTGT ATGAGTTATC CAGAAGTTTA TATCATAGGA AGGCAAGTCG    10140
ATGGCGTTTA TGTTGAATAC TTACATGGAG CAGAGCAAGC CGATTTATTT TTCGATTATA    10200
CGATAGCTCG TGATGAAAGA AATCATATGA ATAAAACCAA TATGAAAGAT GGTACTTGGA    10260
AAATTTTGAA ATATGGTAGA CCGATAACGG TCGAATTAAA CAATAAAAAA GCCAACCGAC    10320
CAATGGTTGA CTAAGAAGAA TATTTTACCA GAAAAGTGGT AGCTTGTGAT ATGTGAGGTT    10380
ACTTTGCCCC AAACATTGGT CACAATAAAA ATATTTTATC ATGAGTAAAG AAAGCTGCCA    10440
ACAAAAAAAG CCGGATTTCT CCGACTGTGG ATAATGTGTG GATAACCTTG TGGAAAAGTC    10500
GTTATCATTA TACCACAAAT CACAAACGGG GGAATCAAAG GATGGTACTT TTTGACGTAA    10560
AGAnATATGA AACACCGGAT GCAAAGGACG TAGATATGGA GCAAACTAAA CATAACGTCA    10620
GTGTGTTCCT GTCTGCCTAT CTTGCTGCTA GATGTCGTGT TGGCCAGCCG AGGGAACCAA    10680
AAGTAACAGC TTCATTCTCT TTGGTTCCAC CATCCACGGC CAAAAACACT TTTGAAGCCG    10740
AGCAAATGTT AATCCAGAAA GAAGAAGCCC AAGAAGAGTT TGATTATTTG CATAAGCTTT    10800
TTGTTAGAGG TTATTCTGCG ATTCAGCATC CGCACAAACC AGATGTTACC GAGCGAAGAA    10860
AAAGAATCTT CTATGATCGA TATATCAACG GTAATCCAAT CTATCTAGCA GCGCAACGAA    10920
ACTGCATCAG CGAAGAATCA GTGAAACAAG AATCTAACAT GATTATTGTT CAATTTGCTT    10980
CAGCGCTGGA ACTGGTTGCT TTTAAGTAGC CATTTATTAC ACTTTTTATA CCTCTTTTAT    11040
ACACTTTATC TACACTTCAT ATACCTTCTA AACGAGTTAT TATGATAGTG TCAAAAAAAT    11100
AAGAAATGCG ACACACTTAC ACAAAATTTA AACGGAACGA TTGCCTACTT ATTTTTTTGA    11160
TTTGAGATTA CAAGGAAGTA AAAAAATTCT ACTTTCTTCG TTTAGTCACT TGTGATCTCA    11220
TTTAAATTCT CTCGCAAACC ACCAATTATA AAACTAAAGA AGTGAGGTGA ATTTCCTCTC    11280
```

TCTTTTTTCT ACAGGTTTGC GAGAGTTAAT GGAGCATAGC TTAATCGGCA GAGCAGCGGT 11340

CTCCAAAACC GTTGGTATAG GTTCGAGTCC TATTGTTCCA GTAAGTGGCA TAAGCTGCTT 11400

AAATAAAATA GATCGTCAAT AAATGTTCGG ACAAACAAAT TGGCGCTACT ACCTTTCACT 11460

AGGGCTGCAT TTATATGCAG TCCTTTTTGT TTTAAGTGTA GTAGATTTTT CATTTTGAAA 11520

GGGGATAAGT AAGACAATGC GTGTATTAAT TAGAAGTTCA GCATCTGGTT CAGAGTATTG 11580

GGATACCGAA GAAAAAAGAA ATGTGTTTGT ACCTAAAGGT CAAGAACCTG ATTTTGAAGT 11640

TACTGAAAAT CCTGAATCAA TGCTAAGTAA AGAAGCTGAT TTACATGTTG GTGGATTACC 11700

AATTACTGTA GGGAATGTAA CGATTGATAC TGATGGAATT AAAGGCGAAC GATTATTAAC 11760

AACTGCAAGT GCTGCTGATG ATGAAGAACA AGATGAGCTT GTTCCGTCTG ATGATGAATC 11820

TGTTGTATTA GAAGAAATGA ATGTAAAAGA ATTGCGTGAA TATGCAAAAC GAAAAGGTAT 11880

TGAGATTCCA AGTGCTGTAC GTGCAAAAGG TGAAATTCTC AATATTATTA AAGAATCTGA 11940

ATAATGCGCT ATTGCCAGTT TGAAGGTTGC TCTAATACAA CAGAAAAAGG AGCTTATTGT 12000

TCCGAACATG CTAGGAAGTC AAGAAAAAAG AAAAAGCCAA TCAATGTTTA TCATCATGAT 12060

AACAAATCAT TTTATCGAAC AAAAGAATGG CAAGATGTCG CTGACTTTGT CTATGAAAGA 12120

GAAGGTGGTT GCTGTCAAAG ATGTGGCCGT TTTGTATTTG GAAGGCAAGC GCATCGGCAT 12180

CATGTGATTC CAATCAAGAA GAACGAAATG CTCAAACTTG ATCCAAACAA TATTCGTTTG 12240

TTGTGTCCGA AGTGTCATGT GATTGAAGAA AATGAAGCAG ATGAGAAAAA AGTTTTTCCA 12300

TCTTATTTTA AAAAATGAAG CCCCCCTATC AAATCTGATT CAAATTTTTT GTCGGGGGAT 12360

AGGGTAGGGG GCAGTCACGC GTGTCGTTAG GTCAAAAATT TTAAAAATAA AAGGGGGGTG 12420

TATAAAAAAT GACCACAAAA GCGCAACGCA AAGCGATTAT TGATGAAAAA GTTAATCACG 12480

AAAAAACGCG AATTTTAGAA ATTATGCGCA AGTCTGATTT ATACACTATT ACTCTTGATC 12540

CATTGATTGA ATCATACTTG GATATTTTTG AAGTTTACCA ATACAAATAC ATGCTGTGGA 12600

AAGAAAAAGG ATTTCCCGAA ACCCAAAAAA CAACAAACAA GGyTGGAGmT AyTAACAATA 12660

GCAAGCATCC ACTAGCGCAA CAAGTCGAAG TTTGGGCCGA TArAAAAATG AAAGCATTGG 12720

ATTTATTAGG ATTGACCAAT AAGTCAAAAA CAGGCAGACA AATTACTGGT GGTTCAACAG 12780

CTAGAGCAGA TGnaGAAATG AAACGGCCAG AAGAAAAGCC TGTAAATGAA TTGGCAGAAC 12840

ATCGGAAAAA ATGGCGTAAA AAGGCAGGGA ATGAAACATG ATTGAACCTG GTGTAAATTA 12900

TGCTGATTTA TTTGCGAAAG AAGTTCGAAA ACATCCTAAG AAATATCCGA AAACGGTTCG 12960

TTtAGCAATA GATCGTTGGT ATCGATGGAA GAAACGAAAA GATATTTGGT TTGATGTTGA 13020

TCGTGCAAAT GAAATGATGG ACTGGGTAGA ATCTTTTATT GTTCaTACAA AAGGCGATAT 13080

GGTAGGTAAA CCATTTCTTT TAGAGCCATG GGAAAAATTC ATTTATTCTT GGATTTATGG 13140

CTGGGTAAAA GAAAATGAAA AAGGGCAAGT AGTCCGTGTT ACTCGTGAGG CATACGTACA 13200

AATACCAAAG AAAAATGGGA AAACATTAAT AGCCGTAGGG GCGTTGGGAT ATGCTATGTA 13260

TGGCGAAGGT GCCTTATCTG TCGATTGCTA TGCATGTGCT TCTGATTTTG CGCAAGCCCA 13320

GTATGCTGCT AAGCCTTTTG CCGCTACTAT CCTAAATAAT CCAGTGCTAT TAGATGGGAC 13380

TAAAATATTT AAAGGTCCAA AAGGCACCGT TTCAAGTATT ACGTATGACT ATTTACATGG 13440

AGATATGGCT TATACAAATA AGTTTATTGT TCAGACAAAA AACATTGATA ACATAGAAGG 13500

TTCCAATCCA TATTTTGTTT TAAATGATGA GCTGCATAAA CAAGAGAAAA TGGAACAGTA 13560

CGATAACTTT AAATCTGCAC AAATTTCATT GCCACAGCCG TTAATGTTTA ATATTTCTAC 13620

AGCTGGTAAA GGAAGTAGTT CGGTTGGTAT TCGTGTTTAT AAAGAAGCAA AAGAAGTCTT 13680

GAAGCGCGAT GATAATGATT CAAACTTTGT TTTGATTTAT GAACCAAATA AGGGATACGA 13740

TTGGACaGAT AGAAAAGTTT GGGAAATGTG CAATCCTAAC TGGGGAATAT CTGTTGATTT 13800

GTCTGCTTTA GAATCAGCCT TTAAAACTGC GCAACGTTCC GCTCACTCGA AAGCTGAATT 13860

TTTaACGAAG CATTTAGATG TGTTTGTGAA TGGTGCAGAT AATTTCTTTG AACAAGATCa 13920

AGTGGAACCG TGTTTGGTTC CCACAAATGA ATTAGGTAAC TTAAGTGGTG AGCCATGTTG 13980

GATTGGTTTG GACTTATCTA AAAGTCGAGA TTTAACTTGT GTATCATTAA ATTTTCCTAC 14040

ATGGGATGCC GAAGGAAAAG CGATACTCAA AGTAAAACAA TTATATTTTA TTCCTAGTGA 14100

AAATATTGAT TTTCGAGAAA AGGAAGATAA TGTGCCATAT TCTGAATTAG CAGAACAAGG 14160

ATTTGTTGAA TTTTGCGATG GTAAGTTAAT TGATCAAGAA CAAATATTTC ATTTTATTGA 14220

AGATTGCATG GATTTTTATG ATGTTCAACA AGTCAATTAT GATCCAGCGA TGAGTGACCG 14280

ATTAGTTGAA AAATTGGAAA ATTTAGGCTT GGAATGTGTG CAAGTTGATC AGTACGCAAG 14340

AGTATTGAAC TCGCCGCTTG AAGATGCCGA GCGATTATTT TATGAGCAAA GGATTATGTT 14400

TGATAATCCT TTATTTTTGT ATTGCGCTTT AAATGTGGTT GTCAAAATGG ATTTTCAAGG 14460

CCGTAAAGTA CCAAGTAAAA ACCAGTCAAA GAGAAAGATT GATGGATTTG TTGCTTTCCT 14520

TTGTGCGCAT AAGGAAACAA TGGATCAAAT GATTGATGTC AACGAAGATG ATATGGATGA 14580

ATATTTAGAT TCTATCTATC GATAATAGAG AGGCGGTGAG ATTTTGAAGC TAAGAGATAG 14640

ACTTTCAAAT GCAGTCTATG GATTTTTGGA AAAGCGTGGC TGGATTGAAG ATATTTATGG 14700

CAATGTAACA AGATATTCAC AACGTTTTGT TAACGATTCT TCTATTATGG AATCGTCTGA 14760

TGTTTATGAA TTGGTACAAG ATATTTCTAA TCAAGTTGCA CTAGCAGAAC CAGTAGTAAT 14820

TGGGCCTGAT GGCGAGGAAG TGAAAAATCA TTTCTTGTTA AACATATTGA AAAATCCTAA 14880

TGATTATTTA ACTGGTTTTG AATTTGCAAA GCTTGAAACA AATACATTGT TAATCAATGG 14940

AGAAGCTTTT CCTATTACAG ATAATGACCA GTTACATTTG GGATATGGTG TTCAAACGAA 15000

ATTAGATGAT CGTTTGATTG AAAGATTTTC AATGAATGGC CAACCAATAC CAGGGAGTAT 15060

GATTCGTCAT ATAAAAAATA TTGGTGTGGA TTCCTTAAAA GGTGCTGGAA TTATTGATCT 15120

TGCAAAAAGC ACACTAGAAG GTGTTTTAAG TGCTGAAAAG GTTTTGACAG AAAAATATAA 15180

GAAAGGCGGC TTGCTCGCTT TCTTGTTAAA GCTGGATGCG CATATCAATC CAAATAACAG 15240

```
CGCTCAACAA AAGATAGTAA AAGCTATTTT AAATCAGTTG GAAGAAACGC AAGATAATGA     15300
TAGTCATTCA GTTAAAATGA TTCCTTTGGG CAAAGGATAC TCAATTGATA CTTTAAAAAG     15360
CCCAATTGAT GATGCAGCTA TTCTTAATTA TTtGGGTGTT tACAAAAAAG ACCTAGGAAA     15420
tTTTTAGGAA TAGATGTAAA TACTTATCAg CATTAATGAG AACaGaTATT GnAAAAGCAa     15480
TGATGTATCT GCACAACAAA GCAaTTAAmC CAATATTAAA AAATAAGAGC GAGCATTACT     15540
CGGCTCTTTT TTTTATGCCT AATTCTGGTT ATCGAGTGGA ATGGAAAATT AATATTTTGG     15600
ACTTTGTACC TTATTCCACC AAAACAAATA TTGGGTACAA CATTGTTCGA ACTGGTATTA     15660
CCAGTCCTGA TAATGTGGCA GAAATGCTTG GTTTTCCTAG ACAAAATACT GAAGCAACAC     15720
AAGCCGTCTA TATTTCAAAT GATTTAACGG AAATCGGCAA AAAGAATGCT ACCGATAACT     15780
CATTGACAAC AGAGGATGAC TTGAAGGGAG GTGGTAAGAA TGAAGAAACA GGAAATTCGG     15840
ACATTTGACA TCACAAACCT TAAAACAAGA AGCGAAGAAG ATAGTCAAAC ACAGATTGTT     15900
ACTGGTTATG CGGCGGTGTT TAATAGTCCA ACAGAATTAT GGGAAGGCCT AAATGAAGTG     15960
ATTAAGCCTG GAGCTTTCAG TCGTGCTTTG TCAAATTCTG ATGTTCGTTG TTTATTCGAT     16020
CATGACTGGG GCAAAGTATT AGGGCGCACA AGAAGTGGAA CTTTGAAACT TGAAGAAGAT     16080
GATAAGGGAC TACGATTTGA AGTTGAGTTG cCCCAATACA ACTGTTGCCA ATGACTkGAT     16140
TCAATCAATG TCACGTGGGG aCATTAwTCA GTGTAGCTTT GGTTTTTATC CaACGGAAGA     16200
AACTTGGGAT TATAGTTCAG ACCCAGTTTT AAGAACTATC CaTGAAGTCG AATTGTATGA     16260
AGTTTCTATT GTTTCTTTAC CTGCTTATGA AGATACAGAA GCAGCACTAT CAAGAAACAA     16320
ACAAGAAATG AAGCAAGATA TTAAAACTAG AAAAAAATTA ATTGAAAAAA TCAAAACAGC     16380
GCTTGAAGCG TAGGAGGAAT TTATTATGAA CAAAGAATTA TTGCGTCAAT TACAAGCTCG     16440
TCACGAGAAA CGATTAAGTG ATTTACAAGG CAAAATTGAA TCTGGAGAAG TGCGTGAAGC     16500
AGATTTAGAT TCAGTTAATG AAGAAATTGA TGGTTTAATC GATGAATTAA AAGCCATTAA     16560
AGCTGAATTA GGGGATGATA ATTCAGAATC TGGTGATGGT GAAGGCGATG ACGGGACCGC     16620
TAAATCTGAT AATACTGATG ATGAAAGTAA AGAAGATCGT GAGAAAGATA CGAACGGAAA     16680
CAACGATGAT AAAAACGAAG AAAATCGTGG CGGCATGATT AGCCAAGAAC AGCGTGATGG     16740
TTTGTTACGC ACAATTCATG AAGGAATGGA GGCTAGAAAT GTGATGTCTA ATGAACAACG     16800
TGAAAAACAA ATTCGTAAAG CATTTGCTGA TTTTGTTATT GGTAATATTT CAGAAAGTGA     16860
AGCACGTGCA TTAGGTATTG AAACAGGCAA TGGTTCAGTG ACAGTACCAG AAGTGATTGC     16920
ATCCGAAGTG ATTTCTTATG CTCAAGAAGA AAACCTATTG CGTAAATACG GAACTGTGAT     16980
TCGCACGGCT GGCGATGTGA AGTATCCAAT TCTTGTGAAA AAAGCAGAGG CTAATGTAAA     17040
CAAAAAAGAA CGTACGACAG ATATTACTGA AACAGCGATT CAATTTGATG AAATTCTACT     17100
TGATCCAGCA GAATTTGATG CATTAGCAAC TGTAACGAAA AAACTATTAA AAATGTCTGG     17160
TGTGCCAGTA GAAGATATTG TTGTAGAAGA ATTGAAAAAA GCATATGTTC GCAAAGAAAT     17220
```

```
TAATTATATG TTTAATGGCG ACGATGCAGG aAATGAAAAC CCAGGAGCTT TAGCTAAAAA    17280
AGCTGTTGCA TTTGAAAAAC CTGTAGATTT AACAGCTGCA GGTGCTGGTC AAAAATTATA    17340
TGATGCATTG ATTGAATTTA AAAATACACC AGTAACAGAA GTAATGAAAA AAGGTCGTTT    17400
TATTATTAAT CGTGCAGCTT TAACTGCTAT TGAAAAAATG AAAACAGATG ATGGATTCCC    17460
ATTGTTACGA CCATTCACGC AAGCAGAAGG TGGTATTGGT TATCAATTAG TTGGTTATCC    17520
AGTTGATTGG ACCGATGCAG CAGATAAAAA AGGTGAACCA GATACACCAG TATTATATTT    17580
TGGTGATTTT TCTGCTTTCA AAATTCAAGA AGTTATTGGA GCGTTAGAAA TTCAAAAATT    17640
GGTTGAAAAA TTCTCTGGAA AAAATCAAGT TGGTTTCCAA ATTTACAACT TGTTAGATGG    17700
CCAATTAGTT TACTCACCAT TCGAGCCGGC TGTTTATCGT TATGAAATTA CAAAACCAGT    17760
AGGTGGTTAA TGTGAATAAC GAaGCTGAAA CATTATCTTT AGAAGAAAAA TTCAAAGCAC    17820
ATATTCATTT TGAAGAGGGG ATGGATGATT CCAtGCTCTC TTTTtATTTA AATATGGCAA    17880
AAAATTATGT AAAAACTGCa ACTGGAGGGC AAGAAGAATA TTtAATTTTG ATGGTTGCTG    17940
GTATgCTTAT GAaTATCGTG TTTCAGAAGA TGAATTAGAT AAGGCGTTGA ATGCGATCAC    18000
GCCATTTATC ATCCAAGGAG TGATTCAACA TGCCGAAGAG GCAGACGAAT AGGTTTCGAT    18060
GGAAAGCGGA CTTGCTAAAT GTAAAAGAAG AAACAGATTC GAACGATAAA GTAGTTACGA    18120
CCTATAAACT TAATAGGCTT TTATGGTACG AAGATATTGG AGTAACTGCA CAAGAAAAAT    18180
ATCTTTCACA GCAAGCCAAA ACAGACGTTG TCAGACGGAT TAAAGTGAGA TTGGATAAAT    18240
CTATCACAGA AAAGTTTAGC GCTGTTAGAA TCGATTCTGT GACCTATAAA ATCACTCGTA    18300
TTTACACAAA TATGGATAAA CGAGAAATGG AGTTGAGTTT GGCTTATGTC GATTAGTTTT    18360
GAAAAATTGA GAATAACTCT GAAATCAGTA GGCGTACCTG TGACACGTGA TAAAGCGGAA    18420
AAAGGAACGG ACTATCCATA TATTGTTTAT TCAAATGTTA GTCAAGGTAA AAAAATGGCA    18480
TCGTCTAAAG TGCATAGACG AATGCCCTAC TATCAAATCT CTTTTTATAC AACAGGTACT    18540
GAAAAGGATT TAATTGCTTT AGAAAATGCA TTGGAGGAAG CTGGTATTCC TTACACTGAT    18600
TTTGTAGGCA TTCAAGGCGA TGAAAATGAT GATACTGTGA CAAATTTTTA CACATATGTG    18660
AGGTGTATTG AAGATGGAAA ATAATAATGG TTTTGCAGAT ATGGCAGACT ATTTAGGAAA    18720
GCTTTCGCAA GTAGATGCGA CAAAATTATC AATAGAATCA TTAACCGCTG CAGCTAATTT    18780
TTATATGGAA AAATTACTAC CCAATATACC TAAATCGCTT CTAAAAAAGA AGCACATGGT    18840
TGATCAAGTG AAAGTAAATA TTAAAGATAA TGAAGTACAA GTAGCTTTTG AAGATACAGC    18900
TTTTTATTGG CGATTCGCTG AAAATGGAAC AGTAAATCAA AAAGCGCAAC ATTTTGCCAG    18960
TGGAACATTT GAACAAAATA AAGATCAAAT TGAAAAAATT ATGACTCAAC AAATATTAGA    19020
TTTATGGAAA GGATGAGTAA ATTGGGAAAA CAAGATGTGT ATTATTTTGA AGGCTTAGAT    19080
GACATCTTAA TTGCCATGAT GACCACAAAA GATGCAGTAG GTACAGAACC AGCATTTGGC    19140
GAGGTTGTTC GTTTGCCAAT AGCCACAAAA TTGGGAATTA AAGGAAATGG AACAGCTTTA    19200
```

GAAAAATGGG CATCAAGTAA AATGTTCCGA CGCGTAAGTC GTGAAACGAA ACATGAAATT 19260

GCGCTAGATC ATGTGGGCAT TCCTATTGCG GTGATGGATG AAATAAAAGG ATTAATCGCT 19320

CAAAGTGGAG TGACTTTTGG TAAAAACACT GCGCGAGAAT TTCCTTATTT TGCCTTTGGG 19380

TTTATCGGAA ATATTGAAAA TGGTGGAAAA AAAGCTGTTT GGTATCCTAA AACGCAGTTA 19440

TCAAATGTTA TTGATGAAGA ATACACTACT GCAGAAGATG AAACCAAAAT TGATGATGTA 19500

ACTGCTAACT TTGTTTCAAC TGGTTTAACA TACAATAATG TTATGTATTC AAGTTTTGAT 19560

TCTAATCGGG ATAGTGCTTC AATAGAACTA TTTGAAAAAT TTATCGCACA ACCTGTTTAT 19620

GACGAAGAAC AATGGAAGAA ACTAGCAGGT CCTTCACGTG GAGGTGGCAG TGAATAATGG 19680

CAAAATTAGC GGATTATGGG ATTGTTATTT CAGATACACC CACAGTTACT ATTCAAGGGC 19740

ATCAGTTTCC CATTTTGTTA ACGATGGAAA CAATGGAGTA TATTGCTGAT GTTTATGATG 19800

ATGACTATTC AAAATTTGAA GCAGATATGA ATGAAATGAT TAACAAGAGT GGTGGTCGTA 19860

TTTCATCAAA AGATTTATCA GCTTCTGATT TAAAAATTAT GCGTGCATTG ATTTACGGCA 19920

TGCTAAGAAC TGGTGGATTG GAAGAAACGC CAGAAACCAT TTTTAAGTTT TTAGGTATGA 19980

GTGCCACAAT TGTTGAAATT TATGGTGCAT GTATGGAAAT ATTTGCAAAG CAGAATTTTC 20040

AAGTTGAAGA CTTAAAAAAA TCCAAGAAGC CACAAGATTA TCAAACTCCG AAAAAAAGGA 20100

AAAACAAAAA GAAAAAGCCT CAACGGAAAT AGGAACGCCG TGGGCTTTTT ATTTATATGT 20160

AGCTCTCACT CTTTTGGGAT GGAGTGAGGA CTTTTTTTTA AAAGCAACTC CCAACTTGTG 20220

GCTTAAGTCA TATATTCAGT GGTTAGTAAG TAATACGGAG TTTGAACCAC CCAGAAGTGT 20280

GACAATGGAT AAAAGTCCTT GGTGGTAGGA AAGGAGCGCT AACGTGTCAA AACAAGAATC 20340

CGATGTTGTC TTAAATTTTA AGATGAATGG AGAAATAAAC TATTCACGAA CAATTAAAGA 20400

CATTAATAAA GAAATGAACT TAGCGGCTAC CGAGTACAAA AACCAGGTAT CCGCAATGGA 20460

TAAAAATGCA ACTCAAACTG AAAAATTAAC AGCTACAAAG AAAAAATTAG AAAAACAATT 20520

ATCTTTAGCT GAACAAAGAA CAAAATTATT ACGTGAGGAA TACGAAAAAT CAGTAAAAGA 20580

AACTGGTGAA TATTCAGAGC AATCACAAAA GCTGTATAAG CGTTTATTGG AATCCGAAAC 20640

AGGTGAAAAT AAACTGCGTT CTGCATTGCA AAGTACCAAT GAAGCTTTGA AAGAGCAAGG 20700

TAATTTATCA ATAAAAACAG CTGAAAAACT AGCCAAAATT GAAAAAGCTG GAGACAAAAT 20760

TAAATCAGTT GGGCAAAAAC TGTCTGTTGG ATTAACAGCA CCAATTATGG GAATTGGTGC 20820

TGCTTCTATT GCCGCATTCA AAGAATTAGA TGAATGTTTG GATAATATAA CAACAGCAAC 20880

AGGGGCTACT GGTAGTCAGC TAGAATCTTT ACAAGCCAGT TTTAAAACAG TAGCAGGTCA 20940

AATACCCGCG GATATGCAAG ATATATCAAC TGGTATTGGT GAAGTAAATA CTCAATTTGG 21000

CTTAATGGAT AAGCAATTGG AAGATACAAC AGGCCGAATG CTTAAATTTT CAGAAATTAA 21060

TGGCTCAGAT GTTTCTCAAT CAACTATTAA TGCAAAAAAA TCAATGGACC TTTTTAGGTT 21120

GTCTATTGAG GATTTGCCAA TGATTTTAGa TTCAGTATCT AAAACTAGCC AAGATACTGG 21180

AGTAGGGGTA GATCaGTTAT TTGATGCCGT AAAtAGAGGT GCGCCCCAAC TCAAAGCTAT 21240

GGGACTTGGT TTTTCTGAAT CAACTACGTT AATAGGTCAA ATGGAAAAAG CTGGTATTGA 21300

TTCAGCAGGA ACTCTTGGCT ATTTGGCAAA AGCTAGTGTC GTATATGCGA AAGATAATAA 21360

AACCATGCAA GAAGGGCTTA GCGGAACTAT TGAATCTATT AAAGGGGCCA CAACTGAACA 21420

AGAAAAACTC ACTATTGCTA GTGAAGTTTT TGGAACTAAA GCTGCTTCAA AAATGGTAGA 21480

AGCTATTGAT AGCGGAgCGT TGTCAATGGA TGGTTTAGCA GATTCAGCAA AAAACGCAGC 21540

TGGCACTGTG GATCAGACGT TTAGTGATAT TCTTGATCCA ATTGACCAAG CAAAGTTGGC 21600

ACAAAATCAA TTTAAAATAG CAATGGGTGA ACTTGGAGAA CAAGTACAAA TAGCATTATT 21660

ACCAGCATTT CAAGCTGCAA CGGATGCAAT AAAAAAAGTT TCAGAATGGT TTGGAAGTTT 21720

AACAGATAGT CAAAAGCAAA CCATACTGAA AATAGCTGGT GTTGTGGCTG CTATCGGTCC 21780

AGTATTAGTA GTTTTAGGAA CACTTGCTAG TTCCATTAGT AGTTTGATTC CAGTTATTGC 21840

TTTTATTGCG TCGCCAATTG GTTTAGTAAT TGCGGCGGTT GCCGCTTGGG tAGCTGCAAT 21900

CGTAGTTGCA TATAATAAAA TCGGTTGGTT TAGGGATTTT ATCAATACCT CCTTTAAAGT 21960

AATTAAAGAT ATTGTGGTTG GTGTATTTAA TGTTTTGAAA GATACGACAA AATCTACTTT 22020

TGATTTCATC ACAGGATTTA TTGGTGGTGC CATGGATGGG GCTGCAAAAA TTATTGGCGA 22080

TTATGTAAAT GAAATTAAGC GTATTTTTGG CGGTATCGTT GATTTTGTAA CGGGAGTATT 22140

TACTGGAGAC *TGGTCAAGAG* *CGTGGCAAGG* TGTTGTTGAC ATTTTTGGTG GTATTTTTGA 22200

AGGCATCGCT GCAGTAGCTA AAGCTCCAAT CAATGCCATG ATTACGTTAA TCAATGGATT 22260

TATTGGTGGA TTAAACAATA TAAAAATACC TAAATGGGTG CCAGGAATTG GCGGTAAAGG 22320

ATTTCATATT GGAAAAATCC CTTATTTAGC AGAAGGTGGA ACTATTCTAA ATGGCCAAGC 22380

CATTGTTGGT GAAGCTGGTC CTGAACTATT AACCGCTAAA AACGGCAAGA CAACAGTAAC 22440

TCCATTGTCA CCAGAAGAAA AAGCTCGTGG AATTGGTGGT GCTTTGAAAG GTGGCAACAC 22500

TATTGAGCAA CATGTTCATA TTGGCCAAGT AGATGCAAAT AATCCGAGTG AGTTAGATCG 22560

AATGAATCGC AAGCTTTATA AAGCAAGTGC GCAAGCTTTC TATGACTTaG GAGGTGTTCC 22620

aACGTGATTT TTATGAATCC TGATGAACCA AATTTCATTT GGAAAGATTT GAATGCAGTT 22680

CGTGATATGG GGTGCATTAT CGAAAATGAG CTGTCAGAGG TTTTACCAAA TAAACGATAT 22740

GAAACGTATT CGATTATCGG AAGAAGTGGT GAATTTAATG AAACGTTCAA TGATTATGAA 22800

CCCTTTGATT ATGAAATTGA AGATGTAACT ATTCCATATG AAAATTTAGC GGCAGTCAAA 22860

AGATGGTTAA CTGGTAAAAG TAAACTTATT ACTCACAATG ATGAAGATAA ATATTTAGAT 22920

GCTATTTGTA CAATGAGTAA ACCAACTTCA TTCAAAAATG AATGGGGTGT TTTTTATACC 22980

TTTAACGTTG AATTTAGATG TCAACCGTTC AAAAGAAAAG TAAACGAACA ACCAAAAGTG 23040

ATTAAAACAA AATCAATTGA AATTACTGAT CACGGTGATG AAATTGCTTT TCCTTATATC 23100

GAAATTAATT CAAAAGGTGG CGATATTACG TTAAACATTG GTAGTAACTC ACTAACGATT 23160

TTGCGTACAC AATCAGGAAT CGTCACTATT GATACCGAAA AGGGAAAAGC AATACAAGAA 23220
GGAAATCCAC TATTTACACG CGGCAGTTGG ATAAAAACGA ATCCCGGTCA AAATAAATTA 23280
AATATATCAG GAAATTTTAT AGAAGCTAAG TTTTGGAATA GGAGCGCGTA TTTATGACAC 23340
AAAATTTTAT TTATGCCTAT ACTGCTATTC CTGAAAATTT AAACGATAAC GGAATGGCTT 23400
TGCCAGATTG GCAAGATTTA CCAGAAATTA ACCGTGTGTT AAATGGTGTG TATCGATTCT 23460
ATGGTAACTA TGCAAGAGAT GGCCAGTATC GCTCATACTT AAAAAAAGGA AACTTTCTAA 23520
AGGCACAAGT TGAAGATGGA TCATATCAAT ATTTTGAGAT TTACAATATT AAAAAAAATC 23580
TGCAGTCAGT TTCAGTGACA GCGAGACACA TTGGTTTTAT GGCAAATAAG AATTTCATTA 23640
TTGATTCGTT CACTGCTAAC GGAAATGGCA CGCAAATTAT GAATAATTTA AAGGCTGCAT 23700
TAACGTTTAA GCAACGGTTT AACTATTTGT CGAATGTCGG TACTACACAT CAATTTACAG 23760
CAAAACAAGT AGGTCCAATC GATGCAATTA TTGGTTCTAA CAATGGCAAT CAAAATTTAA 23820
CAGGTGTTAC TGGTGGAGAA TTAGAGATGG ATAACTTTAA TTTGAAATTA GTAAAACAAA 23880
TTGGAGCAGA TAATGGCTTT AGAATTGATT TTGGAATTAA TTTGGAAGCT ATAGATGAGG 23940
ACTATGACGA CGAATCAATT ATAAACAGTC TTTTTCTTAT CGGTGGCGTA CCAGACAATG 24000
ATTATGACCA AGATAAAGAG CCAATCACGT ATGGCTTTTT AGAAATTGCT GGTGTAAATG 24060
ATAGTAACAG AAGAATTGGA AAACGTGAAA ATTCGGAATG TAAAACAGTT GATGAGCTTA 24120
AAAAATGGGG CCAGTCATTG TTTGATAAAG ACCGCATTCA TGAACCAAAA GTAACGCACA 24180
CTATTAGCAT GGTAGCATTA GAACACACTT TGGAGTATGA AGACATGTAC GAAGAGCTTT 24240
CTTCTTTGCA TTTTGGTGAT GTAGTACATG TTAGAGCAAA AGAAGTCGAT ATTGAAGTAA 24300
CGGAGCGCAT GGTGGAATAT ACTTGGTTTC CGACTTTAGG CAAATTTAAA AATATTGTTT 24360
TGGGGAATGA TTTATCACTT TACACCTCAA CAGTAAATAA TCAAACTCAA GAGCTAAAAC 24420
AAAAAATTGA TAATCGAACA GAAACATTAG TACAAAATGT TTTAAATGCA ACGGCATGGA 24480
TTACTGGAAA CAGTGGTGGA CATGTCGTTT TTCGTCCAGA AAAGGCCCCG TCTGAAATTC 24540
TTATCATGGA CAAAAACAAA GTAGCTACCG CAAAACGTGT GTGGCGCTGG AACTTAAATG 24600
GTTTGGGTTA TTCCTCCAAC GGCGTGAATG GTCCGTTCGA ACTTGCTATG ACTTCTAAGG 24660
GAGAAATCGT TGCTGATTTT ATTAAAGTGG GCATTATTAA CGCGAATGTT TTACAAACAA 24720
GCTTTAATAA AGCAACAGAC GATGTACTAA AATTAGTAGC TGGTGCTTTG CAAATTTGGA 24780
ACAATAAGAA AAAAATCATG GAATTGACTA AAAAAGGGAT GGAATTTTGG AATGGCTCTA 24840
GTCATATTGG GACAATTGGT ACGAAAGGAA ATCCCTTTCC AGGGGTAGTA GATAAAAATG 24900
GAAATCCTGT AGTTTCTGAT GGGAATTCAT TGCTATTAGT CGCAGATAAT CCCAAAAAAA 24960
TTATTGGTTT GTCTAACCAA TCAGGAACAG GACATTTAAT TACTGGTCCT ACACAGTTTT 25020
TTGTTGGAAA TAATTTTAAC TTTTTTGGTC CGAATGGAAG TAAAGCAATT CTGACAGTTG 25080
ATCGATTGAT TGTGGGCGGC AAAGAAGTTA TACCTGGTCA AAATGGTGGT GGCGGTTCTG 25140

GAGCTGGAAC AGGTGGTTAT CCATCAGAAG TTACAAGCGA TGCAGATAAA TTTGCTTGGG 25200

ACTTATGGAG TTACCTATTA GCTAACGGAT ACAGCAAAGC AGCTGCTGCA GGTATCCTyG 25260

GAAATGTACA AGGAGAAGTT GGTCCAAGTA TGAACCCAGA TACCGAACAA ATAGGCGGTC 25320

CAGCTTACGG ATGGGTTCAA TGGGACGGTT CAGCATATCC ATTGGTAGGC GCACCAACTT 25380

GGAATGGCCG AGAATATGTA CAACGCTTAA TCGCAGCTGC AGGTATCAAA CAAGACTATA 25440

GGACGTCATT AGCCCAAGCT CAATTAATTA ATTGGTGTAT GTTCAATGGG CAATGGTTAG 25500

GACAAGTAAG TCCATTAACa GTTGATGaAT TTaAAnnTCT ACGGTCTTTG GTTGGTAACG 25560

TTACTTCTGT TTTTTCAGAA 25580

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1260 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

ACTTAAATAC AAATGTTGGT CGTGCACAAT GGGAACCTCA AAGTATCGAA GCGCCCAAAA 60

ATACGTTTAT TACTGGTAAA TTAGCCAGTG CAGGTGTCTC AGGTTTTGAA CCTTTGGATG 120

AACAAACAGC TACTCGTTGG CACAAAACAA ACATTACAAC AGGTCCCCTT GACATCACTT 180

GGAACTTAAC TGCCCAACAT AGAACTGCTT CTTGGGATTA CTATATTACT AAAAATGGCT 240

GGAATCCCAA TCAACCATTA GACATTAAAA ACTTCGACAA AATTGCTTCA ATCGACGGCA 300

AACAAGAAGT TCCTAATAAA GTTGTTAAAC AAACAATTAA TATTCCGACA GACCGCAAAG 360

GTTATCATGT CATTTACGCT GTCTGGGGCA TTGGTGATAC GGTGAACGCC TTTTACCAAG 420

CGATTGATGT GAACATTCAG TAAAAATACT CACATCTTAA ACTAAATTAA AGGAGTTCTC 480

AAAAAATGAA GCTAAAAAAA ATAATTCCTG CTTTTCCCCT TCTTTCAACC GTTGCAGTTG 540

GCTTGTGGTT AACGCCTACT CAAGCTTCTG CAGATGCTGC GGATACGATG GTAGATATCT 600

CTGGCAAAAA AGTGTTGGTT GGATATTGGC ATAACTGGGC CTCAAAAGGA CGCGATGGTT 660

ACAAACAAGG AACATCAGCA TCACTAAACC TTTCAGAAGT AAATCAAGCC TACAATGTCG 720

TACCGGTTTC CTTCATGAAA AGCGATGGCA CGACACGGAT TCCTACGTTC AAGCCTTATA 780

ACCAAACGGA CACTGCCTTC CGACAAGAAG TCGCACAATT AAATAGTCAA GGTCGCGCAG 840

TTTTATTGGC ACTTGGTGGA GCAGATGCAC ATATTCAATT AGTCAAAGGC GATGAACAAG 900

CCTTTGCGAA TGAAATCATT CGTCAAGTGG AAACATACGG CTTTGATGGT TTAGACATCG 960

ACTTAGAGCA ATTGGCGATT ACTGCTGGCG ACAACCAAAC CGTCATCCCT GCTACGTTGA 1020

AAATAGTCAA AGACCATTAT CGAGCACAAG GAAAAAATTT CATCATTACG ATGGCACCAG 1080

AATTCCCTTA TTTAAAACCT GGTGCCGCTT ATGAAACATA CATTACTTCC CTAAATGGTT 1140

ATTATGATTA CATTGcCCCA CAATTATATA ACCAAGGCGG CGACGGtGTC TGGGttGATG 1200

AAGTTATGAC TTGGGTTGCT CAAAGCAACG ATGCTCTAAA ATACGAGTTC CTCTATnATA 1260

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8188 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

TCCAATGGCT TATCGATTAC AACAACTGGA AGTtCTTTTA AATCAAGCTT GCTTTTTCCA 60

ATTAGAtACC TATGGAAACC ATCAwTAATA ATATATTCGT CTTTTTTATC ATCACAATAG 120

CAAACAATCG GTTGTGTATA TCCGTCTTTT AAAATCGATA ATTCTAAAAG CTCCATTTCA 180

GGTCCAGCGA CATGATTTGG GTTATAACTA TTACTTTTTA TTTTCTCTAC TGGAACCATT 240

TTTACATTTA AAACGGGCAT ATCAATCTTG TTCATTGATT TCTCTCCTAA AAAATAAATA 300

ATTCTTTGTC TCTCTATATT GTTTAAAACA ATAATGTTTC AGATTATTTA AGTACTGGTC 360

TTTTTTTAAA GAAATTCGTA CACACTTTTT TTCACATTTC TTGTCTATGA AAAATAAAAC 420

CTTCTCCAAT AAATACTTTC GAGTATCGTC ATTTTCATCT TCAACAAAAA CATCTCCAAT 480

ATCAATAAAC TCATTGGAAA TTTGAATGGA AGAAAATCCA ATTATCCTTT CCTTATTCAT 540

TAACAAAAAC CAAATAGTAC TAGATGTATT TATTAAATAA GGTAATTTTC TTCTATATAC 600

TCTTTCAGCA AAATATTTCC CCATTAAACT GTAAAAATTT TGTyTATTGT ATGGGGGTAT 660

ATAAACTTTA ATATCTATTT TGCTCATTTC ATTCCTCCTA TTTTTTGTAC AATATTTTCG 720

TATTTCTTTC CTTTTAATTG ATTTATCCCC AAAAAGTCTT TATATTCTTT AAACATTTCA 780

TATAGTTCAG CTTCATCTTT TTTCGTTAAA CTATAATTTC TAGTAGATAA AACAAAGTCA 840

TTTTTTTCTA AAGCTTTAGC TATCCCTTTC CAGTCATTCC AAAGCCTTtC ATCTTTTATC 900

CAATCTTTTC TTTTTGATTC ATCTGGCATA TCTTGGATAT CCATATtATA tGTATtCktG 960

TAATAaGCAA AAaGAATACA TATCTTTCTG TAATAATGAT TCATTAAAAG AGGAGAATAT 1020

AATCCTAATG TTTCTAGTAA AAAAATAGAA TATCTTTGCC AAGTCATAAA TGAAGGTTTT 1080

TGAGAAGTAT TTAATCCTAA AAGACTAGTT TTACCATATA TATTTCCAAA ATTAACTCCA 1140

CTTACTCTTT TTACTAATTT ATCCCATGTT TCAgGTTCAA GTTCAGCAAA TTGATTTAAC 1200

GCTTTTCTTT GTTGTAATCC GAAAGGCTGG CATACTCTCA TGTCCCGGTA CTTATATCCA 1260

ATCTTATACA TTTTCTCATA GATATAATTA TATTCTAGTT CACATTTAAA AGTTGCAGCC 1320

CAAATGTCAC CAACTTTCCA ATCAAATAAA GGATAAAACA AAATATTTCC TTCTAAAGTT 1380

ATATTTATCC AACATCGATT CTTATATTTA TTTTCACCTT TTGTAACAGC ACTATGTCTG 1440

TGTAAACTTT CTTGTGCCCT TATACCTACA CCACATGATA TAGCACTTTG GGGATATTGT 1500

TCTTTATACC AATTTTGAAA ATAGCGAAGA AATCTGTCGG GATTACCATT TGATAATGAA 1560

TAATATTTTT TTAAACCGTC ATCAATATTT TCCAAAGTTA TGGCATTTTT AGGAAGAGAA 1620

TGAATCCATT TTTCTTTTTC GCTTGGATCC CACATTTTCC ATTGAGGTTG AAAAAAACTA 1680

GTGTTGTTAT CCTCAAAAAA AGGTAAGCAA AAATGATAAA TATTTTTTAC TCTAGGTAAC 1740

TGTTCAATTT TTTTTATAAA ATCAACAGTT GCTGTGTAAT TTGCTTCAAT ATTTAATATT 1800

AATAGATCAA AATTTCTATC TAATTTTTCT GCAACTAGAT TAGCTAACTG AACCATTAGC 1860

CCACTATCCT TACCGCCACT AAATGAAAAA TACACCTTAT CAAATTCTTC AAATATGTAT 1920

TCTAATCTCT TTAAAGTAGC ATCAAAGACA TTCATTTTCC ATCCTCCCTA AAAAATAAGC 1980

ATTTAATATC AATGGTAATA CATTAAAAAC TTTGTCAATA ATAAGGATAT ATTCCAAAAT 2040

CCATTCATTG GAAATAGAGA CCTTCTTTAT TAATGTATTA TTACTTACAT AGCGAAgcAA 2100

TaCAAACATT ATCCAATAGT TTCATCTACA AAAAATAAAT TACTTTTAAA ACATCTTGAT 2160

TATGAATTCA TTTCCCAAAG GCCTTTTAAA ATTGATACAG ATGTTCGTTC ACTTTTTTTA 2220

GACTTATTGT ACAAGTCTTT TTAATAGTCC CAGAACAAAA TAGTTTGTTT ATCCAGCTTC 2280

AAATCATTAA AGGCTAATGA ACGGTTTATT TTTTCGATTA CTTTAGATAA AGCTACAGAA 2340

ATTTTTATAT TTTCACAATG CTCAGTTTGA ATTTCAATAA TTATATCATC AATTAGTTTT 2400

TGAATTTCTT ACCTATCAAC ATATTCAGTC TTAGAAAGAG AATCTTTTAA AGCTACTAGT 2460

GATTCAACGA TTTCAGACAT TTTCTTCCTC CTTTCTAGTG GCCTGTAAAA CCGTGTAGAA 2520

CAGAGTCCAT CGAACCGGTA ATTACACTAT CTACaGCGCC ATTCCAATTT ACTCTAcATT 2580

TTCATTTTTC AGATTTGTCA GCCCATACAG ATAGCCTTTT TGACGTTCAT ATTCGGATTA 2640

TACCATTCAG CCGCAGAAGC AGATTCGCTC GACACAAACA ATCCCATACC AACAAAACCA 2700

ATACACATCA CACTATAGCC AACTAGTTTT TTCCTCTCGk ATCCATCTCC TAAAAAATAG 2760

AATAGTtCTT GTAATTATTC ATAAGATATA CTTTTCCGTT TTTTATCCTT TCTGCTTTAT 2820

AACAACTTTT TTCAAGTAAA GAATGACATA CTTCTTCTTA AAAAAACAGT CTGTTTTTTT 2880

CTATTTTTTC GGTGTATAAT TAATACGAAT GTTCATTTTT TGTTTTAATA CGCACTCCGT 2940

ACATAATAGA TCTTGATTGC ATCAATCTTT TGTACTTTTT GATTTATCTT AAGCTGAATG 3000

AATAGAAACC TAGTTACCGC ATTACTTTTT TGTTCTCCAA ATAACAAGTA ATGATCCGTT 3060

TGGGTTATAT CTTGTGAAAC ATTATGAATA CGACCAATGT TTTTTTGTTG TACATGTGGA 3120

TCGTTGTCAA TTCCATTTTC TTTTATACAA CAATCGATTA AATAGTTTTT TACGGATTGA 3180

TTTCGTTCGT ATATGTCCGT GTAATTTACC CATTTTCTTC CAAAATCCTT TAACAAATGT 3240

TGAATTACTT CAAATATGTT GGATTTTATT TTGTTTTCCT GATTTCTTTT ACTTTCTATA 3300

CTTAAACTTT CTGCTTGCAT CATCGTTTCT CTTGATATTG AGGAAACTGC TGTTTTTTTC 3360

TTCAGAAGAA ACCCACCTAC AACTTCTAAG GACAAACAAA AGCAACCAAG AAAAAACAAT 3420

ACTTTTTCAC AATTGTTTCA TCTCCCTTTT CTACTAAAAA TTTTCATTGA GAGAATTCTT 3480

TTATTTGGCT TTACGAATCA AACTTCGAAT TTTCGTTTGG TTATAGGTGC GTTGATACAG 3540

TGCTATGATT CCCCTTGCTT GGCGTTCTGT TCCACAGTTG AAAAATAATT CTCACTATCT 3600

GAAACATCCG TTATGATACC GGTGTGTCCA TAGATTCCAG GAGACAATTG GCTTTACCAG 3660

TTGGCCATGA AAGGATTCTT TTAAACGACA CTGAAATAAA AGAAAAGAAG CCAAATAATT 3720

ATAAGGCTTC CTTCCTTGAA ATTGTTTTCG TGTCATTGAA AAAGCAAATA ACCACGGACA 3780

AAAAATTTGT CGAGTAATCG TATGTTCCAT ATCAAATGGA AAAATTCGAA AAATCCCCCC 3840

CAAAAAAGCC AATCCTAAAC CATACCACAA ATGTTTGTGT TACGGGACGA GGTAACGCGA 3900

CCCCGTCCCA CAATTTATAA AACACTCGCT CTATATACCA AATCTTTGTC CAATTGCGTA 3960

TATTGCTCAT TTTATCGCCC CCATGACCGC ATGCGATCCA GAATCCAATT GAATACTCCA 4020

GACGTATCTC CAACTAACCA AACAGCTACT GCCAGCACAA GAATCAATGG AATGGCTTTG 4080

CTTGCTTTTC GTTCAAAGGC AATATAGGCC GCACCACCCC AAGCATCGCC AATATAATCA 4140

AATAATTAAA ATCCCCAACA AAATTGGTAA ACACATTTTG AAAAAAATCA TTTGCTTTTT 4200

TCCTCCCTGT TAATACACAT TGAACTAATG ACAATACACA TCACCACAGA GCACAACTGA 4260

TAAAAATTTC ACTTGATATT CCTCATTATG CATTTGTTCT GTTCTTTGAT TGCTGTTTTT 4320

ACGTAAGAAT TATAAATAAT TTCTCACACG GCGATACATT ATTTTCACAG GAACATGCGT 4380

ACAATAAATC CGGTTCTGCG TACTTTTCTT TGTTTTAGGA TCAAGTTTCA TAGAATAGAA 4440

ATCATAAAAA TATTTACCAC TTTTTTATCA TATGTTCCTT CGATCACTAG TGAAAAGCTA 4500

GGGTGGTAAA TCCAACAAAT ATAATTGATA CTTCCCTCTA CTGGATAAAG GGACAAAACC 4560

AAGTACCTTT AAATGGTCTG ATACAGTCAT AAACACGATC TATCTCCTCC CAAATAAAAC 4620

AAAAAAAGGC TGTAATCGCA TTTTGATTAC AGTCTTTTTT TGTAAATGTA TTAATTAAAG 4680

CTTTACTTCC GGTCTTTTTA TTTAATGACT TTAGTCAGTT GAAATCTCTA AAGAGATTTC 4740

GACAAAAAAA CCACGCGTAA TGCGCGTGGA GGTCAGAGCT TTAGCGATAA AAGTTTAGCA 4800

CATACTCGTT GGAATTGTAA GTATCATCTT GTGTTTACAC CAAAATATAG AAGAAAAGTA 4860

ATCTATGGTC AATTGAGGCG AGATATAGGG AAAATACTAA GAAAATTGTG TGAGATGAAA 4920

GAAGTCGAAA TCATTGAAGC GCATGCAATG CCCGACCATA TCCATATGTT GGTAAGAATT 4980

CCACCGAAGT TGAGTGTATC AGGTTTTATG GGTTTCTTAA AAGGAAGAAG TGCAGTAATC 5040

ATTCATGAAC GTCATGCGAA TTTAAAATAC AATTATGGAA ATCGATCCTT TTGGTCAAAA 5100

GGTTATTATG TGAGCACTGT AGGACTGAAC CAAAAAACGA TAGCAAAGTA TATACGAGAA 5160

CAAGAAGCAG AAGATCGAGT AAGGGATAGT ATAAACAAAA GAGAGTATAA AGATCCATTT 5220

AGGAAGTAAC AGCTAGATAA GAATAGAAAG GTAGCGGTTG GCGGTCTCTA AAAAAGCTCC 5280

CTTAAAGGGA GCAGCAGGTA GCGAGCCCTT CTAGGGCTAT TaAAAAGCCA CCCGCTATGC 5340

GGGTGGATAT TTACTTAATA AATCAATTTT CCCAAATTAT TTTCTTTAGC TTTCTTAACG 5400

ATTTCTGTTG CAACTACCAC ATCTAATACA GCCGAACCAA CTGTTTTGAA GATCGTGATT 5460

TGTTCTCCAC TTGTCCGACC GACCAATTCT TCATTGATCA GTTGGCCTAG TTCTCCATGA 5520

TAGCTGTCTC TTTGAATATA TCCATCTTGC AGCGGTGAAA TGaAGTCCCC AGCTTCTTTT 5580

AGAACACCGT CCATGGTATC AAAAATAACA ACATCTGCGG CTTTAATTAT TTCTCGTGGa 5640

ATTTCACACA TTTCCGGTGT GTAAGCACCG ACTCCATTGA TGTGCGCACC TTTTTTGACC 5700

CACTCAGCAG AGAATGTCGC ACGTTTTGAA GTGGTCACAC TTGTGATGAT GTCTGCTCCT 5760

TCAACGCATT CTTGACTGGT TTTTGTTGGG CGCATCGTTA CAGAAAATTG TTGCATCATT 5820

TCTTCAGCAA ATTGGTTGGC ACGCTCAAAA TCAATATCAA AAATCCGAAC TTCTTCTAAT 5880

TTACGAACGG TCAACATTGC TTCTAATTGT GACTGTGCTT GTCCACCGGT TCCAATCAGC 5940

GCGCCGATTT TAGCATCTTT TTTTGCAAGA AGTTCTGTTG CTGCTCCTTG AACAGCGCCT 6000

GTTCTTAATT GTGTCAAATA AGTACCATCT AAACAAGCGG ATACCATACC CGTTTCTGGA 6060

TCCAATACGA TCATCGTTGC TGGTACACTT GGTAAGTTTT TCTTGATGTT TTTAGGATAC 6120

ACGGAAACGA TTTTGACACC TAGTGCTCCT TCGCCACCTT CTACATAAGC TGGCATATAT 6180

AAGCTTTGTC CGTTACTTTT TGGCACATCG ATATTGGTGC GAAGAGGCAC AATTGCCTTT 6240

CCTGCTGAAT ACAATGACAA TGCTTTTTTG TCTGCATCGA TGGCTTCTCT CATTGAAAAA 6300

TTTTTGATAA TATCTTCTTT TTTAAGAAAT AGCATTTGAA TCGCTCCTTA ATTAGCTGCT 6360

GTAGCTTTTT TTAATTGACT GCTCAACATC GTATATGCAC TGTATAGTAA GACCAAAATC 6420

ACGATCCATT TCAGAATATA GGTATCTAAA GATTTGACTA GGAATACAGC GACCAATACC 6480

CCTAATACAC CGAAGGTTGA AGTGAAAAGT GTGATTTTTC TGCTGTATTC ATCAAATTTA 6540

ATGAATTGAA CACTGCCGAT CGGTACTGAG AATGTACAAG CCCCCATCAT AATAGGAAAG 6600

GCTGCAACTG GATTCAAGCC TAGCAAGTAG ACAGTTACCA TTGTTAACGC ATAAGAACCA 6660

ATGCCGATAT TGTTCAATGC GCCATAGACA AAGAGTAAAA ATCCTGCCAA GATTAATTTT 6720

CCGCCATACA ATTCACTGGC TGTTCCGTTT GAGgGAAkCC AATTCATTTG TCCCATAAAA 6780

ATTAAGAATG CAGCAATGAC CAATCCTATG CCGACAAACA GTTTGATCGT TTTCTCTGGT 6840

AACTTCACGA CAAATCTTGG TCCTAAATAT GCACCGATAA TTTGGCAGAT AATACAAACA 6900

GCTAATGTCT TGATCCCTAC GTTGATAGCA GTGATATAAG ACAATGCCAT GACGGCTACC 6960

GGTACTACAC ATTGGGTATT CAACGTTCCG GGTAGTTTTT TCATGCTGAC CCATTTCAAC 7020

TTTGGATAAA GGACGGTTCC GATAGCAAAG TCGGAAACGC CAAATGTTGA AAGAAAGAAG 7080

GTGATAAATG ACGTGAAGGG CAACAGACGC ATATCCCCAG GTTCTTCTTT GAAATTTTGT 7140

TTGTTTTTGA AAAGGTCTCT TAGAAAGTAA ATAATAAAAT ATAGATTGAT TGCGACAATT 7200

AGTGCCAGCA AAATATTCCG AATCATAAAA ATTCCTCCTA CTAGTAAATC TTTCTGATTC 7260

CCTCAAAGTT CTCACCATAT TTTGCTTTAA GTTCTTCAGA AGATTCTGCA ATCACTTTTA 7320

GTGTATATGC ATCATGTTTG ACATATTCTT TAGCTAGATT GACCATGGCG TGGTTTTTTG 7380

TCCAAGTGAC ATCGGTTGAT TGCAGTTCTT CATTAATGGT GCCAAACACG ACTTCTTGCT 7440

CATCGGCGAC GATGTTGATC CATCTTGCCC CAAAGTCATC AAGTTTGTCT GTTTCAAAAC 7500

CATGAATGTA ATACTTTTCT AAGGGAAGAT CATATTCATG GGTAGAACTG AATAAAATCA 7560

GTACAAATTT ATCGACTTTT TTTTCGGCCT GCTGCAATGC CTTTAGCAGT GAATCCgTGA 7620

GGTTCTCATG CCAAACTTGG ATGAGCAAAC TTTCTTTTGC ATTTTtCACC AAATGCTCTG 7680

CTTaTCCaGA ACATACTcGA TGCCATCAAC TTTnCCACAG CATGTCTTCT TCATCTTTTT 7740

CCTTAATCAT TcCTAGATTC TTTGTAAGAa AAGaAAGATC aTTTTTAACA CTTTTTTCTA 7800

ATTTCTCTAA AAACTCaTTT GCCGAGATGG CATGATAAAG TTTAGGTTCT GATTTATTGA 7860

CGACAATCAA ATTCTTTTTT AATAGTTTTT CCAAAATGTT ATATACTTTC GAGCGTGGCA 7920

CcCTGAGATT TTACTTACTT CATATCCTGT AAGATTACCT TTTTCTAATA GGGTGGTATA 7980

CACTCTTGTT TCCATCTCTG AAAAATCATA TTTTTTCATG ATTTGAATCA TCGTTTCCAT 8040

TGATTAATCC TCACTTCATT TTTAGTAGCT ACTACGGTAA CAACTATAAT TTTTTTTTGA 8100

TCATCTGTCA AGAGTTAGTT CAATAAAACA CATATAAGAA TTAGCTTAAA AACAAGCTTT 8160

tTATTAAAaA TaAATaATAT AATAGGGA 8188

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3467 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

TGATGAGCAT AAGGAGGACA GGCATGAACA AAAAGAATAA TGGCATGAAA AATGGTTTAT 60

ACTACGTCTT AGTAGTACTT GCCATGGTCA GnGTTGTCTA TTTCATTTTT GGAAATAACA 120

ACCAACAATC GCCAGACATC GATTACTCAA CCTTCCAACA ACAATTGGAA GACGGAAAAG 180

TCAAAGACAT GACCATTCAA CCAACTAATG GGGTTTACCG GATTGAAGGT CAGTATAAAG 240

AAAAACAAGA AGTGAAAGAC ACAGGTGGTC TGTCATTGTG GGGCTCAACG CAAGCGTCGT 300

CAAAAGGCTT CACAACAACC GTTTTACCTA GTGATACAAC CTTAGCTGGC ATTCAAGATG 360

CGGCACAAAA TAACAAGGTG AAGCTAGTTG TCAAAGAACA ATCAACAAGT GGTGCTTGGT 420

TGTCACTGTT GTTTAGTTTC TTACCATTAG TGATTATTTT CTTCTTCTTC TACATGATGA 480

TGAGCCAACA AGGCGGTGGC GGTGGCGGCG GTGGCCGTGT GATGAACTTC GGTAAATCAA 540

AAGCCAAAGA AGCCGATAAA AAAGCTAACC GTGTCCGCTT CTCAGATGTA GCTGGTGCGG 600

AAGAAGAAAA ACAAGAATTA GTTGAAGTAG TGGAATTCTT AAAAGATCCA CGTCGTTTCG 660

CTGAATTAGG CGCACGGATT CCTGCGGGCG TACTATTAGA GGGACCTCCA GGGACTGGTA 720

AAACGTTACT TGCTAAAGCC GTAGCAGGTG AAGCAGGTGT GCCATTTTAC TCAATCTCAG 780

GTTCAGATTT CGTTGAAATG TTTGTCGGGG TCGGTGCTAG CCGTGTCCGG GATTTATTTG 840

AAACAGCGAA GAAAAACGCG CCTGCAATTA TCTTTATCGA TGAAATCGAT GCAGTTGGTC 900

GTCAACGTGG TGCTGGTATG GGTGGCGGAC ACGACGAACG TGAACAAACC CTTAACCAAT 960

TATTAGTTGA GATGGATGGT TTTGATGGCA ATGAAGGAGT TATCGTAATT GCTGCAACGA 1020

ACCGTTCAGA CGTGTTAGAC CCAGCGTTGT TACGTCCAGG TCGTTTTGAC CGCCAAATCT 1080

TGGTAGGACG TCCAGATGTC AAAGGTCGTG AAGsgATTTT aCGTGTTCAT GCGAAAAACA 1140

AACCGTTAGC TGATGACGTT GATTTGAAAG TTGTTGCACA ACAAACACCA GGTTTTGCTG 1200

GTGCTGATTT AGAAAATGTC TTAAACGAAG CAGCCTTAGT TGCCGCTCGT CGTAACAAGA 1260

AAAAAATTGA TGCTTCTGAT GTGGATGAAG CCGAAGACCG CGTAATTGCT GGTCCAGCTA 1320

AAAAAgATCG CGTGaTTAAT AAAAAAGAAC GCGAAATGGT GGCTTACCAT GaAGCAGGAC 1380

ACACAATTGT TGGTTTAGTC TTAAGCCGTG CTCGTGTTGT CCaCAAAGTA ACAATTATCC 1440

CTCGTGGTCG CGCAGGCGGT TATATGATTG CTTTACCAAA AGAaGATCAA TTCTTAATGA 1500

CTAAAGAAGA TATGTTTGAA CAAATCGTTG GCTTACTTGG TGGACGTACA GCAGAAGAAA 1560

TTATCTTTGG CGTTCAATCA ACAGGGGCTT CAAATGACTT TGAACAAGCA ACAGGGATTG 1620

CTCGTAGCAT GGTAACTGAA TATGGAATGA GCGACAAATT AGGACCTGTT CAATATGAAG 1680

GAAACCATCA AGTCTTTGTC GGTCGTGATT ATGGTCAAAC AAAAGCTTAC TCAGAACAAG 1740

TGGCGTTTGA AATCGATCAG GAAGTTCGTC GTATCTTAAT GGATGCTCAT ACGAAAGCGC 1800

ATGAAATTAT CGAAGCGCAC CGTGAACAAC ATAAACTAAT CGCTGAAAAA CTATTAGAAT 1860

ATGAAACATT AGATGCAAAA GCAATCAAAT CATTATTTGA AACAGGTAAA ATGCCTGAAG 1920

GTGCTGACAG CGACTATCCT AGTGAAAAAG AAGCGCAAAC ATTCGAAGAA GCAAAACGTG 1980

CTTTAGAAGA AAAAGAAGCG CAAAAACAAG TTGAAGAAAA ACAAGACTTT GAAGAAGCGA 2040

AAAAAGAATT ACATGATGAA GCAGAAGAAG TCAAAGTAGA AAGCGAACAA ACAGAAAAAG 2100

AAGTCCAATC TGAAGAGAaA AAAGATTCAG ATTCCAATTC TGAATATGAT CGCAATAACT 2160

ACGAAGATCG TTATAAATAA CCAATTAGTA GAAGGTGCAG TATTTGTCGA GTTGACGAAT 2220

ACTGCACCTT CTATAATATT AAGATATTCT ATAGAAGGAA GATTATTTGG TTGAAAAAAC 2280

ACCTGATTCC TAGTATTATC yGACTAATGG aCmAAAAAGC ACCAAGAGAG nGAGGAAAAA 2340

GAATGGAAGA TTATTTAGTA AAAGCCCTTT GTTATAAAGG TTCAATACGA GCATATGCCA 2400

TCAGTGCCAC TGAAACTGTT AGCGAAcGCa aAGAAGACAT GATACATGGA GTTCTTCTAC 2460

TGCAGCGTTA GGACGCACAT TGATTGGAGC CCTTTTACTA GGAGCGACTT TAAAAGGTGA 2520

CGATAAATTA ACTGTAAAAG TACAAGGAAA TGGCCCAGCT GGCGCTATTA TTGTCGATAG 2580

TAATGGCCGC GGAGAAACAA AAGGCTACAT TAAAAATCCA CATGTCAGCT TAAAGTTGAA 2640

TGCAACAGGT AAAATTGATG TACGTGGTGC TGTCGGTAAC GAAGGAATTT TCACAGTCAT 2700

TAAAGATTTA GGATTAAAAG AAACCTTTTC TGGACAAACA CCTATTGTTT CTGGTGAAAT 2760

CGGTGAAGAC TTTACATATT TTATGGCGGT CTCTGAACAA GTTCCTTCAG CTATCGGTTT 2820

AGGTGTATTG GTGGACACGG ATGAATCTGT GAAAGCAGCC GGTGGTTTTA TGATTCAAGT 2880

AATGCCAGGT GCCGATGAAA GTACCATTGA TTTTATTGAA CAGCGTCTAG CGGAAGTGCC 2940

ACCAATTTCA CAATTACTAG AAAATGGAGA AACACCTGAA CAAGTCTTGT ATCGCTTATT 3000

GGGAGAAGAC GAAGTAGAAA TTTTAGAAAA AATGCCTGTA CAATTTAAAT GTGACTGTTC 3060

CAAAGAAAAA TTTGCGACCG CTTTGATTGC TGTTGGGATC GATGAATTAA ATGCGATGAT 3120

TGATGAAGAT CATGGAGCAG AAGCAGTTTG TCAATTCTGT AACAATAAAT ATCACTATAG 3180

TGAAGAAGAA TTAATTGAAc TTCGCGATGA AGCGATTCGC AATACGAAAC AAAAGTAGGA 3240

GGAAAGACTA TGTGGAAAAT TGGAGACGTT GAAATCCCCA ACCGTGTTGT CGTAGCTCCG 3300

ATGGCGGGTA TTAGTAACGC CGCTTTTCGT GTGACTGTAA AAGAATTTGG CGCaGGATTA 3360

GTCGTTTgTG AAATGATTAG TGACAAAGGa ATTAAATTAA GAAACAAAAA AACCTTAGAA 3420

tGCTATATAT TGATGAGCGG GAATAnCCTT TAAGTGTACA ATTTTGG 3467

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21126 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CTTGAnACAA TTGGGGCAAA AAATGTTTTG GCAAATGAAA AAGGCTGGTT ACCAGTGACA 60

GAAGAAGCGG CCATTGCTGC AAAACCAGAG GTCATTTTAA CCAATGTTAA CTATATGAAA 120

GATCCCGCCA AAGAAATTTT AGCTCGTAAA AACTGGGAAA GCGTTCCAGC AGTTCAAAAC 180

AAAGAAGTTT TCGAAATCGA TAACATGTCT AGCTCTTTAC CAAATAACCA CATTACCAAA 240

GCATTAAAAC AAATGGCAAA AGCTGTTTAT CCAGAGGAAT ACAAGGATTT AAAAGATGAG 300

TAAAAAAACG TGTCTCATAC TGATGCTCTT GATTAGTATG AGCATGCTCT TTTTGGGTAT 360

TAAAATTGGA AGTGTTCAGA TTTCGTTAAC GGATTTGCTT CAAATTTTAG TGAAAAAGAC 420

GGTTTCAGAA GATGGGATTT TAGAAGGAAT CATCTGGAAT GTTCGGTTAC CGCGTGTCGT 480

GATGGCCTAT CTCGTGGGCG CAGGCTTAGC AGTAAGTGGC ACGGTGATGC AATCGCTCTT 540

AGGTAATCCT TTAGCTTCTT CGTATACTTT GGGCGTATCA GCGGGAGCAT CCTTAGGTGC 600

TGCGTTAATC ATGGTAACTG GAGTAACTGC TTCGATCTTA GGGGCATTTT TATtGCCTTT 660

AACTGGTTTT ATCTTTGGCT TAGCAACGGT CTTTTTAGTT TTGTTATTTA CGCAAACAAT 720

GGATAGTCAA ATGTCCAATC AAACAGTTGT TTTAGTAGGA ATGATTATGA CCTTGTTTGT 780

CGGTGCTATT TTGACGTTGA TTACAGCACT TTTCCAAGAT TATTTGAAAC AATTAGTCTT 840

TTGGCAAATG GGTAGTTTTT CTGGCAGTAA TTGGCAAAAA ATTGCTATTT ATTGTCCAAT 900

CTTGCTTGTC AGCAGTCTGT TTCTTTGGTT TGATGCCAAC GCATTGGATG TTCTCGGTTT 960

AGGAGAAGAA CATGCGATGT TAGCGGGAGT AGAAGTCAAA ACAGCAAAAC TACGCATTAT 1020

TCTTTTGGCT AGTTTATTGG CAGGAAGTGC GGTTTCATTT GTCGGTGTGA TTGGTTTTGT 1080

TGATTTAATA GCACCACATA TTGTCCGTCG CTATTTAGGG GCGACTCATC GCTGGCTGAT 1140

TCCCGGTTCA GCAATCTTAG GTGGGACCAT CATGGTACTA GGGGACACGA TTGCCCGCAC 1200

GATTTTATCA CCAAGAGAAA TTCCCATCGG GGCCGTAACA GCCTTGATTG GAGCGCCATT 1260

CTTCTTGTAT ATTTACTTTA AAAAAAGGGG CGTTGCTGCA TGATTGAATT AAAAAAAACA 1320

GCTGTAACGC TCCAACAAAC CCCCATTTTA AAAGATATTT CCCTCGTTTT TCCGACAGGT 1380

AGTAAAACTT GTATTTTAGG TCCAAATGGT TGCGGTAAAA CGACACTTTT AAAAACGATT 1440

GCCGGACTTG TCCCTTATAG TGGTTCTGTT TTAATTGATG GTCAAGAAGT TCATGGACAA 1500

AAGCGGAAGG AATTGGCAAC GAAAGTAGCG ATGATGAGTC AATTCACGAC AGTTGCTTTT 1560

GATTATACGG TTTACGAAAC AGTTTTGATG GGGGCGTATC GACAACAAGC ACAACGTTTT 1620

TTACCTATTG TTTCtAAmCA rGAAAArGAA CGAGTCCTTw ATwATTtAGA ACGCACGGGT 1680

TTATTGCCTT TAAAAGATAA mGTTGTCAAT CAGCTTTCAG GAGGACAACA ACAACGAGTA 1740

TTTTTGGCTA AATTATTTGT ACAAGATCCA GAAATTATTT TATTGGATGA ACCCmATAAT 1800

CATTTGGaTA TTCGGTATCA ACAAGAATTG ATTCAGCAGT TAAATGAGTG GAGTGCGCAA 1860

GAAGGGAAAA CGTTAATCGG GGTGTTTCAT GATATCCGTT TAGCGTTAAC GCTAAGTGAA 1920

AAAATTGTGT TTATGAAGCA AGGGAAAGTG GCCGCGCAAG GTGATTTTCA AACGTTGGCT 1980

TCCAAAGAAT TTTTACAAAC GATTTTTGAA ACAGATATTG TCTCCTATTT TCAGAAACAA 2040

CATAAAGTTT GGGAAACTAT TCAATAATTT TGACAAGTTT TTTGGAAACA GGTAAGATAA 2100

AGGCAAAGAA TCAGAAACAC AATGGCGTGT TTTAACAAAT CGGCAAAGGA GCCCAAGACT 2160

AAGTACGTGA TGGTACTTGC TTTGGGCTCC TTTTTTATGG AGGTAACGAT GAAACGAATC 2220

ATTTTAATGG GAGCTATCGG TTGTGGCAAA ACCACGTTAT GTCAAGCTCT GCAAGGCAAA 2280

GAATTAATTT ATGATAAAAC ACAGGCGGTG GAATTTCATA CAGAAATGAT TGATACACCA 2340

GGCGAGTTTA TTTTGCATCG ACAATATTAT AATGCATTAA ACGTCACAGC GGCAGAAGCA 2400

GATGTGATTG GtTTGGTTCA AAGTGCTGTA GAAACACAAC AAGTTTTTTC ACCAGGATTT 2460

GGTAGTATTT TCCCGAAAGA AATCATTGGT ATTTTAACCA AGATTGATTT AGCACAAGAT 2520

TCTCAGCAGT TAGAGATTGT TCGGCAACAA TTAAAAAGTG CGGGTGCAaC AAGAATTTTT 2580

GAAATTTCTT CTGTTGAAAA AATTGGCTTG CAAGAATTAG TAGATTATTT AGAGGAGGAT 2640

GAAGCAGAAT GAAAATTTAT ACAAAAACTG GCGATAAGGG AATGACGAAA TTAGTCGGCA 2700

GCTCAACCGT CGCTAAGGAT TCTGATCGAG TTGAAAGTTA TGGAACGATT GaTGAATTGA 2760

ATTCTTGGGT GGGCTATATT ATCAGCCAAC TTCCACAAGA AAATCAAGAA ATTAAAGAAG 2820

AACTAGAAGC TTTACAACAT TTACTTTTTG ATGCGGGAAC AGATTTGTCA ACACCGATTG 2880

AAGCACAACG ACCATTTAAA TTACAAAAGG CAAGTGTGCA TTGGCTAGAA CAGCGGATTG 2940

ATTTCTACAC TGCACAATCA CCAGATATTG ATCGTTTTAT TTTACCTGGC GGAACCCCAG 3000

CAGCCAGTAT GGTTCATGTG GCTCGGACGA TTGCACGGCG CGCAGAACGC ATCATTGTTC 3060

GTTTAAATTG GACAGCCAAA ATTAACGAAG AGGTATTAAT TTTTACCAAT CGTTTATCAG 3120

ATTATTTTTA TGCGTTGGCT CGCTGTTTAA ATGTTCAAGC ACAACGACCT GATGTTTTTT 3180

ATGAACGAAG CGAAATGGTT TTCCATAAAA TTAAAGAAGA TGGATTGTAA ACAGTTGCTT 3240

TTCTAAAAAA CTTATGGTAA ATTTCAGATA GaTTAAaCGA ATATTTTATG AaAAATATGA 3300

ATGGAAGCCA GTGAGAATCT GGCACGGTCC CGCCACTGTG AAGAAGCAAG GTTGCTTTTA 3360

AGTCAGGTCT TTTCATTTTT TCATTATTGG GCATGCTGTT TCGAGGCAAA ACAGGATGTT 3420

TCTTAACAAC GCTTGTGTTG AAAATCCAGC CCAAGATATT TGTATTAATC CAATTAATGG 3480

CACGAGGAAG TGCTTTTGGG AATTTTGGGC TGGTCTTTCA ATGTTGAAGG AGATTTCAGG 3540

TACAATGACG TACTGCTATA TTCCCTTTTT TTATTGACAA TTAATTAAAG GCGTTGTACT 3600

CTATGGATAA GGTTTCGTGT ACAATGGCGT ATACATAAGG AAGCAAAGAC GCTTCAGACA 3660

GATGATTTTT CGTTTGTTTG AAGTGTCTTT TTGTATTTTA TTAAGGAGGA ATTGAGATGA 3720

AAACAATTCA TTTCCCTACC GAACTATGGG TAGGAGAAGG CGCATTAGCC AATTTGGAAA 3780

CACTTCATGA TAGACGTGTT TTCATTGTCA CAGATCCATT CATGGTTGAC TCTGGATTTG 3840

TGAATGAAGT GACAAAGCAT TTAACAAAAA GTGAATGGCA AATTTTTAGT GATATCATCC 3900

CCGATCCACC AATTGATAAG ATTGCTGCGG GAATTAAGCA TTTAGCGACA TTTCAAGGAG 3960

ATACGATATT AGCTTTAGGT GGGGGATCCG CTATTGATGC GGCCAAAGCA ATGAAATTTT 4020

TTGGCAAACG TACTTTGCAA ACTCAGATTG CTGAGTTTAT TGCGGTTCCT ACAACTAGTG 4080

GAACTGGTTC AGAAGTTACG AATTTTTCAG TGATTACGGT TGCTGAAACT GGCACCAAAA 4140

TCCCGCTTGT AACGGATGAA ATTCAACCAG AAATTGCGAT TTTAGATACA AATTTAGTGA 4200

TGAGTGTCCC ACCAAAAATC ACAGCGGATA CAGGAATGGA CGTCTTAACG CATGTTATTG 4260

AAGCGTATGT TTCAACAGAA GCAAATCCGA TTTCAGATGC GTTGTGCGAA AAAGTCGTTC 4320

GCTTAGTATT CGACAATTTA GAAATAGCCT TTAATGAGGG TAGCAACCAA CAAGCCCGTG 4380

AAAATATGCA TTTAGCTTCA TGTATGGCAG GGATGGCGTT TAATGTTACT tCGTTAGGCT 4440

TGAATCACGG TATTGCTCAT GCGGCCGGTG CACGTTTACA TGTGCCACAT GGACGAATGA 4500

ATGCAATGCT CTTACCTGAA GTCATTGCAT ATAACAGTGG CTTAGCGAAT GGCAAAGTCA 4560

CCAACGAACC AACAGCTAAA CGTTATGCGC AGTTAGCAAA TTGTCTAAAT GATACACAGA 4620

CAACCAATGC ACGGATTGGT GTTCAACAAT TTATTCGACA AATCAAGCAA CTCCGTCAAA 4680

AATTAAACAT GCCTGCCACT TTTAGTGACT ATGGTTTACC AAAAGAAGAA GTCCAGGCTG 4740

CAATCCCAAA AATAGCTGAA GGCGCATTGA TGGATGGCTG CACCAAAACA AATCCTGTCC 4800

AACCAACAGC AGCAGAAGTG ACAAAAATTT TAAATTCTAT CTTATAAGCG ATTGACAGGA 4860

TGAAATCGAT ArCATATAaT GTGGGTAGCA CAACGGCGTG CTTCmAAATT TAAGAGCAAA 4920

GAAGCTCCTT AGTAGAAAGC GAAAGTGGAA CCTTTACGTT TTTTATTGAG GATTTTTTAT 4980

TTGCTAAATT AAAGGAGCGA GTCTGGTTGG AAGAAAAACA ACGAATGATT CAAGAATATG 5040

TTCCTGGGAA ACAAGTAACG TTAGCGCACA TCATTGCTAG TCCTAATAAA GAAATTTATA 5100

CAAAATTAGG ATTGCCAGAA GGCACCAGCA ATGCACTAGG TATTTTGACA ATCACTCCGA 5160

GCGAagnCCA ATTATCGCTG TTGATATTGC GACTAAAAGC GGAGATATCC AAATCGGTTT 5220

TATTGATCGT TTTTCTGGAT CGGTAGTCAT TTCTGGTGAC GTAAGCTCAG TGGAAGCAGC 5280

CTTACAAGCG GTGATTGaAG GTTTACAACA GATTTTAAAC TTTTCTGTAA CTTGTAAAAT 5340

TACCAGAACA TAGCGTTGAA AGGACAAGGG CGTGAAGAAA ATGGATGGAC GAATTGTAAT 5400

AGTCGATGAT GAACCTATTA CGAGACTCGA TATTCGTGAT ATTGTGATAG AAGCAGGGTA 5460

CGAAGTAGTC GGTGAAGCAG CTGATGGTTT TGAAGCAATT GAAGTATGTA AAAAAACACA 5520

ACCTGACCTA GTATTAATGG ACATTCAAAT GCCAATTTTA GATGGTCTAA AAGCAGGTAA 5580

AAAAATTGTT CAAGACCAAC TAGCCAGTAG TATTGTCTTT TTATCTGCAT ACAGTGACGT 5640

ACAAAATACG GACAAGGCTA AAAAATTAGG CGCACTGGGT TATTTAGTTA AACCTCTAGA 5700

CGAAAAATCA TTAATACCTA CAATTGAAAT GAGCATTGAA CGAGGCAAAC AAACGCAGCT 5760

GTTACTAAGT CAAATCGATA AATTAAGTTT AAAATTAGAA GAACGTAAAA TTATCGAAAA 5820

AGCCAAAGGT ATTCTTGTAA AAGAAAATCA TATATCGGAA GAAGAAGCCT ACCAAATGTT 5880

GCGTACGTTA AGTATGAACA AACGCGCACG TATGAGTGAA ATTGCAGAaT TGATTGTAAT 5940

GGATGATGAA TAAATGAAAC GATTAGAGCA ATTATGTCAC CAATATACCA ATCTTTCAGA 6000

ATCAGACATT AAAGAATTAC AACGTACAGC ACGTTATCTT TCTTCGACGA CACTCTATCa 6060

AAGTGCAGAT GTTTTTATTG ATGTGTATAA GGAAATGTCA CAGCAAGCAC TAGTCGTTTA 6120

TCATAAGCCG CCTGCGAAGA CAACTTCTTT ATATAGTGGA GATGTCGTCG GGATGGAAgC 6180

TTTATTAAAA AATGAACCTG GTGTTCTGCG TACCATGCAA ACTAGCtTAA ATAGCATTGG 6240

TTtGTTAGCA GTGACTCAGG AAAATCGTTT GATTAAGCAA AATATCTATC CCATTCGCAA 6300

TGAACATCGA ACGATTGGTG TCATTATTGT CGAGATTGCG GCAGATGAAG AGATTCAAGC 6360

GGACTTACAA AAAGAGGAGC TAAATAACTG CCAGTTAGCA AAAGTAGCGA AATCAACTAG 6420

TCAAGTCGAC GCATTATTTA TTGATCAATT GGCAGAAGCC GTTTTGATTT TTGATGCGGC 6480

GGGTCACTTA TTAATTACCA ATCACAATGC ACAAGAGCTT TATCGCAAAC TTGGTTATCG 6540

GGATAATATC ATTGGAATGA GTTATGACAA TTTGTCCATC GATTATACAA CTTTTGAATA 6600

TGTGTTGTAT CAAATGAAAT ATAAAATGAG TAATCAACCA ATTGAAAGCA AAACGACCTA 6660

CCTTAATTAT TATTTTAAGG TTCGGAAAGT GTGGTTGGCA TCAGAAGAAC AACTGATTAT 6720

GATTATTCAA GACAATACAG AATTTAAAGA AAAAGAAGCG GAAATTATTT CCAAATCTGT 6780

TGcGATTCGT GAAATTCATC ATCGTGTTAA AAATAACTTG CAATCGGTGG kTTCCTTATT 6840

GCGTATTCAA GAACnGGCGA ACGCAAAGCC CCGAAGCAAA GAAAGTCCTT CATGAAAGTG 6900

TCAATCGAAT TATGGCGATT GCAGCCACCC ATGAATTGCT GTCAAAACAA GTCAAAGATG 6960

ACGTTGCTTT ACGGCAAACA TTGGAAGCAG TGATGTATAA TTTTAGACAT CTTTTTCAAG 7020

GCGCACAACC CATTGAGATG ATGATGGATG TTGATCCAGC CATTATGGTT TCTAGCGAAC 7080

AAATGGTTAC CATTTCCCTT GTTGTCAACG AACTATTACA AAACATTTTT GATCATGCAT 7140

TTGAGCCACA AACAAGTGGC GTTGTTAAAC TTAGTGGAAC ATTAGATAAT AAAATGATTA 7200

CCATTACTGT TACAGATAAT GGTAAAGGCT ATGATGTTCA TCAAAGTAAC GAGACCAGTT 7260

TAGGTTTGAT GATTGTCAAA AGCTACGTAA AAGATAAACT AAAAGGAAAA ATTACAATTG 7320

AATCCAATAA ACAGGGAACT AAAACCTGTT TCTATTTCGA ACAGAACACC AGTGATGTTG 7380

TTCATTGATT AAGCAAAGGC GCTTAAAGAA AAGTAGAAAG AAATTTTCTA ACGATTCTTT 7440

AAGTGCCTTT TTGTATGCAA AAGGAGGCGA GTCAGGCGAT GTCGAAGGAA ACACTTTTAA 7500

CCGTAGGAAT TGATTTAGGG ACGTCTACAA CCCAATTAGT CTTGTCAGAA TTAACTGTTG 7560

AAAATTTTGC CTCTGCGTTT ACAGTACCAC GTATCAGCAT TTCTGATAAA AAAGTCATTT 7620

ATCGAAGTGA CATTATCTTT ACACCCTTAC TTAACCAATC AGAAATTGAT GCAGAACCTA 7680

TCAAAGCGTT TGTAGCGGAG CAATATCGCC AAGCAGGCAT TCATAAACAA GATATTCAAA 7740

TGGGTGCGGT AATTATCACA GGTGAAACTG CTCGAAAAAG CAATGCCAAC AATGTGCTAC 7800

GAGCACTTAG TGGCTATGCA GGTGATTTTG TCGTTGCTAC GGCTGGTCCT GATTTAGAAA 7860

GTATCATTGC CGGGAAAGGC GCTGGTGCTC AAACTTATTC TGAAACGAAA CGCAAACCAG 7920

TGGTTAATTT AGACATTGGT GGTGGAACGA CAAATTTAGC TGTCTTTAAA GATGGCGAAg 7980

TGATTGACAC GGCTTGTTTT GACATTGGCG GACGCTTAAT TAAATTAGAT CAACAACAAA 8040

AAATCACCTA TATAGCTCCG AAGATTCAAG AAATTATAAA TAAAAAAGGG CTAACCTTGC 8100

ACCTTGGTGA CCAAGCCACC GAACAAAACT TACTGCCAAT TATTTCgGAA TTAGTTGCCG 8160

TGTTAGAAAA TAGCATTGGC TTGGGGACAC AAAGTCCGTT CTACCAGTTA TTAGTAACCA 8220

ACCATCCACT ACGAAAAGGT GAGGAATTAC CAATTGTAAC ATTTTCTGGT GGTGTGGCTG 8280

ATTGCTTAAA TACGACAAGT ACGAATCTCT TTAAATATGG GGACATCGGC TTACTTTTAG 8340

GGAAATACTT GAGAAAGTCA CTGATTTTTT CGGAAAAAGA AGTCTTAGAA AGCGCAGAAA 8400

CAATCCGAGC GACGGTCGTC GGAGCAGGTT CACACACTGC TGAAATTAGC GGAAGCACGA 8460

TCGCTTATCG TGAGCAAATT TTGCCGGTAA AAAATATCCC AATTCTAAAA TTAGCGCAAG 8520

AGGATGAAAC CCTTACTGTG ACAGAATTAG GCCAGCGAAT CCAAGAAAAA CTCAATTGGC 8580

ATCGTATTGA AGAAACACCC CAGATAGCCC TAGCAATTAG AGGCATGAGC AATCCAACAT 8640

TTGCTGATAT TCAACGCTAT GGTCAAGGGA TTGTGGAAGG ATTAGCTAGT TTAGTTGCTG 8700

AACAAATTCC CATTATTGTT ATGGTGGATG AAGACATGGC GAAAGCTTTA GGGCATGCCT 8760

TAAGTGCCCA TCTGCCAAAA GACTATCCGT TTATCTGTCT TGATTCGGTC AAAGTAGAAA 8820

ACGGTGATTA TGTAGATATC GGTTTACCAG TGGCAGAAGG AGCCGTCCTT CCAGTCATTG 8880

TTAAGACCTT AGTATTTAAT TAAAAAGGAG CGACACAAAG ATGATTTTAA AAACCAAACT 8940

ATTTGGCAAA GTCTATCAAT TTACATCCGT TAAAGAAGTC TTAGCCAAAG CCAATGAAGA 9000

AAAGTCTGGC GACAAGTTAG CTGGCGTAgC TGCTAATTCA GCCGAAGAAC GCGTGGCAGC 9060

AAAAGTCGTT TTATCGGAAT TGTCCTTAAA TGACTTGTTT AACAATCCCG TCGTTGrTTA 9120

TGACGAAGAC GAAGTGACAC GCATTATTAT TGACCAAGTC aATATGCGAA TTTTTGAAAG 9180

TATTAAACAT TGGACAGTTG CAGAATTACG CGAATTTATC TTATCTTCGG AAACaACCGA 9240

TTTTGATATC AAACGAATTT CCCgTGGACT CACTTCAGAA ATGATTGCTG CTGTCTGTAA 9300

ATTAATGTCC AACATGGATC TAATTGTTGG CGCAAAAAAA ATCAATATTG AAAAAACCGC 9360

CAACACAACA ATTGGCCGTC CCGGAACTTT TTCAAATCGT TTACAGCCAA ACCATCCAAC 9420

CGATAATGTC GATGGCATTA TGGCATCAGT TATGGAAGGT CTTTCTTATG GCGCCGGTGA 9480

TGCCTTAATT GGTTTGAATC CTGTTGATGA TTCTACAGAA AGTGTTAAAC GGATCTTAAA 9540

CAAATTTGAA GAGTTCCGCA GTGAATGGGA AATTCCAACA CAAACTTGCG TGCTTGCTCA 9600

CGTAACGACA CAAATGGAAG CCATGCGTCA AGGCGCGCCA ACTGGCTTAG TCTTTCAATC 9660

CATCGCTGGT TCAGAAAAAG GAAACACTGC ATTTGGTTTA AATGCAGAAA TTTTAGCTGA 9720

AGCACAAGAT TTAGCGCTTC ACAGTGGTCA AGCAGCTGGT CCTAATGTGA TGTATTTTGA 9780

AACAGGTCAA GGCTCAGAAC TTTCTTCTGA AGCGAACTTT GGCGCAGACC AAGTCACAAT 9840

GGAAGCGCGT TGTTATGGCT TAGCCAAAAA ATTTGATCCG TATATTGTAA ATACCGTTGT 9900

CGGCTTTATT GGACCAGAAT ATCTTTACGA TTCAAAACAA GTCATCCGCG CTGGTTTAGA 9960

AGATCATTTC ATGGGCAAAT TAACTGGAAT TyCAATGGGC TGTGATGTCT GCTACACCAA 10020

CCATATGAAA GCTGATCAAA ATGATATGGA AAACTTAGCA ATGTTATTAG CAACAGCTGG 10080

TTGTACGTAT ATCATGGGGA TTCCTCATGG CGATGATGTC ATGTTAAATT ATCAAACAAC 10140

TGGGTTCCAT GAAACTGCGA CGATTCGTGA AACATTAGGC TTACGTCCAA TTAAAGAATT 10200

TGAAGAATGG ATGGAAAAAA TGGGCTTAAT GGAAAACGGA AAATTAACAA GTCGCGCTGG 10260

CGATGCATCT GTATTTATTA AATAAACGAG GTGAGAATAC AAATGAACGA AAAAGAATTA 10320

AAAGAAATGA TTGCCGGAAT TTTAACAGAA ATGGTCGCTG ACAATCAAGC GGTTTCGACT 10380

GCCACTGTGA CTGCTGAAGA AAAACCAGTC ACAACACACG TCACTGAAAC GACAGAAATT 10440

GAAGAAGGCC TAATTCCTGA CATTACTGAA GTAGATTTAC GTAAACAATT ACTATTGAAA 10500

AATGCCGTAG ATCCTGAAGC ACTATTAAAA ATGAAAGCTT TCTCGCCAGC TCGTTTAGGC 10560

GTTGGTCGTG CAGGGACGCG TTATATGACT AGTTCAACAT TACGTTTCCG AGCAGACCAT 10620

GCGGCAGCAC AAGATGCAGT CTTTTCAGAC GTAAGCGAAG ACTTGGTAAA AGAAATGAAT 10680

TTTATTTCTA CAAAAACAAT CTGTAACAGT AAAGATGAAT ATTTAACACG TCCAGATTAT 10740

GGTCGTCAAT TTGATGAAGA AAATAGTGAA ATCATTCGGA AAAATACGAC ACCAAAAGCC 10800

AAAATCCAAA TGGTCGTAGG GGATGGCCTT AGTTCAGCTG CTATCGAAGC GAATATTAAA 10860

GAAGTCCTAC CaGCmATTAA ACmAGGATTA AACATGTACm ACTTaGATTT CGACmATGTA 10920

GkTTTTGtTA AAtATTGtCg CGTGcCAGCA ATGGaTaAAA TTGGTGAAAT CACTGACGCT 10980

```
GATGTTGTTT GCTTGCTAGT TGGCGAACGT CCTGGTTTAG TAACAGCTGA ATCTATGAGT    11040
GCCTACATTG CTTACAAACC AACTGTCGGT ATGCCAGAAG CACGGCGGAC AGTTATTTCA    11100
AATATTCACA AAGGTGGGAC ACCGGCTGTT GAAGCAGGGG CCTACATTGC CGAAATTATT    11160
AAAAAAATGT TAGACAAGAA AAAATCTGGT ATTGATTTAA AAGAAGCTGA ATAAAGGAGG    11220
ATTTCTCATG AAAAATGATC GTCTAGGCGC AAATGTCCTA AGTATGAAAG TCATTCCCAA    11280
TGTCGATCCG GCATTAGCAA AAGAATTATC ATTGAAACCA GAACATCGCT CATTAGGAAT    11340
TGTTACTTCC GATTGTGATG ATGTCACTTA TGTTGCATTA GATGAAGCAA CCAAAGCCGC    11400
CGATGTTACT GTAGTTTACG GTAAAAGTAT GTATGCTGGC GCAgCAAATG CTTCAACCAA    11460
ATTAGCCGGT GAAGTGATTG GCATTATTGC TGGTCCGAGT CCTGCCGAAG TCAATAGTGG    11520
CTTGAGTGTC ATTACGCAAG TCATTGAAGA AGAAGCCAGC TTCTATAGTG CTAACGAAGA    11580
TGACAGCATT GTTTATTTTG CACATGTAGT ATCAAGAACG GGTAGCTTTT TATCTGAACA    11640
AGCGAATATT CCTGAAGGTG AAGCGATTGC TTATCTGATt GCACCACCGC TAGAAGCAAT    11700
GGTTGGCTTA GATGCAGCAA TGAAAGCTGC CGATGTTCAA ATGGGTGTCT TTTATGGCCC    11760
ACCTTCTGAA ACGAACTTTG GCGGCGCCTT ATTGACAGGT AGTCAATCTG CATGTAAAGC    11820
TGCTTGCAGT GCGTTTGAAC AAGTCATTCA AAATATCGCT GACAATCCGT TGAGTTATTA    11880
AGGATGTGAC CTGAATGAAA GCATTAGGAA TGGTCGAAGT CAAAGGATTT TTAGGTGCCA    11940
TTAGTGTGGC CGATGCGGCA TTGAAAGCCG CCGACGTTAC TTTGTTAAAA GCTGAAATTA    12000
TTAATGGCGG TTTAACAACT GTGGAACTGA TTGGTGACGT TGCCGCGGTT CAAGCTGCCg    12060
TTGAAGTAGG TACGGAAGTA GCTAAAGAAT TAAATTGTTT AATTGCTCAT CATGTGATCT    12120
CACGGGTTGA TGCTCAAACG GAAGTGATTT TATCAGATCC AGAACCCAAG TCAGCACCTG    12180
AACCGATGGA ACAAGTAGAA GTAATTGAAG AAGAGATTGA AACACCAGAT TTGGAAGAAC    12240
ACACAGGCAC ACGCCAAAAA TTAGAAACAC AACGCGTGGT TGATTTACGC AAACAAGCCT    12300
ACAAAATGAA CCTAAGTAGT TTAAAGAAAA GTGAGATTAA ATTTGCCAAT AAAGAAGCTT    12360
TAATTCAAGC AATCATGGCT GAAATAGAAA GAAGTGAAGA CGAGTGGAAT TAGTAGATAA    12420
AGACTTACGT TCCATCCAAG AAACAAGAAA TTTAATCAGA AAAGCCAAAG AGGCACAGCA    12480
AGTATTAGCA ACATTTTCGC AAAAACAAAT CGATGCGATT GTTCAAGCGG TTTCAGAGGC    12540
TACATTTAAC CAACGAGAAA AGTTAGCGAA AATGGCCAAT GAAGAAACAG GATTTGGTAT    12600
CTATGAAGAT AAAATCATCA AAAATGCTTT CGCTTCAAAA GTAGTATATG ACGAAATGAA    12660
GGACAAAGCT ACAGTTGGTG TGATCCATGA TGATGCTGCC AAAAAAGTCA CTGAAATTGC    12720
TGTCCCAGTG GGTGTAATTG CAGGATTAAT TCCTTCTACC AACCCGACAT CAACAGTCAT    12780
TTACAAAGCA TTAATTTCAT TAAAAGCTGC CAACAGCATT GTTTTTTCAC CACATCCCAA    12840
TGCGTTGAAA TCGATTATTG AAACAGTGGA AATTATcCAA AAAGCGGCTA TTGCTGCAGG    12900
GGCACCAGAA GGTTGTGTCT CGGTCATCAA AACGCCAACC ATGCAAGCGA CCAGCGAATT    12960
```

AATGAAAAAT AAAGAAACAA ATTTAATTTT AGCTACTGGT GGAAACGCAA TGGTTAAAGC 13020

CGCGTACTCT TCTGGAACAC CAGCAATCGG TGTTGGGCCT GGTAATGGAC CAGCCTACAT 13080

CGAACGAAGC GCAAATGTTC CTCATGCAGT GAAACAAATA ATGGATTCAA AAACTTTCGA 13140

TAACGGTACA ATTTGTGCAT CGGAACAATC AATTATTGTT GAAACTGTCA ACCGTGAAGC 13200

GGTCaAAGAA GAGTTAATTA AGCAGGGAGC GTATTTCTTA AGTCCaGCCG AAGCCGACAA 13260

ATTAGCTAAA TTTATTTTAC GACCAAACGG TACAATGAAC CCGCAAATCG TGGGACGTTC 13320

CGTTCAACAT ATTGCTTCAT TAGTTGGTTT ATCTATTCCA AAAGATCGCC GCTTAATTGT 13380

TGCCGAAGAA ACAcATGTAG GCTTGAAATA TCCATTTTCA AGnGrAAAAT TGGCACCGaT 13440

TATTGCTTTC TATACAGTTG AAAACTGGGA AGcAsTTGCG CGTTATCAAT TGAAATTTTG 13500

AAAGGTGAAG GTGCGGGTCA TACAATGGGC ATTCACACAG AAAACAAAGA AGTCATTCGT 13560

GAATTTGGCT TAAGAAAACC CGTCTCTCGT TTGCTTGTTA ACACCTCTGG TACACTTGGC 13620

GGCATTGGTG CTTCAACTAA CTTAGTGCCA GCATTAACAC TTGGCTGTGG CGCAGTTGGT 13680

GGCAGCTCAA CTTCTGATAA TATTGGGGTT GAAAATCTGT TCAATTTACG TCGTGTTGCA 13740

TATGGCGTGC GTGATTTAGA AGAGATTCGT CAAGAATTTG GTCAAACATC AACCACATCT 13800

GTGGCTACTT CTTGCGAAAC AACCAATCAA GAAGAACTAG TAAATGCCGT GGTAGCTCAA 13860

GTTTTAGCTC GCTTAAACTA ACAAGAAACG AGCGAGTAAG AGTAAAACAA ATAAATTATT 13920

GGATTTAGGG TTTGGGACCT AAGTCAAATT GCTAATTGTC CCAAATTTTT AAAAATAAAA 13980

CATACATCTT AGGAGGAAAT TATAATGTCA AGTACAAATG CTTTAGGAAT GATCGAAACT 14040

CGTGGTTTAG TAGGTGCTGT TGAAGCAGCC GATGCAATGG TGAAAGCGGC AAACGTTACT 14100

TTAATTGGAA AAGAACAAGT CGGTGGCGGT TTAGTTACAG TAATGGTTCG CGGTGACGTT 14160

GGTGCTGTAA AAGCAGCAAC AGATGCTGGT GCTGCTGCAG CAGAACGTGT TGGCGAATTA 14220

TTATCTGTCC ATGTAATTCC ACGTCCACAT ACAGAAGTAG ACGCTATTTT ACCACACGCA 14280

ACAACTGAAT TATAATTTTG ATTTTTTGGG GCAGAGGTAG GTTCATTACT TCTGTCCTAA 14340

ACTTTCATTG AGGAGGGGTG TTATGACAGT TATTACAGAA GATATGCTTC GCATCAAAAG 14400

GCTAAAAAAA GAACTAGAAG ATGGCACGGA TTTTGTGTTT CCTGTAGGCA GTTTTCTGAC 14460

GCCAGCCGCC AAAAGCTATC TACGCGAGCA TCGGATTAAT AGCAGCTTCT CTTCAACAAA 14520

CCCATTGCTT TCTGAGGTTC ATTCAGTGAA GGCCTCCTCA CAAAAACAAG AACATTTATC 14580

GAAAGAAaCA ATTTATGAGA TTCGGCACTT AAGCCATCTA TTGTATTTGC CTTTCTTAAC 14640

GAATGAAGCG TTATCAACAG AAGGTTGGTT ATATTTCGAA CAACAACAAC AGTGGTTAGA 14700

GAAATTTATT TTAGAACAGT CGCTTGTAAT GGAACCAGTC ATTTTGGAAT CAAAGGTAAC 14760

TATTCTAACT AACCAACAGA GACAATGGCG CTATTCCTCA AAAGAAATTC AATTACAACT 14820

AGATAAAATT ATTCTCCAGT TAAAAGGAGA ATCACAACTG TTCGGACTTT TTCAAGCTTG 14880

GGCTACAGAA TTAATTAGCA CCATGGCTGA AATATCACGA AACTAGGAGA TGAACAAATT 14940

GGATTCAATG AATGAAATTG TTGAAGAAGT TGTTAAACGA ATTCAGCAAC AACAACAAAA 15000

TACTTTTGAA GTTGAAGCTT CTGGCCGACA TGTTCATTTG TCTCGTCaAG AGATTGATGC 15060

ATTATTTGGA CCAGGATATC AACTAACAAA AGTGAAAGAT TTATCACAAC CTGGACAGTT 15120

CGTTTGTAAA GAACGAATTA CGGTTGCTGG ACCTAAAGGA CTCTTTCAAA ATGTGGTTAT 15180

TTTAGGCCCT GAACGTAGCG AATCACAAGT CGAAGTTTCT ATGACGGATA CGCGTATTTT 15240

AGGAATTAAT GCCCCAGTTA GAGAGAGTGG CAAGACAGAA GGGACACCTG GCGTCACGCT 15300

GATGAACGGA TCAGCAGTTG TGACATTAAG TCACGGCTTA ATTGTTGCCA AACGTCATAT 15360

TCACATGACA CCCGAAGACG CCTTGAAAAA TAAAGTAAGC AATAGCCAAA TTGTGCAAGT 15420

CAAAGTGGAA GGCACACGTC CGTTAATCTT TGATGATGTG GTTGTGCGAA TCAGTCCGCG 15480

ATTTGCAACG TACATGCACA TTGACTACGA TGAAGCCAAC GCTTGCGGCT TAACCAAAGG 15540

AGCTCGTGGC TACATTCTTA AGTAAAAGGA GGCAAGACAA TGaACGTTTC TGrAATgGAA 15600

AATTTGATTC AACAGATTAC TGATCAAATT TGTCAACAAT TGCTGTCGTC GTCTTATTCA 15660

GTGGCTTTAG CAGAAGCAAA AGAAGCATTT CCTGAAGAAA TGTTTGTTCG TTTTCCTGAT 15720

GTAAATTGGG TAACCmAGGA GCAGAAGCTA GCACGAGGCT TGGTTGTAAA GAGGCTAACG 15780

ATTTCCCAAG TAAATGCGAT TGCGCATTTA CAGGAAACGG ATGAATTAGT CAAAAACATC 15840

TTAGCATTCT TATTTGAAGG TAAGCCgGTA TTGGTTTTAA CCCCAATACC TTCAGTGAGT 15900

AAAAATTCAC GATTGAAGTA TCGTTTGAAA CAAACGATTC AGGAAAACGT GGACATGTGC 15960

CAACAGTTTG GGCTGATTTT TTATCATGAT TCAGAAAATT ACGCTGTGTT TCAGGCGGCC 16020

TGTCAAAAAC AAtTCGTTCT TTGGCTGAGA CCAAACGAAC CTACATTACT GAAAAACAAT 16080

TGATAAGAAT GACAGAATCA GGTGTTTCTT TAAGTAAAAA CGCCTATTTA ACCCCGTTAG 16140

CGAAAGATTA TGCTAGGAAA CATCAATTGC TGACGTAAGG AGGAAAAACG ATGCTTTTAG 16200

GAAAAGTGAC TGGTAGCCTT TGGTCAACAA GAAAAGATGA AAAATTAAAT GGCTGGAAAT 16260

TTATGATGGT TGATATTTTG AATCAAGAGG ACGAAAAGCA AGGATTTTTA ATTGCTGCTG 16320

ACAACGCTGG AGCTGGTGTC GGGGACAAAG TGTTAATTAG TCAAGGCAAT GCTGCCCGTA 16380

TTTCGGCAGA AGATCCGAAT GTCCCAATTG ATGCAATGAT TGTTGGTGTC ATTGATTCTA 16440

CAGCGGAAGA ATAGGGAGGA CGAGACATGA GCATTAATGA AATTATTATG TATATTATGG 16500

TTCTTTTCAT GATTATTGGA GGAATCGATA AATGTATTGG CAACAAACTA GGTTTAGGGG 16560

AGCAATTTGA GGAAGGGATT ATGGCTATGG GTTCTCTAGC GCTTTCTATG GTAGGGATTA 16620

TCACTTTAGC GCCTGTCTTA GCCAATCTTT TAAGTCCGAT TGTCGTTCCT GTCTATGAAT 16680

TACTAGGGGC TGATCCAGCA ATGTTTGCCA CAACCTTATT GGCCAATGAT ATGGGCGGCT 16740

TTGCGTTAGC GCAACAATTA GCCAATGATC CGCAAGCTGG TTTATTTGCG GGGGCAATCT 16800

TAGGTGCAAT GATGGGACCA ACGTTAGTCT TTACAATTCC AGTAGCGCTT GGCATyATCC 16860

AAAAAGACGA CCAACAATTC TTAGCAACAG GtGTTTTgAG CGGGaTTATT ACaATTCCTT 16920

TTGGTTTATT AGCAGGTGGT TTAACAGCTG GCATGCCATT GTCATTGATT ATTCCTAACT 16980

wAATTCCAAT TATTATTGTT GCCGCTTTAA TTATTttAGG TTTATGGTTA GCGCCAAAGG 17040

GCATGATTAA AGGCTTCCAA ATTTTTGGTC AAGGCGTAGT GATTGTCGCT ATCTTTGGTy 17100

TAGTTGTGGG GGCGATTCAA CTATTATTAG GAATTACAGT GATTCCTGGC ATTGCACCAG 17160

TTACTGAAGG AATTGAAGTC GTTGGCGGGA TTGCTTTAAC TTTAGCTGGC GCTTTCTGTT 17220

TAGTAGCTGT GATTACGAAA GTCTTTAATA AACCATTGAT GAAACTAGGA AAAGTTTTAG 17280

GAATGAATGA AGTAGCGGCT GCAGGAATGG TTGCTACATT AGCTAACAAT ATTCCCATGk 17340

kCCAAATGTT AAAAGACATG GATAATCGAG GGAAAATTAT TAATATCGCT TTTGCTGtTt 17400

CGGCTTCTTT CGTTTTAGGT GACCACCTTG GTTTTACTGC AGGCGTACAA AAGAAATGAT 17460

TTTTCCAATG ATTGTTGGCA AATTAGTTGG AGGGATTACC GCCATTTTTG TGGCTATCTT 17520

TATGGCAAAT CGAATGCTAG GGAAACCCGA AACAGAAGTG AAATAATGAG GTGTAAATAT 17580

GACAGAATTA AATCGTGAAA TGATTGAAAC GCTTGTTCGA CAAATTGTCA CAGAAAAATT 17640

AATGCCGACT AAACAAGTGG ATCCAAGTGG AATTATGTCC ATCAAATTAC CCGAAATCGA 17700

CGTAACAGAA GACGACCGCT TAGATACAGG CAACCCAGCG GACATTGTTT ACACAAAAGA 17760

TCTGTTTTCT TTAGAAGAAA GTCCGCGCCT TGGTTGCGGT TTAATGGTGA TGAAAGATAC 17820

CACCTTCGAC TGGACACTAG AATACGATGA AATTGATTAT ATTATTTCAG GACAATTAGA 17880

TATACTGATT GACGGACGTA AAATTTCTGC AACAGCCGGC GAAATTATTT TAATTCCGAA 17940

AAGTAGCAAA ATTCAATTTT CAGTGACGGG TGATGCCCGC TTTGTCTACG TGACTTATCC 18000

AGCAGATTGG CAAAGCCAAT AATTATTTTA TCGGCAGTCA GAGAATAATT TTGCCAAAAC 18060

CCTTTACTCA GCGTTATAAT AAATAAGAAC GAAAAAAGGG GTTGAGTGAA TGACATTTGA 18120

GAATCCAAGT CAAAATCAAA TTTTTGATAT TTTAAAACAA GCAAAAAATA TTGCGGTGGT 18180

TGGTTTGAGT AACAAGCCCG ACCGCACAAG TTACAAAATT GCGGAAATTT TACAGCAACA 18240

TGGCTATCGT GTATTGCCTG TCAATCCAGT TTTAAAAGGT GAAACGGTGC TTGGCGAAAC 18300

GGTTTATGGG ACATTAACAG AGATTCCTGA AAAGATTGAT ATTGTTGACG TCTTTCGGCG 18360

GAGTGAATTT TTGCCAGAAG TGGCCGCAGA AACATTGAAA ACAGAAGCGC CTGTCTTTTG 18420

GGCCCAACTA GGTTTGGAAA ACGAAGAGGC AGCGAACTTG TTATTAAATG AAGGTCGTCA 18480

AGTTATTATG GACCGTTGTA TTAAAATTGA ATTGGCAAAA ATGGCAGATG CTGAATAAAC 18540

ACGTGTAATT TTATAAAGAA AGAGCGTCCT GAACGGAAAG CTCTTTCTTT TTTGACCCAG 18600

ATAGAGTATT ATTGTAATGA TGGTTTTAGA AGGAGCGAAC GGAAAACGCA AGAAACCGGC 18660

TCACTCTTCT TCTAATTCTT TAAGrraTAG TTAATGTACG CTTATTAACT AAACGAGAAT 18720

CTArGTATAA GGAGTTTTTT CTTTGGCTAA AAAAATTAAC AATGAGTACA ATGATGCCTC 18780

TATCCAAGTC CTAGAAGGAT TGGAAGCAGT GAGGAAAAGA CCTGGTATGT ATATCGGTTC 18840

AACCGATAGC CGTGGACTCC ATCACTTAGT CTATGAAATT GTCGATAACG CAGTCGATGA 18900

AGCGTTATCT GGCTACGGGA ATGAAATCAA CGTAACTATC CAAAAAGATA ATAGTATTTG 18960

TGTTGCAGAC TCTGGCCGCG GAATGCCAAC AGGGATGCAT GCTTCTGGTA TCCCAACAGT 19020

AGAAGTAATT TTTACTGTCT TACATGCGGG TGGTAAGTTT GGCCAAGGTG GCTATAAAAC 19080

TTCAGGCGGA CTCCATGGGG TAGGTGCCAG TGTTGTTAAT GCCTTATCTA AGTGGTTAGA 19140

AGTTCATATT GTTCGTGATG GCGTAGAATA CATGGAACGA TTTGAGGACG GCGGTAAACC 19200

AGTTGGTACC TTAAAAAAAA TCGGCAAAAC AAAAAAACGC AACGGTACCT CCGTTACTTT 19260

TTTACCTGAC GACACCATTT TCTCTACGAC AAATTTTTCT TACGAAATTT TAGCAGAACG 19320

TTTGAGAGAA TCAGCTTTCT TATTAAAAGG CGTCAAAATT ACTTTAACAG ATGAACGTGG 19380

GGAAGAACCC AAAGAAGAAG TCTTCCACTA CGAAGAAGGT ATTAAAGAAT TTGTTGCCTA 19440

TCTAAATGAA GAAAAAGATA CATTAACACC CGTTGTTTAT TTTTCAGGAG CCAAAGAAGG 19500

TATTGAAGTT GAATTAGCTT ACCAATACAA CGATGGCTAT TCTGAAAATG TTTTATCTTT 19560

CGTTAACAAT GTCCGCACAA AAGATGGCGG TACCCACGAA gTCGGTATGA AAACGTCAAT 19620

GACGAAGGCG TACAATGAAT ATGCACGGAA AGTGGGACTA TTAAAAGAAA AAGATAAAAA 19680

TTTAGAAGGT AGCGACTTCC GTGAAGGCTT AGCAGCCGTC CTTTCTATTC GAGTACCAGA 19740

AAACTTACTA CAGTTTGAGG GCCAAACCAA AGGAAAATTA GGTACACCTT TAGCAAGAAC 19800

CGTTGTCGAT AACGTCGTAG GCGAACAAAT GGGCTTCTAT TTACAAGAAA ACAGTGAAAT 19860

GAGTCAATCA TTAATTCGGA AAGCGATTAA AGCCCGTGAA GCCCGTGAAG CGGCTCGTAA 19920

AGCCCGTGAA GAAAGCCGGA ACGtAAAAAA CGGAAAAAAG GTGAATCGTT ACTTTCTGGA 19980

AAATTAACAC CGGCTCAATC CAGAAATCCT AAGAAAAATG AACTATATCT TGTCGAAGGA 20040

GACTCGGCCG GTGGCTCAGC AAAACAAGGG CGCGACCGGA AATTCCAAGC CATTTTACCG 20100

CTACGAGGAA AAGTCATTAA TACGGAAAAA GCGAAAATGC AaGACATCTT GAAAAACGAa 20160

GAAATCAATA CTATGATCTA TACGATTGGC GCAGGCGTTG GTCCTGAATT TTCAATTGAA 20220

GATTGTAATT ATGATAAAAT CATCATCATG ACCGATGCGG ATACCGATGG TGCCCATATT 20280

CAAGTGTTGC TATTAACGTT CTTCTATCGT TATATGAAAC CATTAATTGA AGCCGGCAAA 20340

GTCTACATTG CCTTACCACC ACTTTATAAA GTCTCTAAAG GAACTGGAAA AAAATCAGTC 20400

ATTGAGTATG CTTGGACCGA TGGTGAATTG GCAGAAGTGA TTGATAAAGT TGGTAAAGGC 20460

TATATGCTCC AACGCTATAA AGGTCTGGGG GAAATGAATG CAGAACAACT TTGGGAAACA 20520

ACGATGGATC CTGAAACTAG AACCTTAATT CGTGTTCGGA TTGATGATGC AGCTCAAGCA 20580

GAACGCCGTG TGACCACATT AATGGGCGaC AAAGTCgAAC CTCGTCGAAA ATGGATTGAA 20640

CAACATGTTC AATTTACGTT GGAAGAAGAT GGCAGTATTT TAGATCGTTC AGAAGAAGAC 20700

ACTTCTGCAC CAACAGGAGA AAGTCTTTTA GATGCGGAAA AAACAAAAGA AGCAGAACAA 20760

ACCGATGACA CTGAAATCAG TTTGTTTGAC ATAGAGTAGG GAGTGACAAT TTTGGAAAAA 20820

CGCCAAGAAG TTCAAGAATT AACCCTTGAA GAAGTAATGG GCGATCGTTT TGGAAGATAT 20880

TCCAAATACA TCATTCAAGA ACGGGCCCTA CCAGATATTC GAGATGGGCT AAAACCAGTA 20940

CAACGCCGCA TTCTCTTTTC AATGAATAAA GACGGCAATA CCTTTGATAA AGGCTTTCGT 21000

AAATCAGCAA AATCTGTCGG AAACATTATG GGGAATTATC ATCCCCATGG CGACAGCAGT 21060

ATTTATGAAG CGATGGTGCG TCTAAGTCAA GACTGGAAAT TACGGGAAGT ACTAATTGAA 21120

ATGCAC 21126

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3222 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

nAGAGAAAnC CGGTAGCTTT TTACGGTTTT CTCTTTTTTG ACTTTCAAAG TCAAATCAGG 60

TACGACTTCT TTTTAAACAG ATAAGaATGT GGTATCATTG TAAAGGACTA TATAATAAAG 120

GAGATGGACG TTTTTTGACA GATAATCAGT CAAAATCTTT GGAATTAATA CTTACTGATT 180

CTGATGATGC ACACATGCTT TTAGGAACAC ATGATAAACA TATCAAATTT TTAGAAGAGA 240

ATACTCACGT AACGATTAAT AGTCGTGGCG AAGTTATTCA GTTAATTGGC GAATCATCAG 300

AGGTAGAGTT AGTCGCTTCC GTGTTAAGAG CGCTACAAAC ATTAATTCAA CGAGGCATCA 360

AAGTGCATAC GCCAGATGTC GTTTCTGCTT TAAAAATGGC CAAAGCTGGA AATTTAGAGG 420

CGTTTATTGC CATGTATGAA GAAGAAATTA TGAAGGACCA TCATGGTCGG GCCATCCGAA 480

TTAAGAATGT TGGTCAAAAA AAATATATCG ATGCAGTTAA AACTCATGAC GTCATTTTCG 540

GGGTGGGACC TGCGGGGACA GGAAAAACCT TTTTAGCCGT GGTGATGGCA GTAGCAGCTT 600

TAAAAAAAGG CGAAGTACAA AAAATTATTT TAACTCGTCC AGCGGTGGAA GCTGGCGAAA 660

GCTTGGGCTT TTTACCAGGT GACTTAAAAG AAAAAGTCGA TCCTTATTTA CGTCCAGTTT 720

ACGATGCCTT GTATCAAATT TTTGGTATGG ATCATACAAA TCGTTTAATG GAACGCGGCG 780

TTATTGAAAT TGCCCCGCTT GCTTATATGC GTGGTCGTAC ATTGGATGAT GCTTTTGTCA 840

TTCTAGATGA AGCACAAAAT ACCACGGTTG CCCAAATGAA AATGTTTTTA ACTCGTTTAG 900

GGTTTAGTTC AAAAATGATT GTTAATGGAG ATACTAGTCA GATTGACTTG CCACGTGGTG 960

TAACAAGTGG TTTAGTTAAC GCCGAGCGTA CGTTAAAAGA CATTGAAAAA ATTGCCTTTG 1020

TAAATTTTGA AGCCAGTGAC GTTGTGCGTC ATCCTGTTGT AGCACAAATT ATTCAAGCCT 1080

ATGAAAAAGA ACAACAAAAG CATTCCTAAG CAGTTAGGGA AGGAGTGAAA GTACCTTTTG 1140

AAAAAAGTAC TTACAAAGTT AGGGGATCGT TTAGGTAAAG CCTATTTACC TATTTTATTA 1200

GCCGTTTTTT CTCTTCTTTT ATTTATGATT ATGTTTGGCA GCGTCCATCA GAAAAGAGTA 1260

GAAATTAAAG AAGGTCAACT GGCCGAGAAA ACCAttCGTG CAAATAAAAA TATCGAGAAT 1320

ACCTACGAAA CAGAACAAAG AAAAAAATTG GCAGCAGAAG CTGTCACACC CGAATATATT 1380

TATCAAGAAG ATACGGCATC TGTCCAACAT AATCGTATTG ATAAATTATT TAAATTGATT 1440

GATTCTGCGA ATGAAAAAGT AGACAAAGAG TACAGCAACA AGCAAGCGAA GGCCAAAAAG 1500

GAAGAAACGA TTCCTGCGCC GACCGTGGAA GAACGAGTAG CTAGCTTAAA ATCaTTGTTT 1560

GAATCTTTAC CGCAAGATGA AGTTACGTTT TATCaAAGTT TTCCTAATGT TTTTTATCaA 1620

ACAATCTTTA CATTAACCTC TGAGCAATTG GATAAGGTTC GTTCAGAAAG TTTGATGTTG 1680

GTGGATGACG CGATGCAAAA TCACGTTCGA GAATCTGATT TGGATAAAAT TCGCCAAGAA 1740

GCTAACGGTA AAATTCAATA CTTAGATATT ACGAGTACCA TGCAACAAGT CATTCGTTAC 1800

ATTGTTAACC AAGGGATTGT GGTAAACGAT ATTGCCAATG AAAAACGAAC AGAAGAgctA 1860

CGCCAAAAAG CGATGAATGC TGTTCAACCC GCAATGATTT ATCAAGGCGA AATTATTGTT 1920

CGTGAAGGTA CTCAAATTGA TGCAAAAGCG GTCGAAAAAC TTGAGCTACT AGGAATGACT 1980

AGTCAAAATA CCTCTATTTT TCCGATGGTG GCATTAGCAT TAGCTATTTT ACTTCAAGTT 2040

GAAGTGTTAA TTTTCTTTAC AAAACAAGTG ACTGAGCCTT CGCGACAACG CTCGTTTATT 2100

ATCTTTTACA CGGGTGCTAT GCTAATCAGT GTTATTTTAA TGAAATTCTT CCAGATTTTC 2160

CAAACAGAGC AGTTGATGTA TATTCCCTTG TTCTATCCGG CAGCCTTTGC GCCGCTGATT 2220

CTAAATCATT TTGTAAATCG TCGTTCAGGG ATAATCGCTG CTATTTTCCA AGTCGTTTTT 2280

GCGTTATTTA TCTTTTATAA TTCTATCGGC ACAAACTCGT TGACGGTTAT TCTGATTATG 2340

TACTTATTCT CAGGATTTCT AGCAACAGTT GTTAAACGGA AACGGATGAG TGAGCAAGTT 2400

TTCCCAGCTT TAATGTGGGT AGTGGTCTTT CCTGTTTTCA TGGCGGTTGT CTTAATGATT 2460

TATCAAGGGA TGAGTTTAAC AGATGGTAAA ACGTGGACAG CTTTAATTTG TGCAAGTGCA 2520

GGAACGGTAC TTTCATTTTT AGCAACAATG GGCTTGCATC CATATATCGA ATTATTAGTG 2580

ACCGATGATA GTATGATTGT CTTGAATGAA TTAAGTAATC CGAACCATCC GTTGTTAAAA 2640

CAACTGTTAG AAGAAGCGCC AGGTACCTAT CATCATAGCA TGATGGTGGC TAGTTTGAgC 2700

GCTAACGCTG TTGCTGAGAT TGGCGGGCGG TCGCTATTAA CACGAGTTGC TTGTTATTAC 2760

CATGACATTG GTAAAATTAA ACACGCCAAT TTCTTTGTGG AAAACTTACC TGCTGGTGCA 2820

GAAAATCCAC ATAACTTTTT ATTACCAGAA GATAGTAAAC AAATTATTTT TGGCCACGTA 2880

ATTGATGGCG CCAAAATTTT AGAAGAGTAC AATATGCCGC AAATGGTGAT TGACATTTGT 2940

CGTCAACATC ATGGCACAAC ATTGATGAAA TTTTTCTATG TGAAAGCAAA AGAACGCAAT 3000

CCTGaAATTA AAGAATCCGA CTTCCGTTAT CCTGGaCCAC GTCCACAAAC aCGAGaAGCa 3060

GGGATTGTGa GTATTGcTGA TACTTGTGAG GCTGCTGTGC GGGCGATGGA TCATCCCACC 3120

AACGAAAAAA TTnCAAGCCT TTGTGCATAA TGTGATTCAA GACAGnATTT CAGATGGCCA 3180

ATTAGATGAA TGTGGGTTAA CGATGAAAGA GATTCGCATC AT 3222

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7812 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GGACATGATT ACGTTAGACT TAAACTTACC TAGTATGGAT GGGATGGCTG TTTGTCGTGA 60

AATTCGCAAA GTTTCAGCCA ATGTTCCCAT TATTATGTTG ACCGCTCGTG aTTCTGAAAG 120

TGACCAAGTA ATTGGCCTTG AAATGGGTGC GGATGATTAT GTGACAAAAC CATTTAGTCC 180

ATTAACTTTG ATTGCACGGA TGAAAGCTCT ACATCGTCGT GCCGAGGTGG CAGAGGCTGC 240

GCATGATACG TCTGAGAATA CAGATGAAAC ATTTGACGTG ATTACGGATC ATTTTAAAAT 300

GAATACGAAA ACACGAGAAA CATATTTAGA TAATCAATTA ATTGAAGGCT TAACACCTAA 360

AGAATTTGAT TTGCTATACA CTTTAGcAAA AAAACCACGC CAAGTCTTTT CACGcGAACA 420

ACTACTAGAA TTGGTTTGGG ACTATCAATA CTTTGGCGAT GAACGGACAG TTGATGCCCA 480

TATCAAAAAA TTACGACAAA AAATTGAAAA AGTGGGTCCT CAAGTCATCC AAACGGTTTG 540

GGGCGTAGGG TATAAATTCG ACGATTCAGG TGTTGCTTAA TGAAGTATTT GTATCAACAA 600

TTACTTGCTT TTATAGGTGT GATCGCTCTG ATTATTCTCA TTGTCGGAAC GTCTTTCACA 660

CAATTGACCA AGCGCACAAT GCAAGAAAAT AACTATGAAC AGCTGTATGG TTATGCAGAA 720

TCGGCTTTAG AAACACGTGA CTTTTTTATT AATGTCGCAG GCGTTTCAGA TCGAGATGTT 780

TTATCCTATT CATTtCAGTT GACTGAGCGC GTACTCCAGA AGCAAGATGT GCAATTTGTG 840

TTTATTAACA AGGACCGCGA AGTGCAGTAC CCACCAGTTG ATAGCACAAA AAAATTAGAC 900

TTTTCCTTAA TCGATAAAAA TTGGGATCAA ATCATGAAAG GCAATCGTGT TaTGCAACAG 960

AAAACATTGA TATTTACGGG GCCCGAAATA CTTCGTCCTA TGTAATGTTA CCGGTGTATG 1020

CATCAAATCA GTCTTCTGAT AAAAAAGTGA TTATTGGTTC ACTAGTTATT ACACAACCTG 1080

CCAAAAACGT TGACCGAAGT GTTCAGTCCG TGACACAAAA CTTGATTAAA GGCTTTATTT 1140

TTTCTGGTGT GATTGCCTTG TTACTAAGCT ATCTATTTGC GACTTTCCAA GTGAAACGAA 1200

TAAATCGGAT GCGTAAAGCC ACCAAAGAAA TTACTAGTGG AAATTTTGAT ATTCAATTAC 1260

CGGTTCATGA CAAAGATGAA TTTGATGACC TAGCAGAAGA TTTTAATAAA ATGGCCGCGT 1320

CTTTAAAAGA ATCACAAGAA GAAATCAATC GGCAAGAAGA GCGCCGTCGT CAATTTATGG 1380

CGGATGCCTC TCATGAAATG CGGACGCCTT TAACAACTAT TAATGGCTTG TTAGAGGGAT 1440

TGCAGTATAA TGCTATTCCT GAAAATCAAA AAGAGAATGC GATTAAACTG ATGCAAAATG 1500

AGACCGCTCG CTTGATTCGC TTAGTCAATG AAAATCTTGA TTATGAAAAA ATTCGGACCA 1560

ACCAGATTCA AATTGTTGTG AAGAAGTTCA ACGGAACAGA AGCTTTAGAA AATATTGTGA 1620

CTCAACTAAC GGCCAAAGCT GAAGCCGCCG GTAATCAATT GTATTTAGAC ACCACCGAAC 1680

CAATTGATGT GTATGCCGAT TATGATCGTT TCGTTCAAGT GGTTGTCAAT ATTGTTCAAA 1740

ATGCGATCCA ATTTACTGAA AATGGCGAAA TTCATATTGC TTTAGAAAAA GGCTATTTAG 1800

AAACCATTGT CCGTATTTCT GATACAGGAA TCGGCATGAC GGAAGAGCAA ATTCTGAATA 1860

TCTGGGATCG TTATTACAAA GTAGATCCAT CCCGTAAAAA TACGAAATAT GGAGAATCAG 1920

GATTAGGCTT ACCAATCGTT CAACAATTGG TCCGTTTGCA TAAAGGAAAA ATTAATGTTG 1980

AGAGTGAACT TGGTAAAGGC ACCACGTTTA TTATTTCATT TCCTGATGTT GAAATTACTG 2040

AAAACTAAAG AAAACGTCTG GTCACAATCA GTGCCAGACG TTTTCTTTAG TTTAATAACG 2100

CATGTAGAGA CAGGGGGTGC ACATAGGGAC TATCAATAAT TTGTGGATCT TTTGTCAGAA 2160

ATAGCAAGCG CTTGTTTCTT TTTTGAACTG CTATTTTTAA TTGAGGTAAA ATTTCTTGAC 2220

GTTGTCCAAT CGTTAATGAA GAAAAGATGT CATCAATCAC TAACGTACTT GTTTTTGAAA 2280

GCAAAAGTTG TATTATGCGC ATTTTAATTT TTTCAAAAGT CGTAAAAGAA GTAAAGGAAC 2340

GACCAAGTAA GGCAGGTTCT AATTGGAACA CCTGGAAAGC TTCGGCTAGA ACTACTTTTC 2400

GATCCCGTTC TTTGATCGTA CTAGCGATAA ATAAATTTTC TTCTAATGGC AAGTAGGGCA 2460

AAAAGGTTTC ACTAGAAGAA AGCCTAGTAA TGGTTAATTG GTTTGTTTGG AAAAAATCGG 2520

CCAATGCGAG TTCTTGTTCT GGAGCAGAAT AAATGATTCC ATAAACTTCT CGCAAAGctA 2580

ACGCGTCGAA TGAGGATTTT TTTTCGGTTA TTTTTTTGAA TGATGGTGTG ACATTCATAG 2640

AAAATCCCTC CCTTATTTCT TATGTGTACG TAAATAAATC AACGTGAAAG GAGTCCCGCA 2700

AACAAGAGGA ATGCCAAGCA AGATTAAGCA GTTAACAAAG CAATTACGGA TTGTATAAAG 2760

TGACTGCATA CTTTTAAAAT CAAAAGAAAA CGCTTGGATC GTTTGAATCC CTAAATGCTG 2820

GATATTTGTA ATCGTGACAT CCTGGTCGGT ACTTTTAGCT AAAAATGCTG GAGTTTCCAA 2880

ATGAGACCAC TGTAAAATAG CATCACGAGC AGCCCAAAGA ACAGTATTAT ACGTTTCTTG 2940

AAACATTAAA GTGAAAAGTA AAAAACTGAA AATTCCTAAT AGAGCCGGTA AACATAACTC 3000

TAGTAAGAGC TGTTTGGTGA CATACATATT TGAAAAGTGC ATGTGATACC ATGTATGTAT 3060

TTCTTCCTTT TTTATGCGTA ATGTCACATA CATCAACAGA AGAAAAGTAA CACCAAAGAA 3120

AATGACACTA CCTAGATAGA GGTGCCAATA GTATTGGATC AGTTGGTAAT GGTAACTAGT 3180

TTGTTTTTCA ATATCGACAA AGTCTGGCAT CGTTCGTAAA AACTCTTTTT CAAAATCAAT 3240

TAAGTTAAAC AGCCCAGTCA ATAAGAAAGA GAACAGGAAG AAGAAGAGTG CGGTGATTAA 3300

CGTGATTTTT TTATGATAAA AACTACTAGC TGTAGCGTAA CGGACACTTC TCATCAACTT 3360

TCACCTCTCT AAACTATCTT TAGTATAGTG aAATAATATG AATTAwTTAk GAaAAAGtAG 3420

CAAAAaGCTT GAGCAAATkG CTCAAGCTTT TTGAAGATTt ATTtACCCGT TTTTtGATAG 3480

CCTtGATAAC TGTCAGtGTG ATATtCATCA ACGGTtGATT TATtATkGTT TTTACTGTAA 3540

ACACTCGTTG ATTTTnCACC nTTTcTTTTT CAATTGCTTC CAGTTGTTGC AATTGGTTCT 3600

TGTAATCATA GTCTTCTGGG TTAACTGGTT TTAAGCCGCT ATTGGTATAG AAACGAAGCA 3660

AGTCACCGTT AGTAATTTGG TCAGACGTTT CTAATTGTTT CGTCGCTTTG GCTTTTAAAT 3720

CAGCGACTTC TTTCTTAACT TCTTCGGTTG GCTCTGTAAT TAAAGTGCCT GTTTTAGTAT 3780

CGTAAATGCT AGAACCTAAA ATAGTATATT TAGGTGTTAC AACATTGCCG TTGCGGAAAG 3840

CAACGATTTG ATTGTGTTGT TTAGAGAATA AATCTTGGCC TAACTGAATG TAGTTTTTCG 3900

TATCGACACC AAGCAAGTGT AGAAGGGTTG GTAAGGCATC AACTTGGCCG CCATACGTAT 3960

GGTTCACACC GCCATTTTCT TGACCAGGAA CGTGAATCAT ATAaGGAACA CGTTGCATAT 4020

TGGCGTTGTC AAAATCGTTC CAATCTGCTT TGGTTTTACC TACTAACTCT GCTAAGTTTT 4080

GGTTTCTTGA ATTAGAAACC CCATAGTGAT CGCCATAAAG AACAATGACT GAGTTTTCAT 4140

AGAGACCTGA AGATTTCAAG TAGTTAAAGA ATTCTTCGAC CGCTTTATCT AAATAATTCG 4200

CTGTTGCAAA GTAACCATTA ATTGTTTCGT CAGAGGTTTT TGCAATTGGG AATCCTGCTT 4260

CATCGTTTGT AAATTGAGAA TAAGGATAAT GATTGGACAC AGCAATAAAT TTAGAATAGA 4320

ACGGTTGTTG TAAATGTTCT AAATATTGAA CGGATTGATT AAAGAATGGT TTATCGTGTA 4380

ACCCATATTG GAATGAGTTA TCTGAATTAA CATCATAATA ACTAGCATCA AAGAAATAGT 4440

CATAGCCTAA ACGTTTATAG GTTTCATTTC GATTCCAGAA GTTACCGGCA TTTCCGTGGA 4500

AAGCAGCGCT AGTGTAACCT TGCGTTTGTT TTAAAATATC GGGTGCTGCT TCAAACGTAT 4560

TTTTGCCACC AACTTGTGTA AAGAGTGAAC CTTGGTCTAA ACCAAATAGT GAATTTTCAA 4620

GTAACGTTTC GGCATCACTT GTTTTTCCTT GACCAACTTG GTGGAAGAAG TTATCAAAGC 4680

TAAAGGTACT TTTACTATTG TAAAGACTAT TGATAAATGG TGTGACTTCG TGTTCTACGC 4740

CATTTTCATC TTTTAATTTG TAATTAACTA AAAATTGTTG GAAACTTTCT AAATGGATAT 4800

AAATTACATT TTTGCCTTTT GCAATACCAA ACTTACTATC ATCGGGTGCC GCATATTGTT 4860

GTTTCACGTA ATCTTCAACT TGTTTCATAT CGTTAGCACT GGCTTCGGCA CGGACTTGGT 4920

TTGTTTGGTA AGTCGTAATG CCATCGTACA CGGTAAAGGC ATTTAAGCCT AAGAATTTCA 4980

CGATATAGTC ACGAGAGAAT TGACGTCCTA ATAATTGGGG ACGTTCTGTT TCAGCCATAA 5040

ATAGATTGAA TAAGAAAAAG ACAACAGAAA GGAGTGTGAC AGAGACTGCT ACACGTGCGC 5100

GGACTGGACG CGCATCTGTT TTAATTTTTT TCGTTAATAA TAATACGCCA ATAATAATGA 5160

AGTCTAAGAA ATAAATGACG TCATAAGGTC TAAAGAGACG TAAGGCACTT TCGCCAAGAC 5220

CGCTAGCTAC TTTGCCAGCA CCAAGCATCG TATTAACTGT GATGAAGTCA GTAAATTCTC 5280

GATAATACAC GACATTTGAG AAGAGCAGTA AACTCATCAA GAAGTAAATA ATCATCATTG 5340

TAATATAAGA AGCTTTTTTA CGGCGAACAT ACAAAGCGAT GGCTAATAAA AAGACCGTTG 5400

TGGCAATTGG ATTAATTAAG AGTATGAAGT ACTCGTTAAT ATTTTCCAAA CGTAAATGGA 5460

AATCAACGGC GTACGCAAAG AGGTTTTTTA GCCACAGTAA GATGGCGATT AGCGAGAATA 5520

ACCCCAGTCG CGTGTTCAAA ATATTTGTTT TAAATAATTT TTTCAAAATA AACGTCCTTT 5580

CTAAACGATG ATTCTATAAA AGTAAAATTA GGTAAATTTC GTATTTTACA ATATCGCTTT 5640

AATTTTACTC TTTCACAAAT TTTATGTCAA TTCTTACCGA AGTGAGGAAT GGTGATGTTA 5700

AAGAAAACTT ACTATTTCCT TAAACTTTTT AGTTCGTTAA ATGACGTGAA GCTTGAATCC 5760

TTTAAAGACA GATGGGCCTT TTTTTGCTAT ACTAGATAGG ACTTTGAAAA GTAGGTGTAC 5820

AGAATGAAAC TTCAAGTAAC TAAAAAAGCA GAACATAAAT TTAAAAAAGG ATATCCTTTA 5880

ATTCAGAAAG AAGATTTGCA ACAAGTGCCA GCGCCGTTGC CAACAGATTG GCTTACTTTG 5940

ATTGATAGTA AAGGCCAGCG TTTAGCCGAA GGATATTTAG GCGAACAAAA TAAAGGAATT 6000

GGTTGGTTGC TGTCATGGCA TGGACCAATT AATCAATCTT TTTTTCAACA ACTCTTTGAA 6060

ATAAGTCGTG AGAAGCGGAC GTCTTTTGAA AAGGATTCGC TAACAACCGC GTATCGATTA 6120

TTTAATGGAG AAGGGGATGG CATTGGGGGG CTAATCATCG ATCGCTATGC AGATTACGCC 6180

GTTTTTTCAT GGTATAATGA AACCCTTTAT CAGAAAAAAG CCGAACTTTT GACGGCGTTT 6240

CGCACAGTTT ACCCTGACAT TATTGGTGCC TACGAAAAGA TTCGCTTTTC CACAAAAGAT 6300

TTACCAGAAT CACAGTTTTT GTATGGCGAA CAAGCACCCG AGCCATTACT TGTCACTGAA 6360

AATGGCGTTC AATTTGCTAC CTATTTAAAC GAAGGATTAA TGACAGGCAT CTTTTTAGAT 6420

CAAAAAGAAG TTCGTGGTCG ATTAGTAGAT GGTTTTGCAG TAGGTAAAAC AGTCTTAAAT 6480

ATGTTTAGTT ATACAGGTGC TTTTTCAGTA GCTGCCGCTA TGGGGGGTGC CGTGGCCACC 6540

ACTAGTGTGG ATTTAGCTAA ACGAAGCTTA CCGAAAACGA CTGAGCAATT TGAAGTCAAT 6600

CATTTAAATC TTGCCCCGCA AAAGrTTATT GTTATGGATG TTTTgACTAC TTTAAATATc 6660

TTCGCGTAAA GGCTTGmGTT ATGACATGAT CATTTTGGAT CCGCCGAnTT TTGCTCGCAA 6720

CAAAAAGAAA GTTTTTTCCG TTGCTAAAAA TTATGGGGAG CTAGTGAAAG ATTCTATTGA 6780

TATTTTAACG GATAAGGGAA CGCTTATTGC CTCAACCAaT GCGGCAAATT TATCCTTAGC 6840

CAAATATCAA AAAATGGTGA TTACTGCCTT GCAAGAGAAA AATGTCCGTT ACAAGATTAC 6900

GGACACGTAT CAATTACCGG CAGACTTTCA AGTGAATCCT AATTTTCCAG AAGGTAATTA 6960

CTTAAAAGTC CTGTTTATCG AAATTGAAAA ATAATTTTTT ACACCAAGCA AAAGTCCCAG 7020

CGCCGCTAAT CGGTATTGGG ACTTTTGCTT GGTnTTTATT TTATAGTAAC TTTTGGGCTT 7080

TTTGAGTGTT TGCAAAATGc AGTTCGCGTA GTTTGCTAAT AAGTCTAGGC CGCCTTTTTT 7140

CAAAATACGG GCGGCTACTT GGTCTGATTT GCTGCCAGCA CCAGAGGGCA CAGAGAAACC 7200

CAATTCTGCA GATTCTTTGT CTAACTCTTC TAAAAAAGCA GCGCGGCTAA TAATAGAGGC 7260

GGCAGCAACG GCAACATGAT ACTGCTCTCC TTTAGTGACG AAGAACAATT TTTCAGTGAC 7320

TTGATTTTTT TCATTACGAA CATATTTCCG ATAATTATTT TCAGGGGTGA ACTGATCAAT 7380

TAAAATCCCT TCTGGCTTTG TGGGAGCCAA TTCTTGTAGC AACAGATAAA TCGCTTGGTT 7440

ATGCAAAGCG ACTTTCATGT GTACAGCATT GTAGTTTGGT TGGATCTCAT TGTATTTTTT 7500

GGGTTCAACG ATTAATAACT TGTAAGGAAT CAGTTCTTTA ATGACATGTG ACAATTGAAT 7560

AATTTGTGGG TCGGTTAATT CTTTGGAATC ACGAACACCT AATGATTTTA GTTTGCTAAT 7620

CATGnTTTTG TCAACATAGG CAGCACAGAC AGTTACTGGA CCAAAATaGC TACCATTaCC 7680

GACTTCATCG GAACCtATCA CAGgACCcAT TTGCCCAAAA TCCAGCAGGG CAGTTGGTAG 7740

TTAGTCTGGG GGAAGAAACT TTTTnTCTTC TTGGGTGTTG GTTACCTTCG GTTTCCCTTT 7800

CCCCCAAAAC CG 7812

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1769 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

TTCAGCGACA GGGATGTTAT TGATGATTAA ACGTTATCAA TCGCAATTCG TGTTTTGGCT 60

ACTAGGAAAT ATTTTTTCAA TTTTGTTATG GTTCCGAGCA GGGACGCACG CTGGTGGTGA 120

TTATGCGATA TTTGTTATGT ATTGTATGTA TACTTTTAAC TCAATTTTTG GTATGTTTAA 180

TTGGTTGAAG ATAAAAAATA AAATGGAAAA AAATAGGTAG GGAGTTCGAT GGCAGTGaTT 240

GTTTTAGCAG GTACGATTGG CGCTGGTAAA TCTAGTTTAA CAGCTTTAAT TGCAAACCGT 300

TTAGGTTCAG AAGCCTTTTA CGAATCAGTT GATGATAATG AGGTATTACC TTTATTTTAT 360

GCGGAACCGG AAAAGTACGC ATTTTTATTA CAAATTTATT TTTTAAATAA ACGTTTTGAT 420

AGTATTAAAC AAGCATTAaC GCATGAAAAT AATGTATTGG ATCGCTCGAT TTATGAAGAT 480

TCATTACTTT TTCATTTAAA TGCAGATTTA GGTCGTGCCA ACGAAACAGA GGTCAAAGTG 540

TACGATGACT TATTACAAAA CATGCTCCAA GAATTGCCTT ATGCTGCGCA TAAAAAAAGA 600

CCTGATTTGT TAGTTCACAT TCGTATTTCA TTTCCCAAAA TGTTGGAACG CATCAAAAAA 660

CGTGGCCGTC CTTATGAACA AATTGAGACA GATCCAACGT TGTATGATTA CTATCAAATG 720

TTGAACGAGC GTTATGATCA ATGGTATGAA GACTATGATG AAAGTCCTAA AATTCAAATT 780

GATGGCGATA AATATGATTT TGTCGAAGAT CCTGAAGCGT GTCAATATGT CTTGGCGTTA 840

ATTGAGAAGA AAATTGAAGA ATTAGAAAGA TAAATAAAAA GAAGCAAGTC CGGACTTGCT 900

TCTTTTTGAC TTAAAATGTG GATAGATAGG CTAAAGTTGC TACTAAATAA AGGAGTGTAG 960

TGCTAAATTA TAGAATCGGA TTTTAACCTT CCCCTAAGTT AAAATCGGAT AATCTATCTA 1020

TCGACAGTTT AAGTTTGTAT AAAATAGCGG TAGCTATTAT CAACGAAGTC ATTCATCGGT 1080

TGACCTGACT TCTCTAGAAG TATACCATCG AAAATAACCG ACAGTTGTTC AAAATTTGGA 1140

CTTAGTAAGT GTAAATTTTG ATTGAAAGTT TCTGTAAATC AATCATTTAT TCAAAAAATA 1200

ATGCTCGAAA GAAGGAAGAT AAATGAGCCA AATTTGGTTT GAACGCATTC TACGTTGGCG 1260

CAAAAAAGCG CAACGATTTT TACAAGGACG TTATGCCCGG TTCGATGAAT TGAATCGAAC 1320

GCTGTTGCTA ACAAGTGTCG TGcTGGCGCT CATCAATATT TTTACTGGCT ATCTTTGGGT 1380

TCGGTTATTG ATTTTAATCA CAGTCGCCTA CGTGTATTAT CGCTTTTTTT CAAAACATAT 1440

TCACCCGCGA TTGAATGAAA ATCAATGGTT TATTCTTCGA AAACAACGCC TAATGCAAAA 1500

GGTTCATGCT TTTCGCAAGC GACATATGAC TCAGAAAGAG TATCGTTATT TTAAATGTCC 1560

AAAATGTCAG CAGTCATTAC GTGCGCCGAA AGGACGAGGA ACAATCAAAG TCACTTGTTC 1620

AAACTGTCAA AATCAATTTA TCAAAAAAGT CTGATCTTCT GTGTTTTTTG GTCTAGGCAA 1680

CTTTCGGACT TGTGGTATAC TTGAGAAAAG AAACCATTGG ATATTCTAAA AGGGGAATAT 1740

CTTGkTAAAA AAGGAGACAA AAAACAATG 1769

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1908 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CnGTCGTTGA CGCATCAACG CCTCnCTCGA GTATATGnGT TCACCTTCAC CGCGTTTAAT 60

AGcAAAcGcT GTTaCCGTTA AATTTTCATT GGCATTAGCT AATCGATTTG TTACTTCTGC 120

AATCGTTGCT TCGACATTTT CCTTTTCTTT GGCAACAATA AGACTGACAG AAGACTTATA 180

GGTAATCACC GCAAAAATGG TAAAAACTAC AAATATAAAG AAAGAACTTG CGAAAGCCCA 240

CTTTATAGTT AAAGATGGGC CTTCCAGTTC TTTTTTAATC GTTCTTTTCA TTAATTAGTT 300

CTCACGAACG CATAACGTAG CCAGTTCCAC GGACAGTTTG GATGTAGCTT TCTTCTCCAG 360

GTACGTCAAT TTTATTTAAT AAGTAGCGGA TATACACATC CACAACGTTT GTTTCTACTT 420

CTGTTTCATA GCCCCAAACT TTATTTaGTA ACACATCACG TGCCAAGACA ACGTTCACGT 480

TTTCCATTAG CGTTAATAGT AATcGTAtTC GCGkTTkGTt AATTCAATCA TTTCAGAATT 540

ACGACGAACG ACACGATTTT CTTTTTCAAT TGTTAAGTCA CGATATGTAA TCGTTGTTTG 600

TTTTGCAACG TTTTTATCGC CCTCAATATC AATACGACGA AgTAACGCAC GTAAACGAGC 660

TAACAATTCT TCAATTGCAA ATGGTTTAAC AATATAATCA TCCGCTCCAT GGTCTAAGCC 720

AGAAACACGG TCAATTACTG AATCACGCGC AGTCATCATA ATAATTGGTG TATTTTTCAC 780

TTGGCGAACA CGGCGACATA CTTCTAATCC ATTTAATTCT GGTAACATCA AATCAAGAAG 840

GATAGCATCC CATTCGTTGT TAAGAGCGGC TTCCAATCCT GTACGACCAT TGTAGTGTAC 900

TTCTGTCGTA TACCCCTCAT GTTTTAATTC AAGCTCAACG AATCTCGCTA AGTTCTTTTC 960

ATCTTCAATA ATTAAAATGT TGCTCATTTG ATTAATTGTC CTTTCTCTTT AAAAGCTATA 1020

CTCAAGCTCA ACCTTTTTAT TCCTTACTAC TTATTAGAGA ATTACTACTC AATCTATCTG 1080

GTTCTTTTTT AAGCCACAGT AAAACCATGA GCATGGTGAG CCACACTCAT GGTCTCCAGC 1140

GTGTTTATTC TTCGTTGTAC CAGCTGTAAT GGTAGATTCC TTCTTTGTCT GTGCGTTCGT 1200

ACGTATGAGC ACCGAAGTAA TCACGTTGTG CTTGAATTAA ATTCGCTGGT AAGCGATCTG 1260

AACGATAAGA ATCGTAATAA GCAATTGCAG ATGAGAATGT TGGTACAGGA ACACCTGCTT 1320

GAACAGCaAT CGCaACAACT TCACGAACGG CTTGTTGATA TTTTyTCGTA ATTTCTACGA 1380

AGTATTCATC TAACAACAAG TTTTCAAGTG CTGGATTTTT TTCATATGCA TCGGTAATTT 1440

TTTGTAAGAA TTGTGCACGG ATAATACAAC CTGCTCGCCA GATTTTTGCA ATTTCACCAA 1500

ATGGTAAATC CCATCCGTAT TCTTCAGAAG CTGCACGGAG TTGTGCAAAA CCTTGCGCAT 1560

AACTCATTAA TTTACTAAAG TATAAaTgTT CGCGAATTTT TTCAATCAAT TCTTTTTTAT 1620

CGCCTGCAAA GTTAAAGGCT GCAGGTTTTG ATAAAATGCC ACTTGCTTTT ACGCGTTCTT 1680

CTTTGTAAGC CGAAATAAAA CGTGCAAATA CTGATTCTGT AATTAATGGT AATGGTACGC 1740

CAAGATCTAA CGCACTTTGG CTTGTCCATT TACCTGTTCC TTTGTTGCCT GCTGCATCAA 1800

GAATAACATC AACGATTGGT TGGTCGGTTC CTTCATCATC TTTACGTGTC AAAATGTCAG 1860

CCGTGATCTC GATAAGGTAA CTGTCTAATT CGCtTCGTTC CATTCTTT 1908

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7515 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

TCTTCTATkG TTAAtTGCaw CTTGTCCAAC TCCTTGyCTT CTTCTACATT ATAACTAAGT 60

TAGGTGACTT TTTCACTAAT AACCAATTGA AACTTTTTCT TGAAAACTTT TTGCATTTTT 120

CATAATTTTG CGTTATTAGC GGATACATTA ATGAACTTAT GGTAAACTAA CATTCTGGAT 180

GGTTCATTCC AGCCAGCCAA TTAAAAACTA TTTTTAGAGC GTTCAACACG CTTATTCATG 240

ATTTTTTTAT AAAGGTGGAC TTATTTTGAT GAAAAACAAA TCGAAGCTTT TATGGTTAGG 300

AAGTTTTTTC TTGCCGTTTT TACTTCTGTT GCTTGTATGG ATGACATTAC AACTTGCGCC 360

TTTTGGTGAC AACAACTTAT TAGTTAGTGA TTTAGGCACA CAATACATGC CTTTCTTGAG 420

TTTTTTAAAA CGCTCTTTTC ACGAGGGAAT AACTACATTC TACTCTTTTT CCAATGAGAT 480

TGGTGAATCG ATTGTTCCTT TGGCAGCCTA TTACTTACTG AGCCCTTTTA ATGTGTTGGC 540

TTTCTTTTTT CCTTATGAAC AATTACCTAT AGCAATTTTA TGGATTATTA CCTTAAAACT 600

TGCCTTGATG GGCACTACGA TGTTTAGCTA TTTAAAATAC ACCTATCAAA AAGTTGATGG 660

CACCACGTTA CTTTTTTCAA CGTCCTATAG TTTCTGTGGC TTTGTCACCG TCTATAGCCA 720

AAACTTTATG TGGTTGGATG CACTGATTCT ATTCCCGCTT ATCTTATTGG GACTCCAACG 780

TTTATGGGAT CAACGCAAAT GGGGCTTATA CAGTATCACC CTATTTTTAG CGATTGTGAC 840

GAATTACTAT ATGGGTTATA TGATTTGTCT CTTTGCTGTT TTGTACAGTA TCTATTGGTT 900

CTTTAAAAAG AATACCAAAG CCCATGCTAT TCGGCAATTT TTTAAACAAA GTCCGTTATT 960

CATCCTCGTT TCTTTTTTAA CAGGAACCGC AACTAGCTTT TTATTATTAC CTGCAGCAGA 1020

AGGGATGCTC TATACAAAAA AAGCAGACTT TGATGTCTCG ACGTTCTTTT TAACGCCTAA 1080

```
GTTCAACACT TCATTCTTCT CGCAATTAGG CTTAGGTTCT ATTAATTATG AGTTGCGTCT    1140
AGACCATTTA CCAACCGTTT TTGCTGGATT ATTTGTGACA CTCCTCTGTG TTGCCTATTT    1200
TCAAACGAAA CAAATTGCGT TAAAAGAAAG AATAGCTTCA GCAATTCTTT TATTCATTCT    1260
TTTTTTAAGC TTTTGGTTAG AAGCCTTTAA TACTGTCTGG CACATGTTTC AAAGCCCAGC    1320
TGGTTTCCCT TACCGAAATG TCTTTATTTT TAGCTTTTTA TTGATTGTCT TTGCCTATGA    1380
AGTTTGGCTT AAAAAAGTGA CCATTCCTTG GACCGCACCT ATTATTTTTA GTCTGCTGTT    1440
AGTCATTGGC TACGGATCTT TGTATTATGG TCCACAAAAG AATCTCTTAA TTTCGATCAA    1500
TTATTTATGG CTAAGTTTAC TTTTTATTTG GTTGATTTTC TTTTGCTTGC GTTTAGCGCA    1560
CAAAAAGGCC TTAAGAAACT ATGTAGTCGT GGCTTTATTC TTACTGGTCA GCACAGAATT    1620
AACAACGAAT TTCTGGATTT CTTTTAAACA CATGCCCTTT GGTAGCCAAG CAACATTTGC    1680
CCAAGATTAC CGAAAACACA GCCAATTAAT CGATGAAAAA ATGGCCTCAG CACCAGAACT    1740
TTATCGCATG AAACAAGTCA TTCCTTCCAA AGAAACAGGG TTCCGTGAAA TCAATAACGG    1800
CTACAATAAT CCTTTACTTT ATGGCTATGC TGGCGTTTCT AGCTATACTT CGACTTTAAC    1860
TGCCACTACC CAAGACACAT TAAGTGCGCT AGGTTTATAT CGAAAAAATG ATCGCCGGAT    1920
TGCTTATGTG GATAATTCAC AACTAACTAA TTTACTTTTA AATGTGAAAT ATGACTTTTT    1980
ACCAATTGAA AAACCTACCA GTGAAAAATT GCTAAAAACa GTTGGTTCCA CGAAAATCAT    2040
GGaAAACGAT GAAGCGATCG GTATGGGCTT TTTGGCGCCA ACCGCATTAA CCAAGTTAAA    2100
ATTAGCAAAA AATAATCCTT TAGATGCCCA AGAAGaACTC CTGCAAACGC TCGTGCCGAC    2160
GGATAAACCT TACTTTAAAA CAGCTTCGTT GATAAATGAA CCTCATCATA CCAACGAAAC    2220
AATTGAAGCA ACCTTTAAAG TGAATAGCAC TGGTGACTTG CATCTTTATA TTCCGAATTT    2280
AAAATGGAAA AAAGTTACAC AATTGAAGGT AAACCAACAA GTTATCTCGA CGCCGATTTA    2340
TATTGCAACC AATCAACTGT TCAATTTAGG GCATTTTGAA AAAGGAACTA CCGTGACCCT    2400
GTCACTAACT GCTGAACAAG TGGTTGATTT AACCAATTGG CAACTTCAAA CTTTAGACCA    2460
AACAGCCTTT AATCGTGCGG TTGACAAATT ACGCCAACAA GCTCTTCACG TGAATGCAAC    2520
TAAAAAAGGC CATTTAAATG GCGCACTGAA TGTACCTGGA AATGACACTC AGTTACTGTA    2580
TACGTCCATT CCTTATGACC AAGATTGGCA AGTCAAATCG TCGCTACAAA AAGAACCCTT    2640
AAAAACACAA CGCATTCTAG GTGGCTTTTT AGCTGTTGAA GTCCCAGCTG GCAAACAGCA    2700
GTTAACCTTT GCCTATCATC CAAGAATGAT TTATCTCGGC ACTGCCGTCA GCGGAACGAT    2760
TTTACTGGGA ACGGCTGGTT ATCTTGGGTT TAAAAAATAC CGTCGAAAAC GTCAGGAGGC    2820
CACTCATGAC TAATCAGAAA AAATTTTTAA ATGGACTCAT TCGGTGGTTA CCCCTAGTTG    2880
GACTGGTGCT TTTCTTCGGC TTAATTCTAT GGGGATATTC CAGAGGAATT TTTCATTCTG    2940
TTGCTTCCTT GCAAGCTTTT ATTAAACAGT TTGGTAACTA TGCGGTTCTG TGTTTCATTC    3000
TTTTACAAAT CGTCCAAGTG ATTATCCCTA TTCTACCTGG AGGAATCTCT TCTGTTGCAG    3060
```

GAATGTTGAT GTTTGGCAAT TTGCAAGGAC TTTTGTATAG CTATCTTGGA TTGATTAkTG 3120

GTGAGTTTAT TGGCTTTTTA CTGGTTCGTT ACTATGGACG TAGCTTTGTA AAAATAATCC 3180

TTTCACCCAA TAAATACAAA AAATTTGAAG AGATTTTGGA CAAAAATGAA CACAATGTCA 3240

AAAAGCTCTT AATCTTTACT ATGTTAGTCC CCTTCGCTCC AGATGATATT GTTTGTTTAG 3300

TGGCGGGAAT TACAGATATT TCTTTAAAAG AATTCATGAA AATTGTTCTT TTATTGAAAT 3360

TCTGGTCTGT GGCTACTTAC AGCTATTTAA TGTTGTATTT GTTTCAATTA TTTGGCAAAT 3420

TATAACAAAA AGACAACAAA ACCAACGTTG GTTTTGTTGT CTTTTCTTTT TTAGATTGAA 3480

CGTAATTGGG CAATCCGTTC TTTGACAGCC TCTTGTTTTT CAAGGTAGTC CTTTTCTTTT 3540

GCACGTTCTG CTTCCACGAC CTCATCTGGT GCATTTGACA CAAAACGTTC ATTTGAAAGT 3600

TTGCCTTGCA CACGTTTGAC TTCTTTTGTC CATTTGTCTA ATTCTTTTTC TAAGCGAGCA 3660

ATTTCTTCCT CGATATTAAT CAACCCAGCT AATGGTAGGA ATAGTTCTGC GCCTGTTAAA 3720

ACAGCGGACA TGGCTAATTC TGGCGCTTCT ATTTCACGAC TGATGACTAA TTCCTCTGGG 3780

TTACAGAAGC GTTCCAAGTA GCTAGTATTA GCTGTTAAAA ATTCTTCTAC TTCTGTATCG 3840

TTTGTTTTAA TTAACAAGGT AATTGGTTTA GAAAGTGGTG TATTCACTTC TGCTCGAATA 3900

TTCCGAACCG AACGAATAAC TTCTTTCAGA ACTTCCATTC CTCGAGCTGC TGCTTCATCG 3960

TTAAATTCTT CATGAACAAC TGGGTACTCT GCCACAACCA ATGACTCACC TTGATGCGGA 4020

ATTTTTTCCC AAATTTCTTC TGTTACAAAC GGCATAATTG GATGCAATAA GCGTAAGATT 4080

TGATCCAATG TGTACACCAA GATGCTCCGT GTTGTTTGTT TAGCTGCTTC ATTGTCTCCG 4140

TAAAGAATTT CTTTACTCAT TTCAATATAC CAATCACAGA AATCATCCCA GATGAAGTTG 4200

TATAATTGGC GACCTGCTTC ACCGAATTCA AAGCGATCAA ATAATTCTGT CACACGGGCG 4260

ACTGTTTCGT TTAAGCGCGT TAAAATCCAA CGATCAGCGA CAGTCTTTTC ACCACTAAAG 4320

TCAATATCTG CAGCAGTCAT ACCTTCTACA TTCATGATAA CGAAACGACT TGCATTCCAG 4380

ATCTTATTGA TAAAGTTCCA AGAAGCATCC ATTTTTTCAT AACTAAAACG AACATCTTGA 4440

CCAGGTGCTG AACCATTTGA TAAGAACCAA CGTAACGCAT CTGCGCCATA TTTTTCGATG 4500

ACATCCATTG GATCAATCCC GTTCCCCAAT GATTTACTCA TTTTGCGTCC TTGCTCATCA 4560

CGAATTAAGC CGTGAATTAA GACATTTTGG AAGGGACGTT CACCTGTAAA CTCTAAGCTT 4620

TGGAAAATCA TGCGACTTAC CCAGAAGAAG ATGATGTCAT AGCCAGTAAC TAATGTGCTT 4680

GTTGGGAAGT AACGTTGGTA ATCTTCGCTT GCTTCATTCG GCCAACCCAT CGTTGAAAAT 4740

GGCCATAAAG CAGAACTAAA CCATGTATCT AAGACGTCAC TATCTTGAAC CCAGTTTTCA 4800

CTGTCGGCTG GCTCTTCCAT CCCAACATAC ATCTCGCCCG TTTCTTTGTG ATACCAAGCA 4860

GGAATTTGGT GGCCCCACCA TAATTGTCGC GAAATAACCC AGTCATGTAC ATTTTCCATC 4920

CAACGTAAAA ATGTTTGATT GAAGCGTGGT GGATAAAATT CAACAGCGTC TTCTGTTTCT 4980

TGATTTTTCA TTGCTTTTTC AGCTAATGGT CCCATTTTTA CAAACCATTG GGTCGATAAA 5040

CGAGGTTCCA CGACAACGCC TGTTCGTTCT GAATGGCCCA CGCTGTGATT CATTGTTTCG 5100

ATCTTAATTA AGCGACCTAG TTCTTTCAAG TCAGAGACAA TCGCTTTACG TGCAGCAAAA 5160

CGATCCATTs CCTcGTATTT TCCAGCCAGT TCATTCATcG TGCCGTCTTC GTTCATAACA 5220

TTTACTCGAG GTAAGTCATG ACGATTACCG ACTTCAAAGT CATTTGGATC ATGGGCAGGT 5280

GTGATTTTTA CAACCCCTGT TCCAAATTCC ATATCTACAT AGTCATCTGC AATAATTGGA 5340

ATTTCTTTAT CGACTAATGG CAAGACAACT GTCTTACCAA TAAGTTCTTG ATACCGTTCA 5400

TCTTCTGGAT GGACCGCAAT CGCCGTATCT CCCAACATTG TTTCAGGACG CGTTGTCGCA 5460

ATTTCAACAA CGCCAGAACC ATCTGATAAT GGATAGCTCA TGTGGTAAAA AGCTCCTTCA 5520

ATATCTTTAT GGATTACTTC AATATCAGAC AATGCTGTTT TCGCTTTTGG ATCCCAGTTA 5580

ATGATATATT CACCACGGTA AATCAAATCT TTCTCGTAAA GAGAGACAAA CACTTTACGA 5640

ACGGCCTCAG ATAGCCCTTC ATCTAAAGTA AAACGTTCAC GGCTGTAATC TAAAGAAAGA 5700

CCCATTTTTG CCCACTGTTC ACGAATGTGA GAAGCATATT CTTCTTTCCA TTCCCACACT 5760

TGATCAACAA ATTTTTCACG ACCTAAGTCG TAACGTGAAA TTCCTTGTTG TGCTAGTTTC 5820

TCTTCTACTT TTGCTTGCGT GGCAATTCCG GCATGGTCCA TTCCTGGCAG CCATAACGTA 5880

TCAAAGCCTT GCATTCTTTT TTGGCGGATA ATCATATCTT GTAATGTTGT ATCCCAAGCA 5940

TGCCCTAAGT GTAGTTTCCC TGTTACATTG GGTGGTGGAA TAACAATTGA ATAGGGctTC 6000

GCTTTTTTGT CGCCACTAGG TTTAAATAAA TCTTGATCTA GCCATTTTTG ATAACGGCCA 6060

GCTTCGATTT CTGTCGGTTG ATATTTCGTT GGTAAGTTTT TTTCTTCTGA CATTTTTTCC 6120

CCTTCTTTCT TAAAATACGC GATTTGGAGC GGTAATTATT TTACAAATGA GTTTTTTAGG 6180

CTAAACACTA AAATACGCCC TAAAATACTT TCGTATTTTA GGgCsTAGaT TCATCTTATT 6240

TACGCGGTAC CACCTAAATt GCATTtGAAT tGAATtCCTT CAAATGCCAC TCATtCGTGA 6300

GATAACGTTC ACGAAACGAA TCGTTCTACT AACGTACTGT GTTCAAACCA TTGACTCACA 6360

AGCTACCTTC TCACTGGTTG TGTAAAAGTC TTGCACCTGA ACGACTTTTT CTCTAAAACG 6420

ACGCCAATGA TACTCCTCTT GTTCGCTGTC TTTGCTTCAT TCTACAATGA AGCGCAAgcA 6480

AAGTCAATTT GTAATCTAGA GGGATTACAC TTTTTTTGCA GATTCTAACG CTTGTTGATA 6540

ATCGGGTTCC GACGAAACTT CTGAGACGAT TTCTTCATAA ACTAACGTTC CTTCTGGGTC 6600

AATCACAAAA ATAGCACGTG CTAAACGGCC CATTTCTGGA ATATACAAAC CATAAGCTGC 6660

ACCAAATGAG TCTTCAGTAT CATGAAGCAT TTCCATTTCG ACACCCTCAG CGGCACACCA 6720

ATTTGCTTGT TCTTCAACTG TGTTATTGGA AATCGTGATA ATCTGAACAC CGTCTAATTT 6780

CGCCGCTTCT TGATTGAAAC GTTTCGTTTG CAGTGAACAA ACACGTGTGT CGATATCAGG 6840

AACCACACTA ATTAAAACCG TTTTACCTTT ATAATCTGCC AAGTTGATTT CTTGGTTGTT 6900

TAAATTTTTC AAAGAAAAGA CGGGAGCTTT TGTGCCAACT TTAGGCTGCT CACCTGTCAA 6960

TTCTAATACG TGCCCTTTTC TTGTAACATT CATTCTGAAA TCCTCCTTTT CCACTTTCAA 7020

TAATAAGTAT AGTAGAGGAT TTTTATGAAG ACAATTTTAT TGCTTTGGGC TATTTGCTAG 7080

TGAAAAAGTT TCATTAAAAT AAATCAGTGT TTGCAGTTCC GTTGTTAAAT CAATATATTG 7140

CACACTAACG TTAGCTGGCA CACTGACACG GTCTGGCGCA AAATTTAAAA TCGCAGTCAC 7200

ACCCGCTTGT ACAATTTTGT CGATGGCTTT TTGAGCATGG TGACTTGGTA CTGTAGAAAT 7260

GGCTACCGTG ACCCCTTCTT GTCGAACGAA TGCTTCTAAC TCAGACATAT CATGAACTAA 7320

TAATCCATTA ATCGTGGTGC CAACTAAAGC CGAATCATTA TCAAAAGCAC AGACAATATT 7380

CAAATTCTCG TTGCGTCTAA AATTATTTTT TAATAATGCC TTGCCTAAAT TGCCACAACC 7440

AATCAGAGCA ATTCTTTTTT CTTCTTGCGT ATTTAAAATA TTACTGAATA CCTCAATTAA 7500

ATACgGCACA TCGTn 7515

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7156 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

ATATGGCCTT CACTTTAGTG AAGAGATTTA ATTGTCTTTA CAGATAAACG CTTAATTTTA 60

GTCGATAAAC AAGGAATTAC AGGTAAAAAA GTTGATTATA AATCGATTCC CTACAAATCT 120

ATTTCACGCT TTTCTGTTGA AACTAGTGGA CATTTCGACC TAGATGCAGA ACTAAAGATA 180

TGGATTTCAA GTACTGAGTT ACCTTCAGTA AGTCTTCAAT TTAGAAAAGA TAAAGATATT 240

GTTGCTATTC AACAAGCTTT AGCAGCTGCT GTATTATCTT GAATACACAA AAAAAGACAG 300

GAAAGCGACC AACTTTCCTG TCTTTTTAAT TktAGATAAA ATAAAAAAGG CTTCTCAGCC 360

TTGATACGTT CCTTATTCAA TCAGACACAT GGCGGCACTT GCTTAGTGCT GCTTCCTTCC 420

GGATCTGACA CAATTCACTA GCCCGCCATT GTCCTAGCCT TTCGGCAAAA CCTAGTATAA 480

CGTATTTTTA TTCTTTTGAC TAGTCTTTCT TTTGATTTTT TAAAAGTTTC TCTTCTTTTT 540

TCTTTGCCCG TAACTTTTTA AAGAACTGAA CTAGCAATAA GCGACACTCT TCTTCTAAAA 600

CGCCTGCTTC GACATAGGCC ACATGATTGA AGCGTTCATC TTCTAATAAA TTCATCAATG 660

TGCCAGCAGT GCCTGCTTTA GGATCAGTTG CGCCATAATA AACTTCTGGA ATTCGTGCCA 720

AAAGCATCGC CCCACTACAC ATTGGACAAG GCTCTAAAGT TACAAATAAT TGGGTTTCTT 780

CTAGACGCCA ATTTTCTATC CCGCGACACG CTTCTTGGAT TGCATACATT TCTGCATGCG 840

CCGTTGCTTC TTGCCGTGCT TCTCTCAGAT TATGCCCGCG GCCAATAATT TCCCCTTGCC 900

GAACAATTAC AGCTCCAATA GGCACCTCAG CCAAACCTTC TGCTTTTTTC GCTTCTGCAA 960

TGGCTTCTCG CATAAAAAAT TCTTTTTCTT CTTGTGTTAA CGTTGTTTCT TTTTTTCCCA 1020

ATTAATTTCC ACCTTTCGGC TTTTTCCCAC TAAACTTCTG TCTTGCTTCA TGGTACAATA 1080

AATCAGTGAG GTGTGTCTAT GATTATTGAA AATTTATCTC AGCTTTATCC CAAAGGCTAT 1140

TTATCAAAAA CAATTGGTTC CGAAGCCACC TATTCAATTC CTGTTGACGA CCAATTTTTC 1200

ATTGTGAATA AAGCTTTCTT AAGTATTAAA GAACAGCATT TGTTGGAAGC GCTTTTCCCA 1260

ATTTCTGAAA aCCCAACAAT TACTGGTAAC CATCCTTGGT TCAGTTACTT ATTTCAACAG 1320

GCAAAATTAC CTGCCGAAGG GACCTTTCGT ATGTTGCAAA TTCAAACGAA TGTCACAAAA 1380

GAATTACAAG CTGAATGGCA ATTTAATTTA ACAAAGATGT TTCCAGATAC AGTTGATTGT 1440

TTCAGTCCAA GTAATAATAT GTATATTTTA GTTGAAGAAC AGTCAAAAAA TACCTTTCAA 1500

CAAGAAGAAA TTCAAGGGAT CTTTTTAACA TTAGATACAG ACTTTGATTG CACTAGTGCT 1560

GTATTTGTTG GCAATTTTTA TTCCTCTGAA GACATTCTCC GGCGTTGTTT CCATGAAGAA 1620

CAACGTATTT TTTCAGAAGA ACTTAATTCC TCTAGTCGGA CGACTGTTTT TAGTTTAACT 1680

GATGTTGCCC TCCACTATTT TACCAAGGAA GCCATGAGCC AAAATGTACT TGTTGAGTAT 1740

TATCGTCGTT TATTAAATAA AGACACCGAT ATTCAGCCAA TCATCAGCGC TTTATGGAAA 1800

AATCAAGGCA ATATCAGCTC CACTGCAAAA GATTTATTTA TGCATCGAAA TACTTTACAC 1860

TATCGGTTGG AAAAGTTTTT TGAACAAACT GGCTTATCTT TAAAAAAAAT GGATGATTTA 1920

ATCTTTTGTT ATTTGTTATT AAGAAAGTAG TACGTCCATT TTTTACTCGT ATGTTTTTTA 1980

TTCCCAATTA GAAAGTAGGC GTTCCAATGA AAATTACCGC ATATGTCGCT AGTGCATTTA 2040

GTAAAAACCA CGAAGGTGGC AACAAAGCAG GCGTTGTTTT TATGACACAG CCCTTAACGA 2100

GAACCCAAAA AATGGCCATC GCCAAAGAAC TAGGATTTGC AGAAACTGCT TTTATTTCAG 2160

ACTCAAAGAT TGCTGATTTC CGCCTTGAAT ATTTTACACC AAAAGAAGAA GTCGACTTAT 2220

GTGGCCACGC GACAATTGGC GCTTTCACCA TTATGAAGCA CTTAAATAAA TTATTGCAAA 2280

CAGCATATAC AATCGAAACG AACAGCGGTT TGCTCGCAAT CACGGCTAAC GGTGAACAAC 2340

TATTTATGGA ACAAACTGTC CCAACTTTTT ATGATACGCT CACTTTAGAA GAATTGAGCG 2400

GCTGCTTTGA TTTAAACGCA GTGgCAACTG CTTTTCCTAT CCAGATTGTG TCAACGGGCC 2460

TAAAAGATAT TTTGGTTCCC ATTAAAAATG CGCAATTATT ACAGGATTTA CAACCAAACT 2520

TTGAAAAAAT TAAGTTACTT AGTCAAAAGT ATGGAGTTAT TGGGATGCAT CTTTTTGCTT 2580

TGTCTGAAGA AAATATCATT TGTAGGAATT TTGCACCCCT TTATGATATC GATGAAGAAG 2640

CCGCTACTGG TACATCAAAT GGTGCGTTAg CTTGTTATCT GCACCATCAC CAAATTTTAC 2700

AAAAAACCAA TTACACCTTT GCCCAAGGTT ATACCTTAAA CGCTCCTTCA GAAGTCTTAG 2760

TGCGTTTGAA CACTACATCC CAAAAGACAA TTGATCAAGT TCTAGTTGGT GGCAGAGGCT 2820

ATTATGTTGA AACAAAAGAA TTATTTTTAT AAAAAAAAGA TGGcCTAGGC CAcTTTTTTa 2880

ACTTCaTTGC ATCTTCAATA AATGTGAAAA AGCGTTCTTT TAAAGAATCT GTATTGTCTT 2940

TAACTTTTTC GCCATATTCA GCTAATTTTG AACCGCCATC TTTCACACGA TCCACAACAT 3000

TTAAAACAGA ATCTAATTCA TCATCGGATA AATCATCGAC AATCGATAAT AATTTTTGGT 3060

```
TTCCATCAAA TTTATCGTCT ACAAACTTTT TAACTTTATA ACGATTGGAA ACATGAGATA    3120
CCTTCTTAAT AATTTTTTCA GAAGCGATGA CTGCAACAGA GACACTTGCa ACTGCAGcAA    3180
TGCTTAAACC GATACCAATT TTTGTTGATG CTTTCaTTTC CTTCAGCTCC TCTGTTATTT    3240
TCTACTTTAA TCATAACATA TTATTACTTT wAkGCGTAwT rGAAaGrCCT TTACTAGCAC    3300
TTAGTCTCGT CGGTTATTGC CACCAAATAA AATAACCAAA CGTAAAAGTT GTAAAAAGGT    3360
CGAAAGCGCC GCTGCTACGT AGGTAAGTGC AGCTGCAAAA AGAACTTTTC TGGCCATTGG    3420
GACTTCTTCT TCTGTTAATA ACCCACCTTC AGAAAGAATA CTTAATGCAC GTCTAGACGC    3480
ATTAAACTCC ACCGGTAAAG TCACCAATTG AAAAATCAAG GCTAATGAAA AAGCCAAAAT    3540
TCCAATGTTG ATTAATGTTT GATTCCAACT AAATAAAACA CCAAGCAAAA TTAGTGGAAA    3600
GGAAATGGTT GAACCGATGT TGGCAACTGG AACAATAGCT GCACGAAGAC GTAAAGGCAC    3660
ATAGTTCGTA TGATCTTGTA CCGCATGTCC GCATTCATGA GCCGCCACGC CAATAGCCGC    3720
GACCGATGTC GATTGAGCAG TGGCTTCGGA TAAGCTCAAC ATTTTGTTTC CAGAATTATA    3780
ATTGTCGGTT AAATCACCCG CAATTTGCTG AACTCCAACA TCATTAATTT GTTCTTTTTG    3840
TAAAATATAT TGTGCTGCCT GCGTGCCTGT AACATGTTTA CTGCTTCGTA CTTGATCATA    3900
TTTACGAAAG GTACTGTTAA CATATGCGGA GGCCGCACCA GAAATCGCAA TTCCGATAAT    3960
TACCAATATA TAGGTTGGAT CGATTCCAAA ATAAAAGGGC ATCATTGTCC TCTACCTCCA    4020
TCTTTTTTTC TTTATTGTAG CATGATAGTG AAAAAAAGGG AGCGACAAAC GTTGTCTTTT    4080
GTTTAAAAAA ATGATGAATT TTTCAAGAAA GTGCCGAAAC TAACAAAATT GTCACTCAAT    4140
TAACGGAAAA AAATATGGCA AATTTCCTAA AAAAAAGGCA TACTATTCCT TGAGGTGAGG    4200
AAAATGAAAA CATTACTAAA AATATTAATT CCCATAGTCA TTTTAGCCGG TGGTAGCTGG    4260
GGCGTTTATA ATTACTATTA CGGTGGAGAA GCCTATTATA CACAAATCAC AACATCTGGT    4320
GAAAAGAAAG ATGAAAAAAC AAATTCTGGT GAAGCAATGA CCATTTATTA CTATCAACAA    4380
CCTGCGTTTA ATAAAAACGG TGAAGAAAAA ACCGTGGAAT TAAACGAATC CCGTGATCAA    4440
CCTTTGCGTA TGAAAGCCTA TCTAAAATTG AAAGTTAATC CCCGTAAAGG CGTGATCAGC    4500
TGGAATGAAG TGACAGAAAA AGAAGTCCCT GAAAAAGCTT TAGAAAAATT AAAATAAAGA    4560
AAAAACCGTT GGCAAGCGAC CAACGGTTTT TTCTTTATAA TTGATTACTA ATTTTTTCGA    4620
TATTTTTTAG CATAATGGAC ACTGTGCCAT CTGGATTATT AACAAATTCA ATTAAATCGC    4680
CATTGCGGTA AACATCTAAC GGAACAATTA ATTCAATACC GTTATCCATT TTTAACTTCT    4740
GTTTGCCAAA TTTCTTTTCA GAAATTTCGC GAGATTCTCT GACTGGCGGT GTTTTGTCGA    4800
CAAACTTCGA TTCATTCACT TCTTCTTGGT AAGCTAATTT AGCTGATACA TTTTCTTTAA    4860
ACACTTGCTC AGCAATCTGC TCATTGTCAA TCGTCCCCGT TTCCTCAATG CTATCATACA    4920
CCGcTTCTTT TACAGAAGCC ATTAGTTCAA ACTCATCTTC GTTAAATTTC TTTCCAATGC    4980
GTTTGACTGC CTTTTTAATT TCTTTGATAT TTTCTTCCAT GGAAGGTGCC GCTTGCGTCT    5040
```

CAATCACTTT TTCAGAGAAA TACCAGATTT TTTCTCCTGA AAATTCATAT CTTTTCTCTA 5100

GTAATTCAAA CGCTAATGTA TCTAAGTCAA CCGTAATCCC TTCATCTGGT TTTTGTGATT 5160

TTGAAGACAA AATTGCTCGG TTGATAATCA GTTTGTTATA AACGGATTCT TCTTCATAGT 5220

CGACGTAATG CGTGTAACTT TCTTTATAGT TTAATTTAAT CATTGCGACA TGTAACACCG 5280

TATCTAATTC ATATAAAATG AACAAAATAT CAGCACTAGG CGCATCTTCG CTTCCTGAAT 5340

AAACATCATA CCAATGTTGG GTAACTTCTT GGCTTTTCTT TATAAAATCA CTAGGAATCC 5400

CACGTAACGT GTCAACAAAG GAACTATTTT CCGCCAATGT TCCAGTCTTT GTTTGTGCAG 5460

TCGAAAGTTT TGAAATTTTT GCAGTTAAAT ACGTTCGAAT ATACTCGGTC GTTAAATCTA 5520

GTTCCTTTTC AGAGAAAACC GGCGTCCCAG TTTCTCGATC AATAATGTGT AAAATTGCTT 5580

TTTTTAAATA AATATCCATT TTTACGCTCT CCTTTTCTCA ACCTTTTTAG TCTAACTGAA 5640

AGAGAGAAAA AAGCCAATAG TTTCTATAAA AAGArGAAGA ATTCGTAGAC AAATGTCTAC 5700

GAATTCtTCT TTACTTCAAA AAGCGATTAC CCTTCTTCGC TGTATAACAT CCCTGTTGCT 5760

TCATCAACTG GATCGACTTT ATTAAAAATG CCTTTCTTTC TGAAAATATA AGTTAAGCCC 5820

ATTGGTACGA TAATGGCAAT CACCATTGCG CCTAAAAACG GCAAGTAATA TTGTGGCACA 5880

ATAGATAAAA TTCCTGGCAA GCCACCGACG CCGATACTGG TCGCACGAAC GCCCATTAAA 5940

GTGGCAAACA TTCCTGCCAT TGCTGAACCA ATCATGCCCG CAACAAATGG ATAAACATAT 6000

TTTAAGTTGA TCCCAAACAT CGCTGGTTCA GTCACACCTA GGTAACAAGA AATcATCGCT 6060

GGAATCGAAA CTTGTtCTTC TTTTTkGTTT CCmCGATGCA TAAAGaTAAC CGCTAAAACA 6120

GCTGAACCTT GTGCAATATT TGATAGAGCA ATCATTGGCC ATAAGTTGGT TGACTGAAAG 6180

TCAGCAATCA ACTGTAAATC AATCGCATTA CTCATATGAT GCAACCCAGT GATGACTAAT 6240

GGTGCGTACA TAAAGCCAAA GACGGCGCCA AATAACCAAT TAAATGTTGA CGTTAATCCT 6300

GCATTTACGA TATCAGAAAC CCAACTACCA ATCGTCCAAC CGATTGGTCC TAAAATGACA 6360

TGTGCAGCCA ATACTGTTGG TACTAAAGCA AACAATGGGA CAAAAATCAT TGAAATTGCT 6420

TGAGGAATCA CTTTTCTAAA GAAAATTTCA AGATAGGCAA GCAAGAAGCC GGCTAACATG 6480

GCTGGGATAA CTTGCGCTTG ATAACCAATC ATGTTAACTT GAGCAAACCC AAAATCCCAG 6540

AAAGGAATTT CTGCTGCTTT AGTCGTCGCA ACTGAATAAG CATTCAGAAG TTGTGGTGAG 6600

ACTAATGTAA TCCCTAGGAC AATCCCCAAG ATTTGTGTGG TTCCCATTTT CTTCGTAATA 6660

CTCCACGTAA TCCCAACTGG CAAGAAGTGG AAAATTGTTC GCCGATTAAC CATAAGAATG 6720

AATTGACTCC ATTCCAAAAT TGAGAAACTT CCACGATTGT TTGACCATCA AGAAAGCCAA 6780

ATGGAACCCC TTCCAAAATA TTGCGGAACC CTAAAATCAA TCCCCCGATA ACAAGTGCAG 6840

GAATCAATGG TGTAAAAATC TCAGCTAAAA CAGAAACAAC TCGTTGAACA GGACTCAAAT 6900

TACTTTTGGC AGCTGCTTTC CCTTCTTCTT TTGAAACCCC TTCTAAACCG GAAATTGCAG 6960

AAAAATCATT ATAAAATTGT GGGACTTCAT TCCCAATAAT TACTTGAAAC TGTCCAGCAT 7020

TTGTAAAGGT TCCTTTTACA GAAGGAATAT CTTCAATCAC AGGTACATCC GCTTTTTTTG 7080

GATCTTTTAA TACAAAACGC ATTCGAGTTG CACAATGAGA AACAGCTGAA ATATTTTCTT 7140

TGCCTCCTAC TGCTTT 7156

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1282 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

ATTCGTTGTA ATTCCAACTG GAATTCCTAA TACGAACTTA GAAACGTCGT ATCCATTTCT 60

TGCTTTAATG TCAGTTGTTT TTGGACCAGT AGCTGGAGGG CTAATTGGTT TGATCgGTCA 120

CACgTTAAAA GATTTTACCA CGTATGGTAG CGCTTGGTGG AGCTGGATTA TTTGTTCAGG 180

AATTATTGGT ATTATTTTTG GCTTCGCTGG TCGTAAAATG GATCTTCAAC ATGGTGAATT 240

CACAACAAAT GACATGGTGC GTTTCAATAT TTTCCAAGCC TTTGGCAATA TCGTCGTTTG 300

GGGCTTAATC GCGCCTAGCT TGGATATTTT GATTTATAGT GAACCCGCAA GTAAAGTGTT 360

TACACAAGGC GTATTCGCGA CTGTTTCAAA TATTGTCGCT GTGGGGATTA TTGGAACATT 420

GTTGATGAAA GCTTATGCAT CAACAAGAAC AAAGAAAGGC AGTTTATCAA AAGATTAAGA 480

AACTGTCGAA ATAATTTAAA TGGTTAAGGG TCTGCCCTAG TAATTACTTT GCTAATGCTC 540

AGACCCTTTC TTTTATGGAG AAACTGAGGT GAGGCCGAAT GAAAAAACCA ATCATTACTT 600

TTAATAATTT TTCCTTTCAA TATCATAGCC AATCTGAACC AACATTAAAG GGGATTCAAT 660

TAACCATCTA TGAAGGAGAA AAGGTTTTAA TCGTTGGGCC TAGTGGTAGT GGTAAATCAA 720

CATTGGCGCA ATGTATCAAC GGGTTAATTC CCAATATTTA TGAAGGTGAG ATTCAAGGAA 780

CAGCTACGGT TGCTGGTAAA AATATCCAAG AAACAAGTCT ATTTGATTTG TCTTTTGATG 840

TAGGAACTGT TTTACAAGAT ACCGATGGGC AGTTTATCGG TTTAACGGTT GCTGAGGATA 900

TTGCTTTTGC TTTAGAAAAC GATGCAGTTG AACAAGCAGA AATGAAAAAA GCGGTTCAGA 960

AATGGTCCGA GATTGTTGAG CTGAATCAGT TATTGCAACA CCGTCCACAA GATTTGTCAG 1020

GTGGTCAAAA ACAACGTGTA TCTATGGCGG GCGTTCTAAT TAACCAATCA AAAATTTTAC 1080

TTTTTGATGA ACCTTTGGCT AATCTTGACC CACGAGCGGG ACAAGAAACA ATGACTTTGA 1140

TTGACACCAT TCAACAAGAA ACAAAAGCGA CTGTTTTAAT AATAGrACaT CGCTTAGAGG 1200

ACGTtCTTTG CGAATCAGTT GATCGAATTA TTGTCATGAA TGAgGGCACa ATTATTcTGA 1260

TACAACGCCA GATGAATTGT TA 1282

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5840 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

TAAGGATTGT TACTGCTTTG CAGGCAAAAC CAGAGTCTTC CATTAACACA ATCAGTGGAG 60

GGATTCTGGG CTTTTTTTAT AAAAAAGTGA GGTTGCATTC ATCATGAGGA TTGAAAAAGT 120

TTTAAATAAT AATGTCGTGC TTTCAAGAAA TGAGCAAGGC GAAGAAATTG TTTATATGGG 180

ACGCGGCTTA GyCTTTCAAA AGAAGATTGG AGACACGATC AATCCAGATT ATGTTGAAAA 240

GGAATTCGTG TTAAAGGACT CCGCTATGGC AGGGCAATTC CAGCAGCTGT TTGAAGACGT 300

CCCCACCAAA GAAGTAGAAG TCGTCAAACA AATCGTTGAT TTAGCAGAAA CAACGCTAGC 360

GATTGAGCTT TCATCAAATA TTTATTTAAC TTTAACGGAT CATATTCACT ATGCGTTACT 420

TCGAGCCAAA GAAGGAATTG ATATTCCGAA TCCTCTTGTA TTCGAAACAC GCAAATTTTA 480

TCCTAAAGAA TTTGAAATCG CAAAGCAAGC ATTAGCGATT ATTGCTGAAA AGCTAGGGGT 540

TCAATTTTCA GAAAATGAAG CGGGCTTCAT TGCTTTTCAT ATTGTGAATG CTGAACAAGG 600

AAATGGCAAT ATGGAAGTTA CCATGGAAGC AACCAAGATG GTTCGTGATA TTTTAACGAT 660

TATTAGTCGC TACTTTGGAC AGGTGTTTGA TGAAGATTCA TTGAACTATC AACGAATTGT 720

TACACATCTT CaATATTTTG CGCAACGTTA TTTAAAACAA GAAGCGCATG aTGAGGAGGA 780

CGAATTTCTT TTtGCATtGG TCCAAGGAAA GTATCCTAAG GCGTTTCAAG CGGTCCAACG 840

TATTAATGAG TATTTATTAA AAAGCTATGA TCGACCAATT GATCAAGCGG AAATGATTTA 900

TTTAACCATT CATATCCAAA GAGTGGTCAA TGATAAAAAA GCATAACAAC CAATAATCGC 960

GGATTGTTAC TGATAATGCA GGCAAAACCT AGATTGAAAA AACCGATAAC GATTTTCGCA 1020

GACTGCGCAC TGTGAAGAGA AGGATCTGGC TCTTTTTCAA TGTAGGTTTT TTTGATATTA 1080

AAAATTAATG GAAATGAGGt CTAGAGATGA GTAAAAATCA AGAAATTGCA GCACGTGTGT 1140

TAAAAGCTGT TGGGGGCGAA GsCAmCGTTA ATAGkGTTGk TCACTGTGCy ACACGCTTAC 1200

GtTTTAAATT AAAaGrTGAA AATAAAGCAG ATACAGCTGC ATTGAATGCT GACCCAGATG 1260

TTATTCAGGT AGTTCAAAGT GCAGGACAAT ATCAAGTGGT TATTGGCAGT CACGTTAGTa 1320

TGTCTATAAG GATTTGATGG CCAACAGTGG ACTAGGTAAT GATTCTGATA ATAGAGAAAA 1380

AGAAAGTGGC GGAAATATCT TTAACCGTTT AATTGATATT ATTTCTTCTA TTTTTACACC 1440

GTTTTTAGGT GCAATGGCCG GCGCTGGGGT TCTAAAAGGC TTCTTAACAT TGGCTGTAAC 1500

TATGGGCTGG TTAGCTGACA CTTCAGGCGT ATATCTTGTT CTCTTTTCGA TTGCAGATGG 1560

TTTATTTACT TTCTTACCAA TTATGTTAGC GTTTACGGCT GCTAAAAAAT TCAATACTAA 1620

TCCCTTTTTA GCGGTTGCTT TGGCGATGGC CTTGGTTCAT CCCAGCATTA CTGCTTTAGC 1680

AGGCAAAACA ATTAGCTTTG CTGGTCTTCC AGTGATTATT GGGCCAAGTG GCTATACTTC 1740

TTCGGTTTTA CCAATTATTT nGGCTGtTTT TGCACAAAGT TATGTAGAAC GTTTCTTCAA 1800

AAAAGTAATT CCGAGTTTTC TACAAATTAT CTGCGTTCCG TTAGCTGTTT TCCTTATTAT 1860

GGCACCTGTT ACCTTCTTAG CTATTGGTCC TATCGGGACG GTCATCGGCG ATTGGTTGGG 1920

ACAAGGATAC AACGCAATTT ACGCTTTTAG TCCAATTATC GCAGGGTTAT TAATGGGTTC 1980

GTTGTGGCAA GTCTTGGTAA TGTTTGGTAT GCATTGGGGC TTTGTACCAA TTATGATGTT 2040

AAACTTAACA CAAGGTGGCG ATACGATGGT ACCGATGTTA TTACCAGCCG TTATTGCACA 2100

AGGCGGGGCT GCTTTAGCTG TCTTTTTCCT AACAAAAAAT GTGAAACTAA AAGGTTTGGC 2160

TTTGTCTTCA AGTATTACGA CTATTTTTGG AATTACTGAA CCAACTGTAT ATGGCGTGAC 2220

TTTACCATTG AAAAAACCAT TTATTGCAGC TTGTATTGGT GGCGGTATCG GTGGTGCATT 2280

TGTGGCTATG AATCACGTGA AAAACTTTAC GTTTGGCTTG GTTAGTATGT TGAGCTTGCC 2340

TGGCTTTATT CCTGCAGAGA CAAAAGATAC TGCACCGATG ATTACTGGTG CAATTGGTGC 2400

CGGAATTGCC TTTATCATTG CGTTTGTCTT AACGTTTGTC TTACGTTTTG AAGATCAACC 2460

TAATCCAGAA ACAGCAACTG AAAAAACAGA AACTGATAAG ATGGTGGCAC CTGTAAAAAC 2520

GAATCAAGAA GACAAAATTA TTTTAGCAAG TCCACTTCAA GGTGAAATTT TACCGCTAGA 2580

AAAAGTACAA GACCCTGTTT TTGCTTCAGG TGCTTTAGGA AAAGGTGTTG CAATTGAGCC 2640

GACTGAAGGC AAACTGTATG CACCCGCAGA TGGTGAAATC ACCACATTAT TTCCGACAGG 2700

ACATGCTGTT GGCTTGACGA CAACAGAGGG CGTTGAATTA TTAATGCATA TTGGCATGGA 2760

TACGGTCGAA TTAGATGGTA AAGGCTTTGA ATTATCAGTG AAACAAGGTG ATTCTGTTAA 2820

AAAAGGAGAT TTGCTAGTTA CTTTTGATAT TGCTGCCATT AAAGAAGCTG GTTATCCGGT 2880

AGTTACACCG ATTGTGGTAA CGAATACGAA TGATTATTTA GATGTTTTGG ATATGAACCA 2940

AACAGACGTG TTACATGGTG AAGACTTTTT AGCCATTATT AAATAAAAGA CAATTGGAGG 3000

AGGAACAAGT ATGACAACAA CGACAAAATT TCCAAAAGGA TTTTTATGGG GCGGTGCGAC 3060

CGCAGCCAAT CAATTAGAAG GGGCTTACTT AGCCGATGGT AAAGGGTTAT CTGTGGCAGA 3120

TGCGATGCCG GGAGGAAAAC AACGTTTCGC TATTTTAGGA GATGAAGCGT TTGATTGGAC 3180

GATTGmCGAA ACGAAATATC GCTACCCTAA TCATACAGGA ATTGATCACT ATGATCGCTT 3240

TAAAGAGGAC ATTGCGTTGT TTGCTAAAAT GGGCTTCAAA TGTTATCGCT TTTCAATTGC 3300

GTGGTCGCGT ATTTTTCCAC AAGGAGATGA AACACAACCG AATGAAGCCG GCTTGAAATT 3360

TTATGATGCG GTTATCGATG AATGTTTAAG CTATAACATT GAACCAGTAA TTACGATTTC 3420

TCATTATGAA ATGCCACTGC ATTTAGCGAA AGAATACGGC GGGTGGAAAA ATCGTCAATT 3480

AGTTGACTTT TATGAACGTT TTGCGGAAAC TGTTTTAACA CGTTACCACC AAAAAGTTAA 3540

CTATTGGATG ACATTCAATG AAATTAATTC AGCTTTTCAT TTCCCAGCGC TTAGTCAAGG 3600

ATTGGTAAAA AGCAATGGAG CCAATGATTA TCAAAACATT TTCCAAGCAT GGCATAATCA 3660

ATTTGTTGCA AGTAGTAAAG CTGTAAAAAT TGGCCATGAA TTAAATCCAG AATTACAAAT 3720

TGGATGTATG ATTATTTATG CCACAACGTA TGGCATTGAT TCGAATCCAG TCAATCAAGT 3780

TGCTACAATG ATTGAAAACC AAGAGTTTAA TTATTATTGT ACAGATGTTC AAGTTCGTGG 3840

TGAATATCCA GCATACGCAG AaCGCATGTA TCAAAAATAT GCTGTGAAAG ACTTAGTTAT 3900

TGAAGAAGGA GATTTAGAAT TATTAAAAGA ATACCCAGTG GATTACATTG GCTTTAGTTA 3960

TTATATGTCG ACTGCTGTCG ATGTTACTGG TACTACCAAC GATACAGCAA ATGGGAATCT 4020

TTTAGGCGGC GTTAAAAATC CCTTCTTAGA AGCAAGTGAA TGGGGTTGGC AAATTGATCC 4080

CGAAGGCTTA CGAATCGCCC TAAATGAATT ATATAATCGT TATCAAAAAC CTTTATTTAT 4140

TGTGGAAAAC GGCTTAGGTG CAATTGATAA AGTAGACGAA AATTTCTATG TGGCAGACGA 4200

CTATCGGATT GATTATTTAC GTCGTCATAT TGAAGCAATG GCAGAAGCAG TTGCAGACGG 4260

CGTCGATTTA ATGGGCTATA CGCCGTGGGG CTGTATCGAT TTAGTCAGTG CTTCCACAGG 4320

TGAAATGAGT AAACGGTATG GGTTTATTTA CGTTGATTTA GACGATGAAG GAAATGGTAC 4380

TAAAAACCGT TATGAAAAAA AATCGTTTAA TTGGTACAAA CAAGTTATTG AAACTAATGG 4440

ACAAAACTTA GACTAGTTAT TGTAATTAAA CAAAACAAGA TAAGCGTAGT AGGTTGTTGC 4500

TTCTGTTTCC GCTGTTTATT AATTTTTAAG CTTGGACCAA AAATCTAAAG TGATTTTTGG 4560

CCCAAGCTCT TTTTTGTCAA TTTTTCAAAA TAGTGCACAA TACGTTATAA TAGAAACAAC 4620

TGAGAGGAGC GTTTCTAATG ATTAAGCAAT TACAACATCC ATTTTTTGTT ATTATGGGTC 4680

CGAGTGGTTC TGGCAAGACC GCAATTACAA GTAAGGTTTT TCCGAAAAAC TATAAAGTTA 4740

TCTCGCATAC TACGAGGAAA AAACGATTGG ATGAAGTGGA TGGCGTGGAT TACTATTTTG 4800

AAACAAAAGC AAGTTTTCAA GCCTTGATTG AAACCAATCA GTTAGTGGAG TATGATTTTT 4860

ACCATGGAAA CTATTACGGT GTCGGTGTAG CTGCCATCGT TGAAACAACG AAAGAGCATC 4920

CAGCCTATAA TGCGTTAACT TTTCCAGGCT TTCAAGCGGT CTTTGAGCGT TTCGGCGAGT 4980

CCGTCATTCC AGTTTTTTTT GACGTTTCAA AAGAAAATAT TTATCAACGT TTAAAACAAC 5040

GAGAATCTGA TCCTAAAATT ATTGAAGAAC GTTTAAACCT TTATGATCAA GAAATTCTTA 5100

TCAAAAACCA ATTGGAACAA TATCCTAATT ATCAACGAAT AGATGCCAAT GGACCTATAA 5160

AAGAGGTCGC TGGCTTGTTA CAAGAATGCA TCAAACACTA TTATTAATCA AGCAACTGGT 5220

GAGACAGAAT CTAACATCAA TTTTGTCTTG CTAGGTGCTT GATTTTTTTT GTAAAAAACT 5280

TGCATTCTGA ATAGAGAGTG CTATAATGAC ATGGTGTCAT ATGACAATGT GTCATAAAAG 5340

GGAGGAACAC AAATATGCCT ACGAATACGT TTTTTCATTT GCCAGAAGAA AAGCAACAAC 5400

GCTTATTAGA TGCTGCGCAA ATTGAATTTT CGCGCCATTC TTTACAAGAA GCGTCAATTG 5460

CAAATATTGT TAAATTAGCG GGAATTCCGC GCGGTAGCTT TTATCAATAT TTTGAGAACA 5520

AAGAAGATTT GTATTTTTAT TACTTTGCGA CATTAAGAAA AAATAGTGAA CGAGATTTAG 5580

AGAAACAAAT TATTGCGGAA AATGGTGATT TAATCGAAGC CATGGATGTC TATTTCTCCA 5640

AAATGATTGT CGAAGTGCTA ACTGGaGAGA ATGCTTCTTT CTATCGGAAT TTATTTGTCA 5700

ACATGGACTA TCGTGCtTCA cGTCGGGTGA CGGGATnAAT TTTAGCCGAC TGGGGGAAGA 5760

GGGAAAAAAA ATAGGGAAAC CAGCCACTGT CCATnAAGCC TAGAnGGGAG GAAAAAGGnC 5820

CATGCCTGCC CCATGGCGGA 5840

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6590 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

CGATCTTTTT GACGACGTTC AATTTTACGT TTTTTCTTAT TGCTGTCTTC AATTGCCCAA 60

TTAATTTTTT TCTTGTATCC TGGTTTAATT TTTTTCTTTT TCTTTTTCAC TAAACCAATT 120

AACGTTGGGT CTAATTCTTC TCGAGATTTT TCTCGTTTTG TTCGACGATT ACGATCATAC 180

GTCTCTACAA TCTCACCATT TTTAATTTCT TTTGGTTTGA AAGAAACACC TAACTGCTCA 240

ATCTGAGTAA tCGCTTCATC ATCTgCTGGT GAATATAATG TTATGGCAGT CCCATTcAAT 300

CCATTACGTC CTGTGCGGCC AACCCGATGG ATAAAGAAGT CCAATTCATG TGGAACTTCT 360

GCATTGATGA CATGTGAAAC CCCTTCAATA TCAATCCCAC GTGCGGCTAA ATCAGTAGCT 420

ACTACATATT GGTAGTCAAG ATTTTGAACT TGACGCATCA CACGTTTTCT TTCGCGTGGT 480

GTAATGTCTC CATGAATTTT GGCAACTTTT AGGCCCTGAT CTTTTAAGTA ATCGGTGATT 540

TCATCAACTC TTTGTTTAGT ATTTGCAAAA ACAATCGCTA AATAAGGATG CCCGATAGTC 600

AATAATTGAT AAATTATTTT ATTGCTGTTT TTTCCTTTGG TGGAAATCAA CCAATTGTCA 660

ATTGTTTCTG AAATAACTGC TTTTGGTTTG ATATGTTCAA TCACTGGATT TTCTAAATAT 720

TTTTTTAAAA ATGGTCGTAA TTTTTCTGGA ATGGTTGCAG AAAAAACGAG CATTTGTAAT 780

TTTTCTGGCA AACGACCCGC AATTTGATCG ACTTCTGCTA AAAAGCCCAT ATCTAATGTC 840

ATGTCTGCTT CATCGACAAC AAAGGCAAAA GCCGTATGCA CTTTCAAGGC TTGTTCATTC 900

ATCATATCTA AAATACGCCC TGGTGTACCA ATCACCACAT GAGGTTGCTG ATGCTTTAAT 960

TTATTTAATT GACGTTGTTT ATCTGTTCCG CCAACAAAAT TTGACACTCG AATTTCTGkT 1020

TGACTGAAAC GAGCCAGTTG tGCGCTTCTT GATAAATTTG ATTGGCTAAT TCACGACTAG 1080

GCGCAGTAAT CACAATTTGT ACTTCGTCGA TAGTTGGCTT AACTTTATCC ATTAATGGAA 1140

GTAAAAAAGT ATGCGTCTTT CCACTTCCTG TTTGCGATTG TCCAATAACA CTTTTCCCTT 1200

TCTTAATGAT TGGGATTAAT TTTtCTTGGA CTTCTGTTGG TTCTTCAAAC CCTTTTTCTG 1260

CGAGTGCTTC GTTGATAAAA GGTTGAAATT GAAATTGTTT AAATGAAGGC ACTTGCCTCA 1320

CCTCTTCTCG TTAAATTCTT TATGCACTTT GGCATTATAA CACAAAAACA ACTTTTACTC 1380

ACCTGCTGAT TCCTTTGTTT TCGCAACTTC TAATGTACTG ACTGTATAAG TGTGGCTGGC 1440

TTTTTCATAA AACAAACGCC GTTTTTTACG AATCAAGGTT AACAAAACGT TGATCAAAAA 1500

TAGTAAATAT AAAATTAAAC AGAACAAATA CAGAAGCAGT GACACCATTT GCAACATCCT 1560

```
ATTAGAAGAA TTCAATAACT CAGTGGCATT GGCTAGATAA ATACTAATTA AACCATACAC    1620

AAAATACAAA TAACCGTAGC GAACAAACAA GGCTTTTAAA CGAATTTTTT CATGTCCTGT    1680

TTCAACCACC CGAATGCGAA CAACTTTTTT CCCTAACGTT TGACCATTTG TCAGTTTCAT    1740

CATGACGATA AAGACCGCAA ACACCATGAT AAAGAACCAC ATCCGATCTG AAAAAATCGG    1800

ATAATTGGTT GCCCAATTTT TATAAGCTGG AATGAAATTT AGTCCAAAAA CGAGCAGGCC    1860

TTGCACAAAG CCAATCACCA ACCAATCAAT TCCCCATGCA AAgAAaCGTC TTAACAGGCT    1920

AACTGTCTGC CCTTTTTCAT AAGAgGTGGC ATCAATTTCT TCTCTTGTTG GCAATAAAAA    1980

GGTAAAGAGC GGCGTCAACC AATAACCAAC TATGCCTCCG ACGGTATTAT TAATTAAATC    2040

ATTGACATCC GCCAAGCGAT ACGAACGTGG ATAAATAAAA TAAAGGCCAG ATAACTGTGT    2100

TAATTCAnAA AATAATGACA ATAAAAAACT CGCAAGAACT GTTTTTAAAA ATGAACATTT    2160

GAAATAATAC CGTAAATACA CACCAAAAGG TAAAACTAAT AAmACATTAA ACAGTGGTTC    2220

TAAAACGGCA TGCTGTTTTA ATGCTGGCAA ATACGTACTA GGATCTTTGA TCGAAAAAAC    2280

AGTTTCGTTT AAAATCCCTG AAATAAAATG AAAAGGCCGC AGTTCCATTG TTGGTCCTGT    2340

GTATTGAGCC ACTTCTGCTC GAGATGGCAA CGGTAAAATT ACCAAGAAAT AGGCACATAA    2400

TAAATAAAAA ACAAATGAAT ACAAAATCAT TGCTCGTGTC CAAACAAATG AACCATATTT    2460

GCGGTATTCA TAGATTAATA ATGCCCCAGA AATAAAAAAT GCTAGAAACG GGAAAATAAT    2520

GATTGCCCAT TTAATTGGTT CGCTATATAC TGCCATAAAA ATCCCTCCAA GATAAACAAT    2580

TTCAAGTATG AATGCCTCAA TAACACTATA CTTGTTTCTA AAAAAATGTA CCAACAATAA    2640

ATTTTATTTT GCAAATCCTC TCCTTTGCAG TATCATTAAG AAAAAGGAGG ACGTTCTCAT    2700

GTTACAGCCA TATTTTGCGT TTGGAGTGCC GCTATTTTTA CTTGTGCTCT ACTTACTCTT    2760

TGCACTGATT CACCGTCAAA CAACGATTCA TTATTTACGG TTTATTTTGT TACTTATTTC    2820

AACATTTCTC ATGGTTTTTA GTTTCCAAGT ATTGCAAGAA TCATGGACCA TCAACCCTGA    2880

AACGTTAAAA GATGCCGCTT ATTCTCCGCA ATGGTTATGG ATTCCTTTAG GTATCGGGTT    2940

AATCCTCACA CTTTACAACG CTTGGCACGG CCTTCGTACA ATGATAAAAT ACAAGACCGA    3000

TAAACATTAG TAAGAGTCTA GGACGATGGT TGTGTCCTAG ACTCTTACTG TCTCATCATT    3060

CAAATGAAAT TTAATTTTTC TTAACAAACT CTGATTTTAA CTTCATTGGA CCAAAGCCAT    3120

CTACCTTACA ATCGATGTTA TGATCACCTT CCACTAAACG AATGTTTTTC ACTTTTGTGC    3180

CTTGTTTAAT TGCGCCACTG GCACCTTTAA CTTTCAAATC TTTGATTACA GTAACACTGT    3240

CACCATCCGC TAAAAGATTG CCATTTGAAT CCTTAACTAC TAAGCCTTCT TCTGCAACTT    3300

CTTCTGTTGG TGACCATTCA TGACCGCATT CTGGACAAAT AAATAATCCG CGATCTTCAT    3360

AGGCATATGC CGACCCACAT TCAGGGCAGT TTGGTAATTG CTCTGTCATT AACGTTCTCC    3420

TTTGTAAAAA ATAATTCTTT TTCATTTTAC CTTATTTTCT AGTGAATGAA AAGTAGTCCT    3480

AATGATTTAT CAACGTCCGA TAACTACTTT CTAATTACTT ATTTATTTTT GAGCCCACTC    3540
```

TTGGCAAGCT GCTTGGATTT CTTGGTAAAT CACTTCAATT TCAGCCGCAT CACGTGCATT 3600

AGCGCCACTT GCCAAAGGAT GGCCGCCGCC ATGATGTTTT TTAGCAATTG TATTAATGAT 3660

TGGTCCTTTG GAACGTAACC GAACGCGATA ATAACCTTCC GGCTGTTCAA CAAAGATTGC 3720

CCATGAGAGA ACATCTTCAA TTTTTCCTGG CATTGAAACA ATGGCTGCCG TTTCAGAATC 3780

GACAATATCA TAACTGTTTA AAATACTTTG TGGTAAAACG ACACGAGCAG CGCCATTTGC 3840

GTCGATTTCT AAATTTTGAT AAATATAACC CGCTAATTTA GCAACCTTTA AAGGCATTTG 3900

GTCTAACTCA CGATTTAAGT CTGCGGCATT AAAGTTAAAC GTACGCAATT GCGCACGACA 3960

GCGAGtGTAT GCGCCGAaGT GGaAGGATAA AGAAAACGTC CTGTATCCCC CACAATCCCA 4020

GCATACAATA AACGAGCTGC ATTATCTGTC ATCGTTAATT CTGCTAAATG CTGTgCCAAA 4080

AATCAACAAT AATTTCACTA CAACTACTTG CATTTGTATT CACCCAAACT AAATCTCCAT 4140

AAGGTTCATC ATTCGGATGA TGATCGATTT TAATTAATTG ATCGCCTAAA GAAAAACGCG 4200

CATCACTAAT TCGTGGAGCA TTTGCAGTAT CTGTTACAAT GACTAATGCA CCTCGATAGA 4260

CATCGTCAGT AATAACGTCC ATTTCAGCTA AAAACTCTAA GCCTTCCACT GGTCCGCCTA 4320

CTTGATAaAT ATTTTTTyCA GGAAAGCTCG CTCTTAAGAG TTCTGCTAAC CCAACCTGTG 4380

AGCCAATTGC ATCTGGATCC GGCCTTTGAT GTCTATGAAT AATAATTGTT TCATATTGTT 4440

TAATTGCTGC CATAATTTCT TTTACTACGT CCATTCTTTG TAGCCTCCTA CGTTCTCTCC 4500

ATCACTTGGC AAACAACAAT TGCTTTTGCC ACAATGACAT TTTCTAAATA AACTTCAATA 4560

TCTAATTTtTG CAGAGCGACG GCCAATTTCT AAAATCCGTG GACGAATGTC CAATTCACTT 4620

TCCAACTGGA TTAGCCGCAA ATAATGTAGA TTCACTTGTT CAATTAGTAC ATTTCTTTTT 4680

TGATTCATCA ACATTGTTCG TTGTGTCACG TTAGAAATAA TTTCACTTAG GACACCGAAC 4740

GAGATGGTAC CAACACTGTT AACCATTTGG GGCGCAACCG AAAAACGAAA ACGGGTTGTT 4800

TCTTTGCCAT TTTCTTCAAC AGGCATCACT TCCCCGGAAA TTTGATCTGA AATTGTATCT 4860

GCAATTTGTG GTTGTCGTTG CACCAATTGC ATAGCTTTCA TCACATCTTG GCGAGAAACA 4920

ATCCCAACAA GCGCTAAATC ATCTTCCACC ACCGGCATGA CTTCTAGACC ATCCCAAATC 4980

ATTTGATGAC TGACACTCGC TACACTCATC ATCTTTTTCA CAACATTGGG ATCCTTTGTC 5040

ATGACTTTAT CCACGGTTAA ATGTTCACTT TTACCAATGA CATCTTTGGC TGTGACAATC 5100

CCCACTAAAC GTAATTTCTT ATTTACTACT GGGAAACGAG AATGATGGGT ACTTTCAGAT 5160

AATGCCTGAT AATCTGCAAT CGTCTGAGTT GTCAATAAAT AATTGGtTTT TTCAAGCGCT 5220

GTATAAATAT CACTGACGAG TAAAATGTCT TTTTTAATCA GCTGATCACT TAAAGCACGG 5280

TTAATCATAG TAGCCACAGT AAATGTATCA TATGTGGTTC TCAAGACAGG CATACCCAGA 5340

CGATCTGCTA AcTGCGCAAT CTCTTCAGAT GTTTCAAATC CACCTGTAAT TAAAACGGCT 5400

GCCCCATTTT CTAAAGCAAG CTTTTGCACA CcTTGGCGGT TACCAACAAT CATTAAGGAT 5460

CCGGGAGTAA TATATCGCGT CATCGCGGTT TCTGTCATCG CACCAATTAC AAATTTATTA 5520

AGGACTTTGT CTAGACCAGA GcTACCGCCT AACACATCCC CTTCAATAAT CCGGACCACT 5580

TCACCAAATG TTAATTTTTC AATATGTTTT TTTAACTTCC GCTCAATTCG TATCGTACCA 5640

ACACGTTGAA TGGTGGAAAC AAGCCCAATA TTTTCTGCAT CTTTGATGGC CCGATACGCT 5700

GTTCCTTCAC TAACACTTAA ACTTTTGGCA ATGCTTCGAA CAGAGATTCG ATCGCCAATT 5760

GGTAAGTCTT CAATATGTTT TAAAATTTGA TCATGTTTCG TCGCCATAAT TTCCTCCTAA 5820

ACTAGGATCG TCTCTCCTGG GTTTAATACT TTCCCGCCAA CTACTTGCTC TACGAATATT 5880

GCTGGATTTT GTTGGATAAC AGGAAAGGTA TTGTAATGAA TTGGAATGGT TATTTCGGGA 5940

TTTAAGTAAG CGATTGCTTC TAGCGCATCC TCTGGTCCCA TCGTATAATT ATCGCCAaTT 6000

GGTAAGAACG CTACATCAAT GGATTTGTCT TTTGCAAAAA GGCGCATATC TGAAAACAAC 6060

GCCGTATCAC CAGCGTGATA AATTTTTTtA TCTTCTGCTT CTAAAATAAT CCCACTTGCT 6120

TCGCCCATGT AGGTCATTAC GCCGTCTATT TCATAACCGG AACTATGCAA AGCTGGCACA 6180

AATTTAACCG AACCAAAAGG AAAGACATAA CGCCCACCTA GATTCATTCC ATGCGCTTTT 6240

ACACCTTGCG ACTGCGCATA CATAGCCACT TCAGCTATCG CAATAATCGT TGCCTTGTTG 6300

GCTTGAGCAA TTGCTAGCAT ATCGCCAATA TGATCGCTGT GTCCGTGGGT GATTAAGATG 6360

TAATCAGGGT GTAAATCTGC AAGTGAAACG TcCGCTAAAG GGTTGCCATT AATAAAAGGA 6420

TCAAATAAAA GGTTCGTGCC GTCGTTTAGA CGAATTTCAA TACATGAATG TCCATGTCCT 6480

ATCAGTTCCA TTTGTATTCC TCCTACGGTA TTTTCTACAC AATCTTATGA TAATTAATCG 6540

kCAAATTCAT yCTaAAAGTA GkaCaGTTaG TACAGTTACC TTCCTTAATG 6590

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7657 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

ATAAAGAGGT AAATGAAGGA GAAATGCGAT GAAAAAAAGA AAGGCAAGAT ATTTGAAGTT 60

GGTATGTATT GGAATTTTAG GGATTTATGC GGGCGCCAGT TACTACAATT CGGAACCATT 120

AATCGCTCAA GAAGAGACCG AAACGAGGGA CGAAAGTCGA CAGCTATCGC TGGAAGAAAA 180

AAGAGCGGAC CAAGCCAATC AAGACCAAGG AACCCTTACT TCTTCAACGC TGAACGAAAC 240

AATAATAAGT TCGTCTGAAC AACAAGAAAG TTCAACACAA AATGACAGCA GTGACCATCA 300

GGAAAATAAT GTGGAACCAT TAGAATCTGA AAATAGTTTA CCGCCACAAG AAAAGACAAT 360

TGCTAGTACT AAACCAACCG AATCGTTACA AAATGAGCAA CCAGTATCTG AAGAGCGTAA 420

AGCGCTCACT GAGAAGTCAC CTCATAATCA GGAACCACAA AATTCTACAT TAACTGAAAA 480

GACCAAAGAG CCGTTGGTAC ATTCAGAGGT AGACAATAGT CAAGAAAAGG AAAAGCAAGC 540

AGGAATTCAA CAACAAGCCG AGGCAAATTT TGTTGTCCAA GAGAATGCTT CTACTTCTGA 600

GTTTATTAGA AAAATCGGTG AAGAAGCACG AGTTATTGGT CAGCGACATG ATCTATACGC 660

TTCTGTAATG ATAGCTCAAG CTATTTTAGA GAGTGGCTCG GGCAATAGTG CGTTAGCGGC 720

GGCGCCAAAT TATAATTTGT TTGGCATTAA AGGAGCTTAC CATGGTCAAA GTGTCAGCTT 780

TCCCACACAA GAAGATGATG GTAAAGGGAA GATGACCACC ATTCATGCCG ATTTTCGTCA 840

GTACCCATCC TATAAAGAGT CATTAACAGA CTATAGTAAA TTAATCATCA ATGGGCTGGC 900

GGGTAATCCA ACGTTTTATC ACGGCGTTTT GAAAGCAAAC ACGACGAATT ATCAGCAGGC 960

AACAAAATTT TTAACAGGTC GTTATGCAAC GGATACATAT TACGATAAAA AATTAAACGC 1020

GTTGATTGAA ACCTATCAAT TAACTGAATA CGATCAAGAA AAGAAGAAAC CAGTAGTGAC 1080

TAATTTGGCA GAAGAAAAGA AACCTTCGTT TGATAAGGAA GCAGTCAAAG AACAGCTAAA 1140

AAAAGAAGCA GTCGTCTATG AAGTAAGCAA AGGAGATTCT CTCGCCACCA TTAGTCAGAC 1200

GTTTGGTGTG TCAGCAACAG CCATCTTAAA ACAAAATTCA AAGACTCAAG AAATGTTTTA 1260

TATTGGCCAA AAGATCACGA TTCCTCAACA TACAGCGGCG ACTTCTCTAG AACCGAAAGA 1320

ACAAGCGTTA CTTCAATCAC TAATAATCAG TAAATCGGTG ACTAATGCTT TGAAGACAGC 1380

AGAGCAATCA AATGGTAAAG CGAATGACTC AAAAACAGCG TCTACTGAGT ATTATGAAGT 1440

AAAAAGAGGA GAGACACTCG CAAAGATTGC TAAAAAAACG GGCTATTCCT TAACCGCCCT 1500

CAAGCAGGCA AATGACTTAA CGTATTCTGT ATTAACCGAA GaCAAACGAT AGCACTTCCT 1560

AAGCTAAAGG ACTAATGCTG ATTTAACTTA TCGGTCGATG GAGTCAATAT ATTATTAGgG 1620

TAAGAGAAGA GGAGAAGTTC ATCCATGGAT GAACTTCTCC TCTTCTCTTA CTATGTTTTT 1680

CATAAATACC TTAATAAATA GTAAAATTAG CAAGAGAATT GTGTCTTTAC ACAAAATAGA 1740

TAGGTTTTCT TGCCGAAGTT TGATAAAATG ATTAGTTGAA AGGATGCGTG AAAATTGAAA 1800

GAATTATTTC GTAAATACCG AGAAGTACTG GCTTATTTAT TTTTTGGTGG TGCAACGACG 1860

GTCGTCAACT TAGTTGTGTT TTTTGTTTGT CAAAACGTGT TAGGGCTGGA TTACAAGATA 1920

AGTAACGCAA TTAGTTGGTT TTTATCTGTT CTATTTGCCT TTTTCACGAA TAAATATTGG 1980

GTATTCGCAA GTAAGCATGA AAGCATAGCT GGTTTTTTTA AAGAAATGGG CTTATTTTAT 2040

TGGTACCGAA TTTTATCTTT TGTGGCTGAT ATGGGTTTAA TGATTCTCCT AATTGATGGC 2100

ATTCATTTTT CATCGTTTTG GGCGAAAATG ATTACACAAG TTGTAGTAGT GATTTTAAAT 2160

TATTTCTTCA GTAAATTTTn CATTTTTaAA GAAAAAGAGG tCTAAGCAGA TTTtAGGCTA 2220

AAAAAAATTT TTTTGTTATA ATATAAACAA GTTTTTAATA CAAAGAATAG AGGTAGAAAa 2280

CGATGGACTA TACATATGCA GATGATAGTT TGACATTGCA TACAGATATG TATCAAATCA 2340

ATATGATGCA AACATACTGG GAATTAGGGA GAGCAGATTT ACATGCTGTT TTTGAATGTT 2400

ACTTTAGAGA GATGCCCTTT AATCATGGct ACGCAATTTT TGCAGGTTTA GAACGTTTAG 2460

TCAATTATCT CGAAAATTTA ACGTTCACAG AATCTGATAT TGCATATCTC CGAGAAGTTG 2520

AAGAATATCC AGAAGATTTC TTGACGTATT TAGCAAATTT TGAATTTAAA TGTACTGTAC 2580

GCTCTGCTTT AGAAGGGGAT TTGGTTTTTA ATAACGAACC GTTAATTCAA ATTGAAGGTC 2640

CTTTGGCACA ATGTCAATTG GTAGAAACTG CCTTGCTAAA TATGGTGAAT TTCCAAACAT 2700

TAATTGCGAC AAAAGCGGCT CGAATTAAAT CCGTGATAGG CGATGATCCT TTATTGGAAT 2760

TTGGAACACG CCGGGCACAA GAATTAGATG CTGCCATTTG GGGCACAAGA GCCGCCTATA 2820

TTGGTGGGGC GGACGCGACA AGTAATGTTC GTGCTGGTAA AATTTTTGGT ATTCCGGTTA 2880

GTGGGACTCA CGCCCATTCA TTGGTTCAGT CGTATGGAAA TGATTATGAA GCATTCATGG 2940

CGTATGCCAA AACACATCGG GATTGTGTTT TCCtTGTTGA TACGTATGAT ACTTTGAAAG 3000

CGGGAGTGCC AAGCGCAATT CGAGTAGCTC GAGAAATGGG GGATAAAATT AATTTCCTAG 3060

GTGTGCGGAT TGATAGTGGC GATATGGCTT ATATTTCTAA ACGAGTTCGT GAACAACTGG 3120

ACGAAGCAGG CTTTACTGAA GCAAAAATTT ATGCATCGAA TGATTTAGAT GAAAATACGA 3180

TTTTGAATCT AAAAATGCAA AAATCAAAAA TTGATGTATG GGGCGTGGGT ACCAAGCTGA 3240

TTACCGCCTA TGATCAACCG GCTTTAGGAG CGGTGTTTAA ATTAGTTTCA ATTGAAGGTG 3300

AAGATGGTCA AATGAAAGAT ACCATTAAAC TATCTAGCAA TGCTGAAAAA GTGACAACAC 3360

CAGGAAAAAA ACAAGTCTGG CGCATTACAC GTAAATCTGA TAAAAAGTCA GAAGGGGACT 3420

ATGTGACACT TTGGAACGAA GATCCTCGCC AAGAAGAAGA AATTTATATG TTCCATCCCG 3480

TTCATACCTT TATTAATAAG TATGTCCGTG ATTTTGAAGC ACGGCCAGTA CTCCAAGATA 3540

TTTTCGTTGA AGGAAAACGT GTGTACGAAT TACCAACTTT AGATGAAATT AAACAATATG 3600

CAAAAGAAAA TTTAGATTCA CTTTGGGAAG AATACAAGCG GGACTTGAAT CCGCAGAAAT 3660

ATCCaGTGGA CCTTTCAACA GATTGTTGGA ACCACAAAAT GAATTTATTG GAAAAAGTGC 3720

GTAAAGACGT TAAACATTTA ACGGAAACGG TGAATAAGGA GGCCTAGAAA TGACAACACT 3780

TCAAGAAAAA ATCATTCAAG AGTTAGGTGT GTTACCGACA ATTGATCCAA AAGAAGAAGT 3840

TCGTAAAAGT ATTGATTTTC TGAAAGCTTA TTTAACTAAA CATCCTTTTT TAAAAACCTT 3900

TGTCTTGGGA ATTAGTGGCG GACAAGATTC AACATTAGCT GGGCGCTTAG CACAATTAGC 3960

GATGACTGAA ATGCGCGAAG AAACAGGAGA TATGTCGTAT CAATTTATTG CGATTCGCCT 4020

TCCTTACGGT GAGCAAGCAG ATGAAGCTGA TGCACAAGCC GCGCTAGCCT TTATCCAGCC 4080

AGATGTTTCT TTACGAGTGG ATATTAAACC AGCAGTGGAT GCAATGGTTG GTTCGCTAGA 4140

AAACGCTGGC GTTCAAATTA GCGATTTTAA TAAAGGGAAC ATGAAAGCAC GTCAACGTAT 4200

GATTACACAA TATGCAGTAG CTGGCGAAAA TGCTGGCGCA GTAATTGGCA CAGACCACGC 4260

AGCCGAAAAT GTGACAGCTT TCTTCACCAA ATATGGCGAT GGCGGAgcGG ATATTTTACC 4320

GTTATTTCGT TTAAATAAGC GCCAAGGAAA AGCTTTGCTG AAAGAACTTG GGCACCAGA 4380

AGCCtTGTAT CTAAAAATAC CAACAGCTGA TCTAGAAGAT GACAAACCGT TAGTTGCAGA 4440

TGAAgTTGCC TTAkGTGTCA CGTATGATGC CATTGATGAC TATTTGGAAG GCAAAAAAGT 4500

```
CTCGGAAACA GATCAACAAA CAATTGAAAA CTGGTATAAA AAAGGACAAC ACAAACGGCA   4560
TTTACCCATT ACAATTTTTG ATGATTTTTG GAAATAAAAA CACAGACCAT CGGCGAATGA   4620
ACGATGGTCT GTGTTTTTTT ACATAACTAC TTTTTAGAAA GTAAACGTTG CTACAGCAAC   4680
TATCCCATGA TAATGAAAGT CGTTTTCATT TAGGCGTTTT CACTTTCTTT TTCAAAACAG   4740
AAAAGGTACG CTCAATAATG TAAAGTGAAT GATGGGAGAG AAAAAGAGAT GAAGCATGTA   4800
ACAAAATtGG GGATTACAAT tATAACAGGA GTTTTGGCAT TATTATTTGA ATTTATTTTA   4860
CATCAGCCGA ATTGGGCGTA TGGCATTATT TTAATAACAG GTTCTGTAAT GGCGTTAATG   4920
ATGTTCTGGG AAATGATTCA AACCTTACGT GAAGGAAAAT ATGGTGTCGA TATTTTAGCG   4980
ATTACCGCTA TCGTTGCAAC CTTAGCTGTG GGAGAATACT GGGCCAGTTT GATGATTTTA   5040
ATTATGTTGA CTGGTGGTGA TTCATTAGAA GACTATGCCG CTGGAAAAGC TAACCAAGAG   5100
CTGAAGTCAT TATTGGATAA CTCGCCACAA AAAGCTCATC GCTTGAATGG CGAAAATTTA   5160
GAAGATGTTT CTGTTGAGGA AATCAATGTT GGCGATGAAT TAGTAGTAAA ACCAGGGGAA   5220
CTAGTTCCAG TTGATGGCTT GGTAAAAACC GGGACATCAA CAGTCGATGA ATCTTCATTA   5280
ACAGGAGAAT CAAAACCAAT TGAAAAAAAT CCTGGGGATG AATTAATGTC GGGTTCCGTG   5340
AATGGTGACG GCTCTTTGAA AATGGTTGCT GAAAAAACTG TAGCAGACAG TCAATATCAA   5400
ACAATTGTGA ACTTAGTGAA AGAATCTGCG GCGCGTCCAG CTCATTTTGT ACGTTTAGCA   5460
GATCGCTATG CGGTACCTTT TACACTAGTT GCCTACCTAA TTGCAGGTGT TGCTTGGTTT   5520
GTTTCAAAAA GTCCGACACG TTTTGCGGAA GTCTTAGTTG TTGcTTCGCC GTGTCCTTTA   5580
ATTCTATCTG CCCCAATTGC TTTAGTGGCA GGGATGGGTC GTTCAAGTCG TCATGGGGTC   5640
GTTATTAAAT CGGGAACGAT GGTCGAAAAA TTAGCTTCTG CAAAAACGAT TGCGTTTGAT   5700
AAAACAGGCA CGATTACGCA AGGACAACTT TCTGTTGATC AAGTCCAACC AATCAATGCT   5760
GGAATAACTG CTGCTGAATT AGTGGGATTG GCAGCAAGCG TGGAACAAGA ATCAAGTCAT   5820
ATTTTAGCTA GATCAATTGT TGCTTATGCC AGAAAGCAAG ATGTCCCATT AAAAAATATT   5880
ACAGATCTAG CGGAAGTTTC TGGTGCTGGC GTGAAGGCAT TTGTGGATGG TGCTGAGATA   5940
CGGGTAGGTA AAAAGAATTT TGTGACACAA GAGTCTCAAG AAACTGAAAA AATTGATAAA   6000
ACGACTATTC ATATTTCACG TAATGGCACA TATTTAGGCC GAATTACTTT TACAGACACT   6060
GTACGCCCAG AAGCAAAAGA GACTATGGAA AAATTACACC AATTACATCT TCAACGAATT   6120
TTAATGCTGA CGGGGGATCA AGAATCCGTT GCAGAAACGA TTGCTGCAGA AGTAGGAATT   6180
ACCGAAGTAC ATGGGGAATG TTTACCACAA GATAAATTAA CTATTCTAAA AGAATTGCCT   6240
AAAGAAAATC ATCCAGTCAT CATGGTAGGA GATGGTGTAA ATGATGCACC TTCGCTTGCT   6300
GCTGCAGACG TAGGTATTGC TATGGGTGCT CATGGAGCTA CTGCGGCTAG TGAAACTGCT   6360
GACGTTGTTA TTTTAAAAGA TGACTTAAGT AAAGTCAGCC AAGCGGTCGA AATTGCCCAA   6420
GATACCATGA AAATTGCCAA ACAATCTGTA TTAATCGGAA TTTTTATCTG CGTTTTACTA   6480
```

ATGTTAATTG CTAGTACCGG GATCATTCCG GCGCTAATCG GGGCTATGCT ACAAGAAGTC 6540

GTGGACACTG TGTCAATCTT ATCTGCTTTG CGTGCTCGTC GAATTGGCCA GTAAGGTGGA 6600

ACAATTAATT GCCTGAATGT CCAGCTTAGT GGTAAAwTAA AGGTACGTTT TTAATATTAA 6660

TAAGTTaCAG GGGAGAGATT GTGCAATGAA AGGAAACCAA ATAGATGGGC AACCAGTCAC 6720

GGAAGAACAA CTATTAAAAG AAGGCTACCG TAAATATACT GGGGAAGGAA TTGATATCTT 6780

TTATAATAAA GATATTTGTG AACATATTGG TAATTGTGTT CGTGGAAATC CAGAAGTCTT 6840

TGAAGTTGGT CGGAAACCAT GGATTATTCC AGATAACGGA AGTGTTGAAA ATGACATGAT 6900

TGTGGTAGAT TCATGTCCAA GTGGTGCTCT AAAATATATT CGGAAAGCAG GGAACGAATA 6960

TGCAAATTAA AGAAGAAAAC AATCGTTTTG CTCTATATAA CGATGAAAAC CAAGAAATTG 7020

GTGAAATGAC TTGGTCAGAT GCTGGCGAAA GCATGATGAT TATTGATCAT ACGTTTGTTG 7080

ATCCAACTTA TCGTGGACAA AAATTAGCTG AAAAACTTGT TTTAGCCGGT GTTGAAAAAG 7140

CGAGAAAAGA GCATAAAAAA ATTATTCCGC TATGTCCATT TGCCAAAAAA GAATTTGATA 7200

CTAAACCTGA ATACGGAGAC GTATTGAGAA AATAAAGAAG AAACTacAAC AGCCGCAAGT 7260

AGAAATTTGA TTCCTACTTG CGGCTGTTGC TATTAGAGAC AAATAAGGAA AGACATCAAA 7320

TTTCACTGGT GTTTACAGAT AAAATATGCT ACATTAAAGT AGATGAATTT CGCATATTTG 7380

TATAAAAAGG AGAATTAACG CATGTATAAT CTAATCTTAA CGtTGGTCAT CATTTTATCT 7440

GTTTTAATCG TCATTACTGT TATGAnGCAA CCAAGCAAAC AAAACAGTGC AGCAAGTGCC 7500

TTCACTGGCG GTGCAGATAA ATTATTTGGG AAGCAAAAAG CGCGCGGTTT TGAAGCTGTT 7560

ATGCAACGCT CAACAGCAGT TCTTGGCGCA ATTTGGATGA TTTTATTATT TGTTTTGGCG 7620

TTCTTATCTT CAAAATAAGT ACGTAATCTT TTAAAAT 7657

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7434 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

TGTCAATGAA ATTTTCTTGA CAGCCCTTTC TTATTTCCTG TTTTTCTTTC GTTTTTTTTA 60

TAATTGTTTT TCTGTTGAAG ATAGCTTAGC GCTTATTTAA TTAAATAACA AAATTTTTGT 120

GTGAAATAAC ATAGATAATT TAAAACTAAG ACATAATTTT ACTATTTGCG CCCAACTCTT 180

TATTTACTGA ATACGTGTTT TCTGAACATT TCGTACTACT GTCTTCCTTG TAAAAAAAGC 240

TGAGACAAAC TGACGAATCG GTTTGTCCCA GCTTTGCGTC CGAGTACATA TTGCTTAAAA 300

TCAAATGATT GAATTAATTA GACTTAGATA GCTGTTGtTG CTCCGCGGtA AACAATACCA 360

CGACGAGGaT CaACAGTAAT CACTTCGTCA TTATTAATTA ATGAkGTkGC ATCTGCAGCA 420

CCTACGATAA CTGGAATATT TkGCGCAATC GCtACaACAG CTGCATGAGA AGTTAAACCA 480

CCTTCTTCAA CAACTAATGC TGATGCTTTT TCAATTGCTG GCATGTATTC TTTATCTGTT 540
GTTTTTGTAA CTAAAATAGC TCCTTCAGTC GCTTTAGCAA CTGCTTCTTC AGCAGTTGCT 600
GCCACAACGG CTTTTGCAAT AATTGCTTCT TCGCCGACAC CTTGACCTTG AACTAATTTA 660
GAACCGATCA TTTGAATTTT CATTAAGTTT GTTGTCCCTT TTTCGCCAAC TGGAACACCA 720
GCAGTGATAA TGATTAAATC ACCTTCTGAA GCATAACCTT CTTCTTGTGA TACTTTTGAA 780
GCTAAATTAA ACATTtCATC TGTGCTTGAT GGkTTGTCAG CAACTGTTGC GTAAACACCC 840
CATGATAATG ACAAGCTACG TGCTTTTTGT TCTGAGAACG TAATTGCAAC GATATGTGCT 900
TnTGGACGGT ATTTAGAGAT CATACGTGCT GTATGGCCTG ATTCTGTTGC CGCAACGATT 960
GTTTGGATAC CCAAGTTACG TGCTGTATGG CCAACAGATT GTCCGATTGC TTCAGTCATA 1020
TCTGTTTTGC TATATAATTT CAATGCGAAT GAATCTTGGT TAACTAATGT TTCTTCTGTA 1080
CGAACTGCAA TACGAGCCAT TGTTTGAACA GCTTCTAATG GGTAATCCCC TGCTGCAGTT 1140
TCGCCTGATA ACATTACTGC ATCTGTTCCA TCGTAGATTG CGTTTGCTAC GTCATTTGCT 1200
TCCGCACGTG TTGGACGTGG GTTACGTTGC ATTGAATCTA ACATTTGTGT TGCAGTGATA 1260
ACTGGTTTAC CTAAAGCATT ACATTTTTTA ATTAAAGCTT TTTGGACAAC TGGAACATCT 1320
TCTGTTGGAA TTTCAACACC CATGTCACCA CGAGCAACCA TTAAACCATC AGAAACTTTT 1380
AAGATTTCAT CGATATTATC GATACCTTCT TGGTTTTCGA TTTTAGGAAT GATTTGGATG 1440
TGTGTTGCGT TTTCTTCTTC TAAAATTTTA GTGATTTCTA AAACGTCAGA AGCACGACGA 1500
ACGAAGCTAG CTGCGATAAA GTCAATTCCT tGACCAATTC CGAAACGGAT ATCGTTAGCA 1560
TCTTTTTCAG TGATCCCTGG TAAGTTTACA GAAACGCCTG GTACGTTAAC GCCTTTTTTG 1620
TTTTTCAAGA CGCCTCGTTT TTAACAACTG TTACGATTTC GTTTGCATCA CGGTCGATAT 1680
CAGTTACTTC TAAATCGATT AAACCATCAT CTAAAAGAAT GTGTGAACtA CGTTTACATC 1740
ATTGATTAAT TCTGGATAAG TGATTGAGAA TTTTTCATTT GTTCCTAAAA CTTCTGTCAT 1800
AGAAATACGT ACTGAGTCAC CAATTTTCAT TGTGATTGCG CCATTTTCCA TGTCATTTGT 1860
GCGGATTTCA GGACCTTTTG TATCTAGTAA GATCGCTACG CGTTTCCCTG TAATTTTTAC 1920
AGCTTCACGA ATGTTTTTAA TACGAGCACC ATGCTCTTCG TAGTCACCAT GTGAAAAGTT 1980
CAAACGGCAA ACATTCATAC CTGCATTAAT TAAGTTTACT AGCATATCAA CGGATTCACT 2040
AGCTGGTCCA ATCGTACATA CGATTTTCGt TTTTTTCATT ACTAAAACGC TCCTATTCCA 2100
AATTTGTTTA TACAAAGCTA ATAAGATTAG CTAATATACA AGATTAAAAA GAAATTTCAT 2160
GATTTAACTC ATAAAGAGAT AAATCTGGTT TATGTTTATT GTTTTCTAAC GTGTCAATAA 2220
TATCAGCTGC AACTACTTGG TTATCTAACA TACCGATACA TAAACCGCCT TTGCCTTCTT 2280
TTAGAAGTTC AACAGCATAT GACCCAAATT TACTTGCTAA TACACGGTCA CGAGCACTTG 2340
GCGAGCCCCC ACGAACCACG TGTCCTAAAA TTGATACACG CGTATGGAAA TCACCGTATT 2400
CAGAAAGTTT ATCAGCAAAT TCATTTCCGC CCATTACGCC TTCTGCTAAA ATGATTAAGC 2460

AATGTTTTTT CCCGCGGTCG CGGCCTTCTT GAATACGTTT GGCAACGTTT TTCATATCAA 2520

AATCATGTTC AGGAATAATG ATTTCATCTG CTCCGCCTGC CACGCCTGAC CACAATGCGA 2580

TATCACCAGC GTTACGTCCC ATTACTTcGA TAACGAATGT ACGAACATGA GAAGTCGCTG 2640

TATCACGAAT TCGGTCAATT GATTCCAAAA CAGTGTTAAT CGCTGTATCA AAACCGATTG 2700

TGAAATCTGT TCCAGGAATA TCATTATCAA TCGTTCCAGG AATTCCAACA GCGGGGAAAC 2760

CACGTTTTGT CAAAGCCATT GCGCCATGAT AAGAACCGTC TCCACCAATA ACAACTAAGC 2820

CTTCAATACC GAATTTTTTC AATTGTTCGA TTCCTTTTAA TTGCCCTTCT TCAGTAGCAA 2880

ATTCAGGATA ACGAGCAGAA TACAAGAATG TTCCACCACG TTGAATCTTA TCGCCAACGT 2940

CAGCAACATC TAAACGACGG ATGTCGCCTG CGACTAAACC GGCAAATCCA TAGTTAATAC 3000

CGTAAACTTC GATTCCATCA AAAATACTTT TGCGTACGAC TGCACGAATT GCAGCGTTCA 3060

TTCCAGGGGC ATCTCCCCCG CTGGTTAAAA TACCAATGCG TTTCATTCGC TATTTCACCT 3120

CATATAATAT ACTTTCAACT GCCAAAAGAA ACCTTTTCGC AAAAATTGAT ATACAAAAAT 3180

ATTACGTCAT TCATTTTACC ATTAAAGAAC CAAAATGTCT TGAGTTAAGA AGAAAAAAAT 3240

GAAAAAAGTT GAAAACAACT AGTATTTTTT ACCGTAAATC ATTTGAAAAT GACGTTTTGC 3300

TCGCCTAAAA CGTAAGCTAA TTGTTGCTGA CTAGCGCTTG TATGAGCGAC CCAATTTTCT 3360

TCATTTAATA CTATTTTCTT ACCTGTTTTT TCAAAATAGA CAATCACTGG AATGTGTCCA 3420

GCGTGACGAC TGAGTACCTC TTTCATTTTC TGAAAAACGT CCGTTTGTTC GACATCTTTT 3480

GTAATTCTGA TAAAACACGT TTCGTCACTA ATTGATGTTT CTACTTGTTG CGCTTTTTCT 3540

ATTCGTTGAA CCAATATTTG CAATTCTTGA TTGTACGTAC TGCGTTCCAC TTTTCCTTCC 3600

ACATAATAAA CTTGCCCTTT CTCGACACTC TGTCTTAACT GGCGAAAAAC GGTTGGGaAC 3660

AACGtTAGTG AAATACTTCC TGTTGtATCA TCCCCATCAA CAAAGGCCAT TTGTTCTCCC 3720

TTTTTGGtCC GAATCGTCCG AATGTTTTTG ACATAAAyCA ATAATTTCgT GGGcTGATTT 3780

TCAACAaGTT CATGAACAAG TTGCGTTTGC TTGGCTAAGC GTGTTTTTTT GAACTCTTCT 3840

GTCGGATGTC CTGATAAATA GACCCCTAAA TATTGTTCTT CTTGTTCTAG TCGCTCTTCT 3900

AGCGGATAAT CGGCAATTTC TTCTTCCTTG GGCTTCAACG TTTCTAACAA ATCCATACTA 3960

CCACCGCTGA GAAGAATACT TTGGATATTA CTCTCTAAAC CTACCGCTAA TTGACGGCGA 4020

TTAGGCGCAA TTTCGTCAAA TGCACCGATG GCAATCAATG GTTCGATATT TTCAACTTTC 4080

AACCACTTAC GGTCAATTCG CAAAAGAAAC TCATCGAACG TTTTGTACGG TCCATTTTCT 4140

CGGCGTTCAT TTAAAATATT TTGAATGAAG TCTCGGCGGA TTCCTTTTAA AGAACTAAAA 4200

CCAAATAAAA TTTGCTGTTG CTGATTTAGG TAAAAGCTGT AGTGACTTTG GTTAATTGAC 4260

GGCTGCAAAA TCGTCTGCTT GTTTTTTCTA GcTTCGCCAA TATATTCTTT AATTTTTGTT 4320

GGATTATGAC GAACAGAATG CAACAATGCG ACATAAAAAG CACTCGGATA ATGAACCTTC 4380

AAGTAAGCCA TTTGAAAACC AATAAAGGAA TAGGCAAATG CATGCGAGCG GTTAAAGCCA 4440

TAATTAGCAA AGCGCTCTAT ATAATCGTAC ACTTGGTTAG CTGTTTCTTG GGGATAGCCT 4500

TGTTGTAACG CACCATTAAC AAAATGATTT CGTTCTTCAT CTAGCACATC TTTTTTCTTT 4560

TTACTAATGG CCCGACGTAA AATGTCAGCT TGCCCTAAAG AGAAGCCTGC CATTTTCGCA 4620

GCAACTTGCA TAATCTGCTC TTGGTAAACA ATAATTCCGT ACGTATTTTC CAAAATGGGG 4680

GCGAGACTAG GCTCTGGATA ACTAATTTGT TCCAAGCCTT TTTTCCGGCG GATGaAAACA 4740

TCAATATTTT GCATCGGACC AGGACGATAC AAGGCATTGA CAGCAGCAAT ATCTTCAATG 4800

CTAGTTGGCC CTAATTTTCT TAATACATTC CGAATTCCAG CAGATTCAAA CTGGAAAACA 4860

CCACTTGTTT CCCCTTTTCT AAATAAAGCC AGCGTTGTTT CGTCATCTAA TGGAATCTGA 4920

TTTAAACGAA TGGTTCGATT ATAGACGCGT TTAACAGCTG TGAGGGTATC ATCawkGATG 4980

GATAAATTTC tTAAGCCCAa GaAATCCATT TTCAGAAGAC CAATTTTTTC AACATCATTC 5040

ATAGTAAATT GGGTCAATAA AATTTCATTC GATCCTGGTT GTAACGGAAC CAAATTCAAA 5100

AGATTTTCAT CaCTAATTAC CACACCTGCT GCGTGCGTAG AAACATGGCG AGGCAATCCT 5160

TCTAGCTGAA CAGCTGTTTT ATAGAGCAAT TGATTATTTG GTGAAAAATT GACCAATTCT 5220

TGCATCCgTT TTGACTCTTG GTATGCTTCT TCTAACGTAA TTTTCAGCTT GTTGGGGACA 5280

GCTGCCGACC AACGATTCGC TTCACTCTGG GACAACCCGA AAACACGTGC AAcGTCCCGT 5340

AAAACCATTT TCGCAGCCAT CGTACCAAAC GTGGCaATTy sCGCCATGTG ATAATGACCA 5400

TATTTTTGAC TGACATAAGC TAAAACTTCC TCCCGCCGAT TATCAGGAAT ATCTAAATCA 5460

ATATCGGGCA TCGTGTAGCG TTCCGGATTC AAAAAACGTT CAAAGAGCAA ATCATATTTG 5520

ATTGGATCTA CATCTGTAAT TGATAGCACA TAGGACACTA AAGAGCCTGC AGCTGAACCT 5580

CGGCCAGCGC CTGTAACAAT TTGATTGTTA TGACAGAAGT CCATCACATC CCAAACGATT 5640

AAAAAGTAAT CATCAAAGCC CATATTGTGA ATGATGGTTA ACTCTTTTTC TAATCGTTTT 5700

TGGTAAACTT CATTTAGATG AGGCACTCTT TCTGGTAATT TTTCTAAACA CAATGTGCGT 5760

AAATAAGAAC CAGCCGTTTG ATTATCAGGC ACAGGAAAAT GCGGTAATAA TTTTTGATGT 5820

AAAGGAATGT CAACTCGGCA ATTTTCAGCT AGCTGAATCG TATTCTGGAC CACTTCTGCA 5880

CCAGCTATTT GAAGCATTTG TTCCATCGTC GTCTCTGCTT CACGCAAGTA ATTCAGACCT 5940

TCTATTTCCG GTGTTTGGCT TGTAATATTC AATTGATTAT TTTCTTTGAT GCTCTCAAGA 6000

ACAGTTACGG CAAAGCCTTC TTCCCTTGTT AAATGACGTA CCTCTTGCAA AGCCACGCCA 6060

GGTAATTGTT CTCTTTGGTA AAAAGAAAAG AGCGCTGGAT TACGTTGTTG ATTCATTGAA 6120

AAAGAATGTT CAACAAATAA CGACTGGGGA TCAAAATATG AAGCAAGTTG TTGGAAAGCG 6180

TGAACCGCTT GTTGTTCATC TTCAAAAAAT GTTTGGGCGA CTTCTCCTTT AGCTGACAAA 6240

GTCACGGCAA ATAAATCACT TAATAAGGAT TGATGTTCCG CTAAGTAAAA GGGACCCTCT 6300

GTCGTCATTT TGCTACTAGA TAGTTGCATT AATTGCTGAT ACCCTTTTAA ATTTTTAGCA 6360

AATAAAAGCA ACTCGGAATG CCCTTCTTTT TCTGCTGGGG TATATTCTAA TCTTAAACCG 6420

ATAATTGGTT GAATTCCCGC TTCTTGGCAT GCTTCATAAA ATTCAACCGC CCCATGCAAG 6480

ACATTGATAT CGGTAATCGC TAATGTTTGA TAACCTAACg CTTTCGCTTG TTTTACATAG 6540

GAACGAATCC TAACCGTACT AGATAAAAGC GAATAGGAAG TAACGGTATG CAATTGTGGA 6600

AATGACATTG GACAGCTCCT TTTTGAATCC TTGCTATTTT TTCCTCTTCT CCTATTTTCC 6660

TATCGCGGAT TCTTTACTTT TAAAGAAAAT ATGCTAAACT AAGAGCTAGA AAAGAGGTGC 6720

TCATATGCGT TATCTTGTTG TATTTTTCTG GACCTTCGCA TTAGGCCAAG TCGTTGGATA 6780

TATCGGCGGT GCATTATCTG GTGGTTCATA CGACTTTAAA ATGACAACGA TTCTATCGTT 6840

GGCAACAGCC GTTATTATTT TATTAATTGC aCACTTTGCA GTACCGAAAG AAACCAAACC 6900

TGTTAAATAA GAAAAAACCA GTTACTAAAG TTATTGTAAC TGGTTTTTTT ATCTTTTCAA 6960

CTGTTTAGAG CTTTTTATTT AAAACTTTTC GGAAATTAAA GAAGGACAGC CCAATGTTTC 7020

CTAAAACGAA ACAAGCGGTT ACCAGAAAAA CAGAAGAATA GCCAAAACCA TGGGCAACTG 7080

TCGACCCTAC CATCGGTCCT AAAACTTGAC CGAAATTACT ACACATCTGA TTGTAACTAT 7140

AAACTCTACT TACGCCCTCT GTTGGCGTTA TTTTACTAAT TAACGTATTA ATTGATGGCA 7200

TTAAAGCACC TGTTGAAAAA CCTAAAAGGA ATCGTAAAAC ACCTAACTGA AAAGGTGTTT 7260

GCACAAACGC CATGGGAATG TAACAAATAA AAGAAAGAAT TAATCCGCCT AATAAAACTT 7320

TTTGGTTCCC AATGCGGTCC CCAATTTTTC CCAAAGTTGG AGAAGAAATA ATTnnTGAAA 7380

CGCCGGCAAT CGAAACAATC AAACCACTAA CAAATAATAC ATTCTCTGTA TCGC 7434

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2732 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

TCGGAACAAG CTTTTAAGTG AAATTGAGAT TTTATATAGT GAAGAATTTG CTATTGCTCA 60

ATGGGCTGTA GAATATTTAA CACAAACCTT AGAGATTCCA TTTAGTTATG ATGAAGCGGG 120

GTATATTGCG ATTCATATCC ATAGTGCTCG CAGCGGGCGT ACTGATAATA GTAAAAGTAT 180

CCGTGAAGTT ACAATCGTTT CTGAAATTAT TCATTTAATC GAGCAGGAAT TGGCTATTGA 240

TATTCATGAT GATAAAAATA GTCTCAGTTA TTCACGTTTG GTGAATCATT TACGTTTGTT 300

TATTCATCGC TTCCAACAAA ATCAATACGC TGTTTTAGAT GAAGAAATTT TGGAAGTCGT 360

AAAGAAAAAG TATGCTGAAA GCTATGAAAT CTCAAAAAAA GTACAAGTAT TATTAATGAG 420

AAATTTTCAT TATCAAGTAC CAAATGAAGA ACTTGGTTAT TTATCGATTC ACATTGAACG 480

ATTAAGAATG ACCAAATAAT AAAAATTAAA TAGTTTATGT GAGGAGTTAC GTATAATGAA 540

AGCGTATATG CAAAGAATGG GACGTTCGTT AATGTTACCA GTTGCTGTAT TACCGGCGGC 600

TTCTTTATTA GTCGGTATTG CTAACTGGAT CGTGGGAACA ATCGGTGCTA GTCCAGCCAC 660

AACTTTTCTA ATGAATGGTG GTTTGGCTAT CTTAAACAAC TTAGCATTAT TATTTGCGGT 720

TGGTTTAGCG TTAGGAATGT CAAAAGATAA AGATGGATCG GCAGCATTAG CTGGTTTAGT 780

CGCTTATTTA GTGCCAAAAA CTGTTTTAGC GCCTGCTTCT ATCCAAGCCA TTAAAGGCTT 840

CAAGGATATT GCTGAAGTAA ACCCAGCCTT CAACAGCATG GACAACAATG TCTTTGTCGG 900

GATTGTCGCT GGTTTAGTTG CCGCGGCAAT GTATAATCGT TTCTCTGGTG TGAAATTGCC 960

AATGGCTTTA TCATTCTTTA GTGGAAAACG TTTAGTTCCT ATTATGTCAG CTATTTCGAT 1020

GTTAGCAATC TCAGCTGTCT TATTCTTCTT CTGGCCAGTT GTTTATAACG GATTAGTTGC 1080

ATTTGGTAAA GGAATCTCTA GCTTAGGTTT CGTAGGAGCT GGCTTGTATG GTTTCTTTAA 1140

CCGACTATTA ATTCCAACAG GATTACACCA TGCCTTGAAT TCTGTTTTCT GGTTTGATGT 1200

TGCTGGTATC AATGATATTG GGAACTTTTT AGCTGGTCAA CAAGCGTTAG ATACTGGTAA 1260

AGCAATTGTA GGACAAACAG GGATGTACCA AGCTGGTTTC TTCCCAGTAA TGATGTTCGG 1320

TTTACCAGCA GGTGCTTTTG CTATTTATCA ATGTGCACGT CCTGAGAAGA AAAAAGTTAC 1380

AGCTTCATTA ATGTTAGCTG CTGGTTTTGC GGCTTTCTTT ACAGGCGTTA CTGAACCTTT 1440

AGAATTCTCA TTTATGTTCG TTGCTTGGCC TTTATATGTT TTACATGCGG TCTTTACAGG 1500

TATTTCTTTA GCATTCGCTG CGTTTATGCA TTGGACAGCT GGCTTTGCCT TTAGTGCAGG 1560

TTTTGTAGAC TTCTTCTTAA GTTTGAAAAA CCCAGTCGCG AATCATCCAA TGATGTTGGT 1620

GGTTCAAGGA CTTGTCTTTG CAGCAATTTA CTACTTCGGT TTCCGTTTTG CAATCACTAA 1680

GTTTAACTTA ATGACACCAG GACGTGAAGA AGGCGATGGC GAAGAAACGC CAGATGTTGC 1740

AGAAGGTGAC AACAAATTTG CTTCTCTAGC TCGTCGTATT TATGATGGTT TAGGCGCTGA 1800

TGCTAATGTA ACTTCTATTG ATAACTGTAC AACTCGTTTA CGCTTAACAG TGAAAGATAC 1860

AGGCAAAGTG GATCAAGCAA AAATTAAAGC AACTGGGGTT CCTGGTGTGA AAGTCATTGA 1920

TGACACCAAT ATCCAAGTGA TCGTTGGTAC GGAAGTACAA TTTGTTGCTG ACGAAATGCA 1980

ACGATTATAT AACCACCAAG CGCCAGCAAC ACCAGTCAAA GAAACACCTG TTTCTCAATC 2040

AGTTGTTGAA GAAAAAGCAC CAGTTTCAAC GAAAGAAACT GAATTATATT CTGTAGCGAA 2100

TGGGAAAGTT ATTCCTATTT CTGAAGTACC AGATGATGTT TTCTCAGCAA AAATGATGGG 2160

TGACGGTTTT GCTGTCGTAC CAACAGACGG CGAAGTTTCA ACACCCGTTG CTGGTAAAAT 2220

TACTAGCATT TTCCCAACAA AACATGCGTT AGGTATCCAA ACAGATTCAG GCATTGAAGT 2280

ATTATTACAC ATGGGTCTAG ATACTGTTGA GTTACAAGGT GGACCATTTA CATTACATGT 2340

TGAAGAAGGC CAAGTTGTAA AACAAGGCGA TAAAATTGCA ACCATTGATT TAGCGGCATT 2400

AGAACAAGCT GGCAAAAAAT CAGATTTAAT TGTTGTTTTC ACAAATCAAG ATATTGTTGC 2460

TCAATACGAT TTACAAAAAG CAGGTCAAAC GACTTCAATG AATGATGTAA TCGGAAATGT 2520

TACCGTAAAA TAAAAGTAAA AAAGGGCTAG CCGTAAGGCA GCCCTTTTTT TGATGAAATT 2580

CGATAAAATT ACTTGTTATT CTATTCTAGG TGAAAACGTT CAATCTTGCG AAGTTACGGT 2640

TTGTTCACAA AAAAGTTACT TTTATTAAAT AAAACTAAAG AAGTAACAAA AACTTCGCnT 2700

TTTTCAGTTA TTTCAAGTAA CATAGAACGT AT 2732

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1150 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

CCAATGGCTG GATGGAACAG AACAAGACGT GGTATCACGA TTACATCCGT GCGACAACTG 60

CACAGTGGAA GGCTACCGTG TAATATCATC GATACTCCAG GACACGTGAC TTCACAAnTG 120

AGTACACGTT CTCTACGCGT ATTAGATGGT GCGGTAACAG TTCTTGACTC TCAATCAGGT 180

GTAGAaCCTC AAACAGAAAC TGTTTGGcGT CAAGCGACTG AATATAAAGT TCCTCGTATT 240

GTATTCTGTA ACAAAATGGA TAAAATCGGT GCGGACTTCT TCTATTCAGT AGAATCATTA 300

CATGATCGTT TACAAGCAAA CGCTCATCCA ATTCAAATTC CAATCGGAGC TGAAGAAGAC 360

TTCACTGGAA TCATTGACTT GATTAAAATG AAAGCTGAAA TCTACACAAA TGATTTGGGA 420

ACAGATATTC AAGAAACTGA TATTCCTGAA GATTATTTAG AAAAAGCGCA AGAATGGCGC 480

GAAAAATTAG TTGAAGCTGT TGCTGAAACT GACGAAGACT TAATGATGAA ATATCTTGAA 540

GGTGAAGAAA TCACTGAAGA AGAATTAGTC GCTGGTATCC GTCAAGCAAC AATCAACGTT 600

GAGTTCTTCC CTGTATTAGC TGGTTCAGCA TTCAAGAATA AAGGGGTTCA ATTAATGTTG 660

GACGCTGTTC TTGATTACTT ACCATCTCCA TTAGATATCG ATGCTATCAA AGGGATCGAT 720

ACTAAAACAG ACGAAGAAAC AACTCGTCCA GCTGATGACG AAGCACCTTT CGCTTCATTA 780

GCATTTAAAG TTATGACTGA CCCATTCGTA GGTCGTTTAA CTTTCTTCCG TGTTTATTCT 840

GGTGTCCTTG AAAGTGGTTC ATACGTATTG AACGCTTCAA AAGGCAAAAA AGAACGTATC 900

GGTCGTATCC TACAAATGCA CGCCAACACT CGTCAAGAAA TTGACAAAGT GTACTCTGGT 960

GATATCGCTG CTGCTGTTGG TCTAAAAGAC ACAACAACTG GTGATACATT ATGTGCATTA 1020

GATGCACCAG TTATTCTTGA ATCtATTGAA TTCCCTGACC CAGTTATCCA AGTCGCTGTT 1080

GAACCTAAAT CAAAAGCTGA CCAAGATAAA ATGGGCGTGG CTCTACAAAA ACTTGCTGAA 1140

GAAGATCCAT 1150

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6729 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GAGCCTAATA AAGAAGTAAA AGCCTTCTTA GACTATATGG TCACCGATGA TGTTCAAAAA 60

ACAATTGTCA AAGACTTAGG TTATTTAGCC ATCACAGACA TGCAAGTAGA ACGTGATGTA 120

AATGGAAAAG TAACTGAAAA ATAAACAAAA AAACTCAGCC AATAAACAGG ATTGGCTGAG 180

TTTTTTTAGG CTTCTTTTTT TGTTAAAGGC AAGGTGrTTG TAAAAGTGGT CCCTAAGCCT 240

AAGTGACTAT CAATGGCGAT TTTACCGCCT AATAATTCTG TGTAATTTTG GACAATAGCT 300

AGACCTAGAC CGGTGCCACC AGAATGACGG AnACGTGCTT TATCGACACG ATAGAACCGT 360

TCAAAGATCC GTTCTTGGTC TTTTTGACTA ATGCCAATGC CAGTATCTTG AACAGATAAA 420

CGTAATTGCT GATTTTCTAT AGCATAAGAA ATAGTAATCT GACCATTTTC TTGGGAATAC 480

TGAATGGCAT TTTCAATTAG ATTCTTCACG ATTGGATAAA ACAATTCGTA TTTGGTGAAA 540

AACTGGCTCT CTGGCAGACC ATGAACGACT GTTTTCAGCT GTTTTGCTGC CAATTGTTGC 600

TGATAAGAAC CGAGAATTTC TGTAATAAAT TTTTCTAAAG TAATTTCTTG TTCTGCATAA 660

GGAATTGTCG CGCTGCCCCG AGAAAGTTCT AAAATTTCTT GAATTAATTG TTGCAGACGC 720

TGGGCATCTT TTTGCATGAT GTGCAAAAAT TGATTAAGGA CTTCCGGATC TTCTTTGGCA 780

CCGTCTATGA GTGTCTCGGT AAAGCCCAAA AGAGAAGTGA CAGGCGTTTT TAATTCATGa 840

GACACGTTTG ATACAAAATC TTTTTGCATT TTTTCTAGTT GGCGAACGCG CGTTAAAtCA 900

TAGGCAATTC CTAACACTTG ATAAtCATTG GcGCTTTGAT CATCCAAGTA ACGAAGAGAA 960

AGATCTAATC GTAAAGGTGT TGGTCCTTGT GTCACAGTGA TTTCTTGATG TAAACTGTTT 1020

TTTTCAGCGA CGACTTGATG GATTAAATGA ATCAAATCGG CGTCTTGAAT CACTTCGAGA 1080

TAATTGCTTT GCATATTTTT AGCAGAGAGT CCTAACATTT CCTGCATTTT AGGGTTCAAC 1140

AGACGAAGCC GTCCTTGCGT ATCAATGATA AAGACACCAA TCATCAACTC ATCTAAAAGT 1200

TCGTGAAATT GCTCATCCGC AGACGTATAC GCTTTATACG TTTGACTTAC TTGTTGaCTC 1260

AGTAGGTTAA TAGTTTCGTA AAGTTCATTC CATTCAGAAG ATTGCTGGaT AATGGATCGA 1320

TTTTCATCTG GATGTTTAAG CATTCGTTTC AGAATCGGTA AGACAGTCAC TAGCGGTTGA 1380

TTTTTTTGCC GTAACAATAA AAAGATAAAC AAGGTAATTA GTAGGAACAA GATGCCAAAT 1440

GTAAAGATAA TATAGCGACG AAAAGATTGA ATATTTTTAG AAAATAGCGC GGTTTGTTCC 1500

GAAATCCGAA TAATGCCAAC GATTTCTCCT GATTTTAAAA CAGGCGTGGC AACGTACAGA 1560

AGTTCTTTAT TTAACGTTTT ACTCTTTCGT AAAGCTGAAC CATAGGTAGC TCCTGATAGA 1620

ACCGCAGCGA TTTCAGGTCG TTTTTTGTGC GACTCTAGCG GTTCATTTTG GCTGCTGTCA 1680

AAAAAGACGG TACCATTTTT TTGCAACAAA GTTACCCGTT CTTCTGCATC ATCTAACGCA 1740

tGGGTGaCAA GGAGTTGTTG TTGGTTTGTC AGTTGCCCTT TTTCGAAAAC GGAAGGGGGC 1800

AACTGGTCTG TTAAAAGGGT ATTTTTTCTT CTTAAATAGG TTTCTTGCTG AGCAACCATT 1860

TCTTTTTGAA AGAAAAAATT GGTCGCTGCA ATACTCCCGA CAAATAATAA AAGCATGACT 1920

GCCGCAACTA AAAAATATTC AATCCGCAGT CTCTTTTTCA TCGTTTTGGC TCCTGAAAAC 1980

GGTAGCCAAA GCCACGGACA GTTACTAAGT AAACGGGCCG TTTGGGATCG GGCTCGATTT 2040

TATCACGTAA ATGACTAATG TGGACATCGA CGATGCGACT TTGTCCAGCG AAGTCATATT 2100

GCCAAATTCG TTCCAACAAG GTTTCCCGAT TAATGACCCG ATCTTTGCGT TTCATGAAGT 2160

AAACGAGTAA TTCAAATTCT TTCGGTGTTA GCTCAATGGG CTGATTGCGC ACAAAGACTT 2220

CATAATTTTG TTCATCCACA CGAATCTCGC CAATCACAAG CGGAgCTTTA GGTGTGTCCT 2280

CTTGAAGCGT TTCGGTCGTG GTAGGTTTTA AACGACGGAA AATAGCTTTC ATGCGTGCTA 2340

GGACTTCTCT GGGACTAAAA GGTTTTGTTA AATAATCATC TGCCCCAATT TCTAAGCCGA 2400

TGATTTTATC TACTTGTTCA TCTTTGGCAG TTAAAATCAA AATAGGCGTG TCAATTTTTT 2460

CTCGACGAAG TGcTTTGGTA ATTTCTAAGC CATCCATACC AGGAAGCATT ACATCTAAAA 2520

TTATAAAATC ATATTGATTA GACAAAGCTA ATTCGAAGCC GTTTTTACCA TCTTCGGAAG 2580

TAGTCACCTG ATAGCCTTCT TTTTCCAAAT TAAAGGTTAA CAAAGTTAAA ATCGATGGTT 2640

CATCATCGAC GACAAGAACT TTTTTCATTT ACATAACTCC TTCAAGAAAT TTCTGAAAAT 2700

AAGTGTATAT TTCACTATTG TTAGTTTAAC ATATTTTTAG AGGAAGAACG TACTTTAAAG 2760

TAAGCGAGAA TGCTACTTTT GAGAAAAGTC AACTCATGGT ATAATAGAGA AATGAAATGA 2820

GGAGAAGCGA ATGATTGGTA TTAGTAGTTG CTTAGGTGGG ATTTGTTGCC GTTACGATGG 2880

AAAATCTCAA GAACAAGAAG CGTTAAAACA ACTTGTAACT AATGGAAAAG CAATAATGAT 2940

TTGTCCCGAA GTTTTAGGGG AGTTGCCTAT TCCACGTCCG CCAGCAGAAA TCaTTGGCGG 3000

GGATGGCTTT GATGTTTGGG CGAATACTGC TTATGTATAT ACGGATAACG GGGAAGATGT 3060

GACTGAGACA TTTAAAGCAG GGGCACAACG CGCGTATCAA AAATTAAAAG CCCTACAAAT 3120

TACAACCTTG ATTTTGAAAG AAAAAAGTCC CagCTGTGGC AGCCAGCTGA TTTATGATGG 3180

ACAATTTTCT GGCATTAAAA AAGCTGGAGT TGGAGtCGCA ACTGCTTATT tTATTCAACA 3240

AGGAATGACC GTTTATTCTG AAGAAACATG GCAACTTGCG GACATTTCTT TTTAAAAGGT 3300

AAAAACTTAG TAAAAAAAGA TTGACTATTC CCTATGAAAT CGGTAaCCTA TAAGGGATTG 3360

AaGATGGGAG CACGGCTCAT GGAAaTGGmA AAAACAAACC GCATGAATGC GTTATTTGAA 3420

TTTTATTCGA CATTATTAAC GGAAAAGCAG ATGAATTATA TGGAGCTTTA CTATGCTGAT 3480

GACTTTTCTT TAGGAGAAAT TGCGGAAGAA TATGAGGTTA GTCGCCAAGC GGTTTATGAT 3540

AACATTAAAC GAACCGAAAA AATTTTAGAA GAATACGAAA AAAAGTTACA TCTTTATTCA 3600

AACTATGTTG TACGGAAAGA ATTTCTTGAA CAACTACAGC AGTATGTTCA AGAAACTTAT 3660

CCCAAAGATA GTGAAATCAA AAAATATATT GAACAAATTC AAGAAATAGA GGAATGAAAC 3720

TATGGCTTTT GAGAGTTTAA CAAACCGCCT ACAGCAGGCA ATGAGTAAAA TCCGTCGTAA 3780

GGGAAAAGTT TCCGAAGCCG ACGTAAAAGA AATGATGCGA GAAATCCGTT TGGCTTTATT 3840

AGAAGCCGAC GTTAATTTAC AAGTGGTCAA AGATTTCACA AAACGTGTCA GAGAACGGGC 3900

AGTAGGAGTC GAAGTATTAG AAAGCTTATC ACCAGCCCAA CAAATTGTAA AAATTGTTGA 3960

TGAAGAATTA ACGAAAACGT TAGGTTCAGA AACGGTTGAA CTGAATAAAT CTCCAAAAAT 4020

CCCGACAGTG ATTATGATGA CAGGGTTACA AGGGGCTGGT AAAACAACTT TTACTGGTAA 4080

ATTAGCAAAA CACTTAATGA AAACTGAAAA CGCTCGTCCG CTTTTAATCG CTGGTGACGT 4140

TTATCGTCCA GCAGCGATTG ATCAGTTGAA GGTTTTAGGT CAACAATTAG AAGTTCCCGT 4200

TTTTGATATG GGaACAGATG CTAATCCAGT GGAAATTGTT CGTCAAGGGT TAGCATTAGC 4260

AAAAGAAAAG ArAAATGATT ATGTCTTAAT TGATACGGCC GGCCGTTTAC ACATTGACGA 4320

AGCTTTAATG GACGAATTGA AACAAATTAA AGAGTTGGCT GATCCCAATG AAATTCTGTT 4380

AGTTGTTGAT GCGATGACGG GGCAAGATGC TGTCAACGTT GCAGATAGTT TTAATGAACA 4440

GCTTGGAATT ACTGGGGTTG TTATTACCAA AATGGATGGC GATACTCGTG GGGGGGCTGC 4500

GCTGTCAATT CGGGCAGTAA CGGGTGCTCC GATTAAATTT GTCGGTTCTG GTGAAAAATT 4560

AACCGATTTA GAAATTTTCC ATCCCGATCG TATGTCGAGT CGTATCCTAG GTATGGGGGA 4620

CATGTTGACG CTAATTGAAA AAGCGCAACA AGATTACGAT GAGAAAAAAG CAGAAGAACT 4680

TGCTCAAAAA ATGAAAGAAA ACAGTTTTGA CTTTAACGAT TTCaTTGAGC AATTGGATCa 4740

AGTTATGGGC ATGGGACCGA TTGAAGACTT ATTAAAAATG ATCCCTGGAA TGAGTAACAT 4800

GCCTGGTATT GAAAATGTCA AAGTCGATCC AAAAGATGTG GCACGGAAAC GGGCGATGGT 4860

CTTATCAATG ACCCCTGCAG AACGTGAAAA TCCTGATCTA TTAAATCCTA GTCGCCGCCG 4920

CAGAATTGCA GCTGGTTCAG GAAATAGTGT GGTTGAAGTC AATCGTATGA TTAAACAATT 4980

TAAAGAATCC AAAAAAATGA TGCAACAAAT GTCCAAAGGG GATATGAACA TTCCTGGTAT 5040

GGATCaAATG TTAGGTGGCG GCGTTAAAGG CAAGCTAGGT AAAATGGCCA TGAATCGTAT 5100

GATGAAGAAA AACAAAAAGA AGaaAAAGAA AAAAAGATAA ACAGGCTTTA GTTAATTATT 5160

GGAAGAAACa GAAAGTGCGT CATCAGTCTC TTCGACTGAT GGCGCACTTT CTGTTTAAGC 5220

GTAGTTTTGG ATAATTGCGT TTAACGCTTC GGCAAAACCA GCGCCGTGTT GTTCATAATA 5280

TGACGTAAAT CGTTCATCGC CAACATACAT AAGGCCTAAC CCTTTGTGTG CTTCTGGCGA 5340

ATAGCTTGAC CAAGAGTACA TTAACCAATC TTTGTGTAAT TGGAAAATTT CTTGCGCAAG 5400

TTGGCTAGTC AACTTGGGTG ATTTTTGATA CATGGCTAAC TTAGAAAATA AATCTTTTTC 5460

CACGTTCTGC ATTGCTTGCA TGTCTTTTTC CGTCAAGTTT AAATATTTTT TGTTGGCTTG 5520

TTCAAtCGTT TCGTTGCCAT ATTTTTCCCG AATCTCTTTA CCATATTTTT CTtCGTTTTC 5580

TTGAACTTTT TGTTGTTTGA ATGCTTCGAA TTTTTGCTGA TCTGACATGT TTATCTCTCC 5640

TTTGtAATGG TGTAAGGTTT GTTGAACAGA GGCTAAAAGA CGATCAATTT CATTGCGCTT 5700

TTCCAATAGT TTTTGTTGAT GCTCGTATAA AGCCTGTTCA ACATTAAAGT CAGGTTGCTC 5760

CAAAATTTCT TTAATTTCAT CTAGTTTCAG ATCCAATTCA CGGTAAAAAA GAATTTGTTG 5820

CAAACGATTC ACCTCGGCTT GCCCGTAAAT TCGATAACCA GAAGAATTAA TTCTGGCTGG 5880

TTGAAGAAGA CCAATTTCAT CATAATATCG TAACGTCCGT GCACTGACGC CAGATAGTGA 5940

AGCCATTTTT TTAATTGTAT ATTCCATCAT TCATTCCTCC TGACAAAATC TATCATAAAC 6000

CTTTACGTAA CGAGAGGGTC AAGCAAAAAA ATAAAAAAAG TGAGACAACC TGTAAAAAGT 6060

TGCCTCACTT TTTCTCGATT GATTCAATTA GGCAGTTAAG CGATTAAAGA AATCCACGTA 6120

AGCATCTACA ATTGATTGTA AGAAGCCAAC AGTGCCATCA ATAATTTTAC CATCTTCACC 6180

AATTAAGTTC GTAATACCGC CAATATAGGC TTCAGGTTGT TGTAAAGTTG GAACATTTAA 6240

GAAGACTAAT GATTGTCTTA AATGATGGTT CGCACCaAAT CCGCTAATkG CaCCTGGtGa 6300

AACACTCaCC aCAAGTCCtG GTTTTTayCC CCAAACGcTT TCgCCATACG GACGTGAACC 6360

GACATCCCAA TGCATTTTTT AAAACAGCTG GTACTGAACG GTTATATTCT GGTGTTACAA 6420

ACATTACGCC ATCTACTTGT TTAACTTCTT CACGGAATGT TGTCCAAGCT TCGGTGGGGG 6480

TTCCTTCaTC ATCCmAATCT TGATTGTATA ATTCTAAATC ACCTATGTTA ATGAAAACAG 6540

GTTCGAAATC AGCTGGAAAT AAGTCCGCCA TTGTTTTTGC TACCAATTTG TTAAATGAAT 6600

CTTTTCTTAA ACTCCCTACA AAAATACCAA TTTTTTTGTC ATGTAAATCC CTCCATTTTT 6660

CTTTTACTTA CTTTATGATA GTACTTTTTC AAAAAAAGCG CAAACGTTTT ATCCTTGTAC 6720

ATTGTTGAA 6729

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 707 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

AACtGTTCAC TTGtTCCATT AACCAGCAAT GTACCTAACG CTGTTTCCAC TTCATATTGA 60

TAGCTTTTCC CTAAAAATGT TCGGACAGCG ACAGTACCTG CGACAGTATT AGTAGCAGCT 120

TCTGAAACAA TCTCAATGTC CTCTGGAcGA AtCGTTGCCA CCGTCGTTGT AACATTTGGT 180

TCAGGATTAG TCGTTTTCAC TATTTGATTT GTTTGTGTCC GATAATGACC GGCTTCCATT 240

TTTTCGACAG AGAAAAAGTT TTCAAAACCG ATAAATTCTG CAACAAACTT AGTACGAGGC 300

AAACGATAAA TCGTTTCTGG TGAATCATAT TGTTCAATCA CGCCATTATT CATAATGGCC 360

ACTTTATCTG AAATAGAAAA ACATTCTTCT TGGTCATGTG TGACAAAAAC TGTCGTAATG 420

CCTAATTGTT GTTGAATCCG TTTGATTTCA ATCCGCATTG CCACTCGCAA TTTTGCATCT 480

AAATTACTCA ACGGTTCATC TAGTAGTAAT AGTTTAGGTT CGATAATCAG AGCACGTGcT 540

AGCGCAACCC GCTGTCTTTG ACCACCAGAT AATTGTTTTG GATAACGGTC GCCCAAGCTT 600

TCTAAACCAC ATACTTTTAG CATTTCCGTG ACTTTTTGGT CCATGGTTTG TTTGTTTTCT 660

TTGCGTAATT TCAGCCCAAA AGCGACATTT TCTGCAATCG TTAAATG 707

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4985 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

nCTCGGTACT TnCCCCATGG GTGGATGTGC CCGGTTCCAT TCnAGGTGAT TTAGCCAAAC 60

GTTCCTTATG CGTAAGCCTC TGAAGTCAGC CACCTTTACG AACGAAAAAG AGAGCGCAGC 120

AACGATTGCT TATTTAAGGG AATTGATTGA AAAAGTTGGT TTCAmCCCTA AATATTATAC 180

TGCCATTAAT TCAAGCTATG ATTTACCTTA CGATTTTTAT CGTCCAAATA AAGACCGTCA 240

TCGTACACAA ATTGAATTGA TGCAAAAAGA TGGCTCATTA GTGGAATTGG CAACGGTCAG 300

TCCCCTAGTT GCAGCACTTG CTGGTCAGAG TCAAGGAGAT GAACGTTTCT ATTTCCCTAA 360

AGAAATGCTT GACCAAGGAA ATAAAAAACA TTACGATTTA TTTGATGAGA CGTATCGTGA 420

GTTTTCAAGT TACATTCATA ATGGTGCGTT AGTTTTAAAA AAATAGGTTT TCTAGGAGGA 480

TTCTATGTCA ATTAAGTTAG TTGCTATTGA TATCGACGGA ACATTGCTAA ATTCACAACA 540

CAAGATTtAC CCCACGGGTC AAAGAAGCGC TCCAAAAAGC AAATGAGCAA GGTGTTCGTA 600

TTGTTTTATG TACAGGCCGT CCTTTGCCAr gCGTGAAAGA ACAATTGGAT GAATTAGCCT 660

TATATGGTGA AAATGATTtC GTGATtACCT ACAACGGTTC GCTTGTCCAA GCAACCAAAG 720

ATAATACGAT TATTTCACGC TATACCTTGA GTTATGAGGA TTTTTaGAAA TTGAAATGTA 780

TTCTCGTAAA GTCGGCGCTC ACTTGCATAC AATTGATGAT TCCGCTATTT ACACTGCCAA 840

TCGCAATATT GGTAAATATA CGATTCACGA AGCATCTTTA GTGAACATGC CTTTAAAATA 900

TCGTACGGTG GATGAAATGA CACCAGAGAT GAACATTATT AAAATGATGA TGATTGATGA 960

GCCGGAAGTT TTAGATCCTG CCATTGCAAA ATTACCATTA CATTTTACCG AAAAATATAC 1020

GACTGTTAAA AGTACGCCTT TTTACTATGA AATCATGAAT AAAAATGCTA GCAAAGGCAA 1080

TGCTCTAGCA AAATTGGCAG ACCATTTAGG CTTAAATAAA GACGAAGTGA TGGCCATTGG 1140

TGACAATGAA AATGACTTAT CCATGATTGA TTACGCTGGG ATTGGTGTTG CGATGGGCAA 1200

TGCGACAGAA AATGTTAAAA CAATTGCCGA TGTGCATACC ACTAGTAATG ACGAAGATGG 1260

TGTCGCTCAA ATTATTGAAA AAATGGTTTT AATTTAAACA TTCTTTTCAA TAACAATAAC 1320

CCTCAATCTA ATCAACATAG ATTGAGGGTT ATTCGTTATT TTATGATGCT GCCGCTAGCC 1380

ATTTTTCAAC GGCCCAACGG TGGCTTCCTC TTTTTAATAA ACGGACTGCT TCTTCTGGTG 1440

CAACCCAGTG TAAAGTATTG GTCCTTTCTA AGGGTTCGCT AAGTTGCCGC CAAGTATTCG 1500

CTACATAAAA ATATCCAGGA TTATAATAGG CAGTTTGGCG ATGGTTCGAA TAAAAATATT 1560

CATCAGCTTC CCCTAAATAA CAGCCAATTT CTACACTAAT TCCTAATTCT TCTAGCACTT 1620

CTCGGTGGAT CGCTTCTTCT TTCGTTTCGG TGCCTTCAAT TTCGCCTCCT GGTAAAAAAT 1680

AAGCTCCATT TGGCGCTTGT ACCAGAACCA TTGTGTTATT TTCTGGCTTT GAAACAATAA 1740

TGTATGCAGC ATATCTTGTT TGATAGGTCA ATGTCTCTTC TCTTTTTCCA AACGTTGGCG 1800

TTTTCATTGG CGCACCTCTT CTCTTTTCAT CTATTTTACC TGAGAAATTC CTTTAGAACA 1860

AGTCAGAAAC CAAGCGTTTT cTCTACATTT CACGTTTATC TATGCTACTA TTAACGTTGG 1920

AAATATAATT AAGTTTGAAA GGGAGTTTTT AACTATGAAA ATGGCTCATA CATGTGTGCG 1980

TGTAAAAGAT TTAGAGGCAT CATTAGATTT CTATCAAAAA GCCTTCAATT TTGAAGAAAG 2040

TCGCCGTCGT GATTTTCCAG AAAATAAATT TACTTTGGTT TATTTAACTT TACCAGGAGA 2100

CGACTACGAA CTAGAACTAA CTTATAATTA TGATCATGAA GCATACGATT TAGGAAACGG 2160

TTATGGTCAT ATCGCCATCA GTACAGATGA TTTAGAAGGA TTACACGAAA AACATCAAGC 2220

GGCTGGCTTT ACTGTCACTG ACTTAAAAGG CTTACCTGGA ACTGCACCAT CTTATTACTT 2280

TGTTGTCGAC CCAGATGGTT ATAAAATCGA AGTCATTCGT GAACGTTAAA AAGTTTCACT 2340

TGTTCACTTA CTGAGCTTTT TGTGTTTGAG AGCTGTCCGA GACAAAACTT GTCTCGACTT 2400

CTTCTCAAAC ACTTTTTTCA AAAAAAGTGC TACAATAGAA CGTATGAATT TATTGAGGAT 2460

GTGATGTTAT TATGACAAAA AAAATTATTG GAATCGCTGG CAATCAACTT TTGCAGGCAG 2520

CTGAAGTGTT TCACGGTAAC CAAGTGACGT ACACCCCACA AGGTTTTGTC AGCGCTGTTC 2580

AAGCCGCAGt GGCGTTCCTC TCGTTTTGCC AATTGGCCCC AAAGAATTAG CCGCTACGTA 2640

TATACAACAA ATTGATAAAC TACTTTTAGC AGGCGGTCAA GATGTTGCCC CTGAATTTTA 2700

CGGACAAGAA CCCCATATAA AATTAGAAGA AACGAATCGT GATCGTGACG AGTTTGAGTT 2760

AGCCCTAATT GAAGAAGCCT TAAAACAAAA CAAACCTATT TTTGCGGTTT GCCGTGGAAT 2820

GCAATTAGTC AATGTTGCTT TAGGTGGCAC ATTATATCAA GATTTATCGA TGTATCCTCA 2880

ATGGTCAGTT AAGCACGGAC AACAACCCAC TCAACCAATC TTTGCGACGC ATCGTATTGA 2940

TGTTGAACCA GACAGTCAAT TATCAAACAT ATATGGTACA ACTGGTCAAG TCAACTCGTA 3000

CCATCACCAA GCACTACATA CTTTAGGAAA AGATTTACGC GTTACTGCTT GGTCCTCTGA 3060

TGGCCTTGCA GAAGCAGTGG AATCAACAAA TGAACAGCAG CCTTTATTAG CTGTTCAATG 3120

GCATCCTGAA CTAATGTACG CACGTGATGC CAAATCCCAA GCACTCTTTA ACTACTTTGT 3180

TCAAAAATTA TAGTAAAAAC GGmGGGACGT CTTTCGCCCC TCCGTTTTTA CTTGCtATCA 3240

TACTTAAAAA TCAAAATTAT CTGGAtCTGC ACCAAAACGA TGATCCTTAT TTAACTGATT 3300

AATTGCGGTT ATtTCTTCTG GTGTTAATTC AAAATCAAAG ACAGCAAAAT TTTCCTGAAT 3360

TCTTTTTTCG TGAACAGATT TGGGAATCAC GATAATATCA TTTTGTAAGT GCCAACGAAT 3420

AATCACTTGG GCCGCTGATT TACCATATTT TTCACCAATT TTCACTAGTA CTGGTTCAGA 3480

TAATAAATTA CCTTGTCCTA GAGGAGACCA AGCCTCCACG GCAATCTTAT TTTCGGCTAA 3540

ATATTTTCGT AGCGGTTCTT GTGTTAACGT AGGATGTAAT TCAATTTGAT CCACCATTGG 3600

CACAACATTT GCTTCTGTTA ATAAATCTTC CAAATGATGT TGATGAAAAT TGGAAACGCC 3660

AATTGCTTTG GCACGACCAC TTGCGTAAAT TTCTTCTAAT GCTTTCCATG ATTCTTTATA 3720

TTTACCAGCA ACAGGCCAGT GAATTAAATA TAAATCAACA TAGTCTAACC CTAATTTTTC 3780

TAAGCTTTTA TCAAACGCTT TAATCGCATT TTCATAGCCT TGATCTTCAT TCCATAATTT 3840

CGTCGTCACG AAAATTTCTT CGCGGGGAAT ACCCGATTGA CGAATGCCTT CGCCCACCCC 3900

TACTTCGTTT TTATAAACGG CTGCTGTATC AATCAAGCGA TAACCCACTT CTAAAGCCCA 3960

TTTTACTGAA TTTGTTGCTT CTGAGCCATC TTCTACACGC CAAACACCTA AACCTAGGCG 4020

TGGCATTTCA TTGCCGTTTG CTAATTTTTT TGTTTTATCC AAAGTCATGT GTTTCTTCAG 4080

TCCCTTCATT TTCATTACTT TTAGTATACT ACGCATTCTT CTTTCACGAA ACAATCACGA 4140

CTTGTCAGGT CATTTTCTTT ATTTTTAATA AATAACGACG CCAATAAGTC AGCACTCCGA 4200

TTAGTGCCAC AACTACAATT GTTCGAaCCG TAACGCTTCC AATCCACAGT TTAACTGCAA 4260

GTAACCAGAT ACAAAAGAGA TAAACACAGG ACCAATTAAA GGAAACTTTC AATGTACATG 4320

ATTCCGCAAT TTCGATTATC TGACTACTAA AAAAACGACT CCCTAATCGT AAAGCGACTG 4380

GATTATGGAA GATGAATGTT CGTCGCTTGT CAGCCGTCAA GCGAGTAGCA CATCGCTCAT 4440

CAATAAAAAT TGTTACAGGT AATTTAGAAA GTAAATCATT TTTTGCCCAT TCAATTGAAA 4500

CTTTCACGCA GCACTCACAA CCTTTAACAA AATTATAAAA TATATTATTT ATTTACAAAA 4560

TTATAACATT CCTAAAAAAA CATGTTAAGT CTTAAAAAAA AACTTTCTAT TTTCTTCCCC 4620

TTTCGCCGAG AGATTTAGTA TACTAAAAGA AGAAATTTTA GAGAAAGTTG GATTTTTATG 4680

TTGATACTTG GTTTAGCCAG TACTGAAACA GTCGATTCTT CTGCTCCTGA TTTAGCAGAA 4740

GCGACCGTAA AAAAAGTGAG CGCGCTACAA CGCTTTTGGA ACAATATTAA TTGGGATGCC 4800

ATTGTCGCTA CCTTAATTGA AAAAAGTTTA TCAATTTTAT TTTTAATTTT TCTATTTTTT 4860

ATTATTCAGC GAATCGGGAA GTATTTAATT GATCGCACCT ATGCAAATTA TAGTAAAAAG 4920

CAACATTTTA GTGAAAGCCG TTTAAAAACG TTACACACTT TAATCATTAA TGCATTTCAA 4980

TATAC 4985

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1346 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CGTCTGTCCT TATCTCGCAC AGCTATTTGG AAAGCGATTA ATGAATTGAA AAAAGATGGC 60

TATCAAATTA CTAGCGTGCA AAACAAAGGG TATCGCCTTG AAAAATCAGA TGTCCTTTCT 120

GCAGAAGGCA TTCAGTTGGC ATTATGGCCA CAAACACCGC CTTTAACAAT TACCGTCCTA 180

GAAACATCTG AATCTACCAT GACCGATGCT AAGCGGGCGA TTTTAGACCA GACGCCCGAC 240

AATACTTTGA TTGTTGCTGA CATGCAAGAA ATGCCTCGGG GCAGGTTCGG CCGGCCGTTT 300

TTCGCTACGC CAGGTAAAGG GATTTACATG AGCATGGTGC TTCAACCGAA CCAAAATTTT 360

GAAGAAATCG CTCAATATAC TGTGATTATG GCCGTAGCGG TCGCTCGAGC AATGGATGCT 420

CTGGCAGGCG TTCACACAGA AATTAAATGG GTCAACGATC TTTATTTAAA TGGCaAAAAA 480

GTTTGCGGGa TTTTGTCTGw AGCCATGAGC AATGTGGAAA CAGGGCAGAT TTCTAATGTT 540

ATCATCGGCA TGGGGATTAA TTTTTCAATT ACAGAAACCG AATTCCCTGA GGACATTCGG 600

CAGAAAGTGA CTTCCCTTTT CCCAGATGGT GAACCGACGG CGACGCGAAA TGAATTAGTC 660

GCCAACATTT GGAATCATTT TTACCAAATT TTAGACGAAC GTTCAACGAA AGACTTTTTA 720

GAAGAATACC GGCAACGCTC CTTTGTGATT GGCAAAGACG TCGCTTTTAC CCAAAACGGG 780

CAAGACTTCC GCGGCGTTGC GACAACCATT AATGGAAATG GGGAATTGAT TGTCCAGTTG 840

CCTGATGGCA CGGCCAAAAC ATTGTCTTCT GGAGAAATTA GTTTGGACCA GATTGGTGAA 900

TGGCGTCGAG GATAAGTGGG GAAGTTTAGA AAAAAGCAAG AAAGCACCTG TCTAAATTTA 960

AACAGATGCT TTCTTGCTCT TTAAAAATCA ACTTCTTTTT GTGCTAACAT TTCGCACAAT 1020

TCTTTGGGGA AATTCGTATA GGGATTTTCA TTGATTTTCA TTTCTTCTAA AGGAGACAAC 1080

CAGTTTTTAT TCGTATTCTG AAAAGCAATA GCTACTTTTT CTTCAAGATA CTTACATACT 1140

TCTGCGCCTT TAAAATTTCG ATAACTGTCT TTTAAATGCA GAACTTTCTC TAATCTCCCA 1200

TATAAATtTT CTCTAGTCAT GTATCGATTG ACTTCCTCAA AGTGAAGTAG TAACCAGAGT 1260

TCAAAGCATT CATTTGAAAA AATTGTCTTG ATTTCAGGAC GATGTCGTTG TGCACATCTn 1320

ACCGCGTTTT CAATCGATTC nGTGGT 1346

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1959 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

CCAAAATATk GaTGCAATGG TTtTGkTTyC CTGaCAATAG nTGayCGATC GTCACAAATT 60

CGTATCCTTC ATTTTTTAAC ATAGTGATCA ATCCTGATAC AGCTTTGACT GATTCATGAT 120

GGATATCGTG GATTAAGATG ATTGCGCCAT TAAAGACATT CTGTTTGACA ACATTATTGA 180

TTGCTTCAGG ATTTCTTGTT TTCCAATCTA GCGAATCAAT ATTCCATTGA ATAATTGGCA 240

TGCCAATTGT TTCAGCACTT TTTTTATCAA TAGCGCCATA AGGAGGACGG ATATTGCGAG 300

GCAAGACGCC CGTAGCCTTA AAAATTGCTT TATTAGCTTT ATTCATTTCA CTTTGTAATT 360

CATCTGTCGA AAGTGTATTC AACTGTGGAT GAGAATACGA GTGACATGCT ACCTCATGAC 420

CCTCGTCGTG GACTTGTTTT GCAACATCTG GGTTTTGATC CACCATCTGT CCTAACATGA 480

AAAAAGTAGC CTTCACGTTG TTTGTTTTTA GTATATTCAA CAAATCAAGT GTTGAGCTAT 540

TATTGGGTCC ATCATCAAAA GTAAGAGCGA CATACTTTTT ATTTTCATCT AATGCAGGTA 600

ACGCATCTTT TATACTTTCT TGCGAAACCA GATCTGTATC AATAAAAGGC GCAATGTCTC 660

TGTATTTTAG AATAACTTTA GTAGTTCCAG TTGTGTTCTT TGGTAGGTCG ATGGTCAATT 720

CATCAGGTGT ATAAGTCATT TTACTCTCGT AAGAAATTCT ATCCATCGTT AATACAGCAT 780

CAATTATTTT TTTAGGCTCT TTGGCGTGAT CAAGAATGGA TTGCTGAATC ACCTGTTGAA 840

TACCTAAGAG ACTTCCTTCA TCCGGAATCA AATCCTTATT TGTGATTTCT TTACCAGTTT 900

TTTCGGAAAC ATAAATGGCT TTCTCAGAAA TACTGTCTGG CTTTTTAAAA TTTTCCTTAC 960

TGCGATCCCA GATAAAACTA TCTGCTTCAA TTCTATACTG ACCAATTTTA GAAGATATTG 1020

CTGTTGCTTT TATTCTGGCA ACTGTAAGCA CTTCTTTTTC ACGATGCTTC TTTTGTTGCT 1080

TTTGAGCAAG CTCGTCTAAG AGCTGGTTTG CATTTTTAAA AGGAACAGTA CTTTCATAAA 1140

CTGGACGGTA AATAAGGCAC TTATTTTCTC CGTCATTTAT AGTATCTATT TTCGTTGAAA 1200

TTCCACTTTG GTCTTGCTCT TCTTGGATTG AAGCCATGAG ATTTTTTACA GCAGTTTCAT 1260

ATTTGGCTTG AGCTTCTACT TCTTGCTGTT TTAATGTATA AACAATATAG CCGCCAGTAG 1320

CGACAATTAC TAGAATAAGT GCGGTAAACA ATGCGAATAT ATACTTGCTA TAATTAGATT 1380

TTTTCTTCGG TATTTTTGGT TGTTTGTTTT CGTTTCCGGT TTTGGGCTTT ACAGCTTCAT 1440

TCAATTGATT TGCATTTTCA CAGATATCAG GTTCTTGTTT GTGACGTTGA CTACGACGTG 1500

CATAGATATT TTTATCAGAC ATATGACTTT ATCCTTCCTT TTATCATCAT AAAGTAAAAT 1560

GAAACAGCGC AATTGTAATA AAAAATAATT TATAGATAAt TTTTACAAGA AAGTTTTCTC 1620

TATTATTCtT TTTCtTTTGG CGTAGTAATT ATCGCTAATT CCTCACCATT TTAGATATTG 1680

TGGCTACTTA ATGTTAAACT CAATTTTGAT ATCATTAGAA TCACTAGTTT TGATTTTAGT 1740

TTTTGGAATT GTACCCTTCG ATAGGTTAAT CACTATGGTT GCAGTCTTTG AATCACCAAT 1800

AATGGTCTAG GATTAAAAAT ATAAACTATT ATCCAATCTT GTTCTATTTG ATTAAATATA 1860

CTACTTTAAT ACCTTTAATT TGTGATCAAT TCCAACTTCA CCAAACTCAT TTACAATTAT 1920

AATAATTTCT TCCCTGGGAT GGGGTGAGAC TTGCAAGGC 1959

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3625 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

TTTTTATTTG GAGGGATTTA TTGTGTTAAA ACCATTAGGC GATCGCGTCG TAATTAGAGT 60

CGCGAAAGAA GAAGAAAAAA CTGTTGGAGG AATTGTTCTT GCATCCGTTG CAAAAGAAAA 120

ACCACAAACA GGTGAAGTTA TCGCAGTAGG TGAAGGTCGT GTGCTTGAAA ATGGCACAAA 180

AGTTCCGATG GAAGTAAAAA TTGGTGACAC AGTAATGTTT GAAAAATATT CAGGAACAGA 240

AGTGAAATAC GAgGCGTAGA ATACTTAATT GTATCAGCCA AAGACATTAT TGCCACTGTT 300

GAATAATAAA CAACTTTTAC AGAaTAATAT AATAGAAkAA AAAAATTATG AGGTGAAGAA 360

ATCATGGCAA AAGAGnTTAA ATTTGCAGAA GATGCACGTG CAGCAATGCT ACGCGGAGTA 420

GATGTATTAG CAGATACAGT GAAAGTGACA TTAGGCCCTA AAGGTCGTAA CGTTGTTTTA 480

GAAAAATCAT TTGGTTCACC ATTGATTACT AACGATGGAG TAACAATTGC TAAGGAAATT 540

GAATTGGAAG ATCATTTTGA AAACATGGGC GCAAAATTAG TTTCAGAAGT TGCTTCTAAA 600

ACGAATGATA TTGCTGGTGA CGGAACAACA ACAGCGACTG TTTTGACACA AGCCATTGTT 660

CGTGAAGGCT TAAAAAACGT AACTGCTGGA GCCAACCCAT TAGGTATTCG TCGTGGGATT 720

GAATTAGCAA CAAAAACAGC AGTAGAAGAA TTACACAATA TTTCATCTGT AGTTGATTCA 780

AAAGAAGCGA TTGCACAAGT CGCTGCTGTT TCATCAGGTT CTGAAAAAGT CGGCCAATTA 840

ATTGCCGATG CAATGGAAAA AGTTGGTAAC GACGGCGTAA TTACCATTGA AGAATCAAAA 900

GGGATTGAAA CAGAATTAGA TGTGGTTGAA GGAATGCAAT TCGACCGCGG TTATTTATCT 960

CAATACATGG TTACTGACAA CGATAAAATG GAAGCTGTTT TAGAAAATCC ATATATCTTA 1020

ATTACCGACA AAAAAATCTC AAATATTCAA GATATCTTAC CTTTATTAGA ACAAATTCTA 1080

CAACAAAGCC GTCCACTATT GATTATTGCG GATGATGTTG ATGGGGAAGC TCTACCAACA 1140

TTAGTATTGA ACAAAATCCG TGGTACATTT AATGTTGTCG CAGTAAAAGC GCCAGGATTT 1200

GGTGACCGCC GCAAAGCGAT GCTTGAAGAT ATTGCTATTT TAACAGGTGG TACAGTAATC 1260

ACTGACGACT TAGGGTTAGA GTTAAAAGAC ACAACTATTG AAAACTTAGG AAATGCTAGC 1320

AAAGTAGTTG TCGACAAAGA TAACACAACA ATTGTCGAAG GTGCTGGTTC AAAAGAAGCC 1380

ATTGATGCCC GCGTTCATTT AATTAAAAAC CAAATCGGCG AAACAACGTC TGATTTTGAT 1440

CGTGAAAAAT TACAAGAACG TTTAGCTAAA TTAGCTGGTG GGGTTGCTGT CGTTAAAGTC 1500

GGTGCTGCAA CTGAAACAGA ATTAAAAGAA TTAAAATTAC GAATTGAAGA TGCATTAAAC 1560

GCAACACGTG CCGCTGTAGA AGAAGGCATG GTTTCTGGTG GTGGTACCGC ACTGGTCAAT 1620

GTAATTGGTA AAGTCGCTGC GCTAGAAGCT GAAGGCGATG TGGCAACAGG GATCAAGATT 1680

GTCGTTCGTG CATTAGAAGA ACCAATCCGT CAAATCGCTG AAAATGCTGG TTATGAAGGA 1740

TCAGTGATTG TTGACAAACT AAAAAATGTT GACTTAGGTA TCGGATTCAA TGCAGCTAAC 1800

GGTGAATGGG TAAACATGGT TGAAGCCGGT ATTGTTGACC CAACAAAAGT AACTCGTTCT 1860

GCCTTACAAA ATGCAGCTTC TGTGTCAGCT TtATTATTAA CAACTGaAGC AGTTGTTGCA 1920

GACAAACCAG AACCAGCTGC ACCAGCTCCT ATGATGGATC CATCAATGGG CATGGGCGGT 1980

ATGATGTAAA AATAAAAAAG TATTCTAAAA GCTTACGATT GAAAAATCGT AAGCTTTTTT 2040

TTACATAAAG nATAAATAAT CTAAATAATA TCTTTGCTTA TGGTGTTTTC TCAAATATTT 2100

ACTCAATTAA CAAAAAAGAT GATAATTTAA AGGCTTCGTT TTAGCTAGAC AGCCTTGTTT 2160

TTATTCGATA AAATTTAGTA ATAACAATCA TTTTTTGAAA ACTAGTAGTT GGTAATTTTT 2220

ATTTATTAAT AGAATATAAA TAAAAATTAT TATATTTTGA AGGGGAATAT TAGGAAGATT 2280

GAGGTGGAAA AATGGAAAAA TATACAATTG GTGTCGTTCC TTTGACAATT GAAACAAAAG 2340

AATACTGTCA AAATTGGTTT GTTAGTAATC AAGTAGAATG TCAGATCATC CAAGGAACAA 2400

ATATACTTCT TCATTTGCAC CATATTGATG GATTAGTGAT TGAAGTTACT GACCACCAGT 2460

AAGTAAATAC TTGTTGTGAG TTATTAATGA CGATTCGCAA GCAGTCAGAT TTACCGATTT 2520

GGTTATTTTC TCGGACAGAA GTAATTAGTA AAGTCAATCG AATTATTTAT TTACAATTAG 2580

GTGCAGATGG CGTTTTTGAT TATTCTTATG ATCGTCAAGA ATGCATGTTA AGCATGAGTA 2640

ATTTATTGCA ACGTGTAAAA AGACGTTTTT ATCCTAAATT AGCTATTGCA AACGAGGAAC 2700

AAACAGTTAC TAAGAATCTT TCTGAGAGAT TGTATTTAAT TGCGCCAAAT TTGAGTGTTT 2760

GCTTGGGAAG TGGAGAAGAA ATTCTTTTAA CAAAATTGGA ATTTTTCACA ATAGAGTATT 2820

TATACAAACA CGCAGGACAA ACAATTACAT ATGAAGAACT CTACAAAAAT GTCTGGAAAG 2880

ATACTGCGAA TGAACGAAAA TATCGTGTCG CTAATGTGAT TTTCCATTTG CGCAAAAAAA 2940

TTGAACAAGA TGTGAACAAA CCACAATATA TCAAAACAAT TCGTTCAAAA GGCTATATGT 3000

TAACGGTGTA AAAGGTACGA AGGAGGAGAA AAAGATGTGG CAATTTTTGA CTAAAAACAA 3060

TGAAGATGAG GAGTATGAGA AGATGCCAGA TTTTGAAAAT GAACAATCGA TGAAAGCACA 3120

ACCAACGGTA ACAGAAGAAG CAATCTTTGA TGACGAGGTC GTTTATGCAA AAGAGCAATC 3180

CGCAAAAGTA AAACAGGTCC AGCAACAAGT GCCGACTGAA CTGTCAGAAG AGTCATTTTC 3240

TTTTGTTGAA AAACTAGAAG AAGAAAATCA ACTATATCAG CAATCCATTA AAAAGCTAAC 3300

GGAGCAATTA TTAATGCAAA CCAATGAAGT GGAAGCATTA CAAAAACAAG TAGTCGAAAA 3360

AGATGTTCAA CTTAAACATG TTAAAGAAAC ATTAAGTGAT AAAGAAACAA CTATCACTTC 3420

TTTACAGAAA CAATTGTCTG AAGAAAAGAT GCAACAGAGA CAGACCAGTG aAGAGrATTT 3480

AGACACAGCC GTTACGCTTT CTCAAAAAGA AATTGGCGAG TGTTATTAGA AGCCAAACGT 3540

CAAGCAAAAG AAACATTAGT CAGCCAACCA CCAGTTGCAC ATTCATGAGA ATGGACACGT 3600

TGCACTTTAC ACGCTGAGCA GTGGn 3625

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3758 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

CCnCCCCCCT TTTTTTTCCG GTCCCnAATn CCTCCGGCCn TTATTTTAGG TTTCCATTTG 60

GTAAAAATAG TATAGGrGsw AAgCatCaGT TATGTATtAt TAGTGamAGA AGGTTGCcGT 120

GAAGTAAAwA TAAAmAGGAG AGCcAAAGGA TGTTAAmAAA TGTAGAAGTG TTCTGGCAAA 180

ACTTTTTAGA TAAACATGAA TTAGATATGT TGATGCCCGA TGTGTGGATG TTTGGTGATG 240

GCTCTTCTGA GATGGGTAAT CGTTTGGGTC AATTGGTTGT TTCTGGTCGT AAGACAGCGA 300

CATGCTCTTC TTTAGATATT TATAAGATGG AAGAAGAACA ACTCCCTAAA GCTGGTCAAT 360

ACGACATAAT TTTAGATGGT CAGTCTCAAC CGCTGGCTAT TATTCGAACA ACCAAAGTTG 420

AAATTATGCC CATGAACAAG GTCTCAGAAT CTTTTGCGCA AGCGGAAGGT GAAGGAGATT 480

TGACGTTGGA TTATTGGTAT GAAGAGCATG CGCGGTTCTT TAAAGAAGAA CTGGCCCCCT 540

ATCAGTTACA ATTTTATCCC GATATGCTGT TAGTTTGCCA AAGTTTTGAA GTAGTGGATC 600

TTTATACAGA AAAAGAATAG ACAGAGAAGC GGTGACTGAC ACGATTGAAA CGAGCAGTCG 660

CCGCTTTTTT TTTACTGATT TTTTGAAAAA GTTGCGTATG TTTATATAGA CAGGGTGATA 720

GAAAAGAGAT GTGTATTTTA AGTGTCGAGG AAAAGATGGT ACAATAACCT AAGGAGAATA 780

CTAACAAGGA GTGAATGTTA TGAAAGAAGA ATTTTCATAT GAAATTTTAG AAGAAGTCGC 840

TGTGTTATCT GAGAATGCTC GTGGCTGGCG TAAAGAATTG AATTTAATTA GTTGGAATGG 900

TCGTCCACCA AAATTTGATT TGCGTGAGTG GGCCCCTGAT CACGAAAAAA TGGGTAAAGG 960

CATCACCTTA ACTAACGAAG AATTTGCTGA ATTAmGCAAA ACAATTAmaT CCATGTAAAA 1020

GGAGCACCCT GTCAATGAAA AGTATGACTG GATTTGGAAA AAAGACCATT CAAAATGAAA 1080

ATTATCAATT GGATATCGAA GTCAAAAGTG TCAACCAGCG GTTTTTAGAT ATTCAATTGC 1140

GTATGCCCAA GGAACTGAAT GCCTATGAAC TGGTGATTCG GCAAGTAATC AAACGCACGT 1200

TAAAACGTGG GCGTGTGGAA GTCTATGTAA ATCTCCAAAA AATCGGCAAC AATCAAAAAG 1260

AAGTGCGCGT GCAGTGGGAC TTAATTGATC AATTATTAAC TTCGGTAGAT CAACATTTAA 1320

AGGAAAAATA TCCGGAAGCA ACGTTTGATG CAGGGGACAC GGTCAACCAT TTATTAAAGC 1380

AAAATGATTT TGTCGAAATA GTGGAAGCAG AAATTGTTGA TCAGACATTT GAACCATTTC 1440

TTGTTCAAGC ATTTGAAGCA GCCATCGCTA GTCTTGACCA AAGCCGTGTG CAAGAAGGAA 1500

CGCAAATTAA ACAAGTCCTG CTTGACTATG TAGCAGTGTT GACCCAATCA ATTCAAGAAC 1560

TGCAAGCGTT TGTGGGCGTT TTTGAACAAG AATACCGTCA ACGTTTTGAA GCGAAATTAA 1620

ACGAATGGTT AGGTAGCCAA GTGGATGAAA CACGCTTATT AACAGAAATG GCCATTTTAT 1680

TGGAAAAAGG CGATATTCAT GAAGAATTGG ATCGTTTGGA TATCCATATT GATAAGTTGC 1740

ATCAGTTGCT TGACGAAACA GAACCAGTGG GCCGAGAACT AGATTTCTTA ATTCAAGAAA 1800

TGAACCGTGA AGTCAATACG ATTGGTTCGA AATCAAGTCC AATTGAAATA AAAAATAGTG 1860

TCGTTCAAAT GAAAACAACC TTAGAAAAGA TTCGTGAACA AATTCAAAAT GTCGAATAGA 1920

GAAAATGGTT TCGTGTTTTT CCAGAAAACG CTATAATATA GAAGTAACGA ATAAATGACA 1980

GAGGTGAAAG CAATGACTAA ATTTTATGAT GTAACCTTTC AAGAATTAAG TGGTCGTTCT 2040

GTGGTAAAAA CAGAAGTCGC TTCAGACAGA GAACCATTTG ACGTCTGGCA AGATGCTTGT 2100

GCAAGCTATT CTGAAACAGA ATTAAATATC CAAATTAACG AAGACACGTT TGTTACTTTG 2160

AATCGTCACT TTGTTGTTCG CATTGACGTG AAAGAAGTTG ACGGTCCTGT CGACAAACAA 2220

GTTCGTCGTC GCGATGAATT GATGAACGTG GTCAATACGC TTTCTAATAT GGGCTTATAA 2280

AAAATACCAC CAAGCATTTC ACGCTTGGTG GTATTTTTTT TAATTAGAAT AAAAGAAACA 2340

AGCCAACGGC GCATAGGAAA GAAAGAACTG CGGCACTTAA AAAGATTGTC CGCATGGATG 2400

CAGTTTTACT GCTTTCTTTG CCGTGTTTTT TCTGACGTTC TAAACGTTTT TGAAGCTGAC 2460

GGGTAATTAC TGAGAAAATC AGCCCGAAAA GGACTAAAAC ACTTGTGGCA ATCGATTTGG 2520

GAATAAAATG ATTCGGTTCT GCAACGGTGG TATCCCCAGA ATGTTCGTTT GATTCTTCCg 2580

CAGCTTCCGC TGTTGTTTGT TTTGGCTTAA CTTCTTGGTA AGCCACTTTG GGTGCTTGAA 2640

CCGTACTATT TAACAATGTC AAGGGCGTGT CGACTTGCAC TTGCCCTGCT ACTAATTTAA 2700

AAGTAGGCTG ACTGCCtTTG GGAACGACAC CGTATAAATC GTTAGCCAAA GTAACTTTTT 2760

TGTGATCAAT GGTTTGTGGT CCTTTGTCTA AAAGTTTTTT GCGTTCGTAT GTCGTATAGG 2820

CATAGTCCAG CAAAGCATTC CCAGCCACAT GCCGTTCGTA TTCACCATCT TGGTTTTGCC 2880

AATTACCAAC GCCTAAAATG ACTTCAATTA AACGGAAATC TCCTTGTTTG ACAGTCGCAA 2940

TATAGTTAAA GCCGCCTGTT GGACTAGAGC CTGTTTTTAA ACCATCCACG CCTTGGTAGC 3000

CATACTTAGC TCCTGGTAAA GAATAATTGT ACGCTTCAAA CGTTTCCTCG TAAGGCGTGC 3060

CGACCATTGT TTTCACCACT GGTTTATTGG TGATTTCTAA AATTTCTGGA TAATTTTTCA 3120

GAAAATGATA GGTTAGTGAA GCTAAATCTC TAGCTGTAGA GACGTTATCG CCATTTGGGT 3180

CAATTCCTTG CATTTGATAC AGGCCGTTAA AGGCACTTGC TTGCGCTCCA CTGCAGTTGT 3240

AATAGGTAGT ATTTGTCATC CCTAGTTCAG CCGCTTTTTG ATTCATTAAA TGAACAAAGT 3300

CAGTAGGCTG GTTCCCTGAA ATTAAGTTTG CCAACATGAG AGTCGCCACA TTAGAAGAGG 3360

GGACAGCAAT CATTTTTAAC AGTTCACGGA CTGGATAAGC AACACCTAAC GTAATTTTGT 3420

TATTACTAAT GGCATATATT TTAGAAATAT CGACATCTTT TTGCGTAGCA GTCACAGTCG 3480

TATCCATTGT AAATTTTCCT TGCTCCATTG CTTCAAAGGC CAAGTACATG GTCATCACTT 3540

TGGCAATACT GGCAGGATTC CACGCTAAAT CTGGTTGCTG CGACCAGAGA ATTTCTCCAG 3600

AATCAGCATC AACAACGATG GCTGCTTTGG GAATGCCATC CGCTTTCAAA GTCGTCCCCA 3660

TTTTTTTAGC AATCGCGGTA ATATCTTCTT CTGCGTGAGT CACAGTAGGT TGGGCAAAAA 3720

AGGCCCCACa ACTGACAACT AAGCCTAGCA GAAACAAT 3758

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

TCTTCTAATC CTGATGCATT aGCTGGCGTG GATGCACAAC GTGTAGCAGC TTTCCAAGCG 60

GCTAATGGGA AAGCATTGGT CAATCTACGA AAAGCAACGC AAGCCAACAA AGTAAGTTGG 120

ACTGTTGTAG CTGCTGCTAG TGAAGGCTGG GCAGCTAAAG TCTTTCCAGA ATTAGCAACA 180

AGTGAAGAAC AAGTTGATGC ACTTTGGAAT GAAATTTTCA AAACAACACG AATTTATGAA 240

GAAAACCCAG TCATTGCTTG GGATATCCAT GATAAAAAAT TACAAGAAAA AGCGGCTGAA 300

TTAAACGAAC AACAATTTAC TGCTTTACAC TACACAGCTC CTGGGACAGA TTTAACGATT 360

GGTTTGCCTA AAAACCATTT ATGGGAAGGC GCTGGTAGTT ATAACGCTCG TGGGGAAGAA 420

TTCATGGCTA ATATGCCAAC AGAAGAAGTT TTCACCGCCC CAGATAGTCG CCGCGTTGAT 480

GGTTATGTTT CTAGTACAAA ACCACTAAGT TATGCTGGCA CAATCATTTC GGGAATGAAA 540

TTTACATTTA AAGATGGAAA AGTTGTTGAC TTCTCAGCTG AACAAGGCGA AGAAGCATTG 600

AAAAATCTCT TAGCTATTGA TGAAGGTGCC AAACACTTAG GAGAAGTTGC TTTAGTACCA 660

GACCCTTCAC CAATTTCACA ATCTGGTTTG ATTTTTTATA ATACATTGTT TGATGAAAAT 720

GCCTCTAATC ATTTAGCCTT TGGTTCGGCC TATGCCTTCA ACTTACAAGG TGGCACGGAG 780

ATGAGCGAAG AAGAATTAGC TGAAGCTGGT TTAAATCGTA GCCAAACACA TGTGGACTTT 840

ATGGTAGGAT CTGACAAAAT GAATATTGAC GGCATTAAGG AAGACGGGAC GATTGTTCCA 900

GTTTTTAGAA ATGGCGATTG GGCATAGTCC TACCAAATGT TGAAACCAAA GTTTTTAGCT 960

CCAAGAAATA AGAAGAAACT GCTTTTATTC AGTAGGTTCT TTTAAAATTT AGAAAGTGAG 1020

GCCTAAGTTG CAGCGACTTA GGCCTCACTT TCTTATGCTA AAAAATCAAT TGCATTCGAA 1080

GGACCTTTAA CGTTATAATC AAATTGAAAA CGGATTCTTT TGCACGATAT AAGTTTACTT 1140

TTAAATAAAA GAATTTTTTT CAAAAAAGTA GAATTGGAGA ATTAACATGG GATTATTTGA 1200

TGGATTATTA GGAAATGCAA CACAAAATAA TAATGAAACA GCTGAaAAAG AaTTACGGGA 1260

TGTTTTAATC CCTAACGrAA AAGTAGATAT GGCCTTCACT TTAGTAAGAG ATTTAATTGT 1320

CTTTACAGAT AAACGCTTAA tTTTrGtCGA TAACCAGGA 1359

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15614 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

ATTAAnATAn ACCAGGTTTT AGATAGTATT TATTCAAGTT TACCTCnAAG GTATTAATCT 60

GAAACCCkGT TATtTtAtTa GAGAGTTCyT TTTCaAATAT ATGTAAATrG cTAACaACrG 120

GGTTCTGTTT GTTGTTGGAC tATTTTTTTT GATCcCCAmC CGTTCTTGAA AAAAAAcAAA 180

AAAATAAACA GTGyTCGCTT GATTAATCGA ACGTTTGTTC GTATAATGAT TTTGCaAGTG 240

AGGTGATGTT TGTGAAATTT GATTATGAAA AAGAACCTAG TCGCGATATT CTTTGTATTG 300

```
ATTGCAAAAG TTTTTACGCT AGTGTAGAAT GCGCAGAACG TGGGTTAGAT CCTTTGAAAA     360
CGATGTTGGT AGTGATGAGT AATGCCGAAA ATGCGGGGGG ATTAGTTTTA TCAGCATCGC     420
CAATGGCAAA AAAAGTTTTA GGTGTTTCAA ATGTTACGCG TAAAAATCAA GTGCCCGCTC     480
ATCCTGATTT ATATGTGGTT CCGCCACGAA TGAATTTGTA CATGAAAAAA AATCAAGAAA     540
TCAATAATTT ATATAAAAAG TTTGTCGCAG ATGAAGATCA CTCTGTGTTT AGTGTAGATG     600
AATCATTTCT TGATGTAACT GCTTCTTTAA GTTATTTTCA TTGTGAAACT GCCTATAAAA     660
TGGCTCGGAT TATCCAAAGA GTCGTTTATA ATCATGTGGG TATTTATGTG ACCGTTGGGA     720
TAGGGGACAA TCCGTTACTG GCTAAATTGG CGTTGGACAA TGGCGCCAAA CATTCGCCAG     780
ATTTTATTGC AGAATGGCGG TATGATCGTG TACCAGATAC TGTCTGGCAG TTGCCTTCTT     840
TGACAGATTT TTGTGGGATT GGTCGCAGAA TGGCTAAACG ACTGAATCGT TTAGGAATTG     900
ATTCTGTATA TGAGTTAGCG CAACGAaTCC GCATCTTTTA CAAGAAACAT TTGGCATGAT     960
GGGGCTTCAA TTGTATGCTC ATTCATGGGG GATTGATCGT ACCTTTTTAG GGAAAAAAGC    1020
ACCACATAAA GCTGAAAAAT CATTTGGGAA TAGTCAGATT TTACCGAGAG ATTATGCGCG    1080
AAGAGATCAG ATTGAACTAG TCTTAAAGGA ACTGACAGAA CAAGTGGCTG CTCGTTTAAG    1140
AAAAGCGCAT TGTCAAACGG AATGTATCAC AGTCTACGTT GGTTATTCCA AAGGTCAAAT    1200
TGACCGTGAA GGACGAACTG GTTGGCGAAA ACAACAAACG ATTCCTGCAA CAAACAATAC    1260
AAAAGTTTTG ATCACCTATG TATTAGCCCT TTTTAGAGAA CATTATCTAG CGGGAACCGA    1320
TGTCCGGCAA TTAGGGCTTT CATATGGGAA ATTAGTTTGG AATGAATCTT TACAATTAGA    1380
TTTATTTTCG GAACCAGAAG AGCAAATTAG TGAAATGGAG TTAAACTATC TGATTGATAA    1440
AATCCGACAA AAATTTGGCT TTCAAGCTTT AATTCACACC TCTTCTTTGT TGGAAGGCGC    1500
GACGGCAATT CATCGTTCAG GATtAGtCGG TGGACATGCT GGAGGAAATG TCGGTTTAGG    1560
AGGATAACAT CATGAAACGA ACAAAAAAAG AmTTTACACC TTATAATGAG TACCATGaTC    1620
GGCCCTTTgA ACTGAAATGG GCGACCGCTT TTGCATTGGG GGAGTTACAA GAAGGAATCG    1680
AAGCTTCTGA CGAATATTCT AAGAGAGCGT TTGAACGCTT ACCGCAACAG ACAACAGAAG    1740
AAATTGAAGA ATACTTAGAG CAGTCTTTAA AAAATAACCA AGTTTTAGAa ATCCAATTGA    1800
ATACATTGGA TGATTTTGGC CGTGTTCAAC CGCCAATTGT TGGTGTTTTT GAAGGGATGA    1860
GTTCAATGGA GTATGTAAAA GTTGGACATA CACATATTTT ATTTGAAGAG ATTCGGCATA    1920
TTAAAAGTCA TGAGTTTACA AAGTGGTCGA GCCTTGATTA AAAAACTATT GTTATACGTG    1980
CGTTTTGTGA AAACATATGA GATAATGTTC ATAGGTAATA GAAATAAAAA TTATGGAGGT    2040
GGCTAGGATG GGACATCATT TAAAAGGACA CCATGGCAAA AAAACAAACG CAAATGGCGG    2100
CTTTTTAACA AAGTTATTGG CAGGTGGCGC CTTAGTGGGC GTAGGAAAAA AGTTATACGA    2160
CAATCGTGAA AAAATTAAAG AGCTGTTAAG TGACGACAAA GAAAAAGGCA ACGACACTGA    2220
AAACAAATAA AGGACTGGAT TGACGGAATG GTAAAAGATG CAGCGTATAT TATTGAAGAG    2280
```

ATTACCACGA TTTTAGAAGG AACCTATTTT GTGATAACCA AGAAAGCGAA TAAGCGCCGC 2340

ATGTGTTTAC ATGTTTCAGA TACCGCGAAA GAATTTTTAC GTTGTCGACT GGAGTATGAA 2400

aCATTCTTAG AGCAACAGTT TGATTATGCC aCAATGTTTT TAGAAACATG TACGGAAGAT 2460

GTTTCTGTTG AAATTCAAAC GGGACTACTT GAAGAAACAA ATTATCAAGC ATTAACCGTG 2520

GAAAAATTAC AACAGATAGC CACTAAAATG TAACACAGCT GTTGTGTTAG TTTTGATTCT 2580

TAATTTGGTT AAGGAGGCGG ATAAAATGTC AGTATATCAA CATATTTTGG TTGCTCTTGA 2640

TGGCTCAGAT CAATCTGAAA AAGCTTTTCA TGAAGCTGTC AGGATTGCCA AAGAGGAACA 2700

AGCCACTCTT TATTTAGCTA CAATTATTAA TGATGCTGAA TTTACGACGA GTCCATTTTC 2760

TTTTGAAGAA TTGTATGATT TAGAAAAACA CAAGTCAGAA GAGATGCTCA CTGGAAAGGC 2820

AAAACAAGCG AGTGAAATTG GTGTAAAAAC AGTCAAAAAA ATTGTTGAGC TTGGGAGCCC 2880

AAAACGTTAT CTTGCCAACA CGATTTCTGA AAATTATGCC ATAGATTTAA TTGTGCTAGG 2940

TGCCACGGGA AGAGGTGCTA TTCAACGAAC CTTGATTGGT TCTACTACAG ATTATGTGGT 3000

TAACCATGCT TTGTGCAATG TCTTAGTCGT TCGCTAATTG TGAGTAAATA ATCAAGGAAA 3060

CAATGTTTTA TCGGGACGTT TTTTTAAATT AAACGAGAGG TGAGTTTATG AAAACAAAAA 3120

TCGGAAAAAC AGTTATCTTG TCAGCATTTT TATTCACAAG TTTCCTTTTA CTGAGTGGTT 3180

GTACCTCGGC TGGCGAAGAG ATGGAAAAAA CAATTGATCG ACAGAAAGAA AAAGTCGATA 3240

AAACGGTCGA TAAGCAGAAA CATAAAAATG AAAATTCCAT GGAAAGTTAC GACGAAAAAG 3300

TTGACCGTTC TTTAGATAGT CAAGAAGACA AAATCGATAC TACTGAGTAA CGGGAAAGCT 3360

TTCAACTATC GGCTAGTAAA TAAATAATAT ACAAAAGGGA GCATCCACTG CAATTATTTC 3420

TCACAGCGGA TCCTCTCTTT TGTGGTTAAA AAGAAAAAGT TAACTGCTCC CCAACAGTTA 3480

ACTTTTAATC ATGTTGCACA TAGTTAAAAC ACTCTATATA TCATTTATAT CATTTGAAAC 3540

GGGTAGTGTA AAAAGAAATC TCTTGAAAAG AGTCTCCATT TATAGGGAAG ATTTCCCATA 3600

AAAAAGAGAA ACGTATAATC ACAGAGAGTT AGAGAAAGCA AAGGGAAATA AGTATGTTAA 3660

AATAAGAAGC AAGAAAGAAT ACAAGTATGC AGAAAATAAC AACTAATTAG GTGAAAACAA 3720

TGGAAATACA TTTTGAAAAA GTAACAAGCG ATAATCGTAA AGCAGTGGAA AACCTACAAG 3780

TTTTCGCTGA ACAACAAGCC TTCATTGAAT CAATGGCAGA AAATCTAAAA GAATCAGACC 3840

AATTTCCAGA ATGGGAATCG GCAGGTATTT ATGACGGCAA TCAATTAATC GGCTATGCCA 3900

TGTATGGACG TTGGCAAGAT GGGCGCGTTT GGCTAGACCG CTTTTTAATT GATCAACGGT 3960

TCCAAGGACA AGGTTATGGG AAAGCAGCCT GTCGACTTTT AATGTTAAAG CTAATAGAAA 4020

AATACCAAAC AAACAAGCTA TACTTGAGTG TTTATGATAC GAACAGCTCa GCAATTCGGT 4080

TGTATCAGCA ACTCGGTTTT GTATTTAATG GTGAATTAGA TACGAATGGT GAGCGGGTAA 4140

TGGAATGGAC ACATCAGAAT AAGTAGATAG AACAGGTGAG ACAAAAGCCA TAGCGCATTT 4200

TGTTTCATCT TTTTCTTTTA ATTGGCTTCA TTTGTTTACT TTTAGTAAGT AAAATGGTAA 4260

ATTTAGTTCA ATCTGTTTTG AAAAAAGAGT TACTGGAGGG TAAGGATGAA AATAGAATTT 4320

TTCCATGATG TTATTTGTAG TTTTTGTTTT CCAATGTCTG ACCGCATGCA TGAGATTCAA 4380

CAAGAATTTC CTGAAATTGA ATTGATTCAT CGGTCGTTTG CGTTAGGATG GTCAGCAAAC 4440

GATTTTGAAA CAATGTTTGG CTCACGAGCA GCAGTAAAAG AAGAAGTTTT GACGCATTGG 4500

GTGCATGCCA ATCAAAACGA TTCAAAGCAT CGTTTTAATA TTGAAGGGAT GAGGAAGCAA 4560

CGATTTGATT TTCCaACATC AAAGAATGTA TTATTAGCAG CAAAAGCAGC TGGATACATT 4620

GGCAACCAAG ACACCTATTG GTTACTGTTT GATAAGTTAC AAGAAGGTTT ATTTGTGAGA 4680

AGTTTAAATA TTGAAGAGCC AGAAGTAATT GAGAAGCTAG TAAAAGAGAC AACGATTGAT 4740

TTTGCTTTAT GGAAAGAAGC TGTTGCTTCC GAGGCTGTTT GGACGGCGGT TCAAGAAGAT 4800

TTTGCGCTAG CTTCTGCTTA TGGTTTACAA GGAGTTCCCG CACTAATCAT TAATCAAAAA 4860

TATCTAATTA ATGGTGCAGT ACCGAAACAG CAAATCAGCC AAACTATTCA AAAAATTCTC 4920

GCAGAAGAAA AACAACAGCA ACCGTTAGTT TCATTAACTC CTTCTTCTGA AGAGTCAGCT 4980

TCGTGTGAGT TTAAGGAAGG CACATGGACG TGTCATTAAG CAGCTTAAAA TAACGTTATT 5040

CTTTGACAAA ACGAAAGACT TCTGTAATAG TTAATGCGGG GGTGATAATA TGTTTAACGC 5100

ATATAGAGAA TACTGGAATA ATATTACAAC AATGAATGCG TCACAACaCG AGCACAATAT 5160

TGGTGGCCAC AATTAATTAA TTATTTGGTG TTAGGGATTT ATTCAGCAAT TACGGGCGTT 5220

TACAGATATA TTGAAGTAAC ACCCAATGAT GGAACAATCA TTAAAGAATG GAATACTGTG 5280

ACCATTATAT TTTTCTTGCT GACTGCTTTA ATTTGGTTGG CAAACTTTAC TGTTCGAGCA 5340

AGAAGATTGC ATGATCGCGA CCACAGTAAC TGGTGGATTT TATTCTATTT ACTTCCTTTT 5400

ATCGGTACTT TAGTGATCTT CATTACATTA ATTTTGCCAA GTAAAAGTCA CACAAGATGG 5460

CCAGTGAATC AATCAGAAAT TTAATTTGCT GGTTTCgTTT AGCTGATTAT CTTAAGAGTG 5520

AAAATAGGAC ATAAGTCATT TCGACTTATG TCCTATTTTC TTTTTTATTT CAGAAAAATA 5580

AAAATATCTT CTGTATGATG GCACCTCACA TTACGCGGTT CTTGATAAAG GGGCTATCTT 5640

TTTGGTAGAT AGGGAAAACC ATTAAGACGT TAAAAGTTTA ATAAATCGAG CAATTCTCTT 5700

TTGTTAGAAA ATAAGATAGC ATCATTAATT AAAATACTTG GAACACTGGT TTCAATAATT 5760

GGTTGATTGG TAATAATCAA GTCAAATTCT TCGAGTCGAT GGTCACCTAA TTGTTCTTCC 5820

GTAATTGCTA ACGTATTTGA ACGAATCCGA TGCAAAGCTA AAAtATCGTG GaCATAGTTT 5880

CCAATATATT GACTATGATT CAGGCCCATG TCGCTAATAA TCAAAATCTT TTTTGAGTCG 5940

AAACTCTGGT TAATTTCTGG AAAAGTATTA AGCAACCAAA AAATATAAGA AGATGAATAG 6000

TTTAATAGAT TTGTTTCTAA AAGAGCGGAA AAGAAATGCA ATTCTGCTTT TAATCGTGCA 6060

TATAAGTGAA GCTGAGTAAA ATGAATTTGT CGAGTGAAAA AGGTGGCCCT TTGTGTCAAT 6120

TCGGAGGTGT CAAAAGGCAC TAATTGAGAA AACAAATAAA GACAGGTTAA CAAGTGAACA 6180

GCGAAGTCGA CGACTTCTTC TGTTAATTGT TCAATCTGTG AAAAAGCATT AAATTTTAAA 6240

TTTTCACGTA TTCTTTGACT GATGACTTCT TGGTCTTCGG AGTAGTATTG CTGAAAAGCG 6300

CCTCTTTGAA AATATCGGAG AACTTCCCGA TAGGTTTCAC GGCTGAGTCC AGTGAAATGA 6360

CGACTGAGTA AAAGATAATC TTCCATTTTG ATTGGACTCT TTTTATCTGT TTTTTTCTGG 6420

AAACGTTGTT GAAATCCTTG ATGCTCTCTT AATAAATCAA CAATAAAATA CATAATAAAG 6480

AAATCGTGTT CATAACTGTT ATCCATaCTA ATAATATTTA ATTGGCATTG ATTACGACGC 6540

ATTAATTGTT TGATTTCCTG CTTATCTAAT TGATAAGGCA ATTCATAATG ACCATAATAC 6600

ATTAAAAAGA AAAAGGTTGT AAAAATACGT AAAGGCAATT CTTCTGTTTG GCTAGTTAAT 6660

TGATAGCCTT GATCAACGAT AACGCGAAGA GAGTAAGTTT CCAAGACCTG GTTACACTGC 6720

TTCAATTTTC GAGCAAAAGA AGAGCGACTA ATATTGATGG TATCGCAATA ATATTCGATG 6780

GTTTTGCCAG GATTTTGGAT AAACTCATGG AGCAATTGGG TGGTCGTTGA CGTTTTTGCC 6840

ATATCCGTCA GCACGTTTAA ATATAAGGAC ACATTAAAGT TTTCTAAACG ATAGCCTAAT 6900

TTATGCGATG TTTTAATGGT TAAACTTTCG CCCCAACGCT GTTTAATCTC CTCAATATAA 6960

CGAATAAAAG TAGACTCAGA ACTACCAATA TATTTTGCAA GTTCTTTTTT AGTCATCCAT 7020

TTATCAGAAA TGGATAAGAC ACGAATGATT TCTTCCATTT TCCGATAATT TTTTTCTAAA 7080

AACCTAATCA TATTGTTTCC TCCCAAGTCC TTTACTTTAA TGAAACAATC ATCGAAACAA 7140

TAACGCAATT TCAGAAATGA AACCGCTTGG TATGAAAGCT GGTGTGATTT AATGAGTTCC 7200

TTCTCTGTAT AATAATTATC AGATTAGAAT AATGAAATTT TACGTGCTAG GCAATTCtaA 7260

TTTAACTGkT AAwTCmTaAA AGgAAAGGAG AGAACAGACA TTAAATTACT AAAGAmATTT 7320

GGGGTTTTCG GCGTAGTCTT TTTACTTGTC TCATCCTATT TCATACCGCT CATTGGCTAy 7380

GCAGAAACAG CCAAAGAAGT TGAAATTACA TCCGCTCAAA TGATAACAGA TGAGAATGAT 7440

AAAACGAATA TTAATATCGA GTTAAATCTT CTCAACCAAA CAGAGCAGCC ATTACAACGA 7500

GAAATTCAAT TGAAAAATGC ACAGTTCATG GATACTGCTG TAATTGAAAA AGACGGATAT 7560

TCTTACCAAG TGACTAATGG TACGCTTTAT CTGACTTTGG ACGCACAAGT AAAAAAGCCG 7620

GTACAGCTTT CGTTAGCTGT TGAGCAAAGT TCGCTTCAAA CAGCTCAGCC ACCTAAGTTA 7680

TTGTATGAAA ACAACGAATA TGATGTTTCA GTTACTTCTG AAAAAATAAC AGTAGAGGAT 7740

TCTGCTAAAG AATCAACTGA ACCAGAAAAA ATAACTGTAC CAGAAAATAC GAAAGAAACT 7800

AACAAAAATG ATTCGGCTCC AGAAAAAACA GAACAGCCGA CCGCAACAGA AGAGGTAACC 7860

AATCCATTTG CAGAAGCAAG AATGGCGCCA gCTACTTTGA GAGCGAATCT GGCACTGCCT 7920

TTAATTGCAC CACAATACAC GACGGATAAT TCTGGGACTT ATCCGACAGC TAATTGGCAG 7980

CCCACAGGCA ATCAAAATGT GTTAAACCAT CAAGGGAATA AAGACGGTAG TGCACAATGG 8040

GACGGCCAAA CGAGTTGGAA TGGGGACCCT ACTAATCGCA CAAATTCTTA TATTGAGTAT 8100

GGCGGTACAG GAGACCAAGC CGATTATGCC ATCCGAAAAT ATGCTAGAGA AACAACAACA 8160

CCAGGGCTTT TTGATGTATA TCTTAATGTG CGTGGGAATG TTCAGAAAGA AATCACGCCA 8220

```
TTGGATTTGG TCTTAGTCGT TGACTGGTCC GGTAGTATGA ATGAAAACAA TCGGATTGGT    8280
GAAGTTCAAA AAGGAGTGAA CCGTTTTGTT GATACATTGG CAGATAGCGG TATTACCAAT    8340
AACATCAACA TGGGCTATGT TGGCTACTCA AGTGACGGTT ATAATAACAA CGCCATTCAA    8400
ATGGGGCCGT TTGATACAGT CAAAAATCCA ATTAAAAATA TTACGCCAAG TAGCACTAGA    8460
GGAGGAACTT TCACTCAAAA AGCATTAAGA GATGCTGGTG ATATGTTAGC AACGCCAAAT    8520
GGACATAAGA AAGTCATTGT ACTTTTAACG GATGGCGTCC CAACCTTCTC TTATAAAGTG    8580
AGTCGAGTTC AAACAGAGGC GGATGGTCGC TTTTACGGGA CACAATTTAC GAATCGACAA    8640
GATCAACCAG GTAGCACTTC TTATATCTCT GGTAGCTATA ATGCGCCAGA TCAAAACAAT    8700
ATCAATAAAC GGATTAACAG TACGTTTATC GCCACGATAG GTGAGGCAAT GGTCTTAAAA    8760
CAACGTGGGa TTGAAATACA TGGATTGGGC ATTCAATTGC AAAGCGATCC ACGAGCTAAT    8820
TTATCTAAAC AACAAGTTGA AGATAAAATG CGTGAGATGG TGTCAGCCGA TGAAAATGGA    8880
GACCTTTATT ATGAATCCGC GGATTATGCA CCAGACATTT CTGATTATTT AGCGAAAAAA    8940
GCCGTTCAGA TTTCAGGAAC GGTTGTAAAC GGAAAAgTAG TTGATCCAAT TGCTGAACCT    9000
TTTAAATACG AGCCAAATAC ATTATCAATG AAAAGTGTGG GTCCTGTTCA GGTTCAAACA    9060
TTACCAGAAG TGTCGCTAAC AGGCGCTACA ATTAATAGTA ATGAGATTTA TTTGGGTAAA    9120
GGGCAAGAAA TTCAAATTCA TTATCAAGTA CGTATTCAAA CAGAGTCAGA AAACTTCAAA    9180
CCTGATTTTT GGTATCAAAT GAATGGTCGG ACAACGTTTC AGCCATTAGC CACGGCCCCT    9240
GAAAAAGTTG ATTTTGGGGT TCCTTCGGGA AAAGCACCTG GCGTGAAGTT AAACGTGAAA    9300
AAAATCTGGG AAGAGTATGA TCAAGACCCG ACAAGTCGGC CAGATAATGT GATTTATGAA    9360
ATTAGTAGAA AGCAAGT4AC TGACACAGCC AACTGGCAAA CTGGGTATAT TAAATTATCA    9420
AAACCAGAAA ATGATACCAG CAATAGTTGG GAGCGCAAAA ATGTAACCCA ACTTTCCAAA    9480
ACCGCGGATG AAAGCTATCA AGAAGTTCTT GGGCTTCCCC AATACAACAA TCAAGGACAA    9540
GCTTTCAATT ATCAAACAAC CCGTGAATTA GCAGTTCCTG GTTACAGTCA AGAAAAAATC    9600
GACGATACTA CTTGGAAAAA CACGAAGCAG TTCAAGCCAT TAGATTTAAA AGTAATCAAA    9660
AATTCTTCCT CAGGTGAGAA AAACTTAGTG GGAGCCGTCT TTGAATTGAG TGGTAAAAAT    9720
GTTCAAACAA CATTAGTGGA CAATAAAGAT GGTAGCTATT CCTTGCCAAA AGATGTGCGC    9780
CTACAAAAAG GGGAACGCTA TACATTAACT GAAGTAAAAG CACCTGCAGG ACATGAGTTA    9840
GGCAAGAAAA CGACTTGGCA AATTGAGGTG AGTGAGCAAG GCAAAGTAAG CATCGATGGA    9900
CAAGAAGTGA CCACCACAAA TCAAGTTATT CCATTGGAAA TTGAAAATAA ATTTTCTTCT    9960
TTGCCAATCA GAATTAGAAA ATACACCATG CAAAATGGCA AACAAGTGAA CTTAGCAGAG   10020
GCGACTTTTG CGTTGCAAAG AAAAAATGCT gCAAGGAAGT TACCAAACTG TGGCAACTCA   10080
AAAAACAGAT ACTACAGGAT TGAGCTATTT TAAAATTAGT GAACCTGGTG AGTATCGAAT   10140
GGTGGAACAA TCAGGACCAT TAGGCTACGA CACTCTTGCT GGAAATTATG AATTTACTGT   10200
```

```
TGATAAATAT GGGAAAATTC ACTATGCAGG CAAAAATATT GAAGAAAATG CGCCAGAATG   10260
GACACTGACA CATCAAAATA ATTTGAAACC TTTTGACTTA ACAGTTAATA AAAAAGCCGA   10320
TAATCAGACG CCACTTAAAG GAGCGAAATT CCGTTTAACA GGACCAGATA CGGATATTGA   10380
ATTACCAAAA GATGGCAAAG AAACGGATAC TTTTGTTTTT GAAAACTTAA AACCAGGGAA   10440
ATATGTTCTA ACAGAAACCT TTACGCCAGA AGGATATCAG GGGTTAAAAG AACCAATCGA   10500
ATTAATAATT CGTGAAGATG GTTCAGTCAC GATAGATGGG GAAAAAGTAG CAGATGTTTT   10560
AATTTCTGGA GAGAAGAATA ATCAAATTAC TTTAGACGTT ACGAACCAAG CAAAGGTTCC   10620
TTTACCTGAA ACTGGTGGCA TAGGACGCTT GTGGTTTTAC TTGATAGCGA TTAGTACATT   10680
CGTGATAGCG GGTGTTTATC TCTTTATTAG ACGACCAGAA GGGAGTGTGT AATCAATGAA   10740
AAACGCACGT TGGTTAAGTA TTTGCGTCAT GCTACTCGCT CTTTTCGGGT TTTCACAGCA   10800
AGCATTAGCA GAGGCATCGC AAGCAAGCGT TCAAGTTACG TTGCACAAAT TATTGTTCCC   10860
TGATGGTCAA TTACCAGAAC AGCAGCAAAA CACAGGGGAA GAGGGAACGC TGCTTCAAAA   10920
TTATCGGGGC TTAAATGACG TCACTTATCA AGTCTATGAT GTGACGGATC CGTTTTATCA   10980
GCTTCGTTCT GAAGGAAAAA CGGTCCAAGA GGCACAGCGT CAATTAGCAG AAACCGGTGC   11040
AACAAATAGA AAACCGATCG CAGAAGATAA AACACAGACA ATAAATGGAG AAGATGGAGT   11100
GGTTTCTTTT TCATTAGCTA GCAAAGATTC GCAGCAACGA GATAAAGCCT ATTTATTTGT   11160
TGAAGCGGAA GCACCAGAAG TGGTAAAGGA AAAAGCTAGC AACCTAGTAG TGATTTTGCC   11220
TGTTCAAGAT CCACAAGGGC AATCGTTAAC GCATATTCAT TTATATCCAA AAAATGAAGA   11280
AAATGCCTAT GACTTACCAC CACTTGAAAA AACGGTACTC GATAAGCAAC AAGGCTTTAA   11340
TCAAGGAGAG CACATTAACT ATCAGTTAAC GACTCAGATT CCAGCGAATA TTTTAGGATA   11400
TCAGGAATTC CGTTTGTCAG ATAAGGCGGA TACAACGTTG ACACTTTTAC CAGAATCAAT   11460
TGAGGTAAAA GTGGCTGGAA AAACAGTTAC TACAGGTTAC ACACTGACGA CGCAAAAGCA   11520
TGGATTTACG CTTGATTTTT CAATTAAAGA CTTACAAAAC TTTGCAAATC AAACAATGAC   11580
TGTGTCGTAT CAAATGCGTT TAGAAAAGAC CGCTGAACCT GACACTGCGA TTAACAACGA   11640
AGGACAATTA GTCACGGACA AACATACCTT GACTAAAAGA GCCACAGTTC GTACAGGCGG   11700
CAAGTCTTTT GTCAAAGTTG ATAGTGAAAA TGCGAAAATC ACCTTGCCAG AGGCTGTTTT   11760
TATCGTCAAA AATCAAGCGG GGGAATACCT CAATGAAACA GCAAACGGGT ATCGTTGGCA   11820
AAAAGAAAAA GCATTAGCTA AAAAATTCAC GTCTAATCAA GCCGGTGAAT TTTCAGTTAA   11880
AGGnnTTAAA AGATGGCCAG TACTTCTTGG AAGAAATCTC TGCACCAAAA GGTTATCTTC   11940
TGAATCAAAC AGAAATTCCT TTTACGGTGG GAAAAAATTC TTATGCAACG AACGGACAAC   12000
GAACAGCACC GTTACATGTA ATCAATAAAA AAGTAAAAGA GTCAGGCTTC TTACCAAAAA   12060
CAAATGAAGA ACGTTCTATT TGGTTGACGA TTGCAGGCCT GCTAATCATT GGGATGGTAG   12120
TCATTTGGCT ATTTTATCAA AAACAAAAAA GAGGAGAGAG AAAATGAAGC AATTAAAAAA   12180
```

AGTTTGGTAC ACCGTTAGTA CCTTGTTACT AATTTTGCCA CTTTTCACAA GTGTATTAGG 12240

GACAACAACT GCATTTGCAG AAGAAAATGG GGAGAGCGCA CAGCTCGTGA TTCACAAAAA 12300

GAAAATGACG GATTTACCAG ATCCGCTTAT TCAAAATAGC GGGAAAGAAA TGAGCGAGTT 12360

TGATAAATAT CAAGGACTGG CAGATGTGAC GTTTAGTATT TATAACGTGA CGAACGAATT 12420

TTACGAGCAA CGAGCGGCAG GCGCAAGCGT TGATGCAGCT AAACAAGCTG TCCAAAGTTT 12480

AACTCCTGGG AAACCTGTTG CTCAAGGAAC CACCGATGCA AATGGGAATG TCACTGTTCA 12540

GTTACCTAAA AAACAAAATG GTAAAGATGC AGTGTATACC ATTAAAGAAG AACCAAAAGA 12600

GGGTGTAGTT GCTGCTACGA ATATGGTGGT GGCGTTCCCA GTTTACGAAA TGATCAAGCA 12660

AACAGATGGT TCCTATAAAT ATGGAACAGA AGAATTAGCG GTTGTTCATA TTTATCCTAA 12720

AAATGTGGTA GCCAATGATG GTAGTTTACA TGTGAAAAAA GTAGGAACTG CTGAAAATGA 12780

AGGATTAAAT GGCGCAGAAT TTGTTATTTC TAAAAGCGAA GGCTCACCAG GCACAGTAAA 12840

ATATATCCAA GGAGTCAAAG ATGGATTATA TACATGGACA ACGGATAAAG AACAAGCAAA 12900

ACGCTTTATT ACTGGGAAAA GTTATGAAAT TGGCGAAAAT GATTTCACAG AAGCAGAGAA 12960

TGGAACGGGA GAATTAACAG TTAAAAATCT TGAGGTTGGT TCGTATATTT TAGAAGAAGT 13020

AAAAGCTCCA AATAATGCAG AATTAATTGA AAATCAAACA AAAACACCAT TTACAATTGA 13080

AGCAAACAAT CAAACACCTG TTGAAAAAAC AGTCAAAAAT GATACCTCTA AAGTTGATAA 13140

AACAACACCA AGCTTAGATG GTAAAGATGT GGCAATTGGC GAAAAAATTA AATATCAAAT 13200

TTCTGTAAAT ATTCCATTGG GGATTGCAGA CAAAGAAGGC GACGCTAATA AATACGTCAA 13260

ATTCAATTTA GTTGATAAAC ATGATGCAGC CTTAACTTTT GATAACGTGA CTTCTGGAGA 13320

GTATGCTTAT GCGTTATATG ATGGGGATAC AGTGATTGCT CCTGAAAATT ATCAAGTGAC 13380

TGAACAAGCA AATGGCTTCA CTGTCGCCGT TAATCCAGCG TATATTCCTA CGCTAACACC 13440

AGGCGGCACA CTAAAATTCG TTTACTTTAT GCATTTAAAT GAAAAAGCAG ATCCTACGAA 13500

AGGCTTTAAA AATGAGGCGA ATGTTGATAA CGGTCATACC GACGACCAAA CACCACCAAC 13560

TGTTGAAGTT GTGACAGGTG GGAAACGTTT CATTAAAGTC GATGGCGATG TGACAGCGAC 13620

ACAAGCCTTG GCGGGAGCTT CCTTTGTCGT CCGTGATCAA AACAGCGACA CAGCAAATTA 13680

TTTGAAAATC GATGAAACAA CGAAAGCAGC AACTTGGGTG AAAACAAAAG CTGAAGCAAC 13740

TACTTTTACA ACAACGGCTG ATGGATTAGT TGATATCACA GGGCTTAAAT ACGGTACCTA 13800

TTATTTAGAA GAAACTGTAG CTCCTGATGA TTATGTCTTG TTAACAAATC GGATTGAATT 13860

TGTGGTCAAT GAACAATCAT ATGGCACAAC AGAAAACCTA GTTtCACCAG AAAAAGTACC 13920

AAACAAACAC AAAGGTACCT TACCTTCAAC AGGTGGCAAA GGAATCTACG TTTACTTAGG 13980

AAGTGGCGCA GTCTTGCTAC TTATTGCAGG AGTCTACTTT GCTAGACGTA GAAAAGAAAA 14040

TGCTTAATTT CTAGCATCAC CGAAGAAATT TTTAGAAAAA CAAAGAGCCT GGGCCAATCA 14100

CTGTCCCAGG CTCTCATGCT TTATTTTTAA GGAGGAAGCA ATGAAGTCAA AAAAGAAACG 14160

TCGTATCATT GATGGTTTTA TGATTCTTTT ACTGATTATT GGAATAGGTG CATTTGCGTA 14220

TCCTTTTGTT AGCGATGCAT TAAATAACTA TCTGGATCAA CAAATTATCG CTCaTTATCA 14280

AGCAAAAGCA AGCCAAGAAA ACACCAAAGA AATGGCTGAA CTTCAAGAAA AAATGGAAAA 14340

GAAAAACCAA GAATTAGCGA AAAAAGGCAG CAATCCTGGA TTAGATCCTT TTTCTGAAAC 14400

GCAAAAAACA ACGAAAAAAC CAGACAAATC CTATTTTGAA AGTCATACGA TTGGTGTTTT 14460

AACCATTCCA AAAATAAATG TCCGTTTACC AATTTTTGAT AAAACGAATG CATTGCTATT 14520

GGAAAAAGGA AGCTCCTTGT TAGAAGGAAC CTCCTATCCT ACAGGTGGTA CGAATACACA 14580

TGCGGTCATT TCAGGCCATC GTGGTCTCCC TCAAGCCAAA TTATTTACAG ATTTGCCAGA 14640

ATTAAAAAAA GGCGATGAAT TTTATATCGA AGTCAATGGG AAGACGCTTG CTTATCAAGT 14700

AGATCAAATA AAAACCGTTG AACCAACTGA TACAAAAGAT TTACACATTG AGTCTGGCCA 14760

AGATCTCGTC ACTTTATTAA CTTGCACACC GTATATGATA AACAGTCATC GGTTATTAGT 14820

TCGAGGACAT CGTATCCCAT ATCAACCAGA AAAAGCAGCA GCGGGGATGA AAAAAGTGGC 14880

ACAACAACAA AATTTACTAT TATGGACATT ACTTTTAATT GCCTGTGCGT TAATTATTAG 14940

CGGCTTCATT ATCTGGTACA AGCGACGGAA AAAGACGACC AGAAAACCAA AGTAGTATGA 15000

CGAAAAGGCT AAACATACTA AAAAAAAGAG TAAAAAAATA GCTTTTCAAT TTTTAATCCT 15060

CCTTATCGTG CATAATTGAA CCAGAGAAAC AGAAGTATTA ACGAAATAAC TAAAAGAGCA 15120

AGCCCTGAAT AAAAAGCGAC AAAGGGCCAA TCAATCGACT GTTTAAATTC CTGCCAAGTT 15180

TGGATTTTTC TGTTTTTTTT CGCGCTATCC TCAAGCGTGA GTAAATAATT CAATAGTAAG 15240

AGGAGTAGCA ACACCGTGAA ATCATTTGTG GTAAAAAGCA CATGTAAAAA TAGAATGACA 15300

AAGACAACAC GGGATAACAC TCGATTCCGC AAAATTAAAA ATAACTTAGC ACGCATAATA 15360

AACCACCATT TCTTaTCAGA GATAATGAAT CTGTTTTTGT CTACTCTTTA GTTATATCAT 15420

AAAATTCTTA ATAATGAAAA AATGACTCGA GAAAATAATT GAAAAAAGTT TTTTTTCCTG 15480

AATCATTATT TTCGTAAATA AAGAATAAAC GTGTTACTCT TGGCTTATCA AATTTGGAAG 15540

GAGTGTTAAA AATGAAATAT CTGGATATTA TTGCTTTAAT TTTATTGATT GTCGGAGGTT 15600

TAAACTGGTT ATTA 15614

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3169 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

TTCACAAAAA AATGAAAACA CAAATTTAAC TTGGAGGATG TTGAGTAGTG AGCAACGGGA 60

AAAAGATTTT nATTAGTCAT AGTTCAAAGG ATcAAGAATA TGTGGATGCG TTTATCCAAT 120

TATTGAAAAA GTTTGGGTTT CGAACACAGG ATATTTTTTA TAGTTCCACT ATTGAAACAG 180

GTGTACAGCC AGGAGAATTA ATTTTTGACA CGATTAAAAG GGAGCTAACC AATCAACCTG 240

TCATGCTGTA TTTTTTATCT GATCATTATT ATCAGAGTAT TCCATGTTTA AACGAAATGG 300

GTGCTTCTTG GATGCTATCA GATAAACACT ATCCAATTGC GTTAAATAAT TTTTCTATGA 360

AAGATATGAA AGGTGTTATT AGTAGTGAGC GTTTGGCGAT TGCGTTTAAT GATAAGACAA 420

GTACTAATGA AATAAACTGc TTATTAAAAA AGTTATCTCA TGACACAGAT GTACAAGCTG 480

AGCCAGATTT TGAACTAAAC GTTGAGAAAA ATATTCAGCC ATTCCAAAAT AAACTGACAC 540

AACTGATTCG GCAAGCAAGT TATTTAAAAC CTGATGAAAA GGGATATTTT GAAACGrTTC 600

tTAGCACTCA TCGTCCTGTG TATGGCACAG CAAAAGGGGT CTATGACTGT TTTAAATTAC 660

CTAGCTTAAT TGAGCCTAAA AGTTTAGGTT TAGACACGCT ATCAGAGGAT GAAAGCCATT 720

GGCTATTCTT TTTTCTTACG TGGGGAACTT TTCAAGAAGG CGAAAAAGTG CGGTTTAAAt 780

TGAAAAAGGa TAAGGcGTAT AACAATCGGG AATTCAGTGA TATTGGTAAA TGTAAAAATA 840

TTTACGTCTC TTATTTAGAG AAGGTAGAGT AATTGGAGGA GGTAATAAAT GAAAATGAAT 900

TGGcGTCAGT TGGTTGGTAC ATATCGTATT TCTGATAATG GTAAACAATA TCAAAATGAC 960

TGGTTAAATA AGCAAGTTGA TTTAGCCACA GAACTTCGTA AATCTTTTGA TAGTGACTAT 1020

CGTAGAGTAA TTAAAAGTGA TGCATTTaGA CGCTTwnCAa GAkAwAACAC cAAGTTTTyC 1080

CTTTwGrAnC GTAATGaTTT TgGTAAGAAC aCGTTTAACA CATAGCTTAG AAGTGGCAAT 1140

GCATAGCAGG GATATACTTA GATTAGTTAT TTCACAGTTG AACGAACGAA ACATTGAGGT 1200

TATGGAATTA TCAGAATGCT ATCGGTTGTT AGAAACAGCT GCATTAATCC ATGATATCGG 1260

CAATCCCCCT TTTGGACATT TTGGGGAAGA AGCCATTCGT ATTTGGTTTG AAAAAAATGG 1320

TCCTTTGTAT GCTTGTTGGT CTGATTTTTC GGAGCAACAA CAAAATGATT TTCTTAACTT 1380

TGAAGGAAAC GCTCAAACCA TCCGACTTTT AACAAAACTG CATCATGATA ATGGGACTTC 1440

TAAAGCAGGA ATGAAACTAA CAGCCACGAC ATTAGATACA GTTATCAAAT ATACGGCATC 1500

TTCAGATCAA TTAAATAAGC AACAGCTTTT GTCTAAAAAA GTTGGCTATT TTTACTCGGA 1560

AAAACAAGTG TTTAACGCCA TTAAATTTGC GACAGGAACA CmAAATAAGC GACACCCATT 1620

AGTCtTTTTG TTAGAAGCTT CAGATGATAT CGCCTATACT TTTTCAGATA TCGAAGATGC 1680

CTATAATTAT GGTTTATACA GCTATCaAGA TTTAAAGCaG TTTGTTGATG CGCAAACAGG 1740

AACAAAAAAA TTTTTAATGT TAGAAAAAGA TAAAACAGAA GAAGCTACCT TACAAGAATT 1800

TTTGAGAAAG ACACAAAGAG CTGTTTGTCA AAGCGTAGCG AAAAATTTTG CTAATCATTA 1860

TGACTTAATT ATGAGAGGTA TCTATCATAA TGAAATTATC ATTGAAGATT GTGATGAAGT 1920

GAAATGCTTT CAGGCATTAA AAAACTTTTC TAGAGTCTAT GTTTTTCAAA CAAAAACAAT 1980

TTTGGATCAA GAGGTATTGG GCTTCAATAT CATCAATCGT TTACTAGATG AATTTGTTCC 2040

AGTCGTACTG AAATATGAAA AAGTTTCCAT GAATAAGTAT GAGGAGCGAA TCTTTAATAA 2100

TATTTCGGAA AGCGCGAAAG CGTTGTATAG AAGAGAAGCA AAGAACGCTA CAGAAGCAGA 2160

AAAAGACTAT TATCGTTTGA AAATGGCGGT GGACTTTGTA TGTAATATGA CGGATGGATA 2220

TGCCAAAAAA GTTTATGATA CATTATTCAC TTAAAATTTG AGAAAAAGCT TCTCAGCGGT 2280

TGCCGCCGCT GAAGAAAACT AGTTCATTGC ACAGGCTGAT CGTTTCAGTA AAAATTTTAA 2340

AGATGCTAGT AATTTCACTG AAAATCGGGT ACACTAAAAT TAGCTAGTGT AACTAGCAAA 2400

AAAGGCAACT GGAATAGATA GGTTGTGCTG TCTATGAAGG GCTTTGTTTG TCAGGTGGTG 2460

CTGACATGCA AAGTGCCTTT TTTCTTTTTC TCAACATTTT AAGGACAAAA AACGATTGCA 2520

ACATGTCACA ATCGCTAACC TCGGAGAAAA AGGAATTTTT CTCCAATGGC CATTTCGATA 2580

ACCTGTTGTG TGTTTCGAAA CATTCAGAAC GGTGGTGCGT TCTTTCAGTG CCTAATTATC 2640

CAATTCGTTT CTTGTTCCTC TATTTACAGT ATACTGAATG AAAGCCTATT CTGCAATGAT 2700

AGAATCAGGT TCCTGTGATT AAATAACATA AAAAATGGGC TCCTTATATA AAGAAAAATA 2760

ACAAAAGAAA TTTTGAATAT TGACGTTTTT TTTAGGTACG TTATGATAAG CGAGAGGTGA 2820

TTTTATGTTA AAAACGCAAA TTGTTGATTA TTTACGCAAT CAAACAGCTT TTTTTGAACC 2880

ATCTTTAGTC AGTGAAATTT TCACAGCAAG CAATATTGCA ACGACATTTT CTATCAAACG 2940

AAATACAGCT AGTCACTATT TAAATCAGTT AAACGAAGAA GGAATTCTAG TTAAGATCAA 3000

TACACGACCG GTTTACTTTT TTCATAAAgA AGCGTTTCAG CAACAAAACT ATCTTTTAAA 3060

AAGGAACTGT TTATCCAAGT TTCCAAGAAA TGATCGATGA ACAGCCTGTT TTTGGATCGC 3120

CAAAGnGGAn TTTTnTTCAG TCGGGTCATn GGGTATCGCG GGTAGTTTA 3169

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17087 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

GTCTATACAA GTTCACAAGG AGGCGCCAAA CATGACCATT TTAGAATTAA CATTCGCAAC 60

GAACCAACTA ACGACGACTT TTTTTAACCA TAAATTGGAA CCGCTTCACC AAAAAACATT 120

CCCTTTTGAA GCCTTGACTA TCGAAGAAAC GGTTCCTGAA CTGTGCCAAT TGATTTTGGC 180

AGAGGCGGCA CCGATTCTGG GGGGGATTGT GGCTGATTTC CCAGCGGACT TCTTTTCGCC 240

AGAGGAAACG TTGGCGCTGG TTCCGCAAGC GGTGATTCAA GCATTCGCGG AACAGTTGAA 300

CACGCCACTT TTAACGGCAG CCGAACAAGC GGCGGCACCG AATCTGCTAG AAGCATTTGA 360

AGAAAAACGG cAGTATCTGG CTTGGCATTT GGACTATTAT GGCTACGTGC CTGGGAAAAA 420

TGAATATGCG GTGGAGTCCC TGTTAACTGT CGGCAATGGC TTTTTAGGCT TGCGCGGCAC 480

GACACCTGAG ATGACGATTT CTGATGATCA TTATCCAGCG ACCTACATTG CGGGGCTTTA 540

TAATACAGCG GCTTCGGAAG TTGCGGGTCA AGTAGTGGAA AATGAAGATT TCGTCAATGC 600

TCCTGACAAC CAACACATCG CTTTAAAAAT TGGCGATGCC ACTGACTGGT TAACGATTTC 660
TCCAGACACG CTACAACAAT TACATCGTCA ATTGAACTTG AAAACGGGGT TGTTTGTCGC 720
AGAAATGATT TTGAAAGATG CGGACAATCA GCAAATTAAA TTAACGACGA AAAAAATCGC 780
CAATATGGCA CAGCCGAATG ACTATCATCT GCAATATACG TTTGAGCCGT TAAACTTTTC 840
AGCCCCGATT ACTTTGAAAA CCGTCACGGA TGGCAGTGTC TACAACTATA ACGTGGCGCG 900
GTATCGCAAC TTAACAGCGA AACATTTTCA GGTGACAGCA TTAAGCGCGC AGGAAAAACAA 960
AACCGTGATT GAGGTTTGTA CCAATCAATC CAACTTAAGC GTCCGAGAAA CGTCTTTGAT 1020
AACCGGTGAT TTTTTTGAAA AAGAAGCCAT TACGATTCAA GAAGAAGCGG AGAAAATCGC 1080
CCAAGTGGTG ACCGTGATGG CGCACCAAGG AACGCGCTAT ACGCTAGAAA AACAAGTCTT 1140
TGTACAAGCC AGCCACGCCG AACAAAGTTG GCAGGTACCG TTTACGCCGA AGGACTCTTT 1200
TGCGGCCGCC GCTCAGGAAA GCGCCCGCGC TTGGCAAACT CTTTGGCAAC AAGCCAACAT 1260
TACCGTGACA GGCGATTTGA TGTCTCAAAA ACTTTTACGG ATTCACAGTT ATCACTTGCT 1320
GGCTTCGGCT TCGCCTTTTA GTAACCAAGC GCAAGCATTG GACGTCTCGA TTACCGCCCG 1380
AGGCTTGCAT GGTGAAGCGT ATCGTGGACA TATTTTCTGG GATGAAATTT TTATTTTACC 1440
TTTTTATATT CAACATTATC CTGACACAGC GAAACAATTG TTGCTGTACC GCTACCACCG 1500
ATTAGAGAAG GCGAAGGAAA ATGCGGCGGC CAGCCAGTAT CGTGGCGCAA TGTATCCGTG 1560
GCAATCAGGG CGTGATGGGC GTGAAACCAC GCAAAAACTT CACTTAAATC CGTTGAATGG 1620
ACATTGGGGC GAAGATCATA GTATTTTGCA ACGGCATGTT TCCTTGGCGA TTGCTTACAA 1680
TGCTTGGTTG TATTGGCATA GTACACAGGA TCATGAATTT ATGAAACAAT ATGGCGGCGA 1740
AATGTTGTTA GAAATCGCTC AATTTTGGAA CAGCGCTGCG ACGTTAGATG ACGCAACAGG 1800
ACGGTTCTTT ATTGATAAAG TGATGGGACC TGATGAGTTT CACGAGGGCT ATCCAGACCA 1860
AGCGGAAAGC GGCTTGAAAA ACAACGCCTA TACGAATTTG ATGGTGGTTT GGCTATTCGA 1920
AGAATTAACC AACATTTTGG CACTTTTCTC TGAGGAAGAA CAGGCACAAC TTTTTGCCAA 1980
AACGCAGACC ACTTCGGCTG ATTTAGCCCG CATGCAGCAA ATTCAAAATT CACTGGAAAT 2040
TGAAGTTAAT TCTGACGGTA TTATTGCGCA ATATGAAGGG TATTTCGGCT TAAAAGAGAT 2100
TGACTGGACA GCGATGAAAG AGAAATACGG CAATATTTAC CGCATGGATC GAATTTTAAA 2160
AGCCGAAGGA GAATCCCCCG ATGATTACAA AGTGGCGAAA CAAGCCGACA CGTTAATGTT 2220
GTTCTACAAC TTAGACAAAA CTCGGGTGGA TCAAATTTTA GAAGATTTAG GGTATCAATT 2280
GCCTGCGGAT TATTTGGAAA AAAATCTTTT GTATTATCTA AAACGGACGT CTCATGGTTC 2340
GACTTTGTCA CGAATTGTTC ATGCCCAATT GGCAGAAATG GCGCAATTCC ATGAGCTATC 2400
TTGGCAACTT TATCAAGAAG CGTTGTATTC CGATTATCGT GACATTCAAG GTGGCACGAC 2460
CGCTGAAGGG ATTCATACAG GGGTCATGGC TGCGACGATC CACGTTACTT TAGCGACGTA 2520
TGCAGGCGTT GACACGCGCC AGAACGAGTT ATCAATTTGT CCAAACTTAC CTGAACATTG 2580

GCAAGCGCTG GCCTTCCAGT TTATTCATCA AGGCGTGACG TATCAATTTT CATTAACACA 2640

AACCAGTGTC ACGATTACAG CAGATAGAGA CACACAATTA TTGGTTCAGG GCGCGTTGAT 2700

TCCGTTAACC GCCGAACGAC CAAAAGAAGT ACATTATCAA TAAAGGAGCG AACCACAATG 2760

ACGATTGGAT TTATTTTTGA TTTGGATGGT GTCATTACTG ACACCGCCAA ATTTCATTAT 2820

CAAGCTTGGA AAGCGTTGGC CGATTCGCTA GGTATCCCCA TTGACGAAAC CTTCAACGAA 2880

ACATTAAAAG GCATTAGTCG GATGGATTCC TTGGACCGTA TTTTAGTCCA TGGCCACCGC 2940

GAAAATGCGT TTACTCCAGC AGAAAAAGAA GCCTTAGCTC AGCAAAAAAA TGAnCaCTAC 3000

GTTCAATTAT TAGAACAATT AACAACAGAA GATGTTTTAC CTGGTGTCGT GCCATTGCTA 3060

CAGCAAGCCC AAGCACGTCA CATTCCCTGC GCCGTGGCTT CGGCTTCAAA AAACGCGCCC 3120

CTGATTTTGG AAAAATTAGG CGTGCGCGCG TAcTTCGCCA CAATTGTCGA TCCCGACTCG 3180

TTAAGTAAAG GCAAACCTGA TCCTGAAATC TTTTTAGCCG CCGCTGACAG TATTGGCGTG 3240

ctACCGCAAA ACGCCATTGG CTTTGAAGAT GCACAATCAG GCATTGACGG CTTGAAAGCC 3300

GCGGGCATCT ATGCCGTCGG CTTGTCGGCC AGCCAACCCT TAATCGGAGC CGACATGCAA 3360

GTTTCTGAAA TGACTGAACT AAGCGTTGAT GCACTGCTCA ATCGCTAAAG AATGACAAGA 3420

ACCTTTAGCA GCAGGCACCT CTTTTCATCT ACATTGTATC TATGCTAGAA TGTAGATGAA 3480

AAGAGAGAGA TGCGTCAACA TACCTCTCTA GTGTAGAGCC GTTTGAGACG GTGACCAATT 3540

TAGTTATAAA AAATAACCGT ACCGGTCAAA GTAGACGGTT ATTTTTTATT GTCATTACAC 3600

TCCATTATTG GATTAAAAAG GATTGAGTTG AACGTGGTTC TGTGGTAGAC TTTTATTAGA 3660

AACAAGTAGA TGTTTCTTGA TGATCAGGCA TCATAATAAC TCAATACTGA AATCGGAAGT 3720

AATAGCCGAT AGTTGGCGAG CAACAGAAAA ACTCGATATA GAATTGGACG TAGAGCCAAT 3780

AGTTGGTGAG CAACAGAAAA ACCTGAATTA TATTTGAAGT TTAAAGAAGG CAAGCTAAGC 3840

TTGGCTTCTT TATTTATAGG AGAAACTAAT GATTGAGAAA AAAGATTTTG ACACAATTAG 3900

AAAGGCGTTC TTAATTTATG ACACCAACTT CAATAATCAT ATGTTCACTT ATACATATCA 3960

ACATCGGAAA ACAAAGAAAT TGTTAGAAAT GAACGTTCAA TTTGAAGCTG GGAATTTTAT 4020

GCATTTATGC GGAGTTGACT ATTATCAAAT AGATAAACAA gGGAGCTGTA AATTTCGTTT 4080

TAAGGCTACT CAATTTTACA GAGCTTTAAA AAATAATAAA GTTTCACTGC GCGGGATTAA 4140

ACCTAAAGAT GATGGAACCA CAGGCCAAAA ATTACAAGTA ATTCCTCTAT TAGAAATGCT 4200

TATTTCTCCA GGTGTACGAA TATGCGATGG AGGAAAATTT TATAATTTAC AGTATGAAAA 4260

AGCAATTCGT TCAGGCAAAA TGATTGTGGC GTTAACATGT AAAGAGAATA ACAAAAAGTA 4320

CGTTCCACAA TCTTTATTAT CGTTAATAAA TCAGCCACGA AAGAGCCAAT CGAAATCTTT 4380

AACTGAATCC CATGAAGTAA TTAAGATAAG TAAGTCTGAG CTGAATAGCA CCTCAGTAAT 4440

TGAAGTTTAC GATAAATTTT AACAAAAAAG CGGGCTAAAC AGACTATAGG CAAAGAGAAA 4500

AAGCCCAGAG GTTTTTCTCC TCTTTAACAG CTTTTAGGAt TTCACGGATT AACGCTAAAT 4560

```
TTATTCGTAC TTTTCGGGAA AATGCGCTAT AATGAAAATG AAAAGAGAGA TATGCTTCAA      4620
CATACCTCTC TGATGTAGAG CCGTTTAAGA CGGTGACTTT TTGAATTATT TAAAAATAAC      4680
CGTACTCGGT CAAAGTTGAC GGTTATTTTT TATTGTCATT TTTAACAGCT TTTTAGTTTC      4740
CAAGGACCAA TACTTTACTA TTAGCAACGA AGTTCACGGT CCTTTGTGCT TGTCTCCACG      4800
TTCATTTCTT TTGCTTATTG TTTAAATCCT ACTAACGGAA AAATGAAATC AACGAATTTT      4860
ATCCCCGCAT AAAAACATAG GGCAAACATC AAAATCTTAA TTAAGATTTT GACAATTGGT      4920
CGTTCAGTTA TCTTTTTCAT TTGTATACTC CTTTCCaATA TTTTTAAAGC TCCAACCTTT      4980
TTAATCTTTT TGACATCTAT ATTTTATCTA ATTATAATAA AACTCACAAT AGAGGAAGCA      5040
TCATTTTTTC ATTTCCTTTT TATTCCTTGA TTAGTAGTTC AATTTGTAAA CATATCATTC      5100
TGACAAAATA GAAAAGGAGG TTGTTATAGT ATCGAAACAA AATTATCTAC AAAGAAAAAA      5160
ACACTTCTTT ATCTCCTGTT CACTTTTTTG ACAGCCCTAC CCAGCATAGA AGCATTAACC      5220
GGAATGAGCT TAATAGCTAT AAAAAATAAC GAAAAAGCCT TGATTATCTT TGATAGTGTA      5280
CTATCTTTAA TTATTATTAT CGTGCTGTCA GTTTTCTATA AAAAAATAGT AAAGAAAAAA      5340
AAGCAAGCTA TCTTTAAAAG TCTATTACTA GCTATTGGCA TGTTCTTATT TTCTATTATT      5400
CTAAATTTAC TTTTCTCACG CATTCCTACT CCAGAAAATC AACATTATAT TAATGAAGTC      5460
CGAAATGTAG CACCTGTTTT AATGGCCCTT CATACTAGAC TCTTTTCACC TATAATAGAA      5520
GAACTCCTGA CTCGTGGCAT TTTTATGAAC TTATTTTTTA TAAAAAACaC CTCTCGAAGT      5580
GTCTTATGTA AAATATTTTT TTCTGGTCTT TTTTTTTGGT CTTAGCCATG GACTATTACC      5640
ATCAATTTCC ATGCTTTTTT ATTGTGGAAT GGGCTGGGTT TTAGCCTCTA CCTATTATTT      5700
CTGTAATAAC AATCTCACGT ATCCTCTTGC TATTCATTTG CTATTAAATA ATTTATAATA      5760
CACAAAAAaC CCGCCCACAT CAGAAAAATG ATGTGGGCGG GTTTTATAAT TATTTATTTT      5820
TkGTTTCAAA TAACGGACTA ACTGGTTTGT TTTCGTGGAT ACGTTTAATC GCATCTCCGA      5880
TTAATTCACC AACGCTGACT TCGTCAATTT TaACAATTTT ACGATCATCA GATAGGTAAA      5940
TAGAGTCTGT TACAACTAAG CGTTCAATTG CTGAATCTTC AATACGTTGT AAGGCTGGTC      6000
CTGATAAAAC GGGATGTGTA CAAGAAGCGT AAACGTCTTT TGCTCCGGCT TCTTTCAATG      6060
CGTTTGCAGC AAGAGAAATT GTCCCCGCTG TGTCAATCAT GTCATCAATT AAGACACATG      6120
TTTTGCCTTC CACATGACCA ATGATATTCA TTACTTCTGC TACGTTCGCT TTTGGACGAC      6180
GTTTGTCAAT GATCGCAATT GGTGCTTTTA AGAATTCTGC CAATTTACGC GCACGCGTTA      6240
CACCACCATG GTCAGGGGAA ACAACAACGA CGTCATCGCC TTGAATGCCG TGTTCGATGA      6300
AGTAATCGGC AATTAATGGT GCACCCATCA AATGATCCAC TGGAATATCA AAGAATCCTT      6360
GAATTTGGAC CGCATGTAAG TCCAATGTCA ACATTCTTGT TGCACCAGCT TTTTCAATCA      6420
TGTTTGCAAC AAGTTTAGCA GTGATTGGTT CCCGGGCACG TGCTTTACGG TCTTGACGTG      6480
CATACCCATA ATAAGGCATA ACCACGTTAA TTGTTTTAGC ACTCGCACGT TTTAACGCAT      6540
```

CAATCATGAT TAATAGTTCC ATTAAGTTAT CGTTAACAGG GCTACTTGTT GATTGAATGA 6600

CGTAAACATG TGCACCACGA ATACTTTCTT CAATATTCAC TTGAATTTCG CCATCACTAA 6660

ATTGTGTTAC AGAGCACTTT CCTAGCTCCA CACCCACAGC ATCCGCAATT TTTTCTGCTA 6720

AGGGACGATT AGAGTTCAAA GAAAAGATTT TCAATCTTGG ATCAAAATAA TGTTTCGACA 6780

TTAGAACCTC CACTTTCTAT TTCAGCCAAT TCGTTAAAAC TGaACAATAC TTnCCTTCTA 6840

AAATAATAGT AGTCCTAGCC mAATAAATCa aGAGaTTTTC tGtATTTCAT ACmAaGATAT 6900

TCmTcAAAAG tCGcTTTCTT CTAACTATTT TATCGTATTT GTCTAAAAAA TGAGGATTAT 6960

TTGAATAAAG AAAGGTACTC GCCATAGCCT TCTTTTTCTA AATCTGCCAC TGGCACAAAT 7020

CTCAAAGCTG CCGCATTAAT GCAATAACGT AAGCCACCTT CTTGGAGTGG GCCATCAGTG 7080

AAAACATGCC CTAAATGTGA ATCCGCTTCT TGACTGCGAA CTTCTACTCG GTGCATGCCG 7140

TGACTAAAAT CAGCTTTTTC TTTGACGCCA CGTTTTTCAA TTGGTTTGGT AAAGGATGGC 7200

CAGCCGCAAC CAGCATCGTA TTTGTCCAGG GAGCTAAACA ACGGTTCGCC ACTAACAATG 7260

TCTACATAGA TTCCGTCTTG GTAAAAGTCA TCATATTCTC CTGAAAAAGG GCGTTCTGTT 7320

GCGTTTTCTT GTGTGACGGC ATATTGAAGA TCCGTCAAGG TTTGTTTTAA TTCTTCTTCT 7380

GTTGGTTTTG TCATTCTGTC CACTCCTTTT CTTCTGTTCA ATTTATTGGG TTCAATGATA 7440

CCACGTTTGC CAAACAGGCA CCTaATAGTT TGCATAGTAA AAGACAGTCC TTTCTCTTTT 7500

TGACGAATGA AAAGGACTGT CTTTTATTAG TTGAATTCTT TTAACAGTCG AGCTACTTTG 7560

GCAATTGGTA AGCCCATAAT GGCGTAATAA TCGCCAGCGA TTGCTTCAAT TAGTAATGCT 7620

CCTTGTCCTT GGATACCATA AGCGCCTGCT TTGTCGGCAT ATTCTGCGGT GTCTAAGTAA 7680

GCATGGATTT CCGTGTCTGT CAGTGGATAA AACGTCACGG TTGAATGAAC TGTCGCGGAg 7740

cGCTcTTTTT CCCCTTGTTT CAACGTCACA CTTGTATACA CGTCATGTGT TTTGCCACTT 7800

AGGAGGCGCA ACATGCGATA ACCGTCTTCC CGAGAAGTGG GTTTGCCGAG GATCTCTCCC 7860

GCCAATGCCA CAATTGTATC GCAGCCAATG ACTAGTGCTT CGGGGGATTG CTCTGCTATA 7920

GCCGCTGCTT TTTGGGCAGC CATTTGCGCC ACATATTCTG CTGGTAAgCC GTCTTTGCCA 7980

ACAGTTTCAT CGATGTCGGC TGGTGCAATC GTAAAAGTTG GAACCACACG CTTTAATAAC 8040

TCCTGACGGC GAGGAGATTG GGATGCTAAA ATAATTTGCA TAAATTGATG ACGCCTCGTT 8100

TCTTTTTAAT GGGGGTCTTG AATGCGTTTA AACATTCGTT CCATATCTGA ATTCGTAAAA 8160

TGAATCAGCA CCGGACGTCC ATGCGGACAA TTAAACGGAT TTTCGCACAA GGCTAAATCT 8220

TTTAATAGCA CGCGTGCTTG TTGCTCATTG AGGTAATGAT TGGCTTTGAT GGATCGCTTA 8280

CAGCTCATCA TAATGGCAGT TGCTTCACGG AATTTTTTCA CGCTGACCGA GCCAGTTGTT 8340

AGCAACATGT CAATCATCTC GCGAACAATG CTTTCCTCTT CGCCTGCAGG ATACCAAGTT 8400

GGGTGCGCCC GCACAATGAA ACTATTTTGT CCAAAGTCTT CCAAATGGAT GCCAACTTCC 8460

GCAAGGGTTT CTTTCTGCTC TTTAATTTTC AAGGCATCGC TATTTGGATA ATCAATAACG 8520

ATTGGTACCA ATAGTTCTTG TAAATCATCA CTGACTTCGC CAATTTTTTC ACGGAAGTAC 8580

TCGTATTTAA TTCGTTCTTG GGCAGCGTGT TGGTCGATAA TAAACAGGCC ATCTTTGCTT 8640

TGTGCGAATA AATAAGTTCC GTGCATTTGA CCGAAGTATT CTAGCTCAGG AAAACGCTCT 8700

GTCGGTCGTT CTTCTTCCAA CTTGTTCAGG GCTTGTTTCA GTTCTCTTTG ACTCGCAGCG 8760

CTATTGGCAT CGAATTCTGG GTGAAGTTTC ATTTCTTCTT GGAAAACCTC GTCAGCAGCC 8820

TCTGATGAAC GCAGAGAAGG TTCTGAATCA GGCACTTCAG TTATTTTCGT TTCTTCGGCT 8880

ATTACCTGTT GCTCATTACT AGCTGTTTCC CAATGATTTG CGGAATCAGC AGCTGCCGAA 8940

GAGGTTTCGT AAGCTGACGT TTCTTCCGCT ACACTCGTTG ATTCTGGCAC TGGCGGCGCA 9000

AATGGATCGT CCATTGGGTC CACTGCTCTT TCTGTCTGGT TGACAGCTTC AGGGGCGCTT 9060

GAGGCTGGTG TTTCGTCTGT CGCATGTTCC GCATAGAAGC GGCCTGTTGC GGGATTATAG 9120

CTTAAGCTGC CCGGTCGTTG TTTTACCGCA GgTTGTTCAG TCGATAAAGG AATTTCCATT 9180

TGTTCGGCTT TAGGTTCTGG CGGTAGTTTC TTTTTAAAAC GCAAATTGTC TGCCGCATTG 9240

GGAATCAACT GTTCTTGGCT CAACACTTCT TGAATCGCCT GTTCAATTAA TGCCATTAAA 9300

TCTTTTTCTT TACTTAAACG AACTTCTTGC TTAGTAGGAT GGACATTGAC ATCGACTAAC 9360

AGCGGATCCA TTTCAATTTC TAACACAGCA ATTGGAAAAC GCCCCACCAT TAGTTTCGAA 9420

CCATAGCCGG CTACAATTGC TTTATTTAAC GCGAAATTTT TAATATAGCG ACCATTGATA 9480

ATCGTTGATA AATAGTTGCG GCTGGCCCGT GTGACTTCTG GCAATGAGAC ATACCCTGTC 9540

AAAGTAAAAT CAAGGTCTTT TCCTTCAATT TTCAACATTT TCTTTGCGGT GCTAATGCCA 9600

TAAATCCCCG CAATCGTTTG TTTTAAATCA CCATTGCCTG TTGTACTCAT CATTTTGTGG 9660

CCATCATGAA CTAAGCGAAA AGCAACTTTA GGGTGGCTTA ACGCCAGACG ATTGACAATG 9720

TCACCGATAT TGGCCAATTC CGTTTGGATC GTTTTGACGT ATTTCAAACG AGCGGGCGTA 9780

TTATAAAATA AATTAGAAAC GGTCATTTTT GTGCCTTTAC GGAGGGCTGC TGGTCGGTTT 9840

TCTTCGACCT TGCCCCCTTT TAGAATCACA TAGCTGCCTT CTTCTTCCTC AGCTGTGGCA 9900

GTTTCGACAA TCATTTCAGA AACTGAGGCA ATACTTGGtA AnGcTTCGCC TCGGAACCCC 9960

AAGCTGCGAA TTCGGAACAA ATCATCTCTT GTATGAATTT TGCTTGTGGC ATGGCGCTTG 10020

AAGGCATTCA GTACATCTTC TTTGGCAATT CCTTCACCAT TATCAATAAT TTGAATGGTT 10080

TTTAAGCCAG CTTCTTCAAT AAAAATATCA ATTTGAGTGC TTCCTGCATC CAGTGCATTT 10140

TCCACCAATT CTTTGACGAC AGAAGCGGGA CGTTCAACAA CTTCACCAGC AGCAATCTGG 10200

TTGGCTAGTT GTTCGGATAG TTCTTGAATT TTCCCCATTT CACGACCTCC TAACCGGTTC 10260

GGTTCCTAAT TAGATACGTT TTTGCAATTG ATGAAGCATG TTTAACGCAT CTAGTGGTGT 10320

CATTTCTAAT AGATTCATTT TCTTCAATGT GTCAATGACA CTTAATTCTT CTGTTGAAAC 10380

TTCTTTAAAT AATGAAAGTT GCTCGGTTTC TTCATGTACT TCTGAAACTT CTTCATGATG 10440

AACTGAAATT GGAATCGTTG TTTCTTCCGC TTCTAAAGCG GATAAAATGG TGGCTGCTCG 10500

TTCTAAAAGC GGGCTTGGTA AACCAGCGAT TTTAGCAACA TGTATCCCAT AGCTCTTATC 10560

AGCAGGTCCT TCCATCATTT TATGAAGAAA GACCACTTCG CCATCTTTTT CGACAGCGCC 10620

CACGTGAATA TTCTTCAAGC CTTTTAAGGT TTCATCTAGA ACAGTTAGTT CATGATAATG 10680

TGTAGAGAAT AGCGTTTTGG CTTGTACTTC ACGATGGATA TATTCAATGA TGGCTTGCGc 10740

TAGCGCCATT CCATCATAGG TTGCCGTGCC GCGTCCTAAT TCATCAAATA AAATTAGGCT 10800

GTTAGGCGTG GCGTGCCGCA GTGCTTGGTT GGCTTCCATC ATTTCTACCA TAAAGGTACT 10860

TTGCCCTGCA ATTAGATCAT CAGAGGCCCC AATCCGTGTA AAAATTTGAT CGAAAATTGG 10920

CATTTCAGCA CTTTCCGCTG GTACAAAACA GCCAATTTGC GCCATGACCA CCGTCAAAGC 10980

CAATTGACGC ATGTAGGTAC TTTTCCCTGA CATGTTTGGT CCAGTAATCA AGAGAATATC 11040

TGTTTCTGGG TTCATGCGGA TACTGTTTGG AATGTATTCT TGATGCCCTA AAACTTTTTC 11100

CACGACTGGG TGTCGTCCTT CGACAATCGC CAAATTTTTC GTGTTACTTC TTAACGTTGG 11160

TCGAACGTAT TGATACCGTT CACTAATTGT TGCAAAAGAT TGTAAAACAT CAACCGCACT 11220

AATGGTTTTA GCCAGTGTTT GCAACCGATC AATGTTCGTT TTCACTTGTT CTCTCACGGC 11280

CAAGAACAAT TGATATTCTA GTTCAACTGA TTTTTCTTCG GCTTCTAAGA TTAAACGTTC 11340

TAATTCTTTT AGCTCTGGCG TAATAAAGCG ctCCGCATTA GCTAATGTTT GctTACGCTC 11400

GTATTTTCCT TCTTCTAAAT TCGCTAAATT TGACTTAGTG ATTTCAATGA AATAACCAAA 11460

GACCCGATTG TAGCCAATCT TCAAGTTTTT AATCCCTGTT TCTTGGCGTT CTTTCGCTTC 11520

TAACTCCGCT AACCATTGCT TACCATTGCG CATGGCATCA CGGTATTCAT CTAATTGGTC 11580

GTTATAGCCG TCTTTAATAA CCTTTCCTTC AGTAATTTGT AAAGGGGCTT CTTCATTAAT 11640

CGCCGTCGCA ATCAATGCCA CTAAATCTTC AACAGGGTTT AAGTCGACTA ACAAATCATC 11700

CCATTCCCCT TGATTGATCC CGACAATTAA CTGTCGAATC GTTGGTACTT GTTCTAAAGA 11760

GGTCCGTAAT TGAATTAAAT CACGCCCATT CACACTGCCG AATGCCACGC GTCCTGCTAA 11820

CCGCTCTAAA TCATAAACGT TCGTCAATGC TGCTTGCAAA TCCAAGCGTT CAAAATAGGC 11880

ATTTAATAAG GATTGCACCA TTTCTTGACG CGCCTTAATT TGACGTTCCT GAATCAGCGG 11940

ACGATCCAAC CATTGTTTTA ATAAACGGCC GCCCATAGCT GTTTTCGTtT CATCTAATAG 12000

CCACAATAGA GTCCCTTTTT TCTGTCCTGT TCGAATCGAC TGACTCAATT CTAAATTGAA 12060

TTTAGAATAA TGATCCATTT TCAAGAAATG ATCTGGCTGA TATTCCACAG CTTTCTGAAT 12120

ATGGGACAAG CTCCGCTTTT GGGTAACGGA TAAATAAGAT AATAATTTGC CTGTAATTTC 12180

AATTTCTAAT GGATTAATCA GTTCATTGGT TAAAAAACTA AACTCGGCAT TTTCTTCCAC 12240

CGTTTCTTGT GTCGAAAAGA TGATATTTAA ACGTAAACTT AAGTTCTCTT TTAACGATTC 12300

AGGGATCCCT GAACCTAGGA CCATTTCTTT TGTTTGCAAC GCAGAAGCTT CGTTCAAAAC 12360

GCCCTCTTCA TCAGCCAAGA CTGCTGTTTT TAATTCGCCT GTACTCAAAT CTACATAGGC 12420

AAAACCGAAC TGATTACCAT CTGTTAGCAC AGCGGTTAAG TAATTATTGT CTTTGGCTTC 12480

CAAGCCTTTA CTATTCATCA CCGTTCCTGG CGTTACCAAC TGGACCACTT CTCGTTTGAC 12540

CATGCCCTTG GTAGTCTTTG GATCTTCCAC TTGCTCACAA ATCGCTACTT TATAGCCTTG 12600

CTCAATCAAC GTATCAATAT AGCCTTGCGC GGCATGATAA GGAACGCCAC ACATAGGAAT 12660

CGGATCATCC GCATTGCGGT TACGACTGGT TAACGTCAAT TCTAAGATTT GCGCTGCATT 12720

GATAGCATCT TCATAAAATA ATTCATAAAA ATCACCTAAG CGATAAAATA AAAACGCATC 12780

CTTGTATTGG GCCTTAATGG ATAAATATTG TTCCATCATG GGTGTATTTT TTGTCTTTTG 12840

TGGCATTTAA TTCCCTTCCT TCTCTAGCTC TTCGTTCACG CGGTATTGAA TGAGCGCCGT 12900

GACATGTTGG ACTAAGTCAT TGGCTTCGAT CAGTTGTTCG CGGTAGGCTA CAACGAGTGG 12960

GTGCTCATCA AATTCTTTTG TTTTCGCATC GGCTCGTTGA ATGGCTTCTT TTTCAGCAGT 13020

TGGTTTGCCA TAGTGAGCAA ATTGAACCGC ATCTTTTTGC GCAGCTTTAA TTTCTTCCAC 13080

TAATTCTGTC AGCTTTTTGT TTTTCTTTAC CTTTTCTTCA ATCGCTTGGT AGCGCTGAAT 13140

GACTTCATCT TGTCCAATTA AATGTAACAA TTTTTCTAAT TCTTTATTAC TTTGTGCATC 13200

TAAATTCAAC GGCTCCATTT TTTTCACCTT CATCTTTCAC CGGCACTTTT TCTATAAGAG 13260

CAGAAAAAGT TAACCTCTTA ATTATCCGAG ATTAACTTTT CTTCTATTAT TCTTGAAACG 13320

GGCGATCTTC GCTAATCTGA ATCGGCTCAA TCATTTTCGC TTGGCCCGTT TGGTCATCAA 13380

TCTCTAAAAT ACAGGCAGAT AAAATTGTCC GCCCTGTTTC GACTACCTCA AAGCGCTTCG 13440

GCAATGCTGT CATAAACTTC TCAATAATCG GTTCACGCTT CATTCCTAAA ATCCCATCAT 13500

AAGGACCGGT CATACCGACA TCACTTAAGA AGGCTGTCCC TTTCGGTAAA ATCCGAGCGT 13560

CATTTGTTTG AACATGCGTA TGTGTTCCAA CAACGGCTGA AACTTGGCCA TCTAAAAACC 13620

AGCTCATGGC TTGTTTTTCA CTCGTCGTTT CGCCATGAAA ATCCACAAAA ATGAGCGGTG 13680

TCCGTTTGCG AGCGATAGcT ACTAACTCGG CGGCTTTCTT GAAAGGATCT TCCAATGGCA 13740

CCATAAATGT GCGGGCCTGT AAATTAATCA CCGCTAGTTC CAATTGATTG ACTTTCACAA 13800

AAACCATTCC TTGACCAGGC ACATCTTCTG GAAAGTTCGC TGGACGAACC ATTTTCTTCG 13860

CATCATCAAT GAATTCAAAG ATATCCTTGT TGTCCCACGT ATGGTTACCC ATCGTAACAA 13920

CATCGACACC ATCTTGCAAA AATTTCTTAT AAATCTTGCC AGTAATTCCC CGACCAGATG 13980

CCGCATTTTC ACCATTGGCA ATGGTCACTT GTGGACGATA TTTTTTCTTT AATTTAGGGA 14040

CATACGTTGC TAAGGCTTCT CTTCCTAAAG AACCGACAAC ATCTCCTATA AATAATATAC 14100

GCAAAAAGGT TCCTCTTCTC TTTTTAACAT CTCTTGTTTA TTATAGTTGT TTTCGCTTTG 14160

AAAAGAAAGT TTTGGTTTTA AAAGATTAAA AAAAGtGTCT CAACGAACAA TCCTTGCAAT 14220

TGTTCGTTGA GACaCTACAC GCGATACTTA TTCGCTTTTT ATTTTATTTC GCATAGTCTA 14280

CGGCACGTGT TTCACGGATA ACCGTTACTT TGATATGTCC TGGATAATCC AAGTCATCTT 14340

CAATTTTCTT CCGAATGTCT CGAACTAAAC GAACAGATTC TAAATCAGAA ATTTCTTCTG 14400

GTTTCACCAT GACACGAACT TCACGTCCTG CTTGAACAGC AAAACTTGAA TCAACGCCTT 14460

CAAAACTATT AGAAATATTT TCTAAGTTTT CTAAGCGGCG AATGTAGTTT TCTAACGATT 14520

CcTACGTGCA CCTGGACGGG CTGCAGATAA CGCATCTGCG GCTGCCACTA AAACAGAAAT 14580

AACTGAGGTT GCTTCCACAT CACCATGATG CGAGGCAATG GCATTAATAA CTACTGAATT 14640

TTCTTTATAT TTCGCAGCTA ATTCTGCGcg AATCtCAACG TGAGAACCTT CAATTTCATG 14700

ATCCAATGCT TTCCCAATAT CATGTAGTAA ACCTGCACGT TTGGCTAATT GAACATCTTC 14760

GCCTAATtCG GCGGCTAAAA CGCCAGCTAG TTTTGCTACT TCGATTGAAT GGTTCAAGAC 14820

ATTTTGACCA TAACTTGTTC TGAAACGTAA ACGGCCTAAA ATTTTAATTA AATCAGGATG 14880

TAACGTATGA GCACCTACTT CAAAGGCTGC TTGTTCTCCG TATTCACGAA TCCGTTCATC 14940

CATTTCTTTA CGTGATTTTT CTACCATTTC TTCGATACGA GCGGGATGAA TCCGACCATC 15000

TTGGATTAAT TTTTCCAAGG TCATTCGAGC GATTTCTCGT CTGATTGGAT CAAATCCGGA 15060

TAACACAACC GCTTCTGGCG TATCATCAAT GATTAAGTCA ATCCCTGTTA ATGTTTCCAA 15120

TGTACGAATA TTACGTCCTT CACGACCAAT AATGCGACCT TTCATCTCAT CATTTGGTAA 15180

AGTGACCACG GAAACAGTTG TTTCAGAAAC AGAATCGGCG GCGCAACGTT GAATTGCTAG 15240

GGAAAGCAAG TTcTTCGCTT TGCGATCAGA TTCTTCTTTG gCACGTTGTT CTGACTCCTT 15300

GACCATCAAG GTTAGTTCAT GATTCAACTC TTCTTCTGTT GACTTCATAA TAATCGATTT 15360

TGCTTCTTCT TTTGAAAGGG AAGCAATTCG TTCTAATTCT TGATGTTGCT GATTAATTAA 15420

GTTTTCTACT TCTTTTTCTC GCTCATCAAT TAATTGCTGT CTAGCACCAA GTTTCTCTTC 15480

TTTTTCTTCC AGTGAGCCTT CACGTTTTTC mAGrGAGTCA TCTTTACGAT CCAACGTTTG 15540

TTCCCGTTGA ATTAAACGAT TTTCTTGAGA TTTAAGTTCT AATCTGCTTT CTTTCAACTC 15600

ACTTTcAATT TCTGAACGAT ACTTCTGGTT CTCTTCCTTA GCTTCTAGCA ATGCTTCTTT 15660

TTTCAGTGTT TCTGATTCTT TcCGAGCGTT CTCGAGAATA cTCGACGCTG AAATTTTTGC 15720

ACCATTTATT GCTTTATCGT GACGTGACTT TGCAATAACA AATCCTAAAC CAAGACCGAC 15780

AATCAAACCG ACGATAGCGA GGAGAATATT TAAAACCATA TTTCCACCTC CATACTATCT 15840

TTTTGTTAAT TCATATCTTG TAGTTTTTTG ATAAACTAAA ATCAATTATC AATATGCCTA 15900

CTTGCATATT TGTGCATACA CACAATTTTA TTCTAAtGTT TGTCGTGAAA GGTGTCAACT 15960

TAAAAAAGCC CTTTCTTTTG GAAAGGGCrG CAAATATTAG ACAACGTCAC ATTTTTAGGT 16020

TTCCGCCGTC ACGAATTGTT TAAATTTTAG AACAAACGCC GCATAGTCTG CTGTTTGACT 16080

AAGCTGCTTG ATGTGGATGG GTTCGGACAT GGACAACAGT CTATTTAACA ATTTAGTTCT 16140

AAACAAAAAA AGACAACACC TGTCCTGTTC AGTAAGTGTT GTCTTTCTTC TTAACGAGTC 16200

GTATTCACTC GTTATCTATT CATCTAAGGG TAATTCTTCT TGACCTTCTG CTTCTTCTGT 16260

AATGGTTGAA CCGTCACCGA TGCCGTATGC ATCACGGACC AATTTAGACA CTTCTGCCAT 16320

CATTTCAGGA TGATCCGCCA TGTATTGTTT GGCATTTTCA CGGCCTTGGC CAATTCGTTC 16380

TTCTTTGTAG CCATACCAAG CACCACTTTT ACTGATAAGG TCTTTTTCCA CAGCCATATC 16440

TAGTAATTCG CCTTCTTGGG AAATACCTTG GCCATACATG ATATCAACTT CTGCCACTTT 16500

AAATGGTGGC GCCACTTTGt TTTTAACGAC TTTaATTTTT GTGCGGTTAC CGACGATGtC 16560

TGTTCCTTGT TTTAACTGTT CTGCACGACG GACTTCTAGA CGAACCGTTG CGTAGAATTT 16620

CAATGCACGT CCACCAGGAG TTGTTTCAGG ATTTCCGAAC ATCACGCCAA CTTTTTCACG 16680

AATTTGGTTA ATGAAAATAG CAATTGTTTT TGTCTTATTA ATTGAGCCTG ATAATTTACG 16740

TAGTGCTTGA GACATTAGTC GAGCTTGTAA GCCGACATGG CTCGCTCCCA TCTCACCATC 16800

AATCTCTGCA CGAGGAACTA ACGCAGCAAC CGAGTCGATG ACAACGATGT CAATCGCACC 16860

ACTTGAAACT AAGGCATCGG CAATCTCTAA GCCTTGCTCG CCCGTATCTG GTTGAGATAA 16920

AAGTAATTCA TCGATGTTAA CGCCTAGTTT CTCCGCATAT TGAGGATCCA ATGCATGCTC 16980

AGCATCGATA AAAGCGGCTG TTCCACCATT TCGTTGCACT TCTGCAATTG CGTGTAAAGA 17040

AACAGTTGTT TTACCTGAAC TCTCAGGACC ATATACTTCA ATAATTC 17087

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11764 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

AAATGAATAT TTATTCTGCT TTATTnTCAA TACTTTGTTT TCATAGTTGC TTTAGCTGTT 60

CCTTTTTTGA ATAAGATTAA AATGAATAAA TAACAAGCAA AATATAAGAT TTTTTTGTTT 120

AACTCTTTTT GTTTCAATAG GAAGTGCTTA GTGTTTTTTC AGAATTGAAA AATGAATACG 180

ACAAAAAATT TATCTATATA ATAGAAGGAA AAATAAAAAT AGTGAATGAT AAAAAAAATT 240

CAAAGAAAAC ACCCGTTCTT TTTAGGCCTG AGTTTTTGAA AAAAACATTC TCATGTCTAG 300

TATCTTTAAC ATTTTGTTAC TTTTTGACTG AAACGAATGT AAAAAAACCG ACATTTAAGG 360

GCTAATCAAG GGTTAGAAAA CGGTTTCTCG AAAATAAAAT CCTGTTTTTT TAACTTTTTT 420

TCGCAAAATA TAAACAGGAT AATATCAAGC TTCCAGCAGC TTTTTATTAC AAAATGTAAC 480

AAAACCGACA TTCTAGTAAT AAATTGTCAC GAGATATTCA CtACGGCCGG CCATTCGATG 540

ATAAAGTATG TGAAGTCGTT TCGATAAAGC GAAGCAGACG TGGTTGTTGC CGAGTGCTAA 600

GAGCAACAAT CAAATATTTT TAATTACAAA GGAGAAACAT ACATGAAATC AATCAAAACG 660

ATTTTACTTG GAACAACTTT GGCCGCAGtT TAGGATTATT CTTAGGAACA GACGCAAACG 720

CAGAAAGCTT ATATACAGTT AAAGCAGGAG ATACTTTATC AACAATTTCT CATCAATTTG 780

CAGGAGACAA TAGCTTAATT CAAAAAATTG CTTCTGATAA CAAATTGCCA AACCTTGATT 840

TAATTTTTGA AGGAGAACAA TTAGTTATTC GTTCAGAAAA AGAAgTTGCT AATACTCCAG 900

CACCAGCTGT AGAAGTTGCA CCAGTTCAAC AAGTAGTTGA ACAACCTGTT GCACAACCAG 960

TACAACAAGA AGTACAACAA CCTGTTGCTC AAGAAGTAGC GCAACCAGCA GCACCTGCTG 1020

CAAGCAGTGA TGCAAAAGAG TGGATCGCAC AACGTGAATC TAGTGGTTCT TACGATGCAA 1080

CAAACGGTCA GTATATCGGT CGTTACCAAT TATCTGCCTC TTATTTAAAT GGTGACTATT 1140

CACCTGCCAA CCAAGAACGC GTGGCAGATC AGTACGTAGC AGGTCGCTAT GGTTCATGGG 1200

ATGCAGCTAA ATCATTCTGG TTAGCAAATG GTTGGTACTA AAATTAAATA AACGArTAAT 1260

CTAATTACTT ATTAAAAACC ATAAACAAAA AAACAGTAAC CTGTAAAGAT TACTGTTTTT 1320

TTTGTCGATT TGTTAAGGTA CAACGCCCAA CGCTCGTCAA TTGGTGGTCT GCGTTATAAA 1380

TTTTTGCTTC CCAAACTTGA AGGGTCTTTC CCGAGTGATC GGGCGTCGCA ATCACGGTAA 1440

GGCTACCATC ATGGACACTA CGAAGATGAT TTACTTGTAA GTCTATGCCG ACAGCAAAAG 1500

TGTTTTCAGG AACGTTTTCG TTCGCACCTA AACTACAGGC AGTTTCAATT AAGACGCCGT 1560

TTAAGCCCCC GTGGACAAGG CCATAGGGTT GTTTATGCAC GTCTTTTACC GTTAAAGTAA 1620

GCTGACATTT CTCAGCAGAA ACCTGTTGAA TCGTGATGCC TAAATACTCT AAAAGATGAA 1680

TAGTAATCaG TCCTTTCTAA TAAAAGAGAG AATAAAAGGT TTTTaTTaTT ATACAGGAAT 1740

GAAGAAAATT kGAAATAGCC AkGTTTTTtC TTAAAATtCa AGCGAmTTTT TCyCtAAAGg 1800

CCGATTCTgk GkATAwTrGA AAaGATGrGA TtAATGAAGG AGGTTTCCTA ATGAGAAACT 1860

GGACAACAAT TAAAGAATAT GATGAAATTT TATTTGAGCA AGCAGGCAAA GTAGCAAAAA 1920

TTACTATCAA CCGCCCGCAT GTTCACAACG CATTCACACC AAAAACAGTC ATGGAAATGA 1980

TTGATGCATT CAACATTAGT CGTGATAAAG AAGATGTCGG GGTTATTATT TTAACAGGCG 2040

CAGGTGACCA AGCATTCTGT TCTGGTGGCG ATCAAAAAGT TCGAGGTAAT GGTGGTTACG 2100

TTGGCGAAGA CAACATTCCA CGTTTAAATG TGTTAGATTT ACAACGGTTA ATCCGTGTCA 2160

TTCCCAAACC AGTCATCGCA ATGGTGAAAG GCTGGTCCAT TGGTGGCGGC AATGTTTTAC 2220

AATTAGTTTG TGATTTAACA ATTGCGGCGG AAAATGCTAA GTTTGGGCAA ACAGGACCAA 2280

ATGTCGGTAG CTTTGATGGT GGATACGGCT CAGGTTACTT AGCGCGTGTT ATCGGCCATA 2340

AAAAAGCAAA AGAAGTTTGG TTTATGTGTA AACAGTATTC TGCTCAAGAA GCTTTAGATA 2400

TGGGCTGGAT TAATACAGTT GTTCCATTGG AACAAGTAGA AGATGTGACA ATGGAGTGGG 2460

CAGAAGAAAT GTTAACCAAA AGCCCGATTG CTTTACGGAT GATTAAAGCC TCTTTAAATG 2520

CCGATACAGA TGGCTTGGCT GGTGTGCAAC AATTAGCTGG TGATGCGACA CTTCTTTACT 2580

ATACAATGGC TGAAGCACAA GAAGGCCGAG ATGCATTTAA AGAAAAACGG ACACCAGATT 2640

TCGATCAATT CCCTAAATTC CCATAAACCA ATGACTTGGT TAAATAAACA AGTACAAAAA 2700

CGTCCCGATC ATCCTGCTTT TTATTTTCAG GATGAATCTT GGACGTTTTT AGAAGTGcAA 2760

CAAGAAGTCA GCCATTGGGT AGCCACATAT CAACAGGTGC TTGCACCAGA AGAAAAACGT 2820

GTGGCTTTAT TCAGTAAGAA CAGCAAAGAG TTGTATTTTT CAATTCTAGC CTTATGGGAA 2880

TTGGGGAAAG AGTTATTATT TTTAAATACG CATTTGACAC TTGCTGAACT GACTTTTCAG 2940

TTAAAAGATG CGCAAGTTAA AACGATTATT GGTGCGCCTG AAACGCAGGC TTTGTTAGAA 3000

```
GAAATTTCTT TTGTCGACGT TCAGCCAATG ATTAAAAAAC AGCATAGTCT TTCACATCAG   3060
GAGTTTCAAC AGCCAAGTGA CTTAGAGTCG GTGGCTTCTA TTATGTATAC ATCTGGAACT   3120
ACAGGACAGC CTAAAGGGGT CTTGCAACGG TTTAAAAATC ATTTGGCCAG TGCTAGAGGC   3180
ACTCAAGAAA ATATGGGGAT TACTGCTGAA GATTGTTGGT TGTGTGCAGT CCCATTATTT   3240
CACATTAGCG GATTGTCGAT TGTTGTTCGA CAACTCGTCT TAGGTTGTAG TATTCGCCTT   3300
TACGATAAAT TTGATGAACA GCAAGTGACA CAAGATTTAC AAGAAGGCCG TGGAACCGTC   3360
ATTTCAGTGG TTGCGACAAT GTTACAACAA TTATTGTCTG TCTATCCTGA AGCTGGCTAC   3420
AGCGCCAgCT TTaAAGGAAT GCTATTAGGC GGCGGACCGA TTGCTCCGGA TAAATTAGCG   3480
CAGTGTGAGG AAAAAGGGAT TCCAGTCATT CAATCTTACG GGATGACGGA AACGTGTTCG   3540
CAAGTTGTTG CCTTGAAATT TGAAGATGCG GCACTGAAAA TCGGCTCTGC TGGACAACCG   3600
TTAAAAGATA TGCAAATCAA AATTGTTGAT GAACTAGGAC AAGAACAGCC AGAAAAGCAA   3660
GTCGGGGAAA TTTTACTTAA GGGACCAAAC GTTGTTTCAG GTTACCTTAA CCAGCGTCAA   3720
CCTGAAAAAT GGACAGCAGA TGGTTGGTTT AAAACAGGCG ACATGGGCTA TTTAGATGCA   3780
CAAAGTTACC TGTATTTGGT TAGTCGTTTA AGTGAACTCA TTATCTCTGG TGGTGAAAAT   3840
ATTTACCCCA CTGAGGTTGA ACAGGTGTTA CAGGCGATAA CAGGAATTAA AGCAGCGGCA   3900
GTTGTAGGAG AACCAGATGC GCAATGGGGa GCCGTCCCAG TTGCATATGT GATTAGTGAC   3960
CAAGAAATCA CCTTAGCGCA AATTCAAGAC CAGTGTTCAC GAAAACTGGC AAAATATAAA   4020
CGACCGAAAC GCATTTATTT TTGTCATTCT TTCCCGCAAA CAGCAAGTGG AAAGATTGCC   4080
AAACATCGTT TCATGACAGA AGAAAGAGAG GCGTTTTTAA TCAGATGAAA TTGCCTGATG   4140
AATTACGACA AGCCTATCAG AAGAATGCCC GTCAGTTCAG TTGGGTTTTT CCATTAGAAA   4200
CAGAACAAAC AGCCTTGGCA ATTTTTGCCG CTGGTGAAAA GGCATACCAA GGAGAACGTT   4260
TTTTTTGGCA AACGCCGAAA AAAGATTTTG CTTTGGTTGG CTTCGGATAT GAAACTATCT   4320
TAAAAGGCGC AGAAACAGAA ACGCATCAAT TGAACAACTT TTTAAAAAAA GAGGCGGCAA   4380
CTCGTTTTCA AAATATGTCT ATTTCTGGTA CGGGGGCCCT TCTGTTTGGC GGTTTGCCTT   4440
TTGATACAGA ACAAACGCCT ACTGAGGCCT GGGGAGAGTT AGGGGAAGGT TGGTTTTATC   4500
TACCAAGTAT TCTATTTACT TTTTCAGGTT CACGTATTTA TGGAACGCTG AACTTTTCAG   4560
GTTCTAGTGA ACAAGAAGTC GAGGAACGTT GGTCGACTTT GGTGGCTCAG TTTGATCGCT   4620
TGTTAGCTAC CTGTGAGGAG TTGCCAGCTG TAGCGGAAAC AAAGATTGAC AAGGAAGAAG   4680
TTGCCGTGAC TGAATGGCTA CAAGCTGTCA ACGAAACAGT AGCAGTCTTA AGAGAAGATG   4740
GCCCTTTAAA AAAAGTGGTT TTGGCTCGAC AACTTGCTGT GAGCAGTCCT GAAAAAATTC   4800
GAAGTAACCA AGTGCTCGTT AATTTAATGG CACAACAACA AAATACGTAT TTGTTTGCTT   4860
TAGAAGCCAA AGACACAAGT TTCATTGGTG CAACACCAGA ACGACTTCTG TTGGGTACTA   4920
AAGAAACGTT TGCGACCGCT TGTATTGCAG GAACCATTAA AACGGGTCAA ACACCAGAAG   4980
```

```
AAACGAAAAC TTTAGGAGCA CAACTTTTAC AAGACCGCAA AAATACCGGA GAACACCAAA      5040
TTGTGGTGGA ACGTTTAGCG AAAGAGTTGG CGAAGATGAC GACTTCTGAA AACAGTATTC      5100
AAGCACCCAT TATTTTGGAA AATCGAGATG TACAACATCT TTATGTGCCA ATCTCGGGTC      5160
AACGTAAACC GGGAATTTCA TTTTTGGAAA GTGTGATGCA GCTACACCCA ACGCCAGCTT      5220
TAGGTGGTGA GCCCAAAGAG TTGGCAGTCG AATGGATTCG ACAATATGAG CCAGGGAGTC      5280
GTGGGTTATA TGGTGCGCCA ATCGGTTGGA TTTCAGGGAA TGATGATAGT GGCGAGTTCG      5340
CTGTGGCCTT GCGTTCAGGT GTCTTTGCTG GTCAGCAAGG CGTTCTTTAT GCAGGTTGTG      5400
GTATTGTTGC GGATTCCCAA GCAGAGCTAG AAAGAGAAGA AACGAAAATA AAATTTCAAC      5460
CGATGTTACG AGGAATTGGA GGTCAGGTTT AATGAATCAT CAAGAAACAA TGACCGATTA      5520
TTTAATGGCC TTTATTGAAG GCTTAAAAAA TAGTGGGGTT GAACAAGCAG TCATTAGTCC      5580
AGGGTCCCGG TCCACACCAC TGGCTTTATT ACTTCATCGG GAAACCGCAA TCCAAACTTT      5640
TGTCGATGTC GATGAACGTT CCGCTGCTTT TTTTGCTTTA GGATTAAGTA AAGCGAGTCA      5700
AAAACCAGTG GTCTTACTTT GTACCTCTGG AACAGCAGCT GCAAACTATT ATCCAGCAAT      5760
TTGTGAAGCC AATATTAGCC ATGTGCCTTT AGTCGTTTTA ACAACAGATC GGCCGCATGA      5820
GTTACGTCAA GTTGGTGCTC CTCAAGCCAT GGACCAATTA CAAATGTATC AAAATCACGT      5880
GAAATTATTT GTTGAAATGG CTTTGCCAGA AGCAACAGAA GAGATGTTGA ATTACGCCTA      5940
TTGGCAAGGC GCTAAAGGAG CAGCATTCGC GCAACAAACG CCTGCAGCAC CTGTTCATTT      6000
GAATTTTCCT TTGCGGGAAC CATTATTACC TGACTTAGAG AGAAAAACAA AGAGTTCTCA      6060
ACAAACGGCT TTATTCGCTG GTCAATCAAT CCTTTCTACA GAACAAGTGC AACAGTTAGC      6120
GGATCAATGG TACCAGAAAA ACGGCGTCTT GGTTGTCGGT GGCAGTCATA CAGAGGAAGA      6180
AGCGACGTTA TTTATTCAAT TAGCCGAAGC TTTGCAATGG CCGTTACTAG CGGATCCTTT      6240
AGCAAACATT GTGACACATG GGCAAAATAG TGAAGTTGTT ATTGCACACA GTGATCTGTT      6300
TCTAAATGTG GCGACCTTGC CACAAGAACC AGAAGTGGTA GTGCGTTTTG GTTCGCTGCC      6360
AATTTCTAAA AACATCATGT TATGGTTGAA ACGCTTAGCA ACGACAGAGA CTGCTTTTTA      6420
TTTTGTGGAC GAAAATGGGC AATGGCAAGA GCAATTGAAA AAATCACAAA CAGTTATTCA      6480
GGCAAAGGAA ACCACGTTTG TAGAACAACT ACTAACGGTT GTCAAACCAA CAGAAGCGAC      6540
TTGGTTAGCA CAATGGCTCC TGCTGGAAAA AACGGTCTCT GAAGTTCTTC TAGAAACGCT      6600
GAACGCGACA GAATTAAATG AAACAACAGC CAGCTTAGCT GTTCACCAAA CAATGAAAGA      6660
AAATGGTCAG TTATTTGTAT CTAATAGTAT GGCTATCCGC TACTTAGATC GATTTATGGA      6720
TAGTCGCCCT TATCGGATGT TTGGAAATCG CGGCATCAAC GGGATTGATG GGATTGTTTC      6780
AACTGCTTTA GGAATGAGTG CCATAGCACC GACACAGCAA AACGTTCTAC TGATTGGTGA      6840
TTTAGCGCTA TATCATGATA TGAATGGTTT ATTTTTGGCT AAACGTTATC AGTTGCCGTT      6900
AACAATTGTT TTATTAAATA ACAATGGTGG CGGGATTTTC TCTTTTCTAT CTCAACGAAC      6960
```

TTTACGTGAG GACGATTTCG AGCCTTTATT TGGAACACCG CTTGATTTAG ATTTTTCGTT 7020

AGTTGCTGAA CTTTACGGTG CTTCGTACCA AGAAGTCAAA ACAATCGCGG AgCTCAAGCA 7080

AATTTTACAA GCAGCGGCTG AAGAACCACA ATTTCAAGTG ATAGAAGTCA AAGGAAATCG 7140

CCAAGAGAAT GTTCATTTGT ACGAATCAAT TTTAGCTGAA ATTGGTAGAC GAGTAGAAAG 7200

ACAGGGGATT TCATGGAACG GTTAATTCGG GGGATGCAGT ACCATTATCA ATGGTTAACA 7260

CCTTTTGATG CAAAGCGAAC GACAGTAGTC TGTTTACACG GTTTCACAGG AACGTTAGCC 7320

ACGTTTGCAG CCGTTTTTCC CAGTCAAACA CCTTATAACG TTTTAGGAAT TGACTTACCT 7380

GGGCATGGCG CGACTGCTAG TTTGGTAGCG CCAGAGCGTT ATACCATGAA ACAGGTTTGC 7440

CATGATATTG CTGAACTAAC CGAATCGCTA AATTTGCCTT GCTTTTGTTT ATTAGGCTAT 7500

TCAATGGGCG CTCGAACAGC TTTAGGTTTT GCATTACATT ATCCACAAAA GGTACAGCAT 7560

CTTTTATTGG AAAGTGGATC ACCAGGTTTG GCCACTGCCG CAGAACGTCA GGCTCGTATC 7620

TGTCAAGATC ATCGTTTGGC GGAACGTCTT TTAGAGGAAC CGCTGGTGGA TTTTATTGAT 7680

TTTTGGCAAG AATTACCGTT ATTTCAAACG CAAAAAGCCT TGTCTGTGGC ACAGCAAATG 7740

GCCATTCGTC AGGAACGTTT GAGCCAATCG GCCTTTGGAT TGGTTAGTAG TTTGTGGTAT 7800

ATGGGCACAG GAGCGCAAGA GAGTTACTGG GAACGTCTAG CAGAATTGCA GCCGATTCCT 7860

ACCGATTTAT TGGTAGGCGG CGAAGATCAA AAATTTATTG GAATTGCGAA AAAAATGCAG 7920

GCACGTCAAC CGTTGTTGAG GCTGACAATT TTTCCTGAAG CAGGACATTG TATCCATTTA 7980

GAACAACCAA CGATTTTCTA TGAAAAGGTG ACAGCGTTAT TGGAAGGAGC CATCTGAATG 8040

AACATTCAAT CAATTGAAAC CTATCAGGTT CGTTTGCCTT TAAAAACGCC GTTTGTCACG 8100

AGCTATGGTC GCCTGGAAGA AAAAGCGTTT GATCTATTTG TAATAACGGA TGAGCAAGGA 8160

AATCAAGGGT TTGGCGAATT AGTGGCTTTT GAGCAACCTG ATTATGTGCA AGAAACGTTG 8220

GTAACCGAAC GCTTCATCAT CCAACAGCAT TTAATTCCGT TACTTTTAAC AGAAGCGATT 8280

GAACAGCCGC AAGAGGTTTC AACAATTTTT GAAGAAGTTA AAGGGCATTG GATGGGAAAA 8340

GCTGCGTTAG AAACTGCCAT CTGGGATTTA TATGCCAAAC GGCAACAGAA GAGTTTAACT 8400

GAATTTTTTG GACCAACGCG GAGGAAAATT CCTGTAGGGA TTAGTTTAGG GATTCAAGAA 8460

GACTTACCTC AATTACTTAA ACAAGTTCAG TTAGCTGTAG AAAAAGGGTA TCAACGAGTG 8520

AAGCTGAAAA TTCGTCCCGG TTATGATGTG GAACCTGTTG CTTTGATTCG TCAGCATTTT 8580

CCTAACTTGC CTTTAATGGT CGATGCAAAT TCAGCGTATA CATTGGCAGA TTTGCCTCAA 8640

TTGCAACGTT TAGATCACTA TCAATTAGCG ATGATTGAAC AGCCTTTTGC TGCAGATGAT 8700

TTTTTAGACC ACGCACAATT ACAAAGAGAA CTGAAGACAA GAATTTGTTT GGATGAAAAT 8760

ATTCGTAGTC TCAAAGACTG CCAAGTGGCG TTAgcATTGG GCAGTTGCCG TAGTATCAAT 8820

TTGAAGATTC CTCGTGTAGG TGGGATTCAC GAGGCGTTGA AGATTGCCGC ATTTTGCCAA 8880

GAAAATGATT TATTGGTTTG GTTAGGTGGT ATGTTTGAGT CTGGTGTTGG TCGCGCATTG 8940

AATTTGCAAT TTGCGTCACA GCCTACCTTT TCCTTTCCAG GTGATATTTC AGCAACGGAA 9000

CGCTATTTTT ATGAAGATAT CATCACGGAA CCTTTTATTT TAGAGCAAGG GACAATGACC 9060

GTGCCTCAGG GACTTGGTAT TGGTGTTACA CTTTCCCAGA CGAATTTGCT AAAATACAGT 9120

CAGTATCAAA AAATCATGTA ATGATAACAA AATTGCGCCT ATGACTATGC AGAGAAGCAT 9180

AATCTAGGCG CAATTTTTTA TTTTTCTAAT TGACTCGCTG CGGCTTCATC CAAGAGAACC 9240

GTTAAATTTG GATGGAGTTG AAGAATAGAG GCAGGATACG TTTCTGTGAC AGGTCCTTGG 9300

AGCACTTGTT TGACCATCTG AGCTTTTTGT TCACCATTTA CAATTAGGAC AAGATGCTTA 9360

ACTTTCATGA TGCTGGCTGG CCCCATCGTA ACAAAGGTCA TGCCTGGCTG CATCATTTCA 9420

GGAACAAACC AAGGCTCTGA ACCAGTGACG GTCACTTGGT AGGTTTCATT TTTAAAATGA 9480

GTGGTAGTGG GCATGTTCCC GCAAAAATGA CCATCTGCTC CTAAGCCAAT TAACATCAAA 9540

TCTAAGCCGC CTGCGTCAGC AATCCGTTGT TCTTGTTGTT GAAAGTTTTC AACAGTTAAT 9600

GGATGGATGT TTTGCTCGGA TATTTCCGCA GGTGTCAAAT ACAGCGTCCG CAAATCTGTC 9660

AGAGTAATGC CTTCCTTTTG ATGAGGAACG GGGATTTCAT CAAAGTTATA ATAATGAACA 9720

TTATCAAAGT CAGATGAATT TTTCACAATC GAGGCTAATT TTTGGTAAAC AAGTGCTGGT 9780

GTTTTGCCAG CGGTAATTGA TAAATTCACT CGTTTATCTT GGCTCATCGT GCTTAGAAGT 9840

AACATTTTTG CCCATTCGCT CATCGTGTCG AAATCTTGTT TAATTAAAAT TTTCATACTT 9900

GCTCACTCCA TTCTTTGAAT AAGCTAAGTA TAATTTGTGG GAACCATCCC ATGTCAAGCT 9960

CTTCACAAAA AGGATTTAAT TAACGGTAGG CAAAAGAAAG AGGAGCAGCC GTCAAAACGG 10020

CGACTGCCCC TCTTTTTAAA ACTTAAGATA ACCTACAACG AGTTAACTTT TTAGCTTTCT 10080

GTTCCACTTG CTCCAACGCT GGTCTAAACG TTCTTTTGAC CAACACTGTT TAACAAAGAA 10140

TTTTATTTTA TTGTTCTTTC AACCATTTTT CTAGTTGTAA ACGTTGCACT TTCATAGTAG 10200

CTGTGTGAGG GATTTCATCA AAAGCCATAA TGATAGGTTC GTTTAAATGT GGTAAATCGG 10260

CTACTTGTGT CCACCAAGCG TCCCAATCCA TTTCTTCCCC AGGGACAACC GCTAAAATTG 10320

GCTGAGGTTC ATTTTCTTTG CCACGGACAA TAATCACTTC TTCTAAGAAA TCTAATGCAT 10380

CCAATAAATG ATCTTCAATA GCTAAATTAC TGTCGATGTT TTCAATTAGA TCTACTTGGC 10440

GATCTTTCAA GAATAAGTGG CCGCGTTCAT CTAACATTCC GTAATCGCCG CTGTCCCACC 10500

ATTCGCCATA GACATTTTCT GCAAAGCGGG CATCTTCTTT GTAATATGTT AAGGCCCGTC 10560

CTTTTGATAA GAATTGGATA TGGCCGTCCG TCATCGCAGG TAAAACTTGT CCTTTTTCAT 10620

CGGTAATTCG AGCTTGTGTT AAATCCTCTA AACCAACGCC CATGTCACGA GCATCTGAAT 10680

CTTTTAGGGA TTCTAAAGTA TGCGCTCGTA AGATCATTGG CCCGCATTCA CTTTGTCCAT 10740

AAACTTGTAG GAAAATTGGT TCATTGGCAG CAGATGCTTT TAAGAAAGCC ACCATTGTAG 10800

CGTTGTTAAT TGCATCAAAA GTAGAATGAT AATAATGGAC ACTAGCGAAA GCTTCAGGTT 10860

TTTCTTTGGC CAAACGAACC CATTGAACGA AGTTATTGGG GTGTGTTTCT ACAGCGATTG 10920

GTTGATAGGT GCTGAACATT TCAGCGACTT TTTCGCTCGA AGCATTTGCT AAAGGCATCA 10980

TAGGGAAGCC CATGGCCATT AAAGAAGACA CGCCAATGTT AAAACGAGAA TGAACTGGCG 11040

AAATATGAAA GCCAACTAAT TTCTTTTCAG CAATTTTTGT AAAAATGGTT TTTTGCCATT 11100

TTGTACGCCA ACCCATAGAG TGATTGGAAT GACAGATAAG TTTAGGAATA CCTGTTGTTC 11160

CAGAGGTGTG GGTCATGTAG GCAATTTCGT CTTTAGTCAG TTCGTCTTGA GGAACTTCCG 11220

TGGCATCAGC TTCTAAAAGA GAAGCAACCG AAAGTTTTTT ATTGGCAGAG CTATTACGAA 11280

CAGCTTGAAC TTTTTCAGCA GTCACATCAT CAAATAAAAT ATAAGGATCT TCTAAACGAT 11340

CAACGAATAC TTCAATTGTT TCAAAAGGTA AATGGTATGA AACCATTACA GGAACGGCTG 11400

CTAGATATGA GGCTGCAACA GCTAGTAAAT ACGTGTCGAA TTTAGGACTT TTATAAATAA 11460

TAATTTTGTC GCCTTTTTTT ACACCTAAAG TGGCTAATTG ATATGCCCGT TTCAAAATTT 11520

CTTGATGGCT TTCACCATAA GTTGTTTCTA AGCCAAGTGC AGGAAAAGCA GGCAAAGATT 11580

CATCAAAAAT AATTGGAACA GTTGGTGTTT TTTCTGCGGC TTCTTGATAG TTTGTATAGA 11640

GGTTTAATGG TTGATAATCA AGTAAGTGTG TCAAGGGAAA GTCTCCTTTG TAAnATAGTT 11700

TGTGTACGAA TGTACAAGTT CACTTTATCC TATCATAGAA TGGGTCTTTT TCTACGTCAn 11760

AAAT 11764

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11307 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

TCCAATATAA AGACACAACA GACAAACGaT GAGGAGTGTG AAwTAAtATG TTaACAATTA 60

TTATTGGGTT TATCTTTTGG ACAATGACAC TAATGTTAGG TTATCTAATT GGTGAAAGAG 120

AAGGCCGTAA ACATGAGTAA TTTAACAAAA CGTAAAAAAG ATTTATTTGA AATGAAAAGC 180

GTTGTATTTA AAGATATTTC AAAGCAACAA AGCGAAAAAG CACAAAAAAG AAAACGACTC 240

TTACAACTAA TGAATCAATA TCCCGATTGG GCAAGTCAAA AAAATAAACT TATTATGCAG 300

GAAATTCAAG AATTAGGACA AGCAATCGGT AATTGGTCGA TGGATCAATC AAGACCCATC 360

CAATCCATCA AGGCCGCATC GTTTACAAAA AGCGAGTATC TCTATTTAAT TTGGCTCGGT 420

TATTCAGATG AAGCGATTCG TCACGGCTTA GACATGTCGA AAGAGTGTTA TTTTATTTAT 480

CGATTAACAC TTTTAAATGA ATAAAAGTAA AGGAGATTAA CCAATGCGTA CGTCAACATT 540

TAATTATATC AAAGATATTT TAGCAGACTT TTATAAAACA GAAGAGTATA TCCGTCAACG 600

GGAAGAAGAA TTACGGCACC CTTATCAAGA AGCAGATTTA AATGCTGGTA TTAGAGGACA 660

AGGACTTCAC TCTGTAGTGA CCGAACGAAT GGCGATTACG ATAGCTATGG aTCGTCGTCT 720

GTGGAACTTa GAGAGAAATC GAGACATTAT CAAAAATTGT TTAGCCGAAG CGGaTGAACA 780

AACGCGCGTG ATTATTGAAG AACTATATAT GAAAAAACGG CCCTCTTTAA CATTAATTGG 840

ACTTGCCCAG CAATTATTTA TTAGTAAAAG CCAAGCCTAT AAATTAAGAA ATCATTTCTT 900

TGAAGCGGTG GCGGATGAAC TAGGGATGTA AACATGGAAA AAGCGTGGAA TTTTTTCAGG 960

TGTCAACATG GTAAATTAAT AGTGTCGAAA GAGATAGATA AACGTGAGGC AACCAAAAAA 1020

ATGAAGACAC GGAATTCTAT GATTTTGACT GCTTTCTTGT GTCAGCTATG AAGGAGCAGA 1080

AAATGCCGGC TACTTTCAAG ATCCTTCATT TTGACTAGAA GAGAGCCAAT TTGTTAACCA 1140

ATCCTGAATT TTTTGAATGG AAAGGTGGCG CTAAAAATGA ATGAAGCGGA ACAAGAGTTA 1200

TATGAAGCCC TTGTTGCAAT CTGCCAGACG TCaGGATTTT TGTTGCTAGA GGAACTGCCG 1260

ACAGATTTAC CAGATCAGCC ATTTGTTTAC TTAGGTGATA GTAAAGAATT ACCTAAGCCA 1320

ACTAAATCAG CTATTTTGGG AGAGATTGAA TTAATAATGC ATGTTTATGG TGCGTTATCT 1380

GAACGACAAC AAATTTCTAC AATTAAAGGA ACGATTTTAC GGCAGGCAAC CAGTAACTTA 1440

AAACGAACGG CTCATTTTAA TTGGGGTATC AAACATCAAG AAGTCAAAGC ACAAATGGTA 1500

AAAGATACCA AACAAATGAA AAAAACAATT TGGCATGCTG TACTACCATT ACACATGCAA 1560

TTTTACTAGG AGGAATTATC AATGGGAGAA GTTATGCAAG GAAAAGACCG TATTTTATTA 1620

GTTCGTCGCT TGGATGAAGC AGCGACAAAG AAAGCAATGA AACCCTTATT TCAAATTGAA 1680

CATGAATGGG AATTCTCACG TGAATCGAGC GGTACGCAAA CAAAAGATGG CGTCGCGAAT 1740

GCTGTTTCTG GTTTAGAAGT TACGTTATCG TTAAGCGGTT TAGCCTCTCG AGATGATGAA 1800

AATTTATACA TGaAAGACGC AgcGAaGATG GCATCTTAAT GGAATTTTGG GATGTTGATT 1860

TAAAAGGTGA AAAAAATGCG GAAGGTAAAT ATCCAGCAAT TTATGCCCAA GGTTATGTAA 1920

ATTCATGGAG TTTACCAGCC AATGTAGAAG AATTAGTAGA AATCGAAACA GAAGCCTCTA 1980

TTAATGGCAA GCCACAAGAT GGCTTTGCAA CAGTAGAAGC AGATATTATT GCAGAAGCAC 2040

AATATGCGTT CCAAGATACC GTTCCAGATA AAGCACCACA ACCTGGCGAA TAATCAAAAA 2100

GTGTTGAATT TTAGGAGGAT AAAAAATGAA TTTAGAGATT AACGGAAAAA CAATTGAAGT 2160

GAAATTTACG ATTGGCGCGA TTCGCGAATT AGATAAACGT TACCAAATTG AAAATGGCGC 2220

TGCCAAATTC GGCATGGGCA TCAGTTCAGC AATGATTTAT TTACGCCAAT ACAATCCAGT 2280

AATCTTAGTT GACATCATGG AAGCTTTACA AAGTGGGCAA TTAAAAATAG GTAAGTCGGA 2340

AATTGAAGCA TGGTTAATGA CCCAAGATGT CAAAAAACTT TCAGATGATT TGCTTAAAGA 2400

AATGGGAAAG CAACCTCTTA CAAAACCAAT GATCGATCAG TTCAGCAAAG AAGCGAAGAA 2460

AGCAGAAGCG CAAssgACCA ACTAATTAAA ACGAGCGATG ACGTGTATCA CGACATCGCT 2520

CTTTCTGCTT TTCGCTACTT AGGCTGTCGT TCATTTGAAG AAGTGGATCA GATGACCATG 2580

TCTGAATTTG AATTACGAAT GATTGCTTTT AATTTAGCAG AAGTAGATGA AGAGCGGAAA 2640

AGGCACGAGC TTGCCTACTT AAATGTTAAA GCGCaAGCGA CAAACAAAAA AGGAAAACCC 2700

GTTTTTGAAA GCTTTAAAAG TTTTTATGAT TATGAAAAAC GAGTTGCTGA AGTTCTGGCA 2760

GCTAACCAGC CACAACGAAC AAAATTAAAT GAGCGGAAAA AAACGCAACT TGCCACTGTG 2820

GCAGAgCGTC TACGCCGCTA TCGAGAAGGG AGGAGAGTAG ATGGAGAATG ACAAAGAAAA 2880

AACGCCGTTA TCGGAGGCAA AGAAAAGCcT TGCAGGCGTC CAACAAGCAT TAAAAAGTAT 2940

GAGCGGTGAG TATGCCTTAT TAAGTGGATA TTTAGGGAAA ATTAGTGCGG GTGTCAATCA 3000

GTCAGCCACG GTCATGAACA CATTTAAAAC CGTCATGCAA CAATCTGGAG AAACAGTGAA 3060

AAAAACAGGA GACGAAACAG CAAAGGCAGC AGATCAAATG AACACAGCGT TAACAGATTC 3120

TGCTGAACAA GCCGGTGaAG CAGCTAAAAA AGCGGGGAAA GAAACCTCTG ATGGCTTTAC 3180

TAATGCACAA AATAATATGC TGAGCTTTGG GACGGCCATG ACTAGTGCCG TTTCCTTACC 3240

TATGCTGAAC GTTTTAAAAA CAGCTATGGG CGTCGGTGCT GGGGTCAGTG GCGAATTTCA 3300

AGGAATGCAA GGACTGATTA TGGCCAGTGC AGGAGGGATT TCTGATTCAT TGCAAGGCGA 3360

GTTGCAAGGG GCATTGACTC AGATGAATCA ATCATTTGAA GCGGCGGCAC AAGTGATTCA 3420

AAGCGTGATG GCTCCAGGAA TGGAAATTTT GGTTCAAGTG GTTATTACAG TCGTCAAAGG 3480

CATTACAGCT TTGGTTAATT TATTTATCAA ATTACCAAAA CCCGTCCAAG TTTTTATTGT 3540

TGCCATTATG GGCATTTTAG CCGCCATTGG GCCCATGTTG ATTATGGTAA CGATGGCTCA 3600

GCTAAAATtT CAACAGTTTA GTGCTGGTTT GGCTCTTGTA CAAGGAAACA TTGGGAAGTT 3660

AGGTGGTGGC TTATCAAAAC TAAGTGCTAG TTTTAGTGCC TTAGGTGGAG GACCATTAAT 3720

TTTAATTGTA GCAGCCGTTT TAGCAGCGGT AGCAGCGTTT ATTTATTTCT ATAAAACCAA 3780

TGaAACaTTT AGAAATAGTA TCAATAGCTT AGCTAGTGCC ATTCmAGGaG CTGTTTCAGC 3840

GGCGTTTGGC AAATTGGTAG GATTGCTACA ACAGATCCAG CCGGCCTTTC AGCAAGTAAT 3900

GGCAGTTTTT AAACAATTTT TTGCAGTAGG CTTAGAGAAA ATGGCGACTA TTTTTTCAAC 3960

AATTGGTCGT GTGCTAGCAG GCGTTTTTGC CAGCGGTTTG CAATTAGGTA GTAACTTATT 4020

AGGGCAATTT GGTGGCACCT TTGACAAAGC TGGTTTAGCG GTTGGTCTTT TGGTAAAAGT 4080

TCTGACAAAG GTTGCACTGG CTGCATTAGG AATTCTGGGC CGTTTGGTCT AATTATTTCC 4140

TTGATTGTTT CATTCGTGAC GGCCTGGATG AAAACCGGTG ATTTGAGTGC GGGTGGTATT 4200

ACCCAAGTCT TTGATAATTT AGGTAACACG ATTACATCGG TTACAACAAT GCTGGCAACT 4260

AATCTACCGA AAGTTATACA ACTTTTTACA ACAGTCTTAA CCAGTATTCT CGGGAAAATA 4320

ACAGAAGCTA TTCCAAGCAT CGTAACCGCG TTATCTAGTT TAATTACGTT AATTGTTGGT 4380

GCGATCGTTG CCAATTTGCC AGTCTTAATT GAAGCGGCAA CACAAATTAT TACTACGTTG 4440

ATTCAGGGGA TTACAACAGT CTTACCAATG TTGATAGAAG TTGGTTTGAG CTTATTAATG 4500

ACTTTAGTTA ATGCGATTGT CACCGCCTTG CCAACAATTA CAACTGCAGC GATTAATATC 4560

ATCACTACAT TAGTGACAGC TTTTGTCACA GCGTTACCAA TGCTAGTTAC AGCAGGTGTT 4620

TCAATTATCA CGGCCTTAGT CAATGCATTT GTTACTATGT TACCGTTGAT TTTGACTGCT 4680

GGTTTACAAA TTTTGATGGC ATTAATCACT GGGATTATGA CGATTTTACC TCAGTTAATT 4740

CAATCAGCGC TGACGATTAT TCTAGCGTTA GTGACAGCGT TGATAGGTGC CTTACCACAG 4800

ATTATCAGCG CAGGTGTCAA ATTGTTAATG GCGTTAATTC AAGGAATTAT TTCGATTTTA 4860

CCAACCTTAG TTGCGGCAgC TATTACCTTA ATTTTGACAT TGGTAAATGC CTTAATTGGT 4920

GCCTTGCCAC AAaTCATCAG CGCaGGCGTc AAATTGCTAA TGGCTTTGAT CCAAGGGaTT 4980

ATTTCAATTT TACCGCAACT GGTTACTGCA GCAATTACGC TAATTACCGC TTTAATGGGT 5040

GCGTTAATCA ATGCGTTGCC ACAGTTGTTA AGTGCTGGGA TTCAACTGAT TCAAGCCTTA 5100

ATTAATGGTG TACTCAGTCT ATTGGGTGCC TTGCTGTCCG CAGCAGGAAC ATTAATCTCA 5160

CAAATGATCA CGAAGATTGG TTCTTATTTT GGTCAACTGT TAGCTTCGGG CGGACAGTTA 5220

GTTGAAAATA TCAAAAATGG GGTTACCAAT GCAGCCAATC AGGTAAAAAA TGCCATTGGT 5280

TCTGTAATTG AAGGTGCTTG GCAAGCAATC CAAGGTTGGT TTTCAAAATT CACCGATGCC 5340

GGTGCGAATA TTGTCGGCAT GATTGCTGAT GGAATTACAG GCGCAATTGG aAAAGCCAAA 5400

GAAGCAATCG ATGGAGTTGT CAGTAAAATT CGTAACTTTT TACCATTTTC ACCAGCAAAA 5460

GAAGGTCCCT TATCTGATTT GCATAAATTG AATTTCGGCG GCACGATTGC CACGGGGATT 5520

TATGCAGGCG AAACAGCCGT TAGTAGAGCA ATGGCTTCTA TTTTAGATTT ACCGCTGTTA 5580

AATGATTTTG CCTTGGACTT AGCTGGTCGA GGAAACTTCA CGGCAACGAT TGACCATCGT 5640

TTAGAAAATG ATGCATACAA TCGACCATTA TTTGTGACAG TAGAGTCAAC GTTAGATGGA 5700

AAAGTTGTCG CAGCAACTAC GGCGCCTTAT TTAGCAACAG AGTTACAACG ACAACAAGTG 5760

AAACAAAATA ACCGCTTAGG AAGGAGAGGA TAACATGTAT AAATTTGTTG ATACCAATCA 5820

AGCAACTCAT TCAACGCCTC TTCCTTCAGA AGCGTTGAAT TTTAACGGCC AATTTTTAGA 5880

AAAAGTCATC CCTGGCTATC AAACATTATC AGTTTCAGGA CGAGAATTAG TTCCAAGCGA 5940

AATTGAAAGC TATCAATTAG GGATTCGTGA TGGTAAACGT CACGTTTATG CGCGAATTCC 6000

AGAACGAGAA TTAACAGTCA AATATCGCCT TTCAGCTGTG AATAATGAAG CATTTCGAGA 6060

TGCATTTAAT CATTTAAACG TTGCTTTGTT TACGGAAAAA GACGTTtCTA TTTGGTTTAA 6120

CGATGAACCG GAAATGCTGT GGTTTGGCAG TAAGTCTTCA GTGAGTGATG TACCCGAAGt 6180

GTTAACCAAG TAACAGGCAC CTTTACTTTA TTGCTTTCTG ATCCGTATAA ATACACACGG 6240

AGkATgGCGA CTAGTGTGAT GTGGGGTTCG CCAACCATTA CATTTCAAGC GAATTACTTA 6300

ATGGGGAATA CAGGCTCAGG TGCATTTGAT TTTCCAATTT TAATTGAAGG CGGGGCTTAT 6360

TGGGGATCAA CCATGATTAC CTTTCAAAAT CGGGCTTACA CGATGGGGGA TTTAGGCAAA 6420

GAAGTTCGGC CAATTGAAAT TTATCCTACG GTTGAAGGAT TAAAAGTCAA ACCGACCATT 6480

ATTTTAACAG GAACCGGACG TGGTGTTTGG ATTAAAACAC GGAACGATAC AATTAACTTA 6540

GGAGACTTTG ATCGTTCGGA AATTATTATC GATACTGAAA ATTTTTATCT GACaAAAAAT 6600

GGTGCACCGA TGATTCGACC AATGAACGAT TTTTATCTAT ATCCCAATGA ACCGCTGTAT 6660

ATTCAAGCCA AAGATAGCGA CTTCCGCTTG ACGATTCGCT ATCCTAACCG ATTTGTGTAG 6720

GAGGGTGATT AAATGTTAAT GGCGCTGGAT TTGAAAAGAA CATATACGGC AATCTTGGAT 6780

AATGCCTATC AAGTCAGTTA TGAAAAAATA GAGAACAAAA TTGGCAGTTT AGATTTTACC 6840

ATGCCACTAG ATGATCCTAA AAATGAATTT ATTGCAGAAA TGCAATGGGT GGAACTGACC 6900

GACAATGAGA ATGAATATAT TGGTTTATAT CGCGTGATGC CAACCACAAT TAAGAAAGAT 6960

GCGAACAATA ATCAAATTCA TTACTCTGCC ACAGAAGCAT TATGTACCTT AGGTGATACT 7020

GTCCTTTTTG GTTGTCACGA AATTAAAAAC AAAACAACGA AAGAGGCCAT TCAATTTTTA 7080

TTGAATAAAC AAAAAACAAA GCATTGGGTC CTAAAAAAAT GTGATTTTTC AAGGAAATTA 7140

ACCTATAAAT GGGAAAATGA AAATGGGTTA GTTGAGCCTT TATTTAGCAT CCCAGCCGAT 7200

TTCGAAGaGG AATATCTTTG GCAATGGAAT ACAGAGGTCT ATCCTTTTGA ACTTTCATTA 7260

GTCAAACCGC CAACAGAACC AGTTGCGCGA ATTCAAGAAG GTTACAACAT GCAAGGATTT 7320

GAAATAGAAC GTAATCCCAA GATGCTAATC AATCGGATTT ATCCATTAGG TTCAGGCGAA 7380

GGTGTTAACA AAGTCAATAT TCGCTCGGTC AATCAAGGGG TTCCGTATTT AGAGAACAAG 7440

GCCGCAATTG ACCGCTATGG TTTATTGGAG TCAATTTGGG TGGAACAGCG TTTTTCTGAT 7500

CCCAAGGCAT TAAAGGAAAA TGCTTTGCGA ATGTTAGAAG AATGGACCAA GCCACAAGTT 7560

TCTTGGGTAG TGACTGCAGC TGATTTAATT AAATTAACAG ATCAACCTTT GGCAATCGAT 7620

CGTTTGCGGT TGGGCACGGT TATCATGATT AATACGAATG AATTTGGGAG TGTCAACCTT 7680

CGTATTAAAA AAGAAAGCAA AAAAGATGTC TTTGGTGCCC CCCAAGACAT TCAGCTAGAG 7740

TTGGGAAACC TGCAAGAAAC AATTCATAGT ACCATGACAG CTTTCAGTCG GAAACAAGAG 7800

ATTAACGAAA CTTACGCACA AGGGGCGACG ACACTTTTAA ATCGTTCAAT ACAAGGAGAA 7860

CTTAGCAAGA CACAGCCAGT GGAGCTGAAT TTATACTTTG ACGAGGACAT TCTTTATATA 7920

AACACCGCAG AATTAACGTT CAAGGCAACT GCTAAAGGAC CTTCGCATTC TGTAACGAAC 7980

ATTGATTTGG TAGTGGATGG CAAAAAATTA CCCCAACTAT CATTGCAACA ACAACGGCTA 8040

AACATTTTGA GTTATTTACG AAAAACAACA GATGAAAAAA TCGAACGCGG CAATCACACG 8100

CTTCAATTTT TCTCTCATCA GCCACTATGG TTGGATGCTT CGGTCATCTG TCGTGTGTAT 8160

ATTCAATCCC AATTGGGTGG CCAGTTTTAA TAAAATAATG AAAACTAGAG GAGTGTGACG 8220

AAATGTCAGT AGAACATATT GAAGAATTAG ATACCCTGAA TCAAGGTCGC CTTAAAATCA 8280

ATGCAATCTT GGATCAGTCG AATGCATCAG CTGAGAAAGT AGATGCTTAC CAAGTCCAGT 8340

TAACGAATGG AATTTCTGAA GCGAAAAACA TAGCAGATGA AGCTGGCAAA GAAGCCGTAC 8400

AAATTGCCAC CGATGCAGGC AATCAAGCmA aTGAAACAGC CAACCAAGCG aTGaACaATG 8460

CCAAAACAGC AaTCATGATT GCAGGaAATG CAGTTTCAAC GGCAAATAAT AATAAACAAG 8520

AATTTGATAC TTTACGAAAT GATTTCGATC AATTAGTAGC AGAAGCGGGT GATAGTAATC 8580

CAGAAATTGT CCAAGCACGT ACAGATACAC AAGGCATCAA ACAAGCTACC TTAGCGAATC 8640

GTCTTCAAAT TGATTTGAAT GACCGTATGA CAAAAGCAGA CGGTATTTCT TTATTGGCTA 8700

AGCCAACTAC TGTCAAAATG AAGTTAGACT TTAACGGTAA AACGGCCGGC AATACAGCCA 8760

CCAATGCAAA CAGTTATTCC ACTGATTTTA CGGCAAAAAT TCTTAAGAAG CCAACAGAAG 8820

TTTGGGAGGA AGTTTCCCAA GCGGACTACA ATAAAATGGC CAGCCGTGAT GATGAGGGCG 8880

TGAAAACAGG TTCCACCCAA AGCGGTGTGA TTCCGCAACA ATTAGCGGCC TTCAATCTCG 8940

TTGAAGCCGC TAAAAAATTA ATTCCACAAA TGTTTGAAAC AGTCACAACT GACGAGGCGG 9000

TGGCATTTAT TCGCCAGAAC GTTCAATTTT TTACGATTAA TCAACGTGTG AAAGCCGCTG 9060

CGCCCAATAA TCAAACGATT AAAATCGCTA CGTATTTACC AACTACGGAT AATTGGGTAA 9120

CTCAAATCCA AGAATCAGCA AAAGAGTTTG GCGATTTTTC AATTCAAATC AATGATCAGA 9180

ATTTTATCAC AGATGAAGGT TTCATTTATT TAATGAGCTA TACAGATTCA TCGAATGGGG 9240

TAACGCCAGC TAGCTTAGAA GTTGATTACG TGGGGCTTCA TATTGGTCTG TCTGTTGATG 9300

CCCAAGCGGT TTTAGCGAAG AGTGGTTTTG TTCAAGCAGA GCAACTCAAT ACCCATATGG 9360

AAAATCAAGA TAACCCGCAC CAAGTAACCG CTGAACAAGT GGGGCTAGGC AATGTAGAAA 9420

ATTATGGnTT CGCATCAGAC AGCGAAGCAG TCGCGGGAAC TTTAACGAGT AAATATATGC 9480

ACCCGAAAAA CGTTGCGGAA GCGATTAAAG GTCAAGCTGT GACACAAACA GGTGATCAAG 9540

AGATTGCTGG GGTGAAGAAT TTTGTAACTA TGCCAACCGT CAATGGTGTG CCTTtGGAAT 9600

CCTCTAGAAT GGCCATTTAT GAAGCTAnTG GAGTTGGTGA AGyCGAGGCA AAATATCAGg 9660

CGGCCTTTAA TAAGGAtAAT ATGAAATTTG TATTAATTAG GGTAGGAAAT CGTGTCGATG 9720

CATTTGTAAG ATGTAATTTG AGTGATCCaA CGAAATTGAA TAATAATTTG GTTAAAGTGT 9780

TTACTGTTCC AACAGGATAT ACATTaTCGA CGAAGATTAC AAAGGGAATA TGGAATyTGG 9840

CGTTAACTGC TATGCAATAT ACATTCCCTC AACCGAATTG TGCTGGTTTA TATGAGATGG 9900

GAAATCAAGG AATTCTTTTT GGTGCTAACC GTGCTGGAAA TATTTACCTA CAAGGAAGTT 9960

GGTACACGGA CGATCCGTTT CCAACAAAAT AACAAGCTAG TTTAGGAGTC TTTTTATGGA 10020

ACGTTATCTC AACACAATAA CAATGCTTTT AAGCATTTTC GGTGGGATTG TCGTACGTTT 10080

ATTAGGCGGA TTAGATCAAT TGTTGGATGT CTTCCTCTTT TTAATTATTG TCGATTTCAT 10140

CACAGGTTGG ATTAAGGCAA TCGCCACAAA AGAATTGTCC AGTCGGATTG GTATGCTCGG 10200

AATTGCGAAA AAAGTGACGA TGTTATTTGT GGTTGCCGTA GCGGTTCGTG TTGAAAAAGT 10260

TGTGGGGAAC AATTTGCCAA TTCGGGAAAT GGTTCTGATT TTTTACATTG CGAACGAAGG 10320

ACTTTCTTTT TTTGAAAACA TTGCGACCTT TATTCCTATG CCGAAAAAGT TAAAAGAGTT 10380

ATTTATTCAG TTAAAAAATA AAGATGATTA AGTAGAAGTG GTCGGGACAA ACGTAGAACT 10440

TTCGcTGATT GCCGAAGAAA TTACTTCTGT CCCGCCATTT ATCTGCAGGT TTAAGCCGTG 10500

GAAGGGAAGT TATTTTGACT TTCCTTTCAT GGCTTTTTTA AGAAAGGAGC ATGCTATGTT 10560

TAAAAAATTA ATGATTCAAC TTGCTTTAGT GATTGGTTTA AGTTTAACGA TTCCGATGAC 10620

GGctTrCGCT TACACCATCG AAGCGGATCC AATCAACTTT ACTTATTTTC CCGGCTCTGC 10680

AAGCAATGAA TTAATTGTTT TACATGAATC TGGAAACGAG CGGAACCTAG GACCACACAG 10740

TTTAGACAAT GAAGTGGCCT ATATGAAACG AAATTGGTCA AATGCTTATG TCTCATATTT 10800

TGTCGGATCT GGTGGACGAG TGAAaCAATT AGCTCCTGCT GGCCAAATTC AATATGGCGC 10860

AGgTTCTTTA GCTAATCAAA AAGCCTATGC GCAAATCGAA TTGGCTCGAA CGAATAATGC 10920

GGCGACATTT AAAAAAGATT ATGCTGCCTA TGTTAATTTG GCCCGTGATT TGGCTCAGAA 10980

CATTGGTGCT GaTTTTTCTC TGGaCGATGG AACAGGTTAT GGCaTAGTCa CTCATGATTG 11040

GaTTACAAAA AATTGGTGGG GAGATCATAC AGATCCTTAT GGtTATTTAG CGCGTgGGGG 11100

ATtAGTAAAG CGCAtTGGcA CmAGATTTaC aAcGGGCGTT TCyGrAACAg GTGaGACTGG 11160

TCATTATTCA GCCAGGTAAA CCTAATGCGC CAAAATATCA AGTAGGACAA GCAATTCGTT 11220

tCACTTCAAT CTATCCAACA CCAGATGCTT TAATCAATGA ACATCTATnA GCAGAGGCAC 11280

TTTGGACACA AGTAGGAACA ATTACAG 11307

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2866 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

AAAGCACAAA TGCAACAAGT TGTGATAGCT ATTCAAAATA CGGAAGCCTT TCAGTTATTG 60

AATGTAATGA AGGTTATAGG AGATGAGTAA ACATGTATGA AGGTTTGTTA AAACAGTATC 120

AAGCGTATTT GCCCGTAACA GAAAAAACAC CAATGATTTC TTTAGCCGAA GGAAATACGC 180

CACTGATTCC TTTACCTAAT TTATCCAAAG AATTAGGTAT TCAATTATAT GGAAAATACG 240

AAGGCTTAAA TCCCACTGGC TCATTCAAAG ATCGCGGGAT GGTGATGGCT GTAGCAAAAG 300

CGGTAGAAGA AGGTGCCAAA GCGATTGTTT GTGCTTCTAC AGGAAATACT AGTGCCGCCG 360

CTGCTGCGTA TGCCACAAGA GCAGGCATTA AAGCGTATGT AGTGATTCCA GAAGGGAAAA 420

TCGCACTGGG AAAATTGGCG CAAgcTATTA TGTATGGGGC GGACATTATT TCGATTCCAG 480

GTAACTTTGA TGAAGCATTA AAAGCGGTGC GAGAGATTGC GAAAACTGAG GCAGTTGCGC 540

TAGTTAATTC AGTGAATCCT TATCGCTTAG AAGGGCAAAA AACAGCAGCT TTTGAAGTTT 600

GTGAACAATT AGGTCAAGCG CCCGATGTCT TAGCGATTCC AGTTGGAAAT GCCGGCAATA 660

TTTCCGCTTA CTGGAAAGGC TTTAAAGAAT GGCATGAAAA ACAAGGAACA ACTTTACCAA 720

GAATGCATGG TTTTGAAGCA GAAGGTGCCG CAGCGATCGT GAAAGGCCAA GTGATTGAGC 780

AACCTGAAAC CGTGGCAACA GCGATCCGGA TTGGGAATCC TGCTAGTTGG CAATTAGCAG 840

AACAGGCACG GGATGAATCC GGTGGTTTCA TTGATGCCGT GACAGATCAA GAAATTTTAA 900

CGGCTTACCG AAAAATTGCT GCTCAAGATG GGGTTTTCAT TGAACCTGGT TCGGCTGCTT 960

CATTAGCTGG AGTTATTCAA CATGTTAAAA GCGGTAAAAT CAAAGCAGGT GAAACCGTCG 1020

TGGCAGTCTT TACTGGTAAC GGACTGAAAG ATCCTGATAC AGCAATGGAG ACAGAAGTAG 1080

CTATTTCAAA AATGAGTGAC GTGGAAGaAA TGCGTCTGcA CTTACgAAAA GGAGTCGCAA 1140

CACTATGaAA ATAAGAGTAC CTGCCaCtAG TGCCaACTTA GGACCTGGCT TTGATTCATG 1200

TGGCATTGCA TTATCCGCTT ATTTAACAAT CAACGTTTTA GGTGAAAGTG AATTTTGGGA 1260

AATTCAACAT ACTCTGGGAG AAGAGATTTC TACTAATGAA GAGAATTTAT TAATTCAAAC 1320

GGCATTGAAA ATTGCACCAG AATTAACACC AAAAGTTATT CGGATGGTTT CTGACATTCC 1380

CCTTGCAAGA GGACTAGGCA GTAGCTCAAG TGTGATTGTA GCAGGCATTG AATTAGCTAA 1440

CCGACTAGCA CATTTAAATT TGTCCCCTAA AGAAAAAGTC CGTTTGGCCA CAGAGATGGA 1500

AGGACATCCG GATAATGTGG CCCCAGCTAT TTTAGGAGAT TTCGTGGTGG CTAGCCATGT 1560

AGAAAACCAA GTCTACCATG TCAAACATCA CTTTcCAATG tGTGATGtCA TTGcATTTAT 1620

TCCAGAGGAA CCATTATTTA CCGAAAAAAG TCGCGCTGTT TTGCCCGAAA AATTGGCTTA 1680

TAAAGAAGCG GTCGCGGCCA GCTCAATTGC CAATGTTATG ATTGCCGCAA TTTTAAACGG 1740

GGATTTACCT TTAGCAGGCA AAATGATGGA ACAAGATAAA TGGCATGAAA CGTATCGTCG 1800

TTCATTGGTT CCTCACTTAA AAGAAATTCG TCGTTTAACT CAGCAAAAAG GTGCTTATGG 1860

CTCTTTCTTA AGCGGTGCTG GTCCAACTGT GCTAATCTTG TCGCCAGAAG AACGAACCAA 1920

CGAAATTGTT CAAAGCTTGG AGAAATTATC TACCAAAGCG AGCATCCAAA TTTTCAATAT 1980

TGATCAAGAA GGTGTTCAAG TCTTTTAAAA AAAACGAGAA AAAATCACTA GTGATTTTTT 2040

CTCGTTTTTT TTAATCCTCG GATGTTTCCG TGGAATAATA ACTATGATCT TCTAAGTTCA 2100

AAGCATTAAG AATCATTGAG GCAGTTTCTT CAATTGAAAG AGAAGCGACA TCAATAACGA 2160

TACAACCTAA TTTTTGATAA AGATCATTAG CAAAAGCCAG TTCTCGCCGA ATTTTTTCAA 2220

TATCTGAATA AGAAGTATTT TCAGGCAATC CATATGCGCG CATTCGTTCT TTACGAATCC 2280

CGTTTAAAAC ATTGGGATCA TTGGTTAAGC CAACAATTTT TTTAGGGTCC ATTTCGAATA 2340

ATTGTTTCGG TAAATGTGCT TCAGGAATTA GTGGCAAATT AGCAACTTTT AAGTTCTTGT 2400

TGGCTAAAAA GAGACTTAGC GGTGTTTTAG AAGTTCGGGA AACGCCAAGA AGTAACACAT 2460

CCGCTTCTAA AAAGCCTCGT GGGTCCTTGC CATCGTCATA TTTGACAGCA AATTCCATTG 2520

CTTCAATTCG TTTAAAATAA TTTTCATTCA GATGATGTAA GGCACCTGGT TCCCGCATGG 2580

GCGCGACACC TGTTAATCGT TCGATTTCCG CAACTGGTGG TGTCAAAATG TCAAAATGGA 2640

ACAATTGATG CTCATTGAAA AATTGATTAG CAATCGCCAC GAGATGATCA TTAATGATms 2700

tATGCAGTAC CATGGCCTCT GCTTTTTTGG CATCTTCTAA CGCTTGCACC AGTTTATCTT 2760

CGTCTTTAAC GAATGTTCGT CTAATCAAAG TAAAATCAAC AGTGGGGTAC TGGGCCATTG 2820

AAGCAGCCGC CAATTTTGAA GCGGTCTCGC CAGCAGAATC TGAGAT 2866

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1663 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

GATTTnAAAn nAATTAATTG AAAATACCCG CTGsCAtGaC ACGGGAAGGG aTTACGAAAG 60

AAATTCCAAT GGACGAAATT GTTCCAGGAG ACATTGTAAC TTTAGCTACA GGAGaCATGA 120

TTCCTGCAGA TGCCGTGTTG ATTTGGACCa AAGATTTATT CGTTAATCAA TCgTCGCTAA 180

CAGGCGAATC CATGcCyswC GAAAAATTTG TGGATGCAGG TGTGGATCGT CAACAAACAG 240

AGGTTTCTGC ATTAGATATG CAAGACCTAG TGTTTATGGG CACAGATGTC TTGAGTGGAC 300

AAGGAAAAGC GATTATTTTA AAAACTGGGC AACATACATT TTTTGGTGAT ATTGCTAAAA 360

ATGCTACAAC ACAACGAGGA AAAACTAGTT TTGACTTAGG CTTAGCAAAA GTCAGTAAGT 420

TTTTATTACG AATGGTAATG ATTTTGTTTC CAATCGTTTT CTTGATTAAT GGTTTAACTA 480

AAGGTGCTTG GGGCGAAGCG TTTTTCTTCG CGATTGCCGT TGCAGTTGGT TTGACACCAG 540

AAATGTTGCC AATGATTGTT ACGAGTAACT TAGCCAAGGG TGCATTATCC TTATCGAAAC 600

ATAAAGTCAT TGTTAAAGAA TTAGCGGCTA TCCAAAATTT AGGTGGTATG GATGTGCTCT 660

GTACAGATAA GACTGGAACG ATTACCGAAG ATCGTGTAGT TTTGGTTCAA CATTTGAATC 720

CATTAGGTGA TTTGAGTGAC GAGGTTCTAA ATTTAGCCTA CTTAAATTCT TCTTATCAAA 780

CAGGTTGGAA AAATTTAATG GATATTGCCG TTATTAATTT TtATGAAGAG CATCAATGGA 840

AAACGCCGTT TAAAAATGTC ACTAAAATTG ATGAAATTCC TTTTGATTTT TCTAGACGCC 900

GCTTAACGGT TGTCGTGAAT GCAGATGATC ATCAATTGAT GATTACGAAA GGCGCCGTGG 960

AAGAAATGGA AGAAGTGTGC ACGCATGCGC AAATTAATGG CGAAATTGTA CCATTAAGCA 1020

CTGCTGTACG AGAAGAACTA CGCCGAGTGA ATGTTCAAAT GAACAAACAA GGGATGCGTG 1080

TGTTAGCGGT AGCTGTAAAA AAAGATGTCC ATAAAGAAGC AGTTTATTCA GTTGAAGATG 1140

AAAAAGAGAT GACATTGATT GGGTTTATGG GTTTTCTTGA TCCAGCAAAA GAGTCAGCCG 1200

TCAGCGCTAT CCGCTCCTTA CATGAACATG GTGTCAACGT GAAAGTGCTA ACGGGCGATA 1260

ATGACATTGT CGCCAAAAAA GTCTGTAAAG ATGTAGGCAT TGAAGTTTCC CATGTGTTAT 1320

TAGGTTCTCA AATTGAAGCA ATGACAsnTG AAGAGTTGCG GGCGCAAGTA GAAGAAACTA 1380

ATTTATTTGC CAAATTAAAT CCGATGCAAA ATCTAAAATT ATTGAATTGC TACAAGCAAA 1440

AGGTCATACG GTTGGTTTTA TGGGGGATGG CATTAATGAT GCACCGGCAT TACGAAAAGC 1500

TGATGTGGGC ATTTCAGTAG ATACTGCAGC TGATATTACT AAAGATGCTA GCTCGATTAT 1560

TTTACTGGAA AAAAGTTTAA ATGTACTGGA ATCTGGTGTG ATTGAAGGAC GAAAAGTATT 1620

TAGCAATATG ATGAAATATA TTAAAATTAC CATCAGTTCT AAT 1663

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5663 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

CTTAATGGnG ATGGnAAGGG GATATTATTG GTTATTTTTT GAGnATATAT GATATCCTTT 60

ATTATGTAGG AATCnTATTA TATGGsAGGm ACGAGCGATG AAAGTTATTG TTTTAGGATC 120

ATCACATGGA GGTTATGAAG CGGTAGAGGA ATTACTAAAT CTTCATCCTG ATGCAGAAAT 180

TCAATGGTAT GAGAAAGGTG ATTTTATCTC ATTCTTGTCT TGTGGCATGC AGTTGTACCT 240

AGAAGGAAAA GTGAAAGATG TTAATTCTGT TCGCTATATG ACTGGCGAAA AAATGGAGAG 300

CCGTGGTGTA AATGTCTTTT CTAATACTGA AATTACAGCG ATTCAACCAA AAGAACATCA 360

AGTGACAGTG AAAGATTTAG TGTCAGGTGA AGAACGTGTT GAAAATTATG ATAAATTAAT 420

CATCAGTCCC GGAGCTGTCC CATTTGAATT AGATATTCCA GGTAAAGATT TGGATAATAT 480

TTACTTGATG CGTGGTCGTC AATGGGCCAT TAAATTAAAA CAAAAAACAG TAGATCCAGA 540

AGTCAATAAT GTGGTTGTGA TTGGTAGTGG TTATATTGGG ATTGAAGCTG CCGAAGCATT 600

TGCAAAAGCC GGCAAAAAGG TTACTGTTAT TGACATTTTA GATCGTCCAT TAGGGGTATA 660

TCTAGATAAA GAATTTACAG ATGTTTTAAC AGAAGAGATG GAAGCTAATA ATATTACCAT 720

TGCAACTGGT GAAACAGTTG AACGTTACGA AGGCGACGGT CGTGTGCAAA AAATCGTTAC 780

AGATAAAAAT GCGTACGATG CTGATTTGGT CGTTGTAGCG GTTGGTGTCC GTCCAAACAC 840

TGCTTGGTTA AAAGGTACCT TGGAATTACA TCCGAATGGC CTAATCAAGA CGGATGAATA 900

CATGCGGACA AGTGAGCCGG ATGTATTTGC AGTAGGGGAT GCTACGTTAA TTAAATACAA 960

TCCTGCAGAC ACAGAAGTAA ATATTGCCTT AGCAACGAAT GCTCGTAAAC AAGGTCGCTT 1020

TGCTGTGAAA AACCTAGAGG AACCAGTTAA ACCTTTCCCT GGTGTTCAAG GATCTTCTGG 1080

CTTGGCCGTC TTTGATTATA AATTTGCTTC AACAGGGATT AACGAAGTCA TGGCTCAAAA 1140

ATTAGGAAAA GAAACAAAAG CGGTGACAGT AGTAGAAGAC TACTTGATGG ACTTTAATCC 1200

AGACAAACAA AAAGCTTGGT TTAAATTAGT GTATGATCCT GAAACAACAC AAATTTTAGG 1260

CGCTCAATTA ATGTCGAAAG CAGATTTAAC TGCAAACATT AATGCTATTT CATTAGCGAT 1320

TCAAGCCAAA ATGACGATTG AAGACTTAGC CTATGCGGAC TTCTTCTTCC AACCAGCGTT 1380

TGACAAACCT TGGAATATTA TTAATACAGC GGCTTTAGAA GCGGTGAAAC aAGAACGTTA 1440

ATAGAAGAGA mAATAAAACC CATGAGTATT TATAGTACTC GTGGGTTTtA TTTTTATTTT 1500

GTATTCAATT GATTTTTCAA CTCTTGTTGG ACACGAGTTA ATTCTTGTTC ATCCACACGT 1560

TGATAGGAAA CACCATCTTG CATAAAACCA GTACCTTGAA GTTGGTCTTG TTTCACTTTA 1620

CCAAAGGCAC TGCGATAATC TAAGGCAATT TTTTTCATGT CATCAAAACT TAAATCTGTC 1680

TTCATATTAT CAGAAACAGC TGTTAAAATT TCTTGaTAGT TGCTTACGCT GTTAAGACTT 1740

AAGACTTTTT GGACGATGCC TTCAATCACT TTTCTTTGAC GTTCTTGGCG GCCGTAGTCA 1800

CCATTAGGGT CTTCGTAACG CATTCTTGAA TAGGAGAGTG CTTGTTCACC ATCCAATGAA 1860

ATTTTACCAA TCGTAAAATC ATATCCGTCT TGAGAAAAAG TCAGATTATT GTTCACTTCG 1920

ATTCCGCCAA CCGCGTTGAC TAATTCTTTT AAACCAGCCA TATTAATTGA AACATAATGA 1980

TTAATAGGTA TGTTTAAATA GTTTTCAACT GTGTCCATTG CTAAAGATGC GCCACCAAAA 2040

GCATAGGCGT GATTCAATTT ATCTTGTTTT CCTTGACCTG GAATATCAAC ATAGGTATCG 2100

CGAGCAAGAC TGACTAAGGT TGTTTGCTTG TCACGAGGAT TAACTGTTGC AACAATTGTT 2160

GTATCCGAAC GACCTTGCTC GACACGCCCA TCATCGCCTG TATCAATCCC TAATAATAAA 2220

ACAGAAAAAG GCTCCTTATT GTTTAAATTG ACCTGACTTT TTTTAGATCG TTCTACTGTT 2280

TCATAGGTTT TATCCATTGA TTTAGAAACA TCCCAATAAA GTTTTGCTCC CATtCCAaCA 2340

ACTGCAaGAA aTAGAmCCAa GATAATTCCA ArGATAATGG CAAAAATTTT CTTCCCTTTT 2400

GACATATTTT yCCTCCAATA ATTTATCACA GGTTGTGATT TTTTTCTTAT AGGTTAATCT 2460

TACTAATTTT TAGTGAAGAA GTATATAATA ATTGTTGAAA GACTATAAAA CTTTAAGGTA 2520

TTTTGCATTA TTTTTGCTTA ATCTTTTGAA ATATCTCAAT TCTATTAAAA TATGCGAATT 2580

AGTATATTAC ACGTTTCCAA TGTGTTTCAT TGTTGATTAA TTAGAATTTT AGTGGTATTT 2640

TTGTATTTGT AATTAGTACA GATTTTAATT GGAGGTATAA CAGTGAGTAA AAAAGGATCA 2700

GATATCATAG TAGAAAGTTT AATCAATCAC GAAGTGCCCT ATATTTTTGG GATTCCCGGT 2760

GCAAAAATTG ATGGGGtTTT TGATGCATTA GTtGACAAGG TCCAGAATTA ATTTTAGCTC 2820

GACATGAmCA AAaCGCTGCC TtTATGGCAC nAGGGaTTGG TCGTTTGACG GGAGAGCCTG 2880

GCTTAGTTAT TGCAACTAGT GGACCTGGTG CCAGCAATTT AGCAACTGGG TTAGTAACTG 2940

CGACTGCAGA AGGGGACCCT GTTTTAGCAT TGGCTGGACA GGTGAAACGG TCCGATTTGT 3000

CAAAATTAAC TCACCAAAGC ATGAATAATG CTGCATTATT TAGTCCAATA ACAAAATTCA 3060

GTACAGAAAT TCAAGATCCA GAAACATTAT CTGAAAATAT TGCCAATGCT TACCGTATCG 3120

CCAAAACAGC GAAAAAAGGC GCAACATTCT TATCGATCCC ACAAGATGTC ACAGATAGTC 3180

CGGTGACAGG CGAAgcAATT AAGCCTTTGT CGGCTCCGAA ACTGGGCCAT GCATCAGCAG 3240

GAACAATCGC AGCTTTAGCT GAACGAATCA AGCAAGCGAA ATTACCTGTT TTACTTTTAG 3300

GAATGCGGGC CTCTGCACGA AAAGTTACCG CAGCCATTCG TGAATTGGTT GCTATAACTG 3360

GTTTGCCAGT CGTGGAAACG TTCCAAGGAG CAGGCATTAT TTCAAAAGAA TTAGAAGAAC 3420

ATTTCTTTGG GCGTGTTGGT TTATTTCGAA ACCAACCTGG GGACCGTTTG TTAAAAAGAA 3480

GTGATTTAGT GATTGCGATT GGTTATGACC CAATTGAATA TGAGGCACGT AATTGGAACG 3540

CAGAAAAAGA TGCGCGCGTT ATTGTTATTG ATGAAGTTCA AATGGAAATT GATCAATATA 3600

TGCAACCGGA AGAAGAGCTA ATTGGTGATA TGTCTAAAAA TATTTTGAAA TTATCAGAAG 3660

```
CATTTAGTGA ACCAATATTA ACAGAAGATG CGCAAGATTA TTTAGAAACG CTTCAAGAAA    3720
AATTAACGAT TAAAGAAGTA AAAACAAGTA CAATTGAAAA TCGATTACAT CCATTAGAAA    3780
TTGTTCAAAC GCTTCAAGAA AAAACCACGA ATGAAATGAC AGTTACTGTT GACGTAGGGA    3840
GTCACTACAT TTGGATGGCT CGCCACTTTA GAAGTTATGA ACCGCGTCAT TTATTATTTA    3900
GTAACGGCAT GCAAACATTA GGTGTCGCTT TGCCTTGGGC AATCTCAGCT GCACTAGTTC    3960
GTCCAAATAC ACAAATTATC TCTGTTTCTG GGGACGGCGG TTTCCTATTT TCGGCGCAAG    4020
AATTAGAAAC AGCGGTTCGT TTAAAGCAAA ATATCGTGCA TATTATTTGG AATGATGGTA    4080
GTTATGATAT GGTTAAATTC CAAGAAGAAA TGAAATATAA TCGATCCTCT GGCGTAGATT    4140
TTGGTCCAGT TGATTTTGTG AAGTATGCTG AAGCTTTTGG AGCCAAAGGG GTTCGAGTTC    4200
ATTCACAAGC AGAATTTGCC GCAGCTTTAG AAGAAGGAAT GCAAACAGAA GGCCCAGTGA    4260
TTATTGATGT TCCTGTGGAT TATTCAGATA ATCmAGAATT AGGTAAAaCA TTGTTACCAG    4320
ATCAACTTTA TTAGGAGATG GAAAAATGAG CGAACAATAT GTTTATCAAC ATGGTACGCT    4380
AGGCGGTTTA ATGGAAAGTT TGATGGCAGG AACGGCAGAA ATTGGTACGT TACTCACGCA    4440
AGGTGATTTT GGAATTGGGA CATTAGAAGG CTCAAATGGC GAAATTATTT TATTGGACGG    4500
TACATTGTAT CATGCAAATC AAACTGGCGA AATTACTATT CTAGAAGGCG AAGAATTAAC    4560
GCCATATGCC GCAGTTACTC GTTTTCAAGA AGATGGCGCA TTCCCTGTAT CAACGGAAAC    4620
CGATGAAAAT ATTAAAGCAC AAATTTTAGA AAAAATTAGT CCTAATTTTT TTGCGGCAAT    4680
TAAAATTAGC GGTCTTTTTG CGAAGATGCA TGTCCGTGTG GCACCTAAAC AAGAAAAACC    4740
GTATCCACCA TTTGTAGAAG CAGCACGCAA TCAACCTGAA TTTACAGCGG AGAATATTCA    4800
GGGAACAGTT GTAGGGTTCT TTACACCTAA ATTATTTCAT GGTGCCTCTG CCGCAGGgTT    4860
TCATTTGCAC TTTATCAGTG AGGATCACCA ATTTGGTGGG CACATTCTTG ATTTTGGCAT    4920
TAAACAAGGG ACTGTTTCGT GGATGGAAAC AGCAGAATTG CGACAGCATT TTCCAGTTCA    4980
TGATGCTGAT TATCGGAATA AAGAAATTGA TATTGCAAAA GCTTTGTCCG CAATTGAAGA    5040
AGCGGAATAA CTGAATAACA AGGCACATTT CTCCACGscC ATTCAAGTTG ACTAACTAGG    5100
TCAACTTGAA TGGGCTTATT GCTGTTCGTC GCTTTTTGTT GTTATTTTTC AAAAAAAAAT    5160
AGTATCCTTT TTATAATAAG CAAATAACAA TTGGAAAGGA TGGCGTAATG GACACACTTT    5220
ATAGAAGTTG GCAATTGTCT GGTTGGCTCT ATCATGATAT TTTTGTTATC ATTGTAGCAA    5280
TTATTTTTAT AGTAATTAGT GGTATTTTGG TTATTTCCTT AATAAGAAGG CGGAGTACAA    5340
GGAGACTGGT TCCCTATGCG TTAATTTTAT TGGTTTATCT TGCAGTGGTA CATTTTGCTG    5400
GTTTAATCTT TTTCGGAATG TTTCGTTCAG TAACGATTGA AGAGAAAAGT GCGACGTTTT    5460
ATTCTGAAAA AACAAAAGGC TTAACTTCAA TCGAGCGAAT GATTATTCCA AATGGACGAA    5520
CGAATGGCAT TAGCACCAGT AACTCATTGT TCCCAAGTGG ATTTCnGTTA ATAGTCAGAC    5580
GGGCCGAACG AATGTGGnAG TAAACGTTTA AGTTGGCCGT GATTAATTTA ATGGnCCAAC    5640
```

AGATCCATAG GGGGTTTTGA nAA 5663

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1799 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

AAACAGCATA GACAGCAATT GAGTTATGGC CmCgGTTtGA AGcwTwAAGG rAACgGCCaT 60

CAGArGAAaT ACGrATCGsA GcmCCmCCmT TAAACsCATC ATAATCAGCA GGTAAAGTAG 120

AAACTTTTTG TAAtTCaGTG AAAGaACCAT CTGTTyCATC ATATGAAAGA ACTGAAACAG 180

TGCTATCTAA TTCACCAAAT AAATAAGCCA ATGATTTTGT TGGGTGGAAC ACTAAATGAC 240

GTGGACCAGT CCCTGGCTCA GTCACGAATT GAGCGATTTC AGAAAGCTTT CCATCTAAGC 300

TAACATTATA CGTATACACA CGATCGGTTC CTAGATCACA GACGACTAAG CGTTGATCAG 360

GCGTCAAATC TGTGTAATGA ACATGGGCAT TGTTTTGGTT TTCATGAGGA CCAGTTGGCT 420

CTGTATGAAT CACTTTGTCA GCATCTTCCA AAGAACCATC TTCTAAGATA CGATACACGT 480

GAACGATACC TTGGTGATAA TTTGCTGCaT AGACAAGTTG TCgGTtyTCG TCaACAGCTA 540

cATAAcATGG CGGCGCACCT TCCtCaGTTA CTTTAtTCaA TAAGTTAAAA tGGTCCATTT 600

CTGGTTGATA AGCGGCAGTG CCACCTTCAC CGTCAACACT AGTGACGCTA AAAAGTTCGC 660

CCTTTTGACT CAGAGCAAGA TAGGTAGGAC TCGTTTCTTC TAACAAAAGT GTCGCTTCTG 720

TTAAACGACC TTGTGTCGTA TCTAATGCAA TTTTATAAAT GCCTTGGCTG GCTCTTCGAG 780

TATAGGTACC TAGTAAAATT TGTTCTAACA TCACAATTCC TCCAATTTGT TTCAACACAA 840

TCACTTTTTA TGAAAGTGCT TTTATATAAT GTAGTATACC aTAAGAAGTC TAAGAACAAT 900

AAAACTTTCT ATCTGAGATG TTCgCaTTTT AAAgtGGTTT GtGCtATTAT GaTGaGaGAC 960

tGTGaTGACa GTCaTTTTTT gaTTTTTAGA GGGGTGACGA ACATGTACGA CaATACgGAT 1020

AATAACGGgA CCGTTCGTTG TTCATTTTGT GGGAAAACaC AAGAAGAAGT AAAAAAAATT 1080

GTTGCTGGTC CGGGTGTCTA TATTTGTAAT GAATGTATTG ATCTATGTAA AGAGATTATT 1140

GATGAAGAAT TTTATGATGA AGCAGTCCGA GAATTAACGG ATGTGCCAAA ACCCCAAGAA 1200

ATTTTAAATG TTTTAAATGA ATATGTTATT GGACAAGAAC GTGCAAAACG CACACTTTCT 1260

GTTGCTGTTT ACAATCACTA TAAACGTGTG AATCAATCAG AAACAGCTGC AACTCAAGAT 1320

GACGTTGAAT TGCAAAAAAG TAACATTTGT TTAATTGGAC CAACAGGTTC AGGTAAAACT 1380

TTCTTGGCTC AAACGTTAGC AAAAACATTA AATGTACCTT TTGCAATCGC TGATGCGACC 1440

AGTTTAACAG AAGCAGGTTA TGTAGGGGAA GATGTTGAAA aCATCCTTTT AAAATTACTA 1500

CAATCTGCTG ACTATAATGT TGAACGTGCG GAAAAGGGCA TTATCTATAT TGATGAGATT 1560

GATAAAATTG CCCGTAAGAG TGAAAATGTT TCGATTACTC GAGATGTTTC aGGTGAGGGT 1620

GTTCAACAAG CCTTACTAAA AATTCTTGAA GGAACAGTTG cAAGTGTACC ACCACAAGGG 1680

GGCCGCAAAC ATCCACATCA AGAATTTATC CAAATTGACA CAACGAATGT TCtCTTCCAT 1740

CGTTGGTGGG AGCATTTTGA TGGGAATTGG AAACCATGGT TTAAAAACCC GCCTTAGGC 1799

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11266 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

CGCCAAGTTC GTCAATAATC GGTCAGATTG TTTTCAATCA GGAAAtcAcG TTTTcGCCGC 60

TTCTTTCAGg TgAAAATTTT CGCCCAGCGT TCATAGCCTG CGCCnTTTtC CGGCGGCAAG 120

TTTCGCCTGA ATGTCAATCA GGAGATTGAC CTTTCTGCCG GTTCGCTTCG GCGCTGTTTT 180

TCCTGACGTC AAACGTCTGC GCTGGCTGGT AATTCTGGCG CGCAAGGCGG TCTTGGTATA 240

ATCGTCGCCC AGCCGGTTCA CCCTTGTAGG GCGATCCCAT CCGGGCAGAC GGAATCCGGG 300

ATTTTTGCCG TCCACAACCT CATATCCCTC AGCCCGCATC TTTGCCAGAA ATTCCTCATA 360

ATGATTGCAC TCTGGCAGCA GCCGGTCAAT GGCTTGCTGT AACTGCTTTT TGTGgCTGAC 420

GCCGTACTGC GCCGCCGCCA TTTCCTTTTG GGGCCTGGCT TTTTTCTTCG GATTCTCAAT 480

CACCGAAAGG CCGTGGGCGC GGCACAGTTC ATCATTCATG CGTCGCAGGA TAAAAGCGGA 540

GTTTTTCACG TTCTGGAATT TTGCGTCACA GTTCAGGTTG GTGCTGTTAA ATTCAATGTG 600

GGTATGGATA TGCCGCTTAT CCGTGTGGGT GGCCACCACA AACTGATGCT GCCCGCCGGT 660

AAATTTCATC GCCAATTCAC AGCCGATTTT GTGCGCCTCC TCCGGGCTGA CTTCATCGGG 720

TTTGAATGAC TGGATAATGC GGTACATGAT AATGTCATGC TTTTCTGAmC TGCTGCGCCC 780

GGTAAGCTGG GCGTACAGTT TTTTGCTGAT TTCAAATTCT TCGGCGGCAG TTTCCGGGCT 840

GCACATATAG GAGGAAACCA GTTCGCCGCC GTCTGTCTTT TCTTCTTTCT GGTCATAGTC 900

GTGGCGTTCT GCCATTGACT GCTGACGGGT GCGGTTTTTC ATAATCTCTC GTGGTAAAAT 960

ATTAGAGATA GCCCAATGGT TCCTCCTTTC AGAAAAAAGG CTATCCATAG CCTAAATAGT 1020

AGCATGGGTA GCCTACTTTT GCAAGCCATA TTTTGTATGC TTGATACnAA AGGGGGGCAG 1080

GATACAGCTA TGGATACGAT TTCAGCGATT AAGACCCGGA TACTCCGTTT ATGTGCGGAG 1140

CGCAACATTA CCATCAACAA ACTGGCTACG ATGTCGGCGC TGCCGCCATC CTCTGTCAAA 1200

AACATCTTGT ATGGAAAGAG TACCGATCCT AAAATTTCAA CCATCAAGAA AATATGTGAC 1260

GGTTTGGATA TGACGTTGAT AGATTTTTTC GATGCGCCCG AATTTGAAAA TTTAGArCCG 1320

GAAATTAAAT AGCGTTACCG CGCCGCGTCC CGCTGCCGCT GTGCCGTCCG TACCGCTTTA 1380

ACGTGGGTGC AGAGAAACTC GTTATAGTCC TTGCCCTGCT TCGGCGGGTT GGGGTAAACC 1440

AGCTTCACCC GCCTGGAAAG TTCCGCATCG TTCTGGATTT CCTGGGTGAG CCGTTCCATG 1500

CCCTCCAAGC CGGCCTTGTC ATTGTCGAGG CACAGACTGA CTTGGGTAAC GTGGGGGTGG 1560

TCGTGAAGAA ACTGTATCAG GGCGCGGGGG CCGGTTCCGC CCAGGGATAA ATAATGGCTG 1620

TCCCGTGGTT CGCCGCCGGA CAGCTTTACT AATGTGGCGA GGGAGAGCGC GTCAATAGGG 1680

CTTTCCGCCA CCGCCAGCCG GGGGCAGTCT GCCTTGCCTG CCGGCAGACA GAAGCTGTAC 1740

CGCTTGTCGC TTCCCTCCAC ATCCATGCGG AAGCCGTCCC GCGTCCCCCG CArGCTGGCG 1800

GCGCGGGCGT TCCCTGCTGT GTCGCGCCCA ACAAAGACGC AgTTCTGGTA TTTGCAGCTC 1860

TCGTATAAAA TCCCGGCTTC CATGCACACG CCAAGCAGTT CCGGGTCAAT GCCGCGATCC 1920

TGCAAATAGC CCACAACAGC GGATGCACAT CGGCTGGCCT GGGGCAGTCG GAAGGGCTTT 1980

GGCTGTTTTT CTTTTATGGG CGGCGCTTTT TCGGGCGGTG GCGCACAGTA GCCGCACAGG 2040

GTTTCCACCG CGCCTACAAA GTCCATGCCG CGCACCTTGA TCAGATAATC AAGGGCCGTG 2100

CGCCCACCGA TGCCCCGGCT GTTCCAGCAC CATTTCCCGT TGGAGATTTT TAAGCTGTCA 2160

TGGGTTCTGG TGCAGTATTC GCGGGGGCCG CTTTTTTTCA GTTCCTGCGG CTCATAGGCC 2220

TGCAAATAAG AGAGCAGATC CCATTCCTTT GCTTTTGCAA TCTGCTCTTT GGTTACGCCT 2280

GCCAAGTAAT CACCTCCAAA TATGAAAAAA GGGCGGCGGG AGCATCTTTC TTTTGCCCCC 2340

GCCGCAGTTG TTTACAGTGc TTCAAATTTC TTCAGCAGGT CGTTCACGCC CTGCCAGATT 2400

TCAGAGTAGC CCTGCCGCAA GTCCTCCACA TCCTGCTCAT AGAGGTTGTG GGTGCTGTTG 2460

CACsGGCGGG CGATCTGGTT GACGTTGCTG GAAACGGAAc GCAGCAATTT TACCAGTTCC 2520

AGCACCGCGC CCATGTCAAG CTGCACGATG TAGCCGTCCA CCGCCATCTT GCGTAGGTAA 2580

GCCCGCAGgT TGGAGGTTCC CAGCAATTCC ATTTTGCGGT CAATCGCCGC CTTTTCTTCC 2640

GGGGAAACTT TGGTGAACAG CCCCACCGTG CGCCCCTCCT TTATCCTCAC AGCACCGCCT 2700

CCTTTCGCGG GGTGCGCTCC ATGCCGGAAT GTACCGGCGT GGGCGGTTCA TAGTTACGGA 2760

TTGCCTGACG GATAGAGGGT TTATAGCCAG CCGCCACAGC GGGCATCGGC ACAACCGGCG 2820

CGGGTTTTCC GGCGTCCTCC AAATCAGGGC TTTCTTCAAC CTCCTGCTCA TCCTCCTTTT 2880

TGTCCGGCCT GGACAGTTCC CGGTTAAGCT GGTTTAAGCG GGCGGATTTT TCCCGCAGTT 2940

CTTCTTCTTT TGGGAATGGC CGGTCGGCTT CTTCTTTGGC GTTTTCCAGT TCGCTTTCCA 3000

GTGTGGCAAG CCGTGCTTTT TGTGCGTCCA ACCGCTCCGT CAGGCTTTCC AGCGCATTGT 3060

TAATACGGGT GACGTTGCCC ACCGCATCGT CAGACAGTTC GGCGGAATAG GTCAGGTTCT 3120

GCTTCATGGT GACGCGGAAT TTGCTGCCGT CACACCTTAG CTGCATGGGA AAGCCGCGAA 3180

ACTCGCCCAA ATCCACGGTT TTTTCCGGGT CGTCCATCAG CATACACGCT TTAATCACCG 3240

cTTCGCCCGC CGCCTTGCGC TCCGTGTAGG TCTGCCCCAT GACAGTCATG GAAAACGCAT 3300

CGTCCTTGAC CGGGTGGGCC TGCACGGTTG GCAAATCTGC CGTCAGCGCT TCGATTAACA 3360

GCTTGGTTTC CGCGATTTTC TGCGGGTAGT ATTTCAGCGC TTTGTCCTGC ATTTCGTACT 3420

```
GTTGGGAGGT GTGGTTGGCT TTGAGCATTT TCAGCTTGGA AACCTGCACA TCCAAATCCA    3480
TTTTTTCGCG GATACGGTCG TCGCCGGTGG CAAGGGCCTT GACCTCAGCA TAGGAAAGGG    3540
CAGTGGCGTC CACATCCTCG GCAGAGCGGG CAGGGCTTTT AGAGGTCATG ATCTGACCGA    3600
TGAATTTCTG TTTGTTTTCC ACCAGCCCCC AGTTGTAAGC GTCGAATGTA CCCTTGGTGA    3660
CATACTTGAA CATTTTCACC GACGCATTCA TGTTGCCCTG GCGCAGAGAT CGGCCCGCCC    3720
GCTGTTCCAA ATCGGCAGGA CGCCAGGGGC AATCCAGGTC ATGGGACGCG ACAATGCGGT    3780
TCTGCACATT CGTGCCTGCG CCCATTTTTG CCGTGGAGCC GATTAAGACA CGCACCTGCC    3840
CCCGCCGCAC CTTGGCGAAC AGTTCTGCCT TTTGGGTTTC CGTGTTGGCG TCGTGGATAA    3900
AGGCGATTTC CGCTTCGGGG ATGCCTTTGG CAATCAGCTT TTCCTTGATG TCATCATAGA    3960
CATTGAATGA GCCGTCCCCC TTGGGTGTGC TGAGGTCGCA AAACACTAAT TGTGTTCCTA    4020
AAATCTCCGT GCTGGCCTGC CATTCCTGCA CGATATTTTT CACGCAGGCG TTGACCTTGC    4080
TGTCCGGGTC GTCGGGCAGG AGCGGATTTT GCAGCCGCTG ATCCAGCGCC AgCTTGCGCC    4140
CGTCCGAAGT AATCCGCAGC ATATTGTCGA TATGGGGGTC AACCTGCCCG CCGCGCACAC    4200
TTTCCGCACG CTCGCCAAGT TCCTCCACCA TTTCTTTCTG GAATCCGCTG GGTTCCGTGG    4260
TGACGGTGAT ATATTCCGCC TTTGGCACGG GGAGATTCAG CATATCGGCG GTCTGGATGT    4320
CGGCGGCTTC TTTCCAAATA GAAATAAGTT CCGGCAGGTT ATAGAATTTC GCAAACCGGG    4380
TCTTGGAGCG GAAGCCGCTG CCCTCCGGCT TTAATTCCAT CGCCGTGACC TTTTCGCCAA    4440
AGTCTGCCGC CCATGCGTCG AAATGGCGGT GGCCGTTCTG CTCCAGGGTG TCAAACTGGA    4500
GGTAGCGCAT CATGGTATAC AATTCTACCA TTGAGTTACT CACCGGGGTT CCGGTAGCGA    4560
ACACGATGCC GCGCCCGCCG GTGATTTCGT CCATGTAGCG GCACTTGGCA AACATATCGC    4620
TGGATTTCTG CGCGTCCGTC TGGGCAATCC CCGCCACATT CCGCATTTTT GTGTGTAAAA    4680
ACAAGTTCTT AAAGCCGTGC GCCTCGTCCA CAAACAGACG GTCAACGCCC AATTCCTCAA    4740
AGGTAACAAC ACTGTCTTTT TTCTTTCCGT CCGCCAGCTT TTGCAGCTTG GCTTCCAGGT    4800
TTTTCTTTGT TTTCTCCATC TGCTTGATGG TGTAGCGCTC GCCGTCCTCC GCCTTTGCTT    4860
CGGCAATGGC GGTGACGATT TCATCAATCT GCCCCTTAAT GACGGCTTTC TGCCGTTCCG    4920
GGGAGAGGGG GATTTTTTCA AACTGTGAGT GTCCGATAAT TATCGCGTCA TAGTCGCCGG    4980
TGGCGATCCG GGCGCAGAAT TTCTTGCGGC GGGCAGGCTC AAAGTCCCGT TTCGTGGCAA    5040
CCAGCACCTT TGCGCCGGGG TACAGGCGCA AAAAATCGCC GCCCCATTGT TCCGTCAGGT    5100
GGTTGGGAAC GACGAATAGG GATTTCTGGT TCAGACCCAG CCGCTTGCCC TCCATTGCGG    5160
CGGCAATCAT TTCAAAGGTC TTGCCCGCGC CGACACAGTG GGCAAGCAGC GAGTTTTTCC    5220
CATAGAGCAT CCGCGCCACG GCGTTCTCTT GGTGGGTACG CAGGGAGATT TCCGGGTTCA    5280
TCCCAGCAAA ACGGATATGG GAACCGTCAT ATTCGCGTGG ACGGATGCAG TTGAATTTCT    5340
CGTTGTACCG CCTGCAAAGG GTTTCCCGAC GTTCCGGGTC TTTGAAAATC CAGTCCTTGA    5400
```

ACGTTTCGCA CATGGCTTCC TGCTTCTGCT GGGCAATGGC GGTCTGCTTT TCGTTGAGGA 5460

CGCGCACCTC CTTTCCGTCC ACTTCTGTCT TATCAAATAT CCGTACATCC CGCTGGTTTA 5520

AGGTCTGCTC AAAGATTTCA TAAGCGTTGA TGCGCTTGGT GCCGTAAGTA GCATAGACGC 5580

GCACGTTATC CCGGTTGTCT GTGCTTTTTC CCTGCACGTT CCATTCGCCG GTGGTAGGGG 5640

AATATTGCAG CTTGATTTTT CTGTCCTGTA AATACACAGG TGTATGCAGC AGTTCAAACA 5700

TGAATTGCTG GTAATACTCC GGGTCAATCC AAGACGCGCC GATACGGACG CTGATTTCAG 5760

ATGCCGTCAG GTCTTTGGGC TGCACCTGTT CCAGCTTTTC GGCATTGATG GCAAACTGCG 5820

GGTTTTCCTC AGCGGCGGCG CGGGCAATCG CTAACTTGGC CCGGACGTTG CCGGAAAGGT 5880

ATTCGTCCGC CGTCTGCCAG CCCTGATGCC AGTGTGCGCC GCCGTCCTCC AGGTCAAAGG 5940

GGCCGGTGGC GGGGTTTTTG AAGATGATGC CTTTTAAGTC CTCCACAATC TGCGGGATTT 6000

TTTCAGGGCC GCCCATGAGG GATGCCATAT AGCCCAAATC CACGCAGGCT TTTTCACCGA 6060

TAGATACCGC CAGCGCCTCC ACAGCCGTGT CCACGCTGGT AACGGCGCGG TGGGTCTGGA 6120

TGGTGCGCCT GGTGAACATA TCCGCTTTAC GCTCCAGCTT CCCCTCCTCG TTGATGATTT 6180

CAAGAGAGCA GAGCAGCGGA TAGGAGGAAT CCTGCTCAAA GGCCAGCTTG TTGCCAGTGC 6240

TGTTAAGAAG CCCGTACTTT GCGCTGAACG CATCATAGAG CCGGTTTAAT TTTGCCTGCT 6300

CGGACTGGAT TTCTGTGTCG TCTGCGCCGG CAAGCTGTAA GTCGATCAGG CGGCGGGCGC 6360

TGTCCCGGAT GCCGATCATG CCCCTGACGC GCTCCGTGGG GGTTTTGCCA AGGACGGCCT 6420

GGGTCATGCG GGAGTTTTCC CGGAAGTACA GCTTATCATT GGATAGGGTG TAACTGAAAT 6480

TGCGTACAGT GGGGTCGGCG GGGATGCTTT CCAGTTCCTC GCCGTCTTGC CCGGAATCCA 6540

CTTCTAACAG TTCCCGGTCA GGCGGTGCGA TATGCCGGAT AGCTTCTGCA AGCTGCCCGG 6600

AAAGGTCTGC GCCGGGGATG GGCTGGCAGG CGGTTTCAGT TTCATTGCCG TACATACTGC 6660

GTCCCATTGT CATTGTCCCC AGCACCATTT CCGGGTGGTC TAAAAAGTAG CGGTTCAGGG 6720

GGATGCCGTC CTCTGTCTGC CCGACCTCCA CCCATTCCGG GAGTTTTTCC GGGATGCGGT 6780

CGCGCTTCTG GAAAAAGAGG ATGTCCGTTG TGACTTCGGT TCCGGCGTTG GCCTTGAACG 6840

CATTGCTGGG CAGGCGGACT GCACCCAATA AATCGGCTTT TTGTGCCAGC GCCTCCCGCG 6900

CCTTGCTGTT TTTCTTGTCC ATTGTGCCTT TGGTGGTGAT AAAGGCCACA ATGCCGCCGG 6960

GACGCACCTT GTCCAGCGTT TTGGTGATGA AATAGTCGTG AATCATCAGG TTTTGACGGT 7020

CGTACCGCTT GTCATGCACC TGATAGCTGC CAAAGGGGAC GTTGCCCACC GCAAGGTCAA 7080

AGAAATCGTC CGGGTGGTCG GTGCGCTCAA ATCCGTCTAC CGTTATATCC GCCTTTTGGT 7140

ATAGCTGTCT GGCGATGCGC CCGGTCAGGT CGTCCAGTTC CACACCATAT AACTTTGCTG 7200

CCGCCAGTTT TTCGGGCAGC ATCCCGAAAA AGTTCCCGAT GCCCATAGAC GGCTCCAAAA 7260

CCGTCCCCGG CGTAAAGTCC ATCTGTTCCA CCGCATTGTA GATTGCCTGG ATAACGGTGG 7320

GGCTGGTGTA GTGGGCGTTG AGAACCGTGC TTTGTGCAGA CTGGTACTCC CCAGGGGTGA 7380

GCAGTTCTTT CAGTTCGGCG TATTCCTTTG CCCATTTTGG GTCGTTTGGC TCAAAGGCGT 7440

GGGAAATGCC GCCCCAGCCC ACATAACGGG ATAAAATGGC CTGTTCTTCT GGTGTGGCAA 7500

GGCGGTTTTC GGCTTCAATC TGTTTCAGGG TGCGGATGGC TTCCACGTTG TACTGGTACT 7560

TGGTCTTTTC GCCGCCTGCG CCTAAATTTT CATCGGTGAT GCGGAAGTTG TGGCGTTCCG 7620

GCTCCAGACG GAGCGGCTGT AAAATAATCT CGCTGGGGAA TACCCGGCTG TCCACTTCCT 7680

CCTGGGCGCG GCGGCGGGTC TGTCTTTGCG CCGGGGTTTC CGTAATCTGC CCGCCGTCAA 7740

TCTCCACCAG TTCCGGCGGT GCTTCCTGTT CTGGAGAAAC GTCCTTCTCT GGAAAAATTT 7800

CCGGCTCTGT TGGCTCCACG CCCTCTCTGG CACCTGCCAT TTCTGACCAC CTGATCCGCT 7860

GGTTCAGTTC GGCTTTTAAC CGTTCTTCAA TATCCGGTTC CACGGAAACC AGCATATCAT 7920

AGTCAAGGAA AGCAAGTTCC ATATCAATCA GTTCCCGCAG GCTGCCATAC TGCCGTGTTT 7980

CCACGCATTT CCCGTCTACG GACTGGCTGA TGGAGAAATC CAGCAGGTTT ACGTCCACCT 8040

GGACTTCTAA ACCGCCGTCC TCGGTCGTGC CATAGGCAAT CCCGATATGC TCCAAATCGG 8100

AAAAATCCAC ATCATCGCTG CCATATTCGG CGTAGGCAAA CTGTTCAATT AGCTGTTTTG 8160

CCGTATCTAA ATCGGAGGTT TCGGGGCGTT CCGGCTCCTG CATGGGTTCC GCTGTAAAAT 8220

GCTCCTGCTG GACTTCCTCT ACAAACTGCC GCACAAAGGG GATGCTTTCA CTGCGGAAAA 8280

TGGGATACCA GCCCCGCAGT TCCATATCTT GCAGGCTGAC GGTTCCGGCG TTATAGTCCA 8340

CGCTGTCAAT CTTCATCAGC CGCCCGTCAA TGGTCAGTTC CATGCCGACA GGGATATAAT 8400

CCGCCGCGCT GCGTTCCTTG ATGATTTCAT AGGGGGCATC GCCGCCGCGT TCCGTGTCAG 8460

GCCGCGCCAG CACCACGCTT TTTCCATGTT CCAGCAGCTT TTTTAAGACG GACTGCCAGC 8520

CATTGGCTCC GGTGATGGAC ACCATCCCCA GGTCGGGAAT ATCGCGGGTG ATGGTTTTCG 8580

TTCCCAGCGC AGGGGCGGCT TCCTCTGCGT CTGTGCCGTA AAACAGCATG GTTTTCCCTA 8640

CCCGGACGCC TACCAGTTTG TCAGGGTATT GGGCTTTCAG GTTCAGGTAT TCCTCCACGT 8700

TGGGCGTAAG GTTCAGACCG TCTGCGGCGG CTGTTTTTGG GGCTTCCTGG GTCTGCGGCG 8760

AACCTGTGCC GGTGGGTTCT GCGCCCGCCT TTTGCGCGGG GGATGGTACA ATATTGCTGT 8820

CCCCGGTTTC CGGCTCTGGA TTGACCGCAG CGGGCGCGTC CTGCACCTTT TCCGGCTCTG 8880

GCTCCGAGTT CCCaGGCTGC ACCCGTTCCG GCTCCTGGGG CGGGGTGGAA AATACTGGCG 8940

TGTCCTGGGT GGCAGAAATG GACATTTCAG TGGCAGCGGC AACCTTTTGA ATTTTTTCCC 9000

TGATTTCCTC CGGCATATCC GCGTCGTAAA AAGTCACGGT ACGGTCAGGG GCGATATGGG 9060

CAACTGTTTT GTAGTCGCCG TGTTCTTCCT CCAGACGGTT CCAGACGGTC AGGCCGTTCC 9120

CCATATGGCC ATAGCCTAAA TCATAGCGGT CAGCGGGGCT TTCGGCCATT ATGGGCGCAG 9180

CCCCTTCTGG TATGTTCGGC GTTTCCTGCG GTGGTTCCGG CTTTGATGCG ATAGGGACGG 9240

CCGTTTCAGA TGCGCCGGTT TCTGCCTGCG AAAGTTCCGG CTCCGGTTCA GGAGTTGCGG 9300

CTGGCTGTTC TGGTTCGGAC GAAAGCCCGA ACATATCTAA AGTAAGCTGC CCGTCTGCAT 9360

CCTCCCGTGG GGCAGGCTGT TTCTTGCGGG AAAGGGCTTT TGCCGCAGCA GCAAGGCCGG 9420

TTGCAGGTTC AGCCGTTTGT CTGGCGGAAT CCTGCCTTTC CTTCTGTTTC TGTTTCTGCT 9480

TTTTCTGCCG TTCCTCCTGC AAGACGGCAT CAGGCAGAGG GGTGAATAAG GAGTATTCGC 9540

CCCGCAGgTA TGCGCCCATT TTTTGAAGCG TCGGCTCCAT CAGGGAATAG GCGCGCTCAT 9600

CCGGGGATAC CATTGCCAGG GCTTTCACCA TTTTTTCATA GTGCGCCGCG CATTGTTCCG 9660

GGTCATCTAA CAGCGGGCGG AATACCTTTT TTGCCGCGTC AACATCCCAG TCGTGGGGGA 9720

TTGTCTTGCT GGCATCGTTG GGGTCGTAGT ATTGGAAGGT ATATAGACTC CGAGCAAGCT 9780

GCATTTTTTC ATATTCCGGC AGATATGCCT TTTCCcTGCG GTTCAGATAC CGATCCGTTT 9840

TAATCAGTCC GGCAATCCGC TTTTGTACCT GCTTCCAGTT CAGCGTAACC GTATCATAAT 9900

TGCCACCGGA AAAATCGTCG GCGCGGGACA GCTTGATGCC TTTGCCGTCG TGCCATTCAT 9960

CATAGCCAAT GTAGCCATGC CCGCCAATGC CGTATTCGTT TTTCAGAAAT TCGATGCACT 10020

CTTTGGCATC ATGCCCCTGC ATAAAGTAGG CGTAAATACG GAATTTCCCC TCGGAAACGC 10080

CAGAGCCGCC AGTGAACATC CGGGTAATCT CGTCCTCGGT GACAAAAGCT CCACGGGCAG 10140

GCGCAAAGCC CTCCACCCCA TGAAATGCCG TCACCGGCTT ATCCATATCG GTCAGGCGGG 10200

TAAACAGTTC GGAGGGCGCG GGCCGCTGGC GGAAGCGCAG CAGATTCGGG TTCTCCTGAT 10260

GGGAGATGGC AAACTGCCGG ACTTCCTGTG CAATTTTGGC TCGTTCTGCC GGGTCTTTCA 10320

GGAGCGCGGC AATTTTAGCC GTGCTGTCCG GGAAACCACC GTACAGCGCG TCAATGGTGG 10380

GCAGCAAGCC CTTTTCCTTT GCTTCCTCGC TGAAATTCTG CCGCAGAAAC CACAGTTTTT 10440

CCGACAATTC CCGAAATTCA TTGTCACGGG CCGCGTCGAT TTTGTCCTGG CTGGCAAACT 10500

GGCCGCTGTT CAAAAGCTGC CGGATCTGTG CCGCCGCCTG CTTCCAGGTG ACAAGGGAGC 10560

CGCCCCCGGC AGAGCGCCCC GGCGCGATGT GGATACCGCT ATGGCTGAAC CACAGCGAAT 10620

ACTCATGCCC GGCAATTTTA AGCCCCTTGC CGCCGCTGCC GTATTCCTTT TGCAGAAAAT 10680

CGGCGGCAGA CTGCTCCGTA GGGTCTTTCT GGAAAAAAGC GACGATACGC AGGATGCTGT 10740

CCGGCTCGTT GCCGCCGCTT GTGAGGGCGC GGCCAATGAC CGCATCCGTG AcCTTGCCGG 10800

AGGAAAcGGC GGrGCGCAGT TCCTGCTCGG TTTTCTGTGC CTGAGCAATA GTTTCAATCT 10860

GtGTTCCACA GTGGGGAATA AAGAAAACTG TGTAAATGCC GGCTTTTCAG AcGCAGAAAm 10920

GcGGACTGTT CGCCCGCCGC TTCCTGCTGG TTCTCCTGAT TTACTTGTAA ACGATCTCGT 10980

TCAGTACGGT TTCCTCCGCC TGGGCGAGAA ACAGGTTGAC TGCCTGCACC CATCCCATCT 11040

GGTCGCGGGC TTTCATCCCC TCGTmCACGC CGTTCTGTTT CTTTAATTCC GCCACGATCC 11100

GGTCTGCCTG TTCCTGCGCn TGCCGGTCGA TCTCCCGCAG GTGGGGTTCC AGCCGCCCGG 11160

TCAGCAGCAT CGAGGTGTAT GTCCCCTTGT GGTGTTCCTG AAGAAAGGTC TGCCGCAGCC 11220

TGCCGTACTT GCCCAGCGGC ATTtCCTCCA TCTCGTCCTC CCCGTC 11266

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7528 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

AATTTTAGTG TAGTAAAGAA GTACGTGAAT TTGATTTTGG ATATTGTATT AGGGATTTTA 60

GGAGTTTTCT TATTAGTATT TATGGCTAGT TATTTAGTTG ACCTCGCTAC ATATCTCATG 120

AAGCCAATGA CGCCGGATAA TTTTTCAGTT ATCATGCAAG AAATTATTTC ATTCTTTATG 180

TTATTTGAGT TTATTATGAT GATTCTTCGC TATATTCAAG AAGGACACCA TATTCCGATc 240

GTTATTgATT TTAATTgTTa ACCGCAATTT TACGTCAATT AATGGkTGCG CATGGCGATG 300

GCyTGCAGAC TTwATkGTtA wCTTTAwCAA TTCTTTTACT AGTGGTTGTG CTATTTATTC 360

TCGGGTTAAA TGGCAGCAAA TTTTAtACGC TAGGAAAAGC GAAAGTAGaA gAAGaTCACG 420

AGCaAGATTT TCayCATTAA TCgAACaTAA TCaAAGACCG TCCAACGTGA GAATTTTTCT 480

CTAACGTGGG GCGGTCTTTT TAATGCATAG TTAAATAGTT AACTCAATCA GTTCAATTGT 540

TGGTAAGACA CCAAAGCGTG CGCTGATTTT AGTGGTTCCG ATTCCTGGGT GAACAATTAG 600

CGGCGTTTCA TCTTCTAAAT GATAAAGCCC ATGTGTATAT TTTTTAGCAA GCTGATTTTT 660

GTAAGTGAAA AAAGGCAAGC GAATCTGTCC GCCGTGACTA TGTCCTGCTA AAATGAGGTC 720

AATGTTTTTG TTGACTGCTG AGTCTGCGAC ATCTGGCTCA TGGGTTAGTA GCAGAGTGGG 780

CGTTCCAGTA GGGCGGTTGA CTAAAGCTGC AGCTAAGTCG GTTTGTCCAT ATAACGAATC 840

TTCTAAACCA GCGACGTACA AGGAGATACG GTGTGAGACA GGCAGTAATT TGCCCGTGtT 900

TTTTaGTACG GTAAAGTCGG CAGCTTCTAA GATTGtTTCG TAAAAATTGG TATCTAAGCC 960

GCTATACTCG TGGTTGCCAT AAATAGCGAA TTTGCCATAA GGGGCGTTAA TTTTACGGAA 1020

AAGTTGACTC ACTTCATCTT GAATTTCGAT ACCGAAACGG GCATAATCAT CAAATAAGTC 1080

TCCTGTAAAA ACGACTATGT CTGGTTGTAA TGTATTTGTT TTTTCAACAA TCTGTTTTAA 1140

GGCCTTCACA GAATAATTCT TTTTTAAATG AATATCAGAG AGTTGAACAA TCCGTAAGTT 1200

TTCCCTGATT CTGTTACTTt AATTTGATTT TTGTGtATGA CTAAGCATTT AGGTTCAATA 1260

AAAAAACTAT AGCTTAAAAT GCCGATAATA ATGAAAAATA GTGCCAGAAA TAAATAAACG 1320

TTCATAATTT CCTCCTTGAA GTCGCTAAAG TAGTTTACTA CAAAAGTGTT AAGCAAAAAA 1380

TAAGGAAAGG TTAAATTTTC AGTAAACTTT CTCGATGAAG TTTTTTGGTG ATTCATGCTT 1440

TTCTGAAGTA AATTTGCTAT AATGAAAAAT GAGAAAAAAG AAAGAAGGTC GAACCAATGA 1500

TGGCAAGAGT CGAAAAGCTA CGGGAGTTAA TGAAAGAAAA TAATTTACAA GGATTTTTAG 1560

TGACAAGTCC GTATAATTTA CGTTATTTAA CCAATTTTAC TGGGACmACA GGCTTGGCAA 1620

TGATTACATT AGATAAAGCT TTTTTTGTAA CGGATTTTCG TTATaCGGAA CAAGCTGCGG 1680

```
AACAAGCAAa CTGGTTTTAC TATcGTTAAA AACACAGGAC ATaTTTTTGa CGAAGTAGCA   1740
GATTTAGCTG AACGTTTGCA ATTAGATAAT TTAGCTTTTG AAGAAACGCA AGTTAGTTTC   1800
GCTGATTATA GCTTATTGGA AGAAATTTTA CCATGTGAAT TAGTTCCTGT AATGGGTTTG   1860
ATTGAAGAAT TAAGAGAAGT GAAAGATGAA GAAGAAGTAG CGATTATTGA AAAAGCGTGT   1920
GCCATCGCGG ATCAAGGATT TGCCTTTGTC TTAGAGATGA TTAAACCTGG AATGACAGAG   1980
ATTGAAGTGG CTAATCAATT AGATTTCTTC ATGCGTTCAA AAGGAGCCAG TGGTGTTTCT   2040
TTTGAGACAA TTGTGGCCAG CGGTCTACGT TCAGCGATGC CACATGGCGT TGCTAGCCaT   2100
AAAGTCATCG AAAAAGGTGA GTTAATCACT TTAGATTTCG GTTGTTATTA TGAAGGCTAT   2160
GTTTCTGATA TGACACGGAC ATTCGCTATT GGTAGTATTC AGCCAAAACT AAAAGAAATT   2220
TATGACATTG TTTTAGAAGC ACAATTAAAA GTGTTAGCAG AAGCTAAACC TGGGTTGACG   2280
GGCATTCAAT TAGATGCGAT TGCACGGGAT CATATTGCTT CCTATGGCTA TGGCGATGCG   2340
TTTGGTCACA GCACTGGTCA TGGCATTGGG TTAGAAATTC ATGAAGGACC AAATGTTTCA   2400
TTCCGAGCAG ATAAACAGTT TGTACCAGGG AATGTGATTA CGGACGAACC TGGAATTTAT   2460
TTGCCAGGAA TTGGTGGCGT GCGTATTGAG GACGATTTAT TAATTACTGC TGAAGGTAAT   2520
CGTGTATTAA CACATGCGCC AAAAGAATTA ATCATTTTAT AAAGGTTGGG AGAACATGGA   2580
TATCCAAGAG TATCAGCGGT TTATTAGTGC GTTTTATAAA AAACGTAACT GGTATCAATA   2640
TGATCCGTTT ATTCGTTCCA ACTTTTTAAC AGAAGAAGTC GGCGAAGTTT GTCGGGCCAT   2700
TCGGACAATT GAAATTGGCC GTGATCGATC GGATGAAACA GTTTTGGCGG ATGAAGAGAA   2760
TCTGGAAAAT TTAACGGAAG AGTTAGGCGA TGTGTTGGAT AATTTATTTA TTTTAGCCGA   2820
TAAATATGAC ATCTCTTTAG AAATGATTAT GCAAAAGCAT CGGACGAAAC TAATGGAACG   2880
TTATCCTGAA GAATAAAGAT GTAAAAAATAA GAAAGCCAAA ATTACTTGGT CTTTCTTATT   2940
TTTACGTCTA AAGGCTTGAA TTAGAAATTT ATGACAAGAA ATTATCTGTT CTTTTAGTCA   3000
GAACTTAGCT TTTCTGGTAG CGAAATAGGC TAATTAATGA TAAACTAAGA ACGAGTATTG   3060
ATAAGCAAAA TTTGGACGTT CTTTCACAAC TAACTGAAGA AACCAGTGtT GTTTTAATTT   3120
TtAAGAATCA TCCGTTGCTT AAAAGAGAAT CCTTAAGAGA CTTAGTGATT CATCAAATAA   3180
ATCGTTTATT TAACGATTTT GTGCGTAATA GATAAGGAGG ACAACCAAAT GTCAGAAGAA   3240
AAAAATCTTG TAATCAATGC AAACGATTCT TTAGGAGAAA TCGTGATCGC TCCTGAAGTC   3300
ATTGAAGTAA TTATCGGAAT CGCTGCTTCA AAAGTAGAGG GCGTTTATGG GATGCGCGGC   3360
ACGTTTGCCA ATAATGTGAA TGAATTATTA GGCCGTGCAG CACACGGCAA AGGTGTATAT   3420
TTACGAACAG AAGAAGAAGG CTTAAAAGTC GATATCTATT GTTACTTGAA CTACGGTGTT   3480
TCTGTACCAA AAGTTGCGTT AGAAATGCAA GACCGTGTGA AACAACAAGT CTTATTTATG   3540
ACAGATATTG ATTTAGTTGA AGTGAATATT CATGTTGTAG CGGTAGTACC TGAAAAAATC   3600
CCTCAACCAG ACTTAAATGA ATTATTCCCA GAAGATGAAG ATGGAGAAGA GAATGAGTAA   3660
```

AACAGAATTA ACCCGTCACG AGATTCGTGA AAAAGCGCTC CAAGCCCTTT TTCCTTTAGA 3720

TTTTAATGCT GATTTAACCA AACAAGATGC CATTGATTAT GCATTGGCTT ATGATAATCG 3780

GGAAATCGTC AGTGAAGATG GCGAAGACCT AGTCCCTACG TATTTAGATT TACTTGTCGG 3840

CGGTGTGTGT AGCCGCAAAG CAGAACTAGA CGAAGTGATT ACGAACCATT TAGGTAATAA 3900

CTGGTCTATG CAACGCTTAG CAAAAATAGA TATTGTTATT TTACGTTTAG CTATTTTTGA 3960

AATGCTATAC GTTTCAGATG TGCCAAACAT TGTGGCATTA AATGAAGCCG TTGAGTTAAG 4020

TAAAAAATAC AGCGACGACC GTTCACGTAA ATTTGTTAAT GGTGTGTTAT CAAATGTAAT 4080

GAAAGAAATT GATTCAGAGG CGTAAGCTCT GAATCTTTTT TTACTATCTA ACTGGTGTTA 4140

AGATACTGGG TTATATGCTA AAATAAATAC TATTAAAGAA GTAAAGAAAC AATAGAGGGA 4200

GCGAAGGAAT GAGTACAGTA ATTAACGGTC GTGAATTAGC AGATCAAATG CAAGCAGAGA 4260

TTCAAAAAGA TGTAGAAAAG ATGACACAAC AAGGCATCCA ACCAGGATTA GTTGTTTTAT 4320

TAGTTGGGGA AAATCCTGCC AGCCAAACGT ATGTGAGAAA TAAAGAACGT GCAGCAGCCA 4380

AAATTGGCAT TCTGTCAAAG GTCGAAAAAC TGCCAGAAAC TATTTCAGAA GAAGAATTAT 4440

TGGCTGAGAT TGACAAATAT AATCAAGATT CACGCTTTCA TGGCATTCTT GTACAACTAC 4500

CTTTGCCAAA ACATATTGAT GAAGAAAAGA TTTTATTAGC GATTGATCCC AAAAAAGATG 4560

TAGATGGGTT CCATCCAATG AATTTAGGCC GTTTGTTTGT AGGCAAACCT GAAATGATTC 4620

CTTGCACGCC GTATGGAATT ATGAAGATGT TTGAAGCTTA TGATATTGAT TTAACAGGTA 4680

AACGTGCGGT GGTTATTGGT CGAAGCAATA TTGTTGGCAA GCCAATGGCA CAATTATTGT 4740

TAATGAAAAA TGCGACCGTG ACGATTGCCC ATTCTAAAAC TGAACATTTA GCAGAAGTTG 4800

CAAAAGAAGC GGATATTTTA GTTGTAGCAA TTGGACGCGG GCATTTTGTC ACCAAAGAGT 4860

TCGTGAAACC AGGTGCGGTA GTGATTGATG TGGGCATGAA TCGCAATCAA GAAGGAAAAC 4920

TGATTGGTGA TGTAGCATTC GATGAAGTTT CTGAAATTGC AAGCTATATC ACGCCGGTTC 4980

CCAAAGGTGT TGGCCCAATG ACGATTACCA TGTTAATGTA TCAAACAGTC GAGGCAGCAA 5040

AAAAACAAAA GTAAGGTGAA AAAATGACGC AACAGTATTT AACAGTAACA GCCTTAACGA 5100

aGTATTTAAA ACGTAAATTT GATGCAGACC CTTACTTAGG ACGTGTCTAT TTAACAGGTG 5160

AAATTTCGAA CTTTCGCTTC CGAGCGAACG CGCATCAGTA CTTCAGTTTA AAAGATGATC 5220

ATGCGAAGAT TTCTGCAATT ATGTTTAAAT CAGCCTTTCA AAAATTGAAG TTTCAACCAA 5280

AAGAAGGCAT GAAAGTCATG GTGGTGGGGC GTATTTCGCT TTATGAAAAC AGCGGCTCCT 5340

ATCAAATTTA TATTGAACAT ATGGAACCAG ATGGTGTAGG TGCACTTTAT CAAGCATTGG 5400

CAGAATTAAG AGAAAAACTA GGAAAAGAAG GACTTTTTGA AGGGCCTAAA CAACAGTTGC 5460

CTCGTTATCC CAAACGAATC GCTGTGGTTA CAAGTCCTAG TGGTGCTGTT ATTCGGGATA 5520

TCATTACGAC AGTCAAAAGA CGGTATCCGA TTGCACAACT GGTCCTTTTT CCAACCTTGG 5580

TCCAAGGCGA GCAGGCGGCG GATGACATTG TGCGGAACAT TCAGCGAGCC GATGCGCAAA 5640

GCGATTTTGA TACGATGATC ATCGGCCGTG GCGGTGGTTC AATCGAGGAT TTATGGCCAT 5700

TTAATGAAGA AAAAGTGGCT CGAGCCATTC ATGCTGCCAC CACACCGATT ATTTCATCGG 5760

TTGGTCACGA AACAGATGTG ACAATTGCCG ATATGGTAGC AGATGTTCGT GCAGCAACGC 5820

CAACGGCAGC AGCGGAACTA GCGGTACCAG TTTTAAATGA AGAACTATTA AGAATTAGTG 5880

AACGGCGTAG TCGTTTAGAA CAAAGCTTTT TATATCTCTT ACAACAACGA ACAGAACGTT 5940

TCCAACGACT ACAAAATTCT TATGTGTTTA AGCAACCTGA GCGACTTTAT GAAGGTCAAA 6000

CGATTAAATT AGATCGAATG ACGCAACGTC TTTTCCAAGC AATGACAACA ATTCATCACC 6060

AAAAACAACG TCAAGCCCAA GGAATTATCG CTCAATTGCA ACAACAGACA CCAAAGGGAC 6120

AATTAAGAGA AAGTCAGCAA CAATTGGCAT TTTTACAACG AAACTTGCAA ACACAAATGA 6180

CACAATTGTT CCTAAACAAA CAAAAGCAAT TCACGTCTGC TGTTCAACAA CTAGACTTAC 6240

TCAGCCCACT CAAAATTATG GGTCGAGGAT ATAGTTATAC TACAAAAGAG GACCGTGTCG 6300

TTAAAACTGT TACCGAGTTA CAGCCAGCAG ATCAATTAAC GATTCATTAT GCTGACGGTA 6360

CAGTTCAAGC AAACGTCGAA ACAATAACAG CGAAAAAAGA GGAGTTTTAA CATGCCAGCA 6420

AAAGAAAAAA CGTTTGAAGA ATCATTAAAT GCCTTAGAAG AAATCGTTCA GCGCTTAGAA 6480

CGAGGAGATG TTCCTTTGGA AGAAGCCTTA GCTGCCTTTC AAGAGGGCAT GGCATTAAGT 6540

AAACAATGTC AAGATACATT GGAAAAAGCG GAAAAAACTT TAACGAAAAT GATGACAGAA 6600

AACAATGAAG AAATTGTTTT TGAAGAGAGC GAGGAAGCGT AATGACGAAT TTTAGTCAAC 6660

AGCATTTACC GTTGGTTGAA AAAGTCATGG TCGATTTTAT CGCAGAATAT ACTGAAAATG 6720

AGCGTTTGAA AGAAGCAATG CTCTACTCAA TACACGCAGG TGGCAAACGA TTACGCCCGC 6780

TATTAGTGTT AACAACAGTG GCCGCTTTTC AAAAAGAGAT GGAAACACAA GACTATCaAG 6840

TGGCTGCCTC TTTAGAGATG ATTCATACGT ATTCATTAAT TCATGATGAT TTACCAGCAA 6900

TGGACGATGA TGATTTACGT CGTGGCAAAC CAACCAATCA TAAAGTGTTT GGTGAAGCGA 6960

CTGCCATTTT AGCAGGGGAC GGCTTATTAA CAGGTGCATT TCAGTTGCTT TCTTTGAGTC 7020

AATTAGGCTT AAGTGAAAAA GTTTTACTGA TGCAACAACT GGCAAAGGCA GCGGGGAATC 7080

AAGGCATGGT TTCCGGCCAA ATGGGTGATA TTGAAGGAGA AAAAGTCAGC TTGACTTTAG 7140

AAGAGTTAGC GGCTGTTCAT GAAAAGAAAA CCGGAGCGCT AATTGAATTC GCCTTAATTG 7200

CTGGGGGCGT ATTAGCCAAT CAAACAGAAG AAGTCATTGG TTTACTGACA CAGTTTGCGC 7260

ATCACTATGG CTTGGCTTTT CAAATTCGTG ATGACTTATT AGATGCCACA AGTACAGAAG 7320

CAGATTTAGG TAAAAAAGTA GGCCGAGATG AAGCGTTGAA TAAAAGTACG TATCCAGCTC 7380

TTCTAGGGAT TGCTGGCGCG AAAGATGCGC TAACACATCA ATTAGCAGAC AAGCTGTGAC 7440

CGTCTCCGGG AGCTGCATGT GTCAGAGGTT TTCACCGTCA TCACCGAAAC GCGCGAGACG 7500

AAAGGGCCTC GTGATACGCC TATTTTTA 7528

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6682 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

TTTTTTTGCA ACGTTTCCAT GAAAGAATGC GGGAAGATAT CCCTAAAAAT TGATCTTCTT 60

ATTTCTCTAT TCAAAATGAT TCTAGCCTAA CCTTTTTTAC TATTTCATGA TATGATGAAG 120

CAAAGAATAT TTTAGGAGGT TTTTACATGG CTGCAAAAGA AGCAAGTTTT GATGTAnTTT 180

CAGAGGTCAA TATGGAGGAA GTCAAAAACG CAATTCAAAT TGCCTTGAAA GAATTGAAAA 240

ATCGTTTTGA CTTTAAAGGG TCAATTGCTG ATATTAAATT GGAAAACGAT AAATTAGTAG 300

TCGTAGCTGA AGATGATTAT AAAGTCGAGC AAGTGAAGGA TATTTTATTT GGaAAATTAG 360

TCAAACGCAA TGTTCCGATT AAAAATATTC ATTTTTCAGA AAGTGAAAAA GCCTTAGGCG 420

GTACTGCTCG CCAATATGGT GATTTAATCA GTGGTATTGA TAAAGAAAAT GCAAAAAAAA 480

TCAACACAGC TATCAAAAAC TCTGGGATTA AAGTTAAGTC ACAAATCCAA GAAGACAAAA 540

TTCGGGTAAC TGGAAAAAGT CGGGATGATT TACAAAAAGT CATGGCTCTT TTACGAGAAT 600

TAGATTTACC AATGGcTCTT GAGTTTAATA ATTATCGCTA AAAAAATCCT TCAACTGTTA 660

ATGAACAGTt GAAGGaTTTT TTTGTTTtGG TACCGGCCAT TCTTGATTAT TCAATAAAaC 720

GATTGGTCGG CGTTGGATTT TTCCTGAAAA AGCGGCAATT GCTTTGAAAA CCTTTGgCAT 780

AAAATAAAAC GCATTtAATT CTTCTAAATC CAGTTCTTCT TCTGCCGTAA ACGAGACACA 840

AACCTCTTGT TCTGAAAGCA TAATTTCAAT TTTCTTTAAA AATTGAAATT GCCGCATTCC 900

TGGCACGAAA TAACCAGCCT TTGCAACACG AGAGGTTGTA ATATCTAATT TATCACTAAT 960

GATTAGCGCT AGTGTCATTA ACTCTTCCGA ATCAAAGCCG TTACTATGAT TTTCTATAGA 1020

AGATAAAATT TCTTCTTGAT AAGGCAAACA AATCTTTTGc TTACGAAAAt ATGCTTCTGC 1080

AAATtGTTTC CCACGAAGTG CATGTCCTGC TTTTCCTTGT AGCGCACCCA CATCATGGAG 1140

CAGTGCCGCA ATTTTTGCCG CCTCTAAATA CTCCTCTGAA ACAGCAAGTT GTTTCAAAAT 1200

CATTTCCGTC ATGGCTACGA CATTGGTTAC ATGAAACCAA TTATGGTATG CCCACCCAGC 1260

TTCTTCATCT TCCATTTGAC CAATTGCATC ATAAATTTGT CGAATTTCAG TATCTTGATG 1320

AATGCGTTCA AACAGCTTCA TTTATTTTCT CCTTAAAATT AGTTACTGTT CCAAAAATTC 1380

TTTTAACTCT GCAGGATTTT TTTCGCTGTT TAAACCAACA ATTAAACGAA TCCGCGCTTT 1440

CGGTCCGTTT AAACCACGGG CAAAAACGAC CCCCATTTTT TTCAAGCCAA CGCCACCACC 1500

AGCATAATCA TAGATATCTT CAGCAATGCC ATTGGAACAA CGAGAAACTA AAACGACTGG 1560

AATCCCGTTG TCTAGCATTT TTTGTAATGC TGGGAGCGTT TCTGGCGGTA AATTACCCGC 1620

GCCTAGCGCT TCAATTACTA ACCCATCCGT TTCTGGGGTG TTTAATAATT CAAACATCCG 1680

TTCACCCATC CCCGCATAGG CCtTCcTACG TGAACATTTC CTTTCACAGA aGAAACATCG 1740

CAAACTTCTT GCGGTAAGAC CTCTTTCGCA AAAAAAGCAC GTTCTTTCGC AATGGTTCCT 1800

ATTGGTCCAA AAGTTGGCGT CCGAAACGTC GCTACGTTCG TTGTATGTGT TTTTGTCACA 1860

TAACGCGCCG TATGAATTTC ATCATTCATC ACGACTAAAA CACCTTTATC GTACGATTCG 1920

TCTGAACAAG CAGTCCAAAT AGCACTGATG AAATTATACA AGCCATCACT ACCGATTTCG 1980

TTACTTGACC GCATTGCTCC TGTTAAAACA ATGGGGATTT TTTTCTCTAA GGTTATATCT 2040

AAAAAATAAG CAGTTTCTTC TAATGTATCT GTTCCATGCG TAATAACCAC GCCATCAATT 2100

GGTTCAGAAT AAGCTTTTTG AATCCGTTCT TTTAATTGAA ACATCCGTTC TAAGGTCATG 2160

TGTGGTGAGG GAATATTAAA AATATCTTCA ACAACTAAAT GAACCTTGCC aGAAAATAAT 2220

GCTTCTTGTT CCATCAACGG ATTTTTTTCA TTCGGGGCAA CACTTCCTCC TGCTACTTTC 2280

GACATGGAAA TGGTCCCACC TGTATGAAGG ACTAAAATGG TCTTCATTTG TTTCCTCCTT 2340

TGTTTTCTTT ATTTTTTAAC CAAAATTGTT TCACTTGTTG AATGAGTTGT TCATTGGTGT 2400

TTGCCTTTCG ATGCGCCAAC CAATGCGGCG GAAAGAAACG CTGGACACTT GGTAATTCAA 2460

ATCGTTCGTC TAATAACAAC ACAAAGCCTT TATCTTGCGC ATCTCGAATG ACACGTCCAG 2520

CCGCTTGCAA GACTTTATTC ATGCCaGGAA TTTGATACGC AAATTGAAAC CCTTGCCCTC 2580

GCTCTTCTTG ATAAAAAGTG CGAATCAGTT CTTGCTCTGG GTTCATTTGT GGTAGCCCGA 2640

CACTAACAAT TATCGTACCA ATCAATCGGG TACCTTTCAA GTCAATTCCT TCTGAAAATA 2700

TACCACCTAA AACACAAAAA CCGATGAAAG TTTCTTGAGG ATTCTCTTTA AATTCAGCTA 2760

AAAATGCTTC TCGCTCTTTT TCGTTTAAGT CCGTTCCCTG AATTTGAGTT TTTACTTGCG 2820

GGTAACGTTG GCTAAATACT TGATAGACAT CATCCATATA GGCGTAGGAA GGAAAGAAAA 2880

TTAAATAATT ACCTGTTTTT GTTTCACTTA ACTTCCCTAA AAGTTCACTA ATCTGTGGAT 2940

AACTTTTTTC ACGGTTTTTA TAGGTTGTTT GAAGATGACT ACCAATCATC AACAATTGAT 3000

TTTCTTCCGG AAAAGGATAC GGAATCCGAT AACGTAGGCT TTCTTCACCT CCACCAAGCA 3060

CTTCTTGATA ATAATTCAGA GGTGTTAAAC TTGCCGAAAA TAAGATACTC GCTTTCCCTT 3120

TATCTAATCG TTGTGACAAT AAATAAGCCG GATCGACACA AAATTGTTTC ACAACGCAGT 3180

CATAGTTTCG CAAAGAAATA GACGTATAAA AGCCGTTGTC GTAATACTCG CTGATTCGAA 3240

GATAACTTAA ACAATCAAAA TAAAAACTGA GAATCGACGC CAATTCCTCA ATGTGTTGAT 3300

TTTCTGGCAA CCATTCTGCA ATTACTTCTG TTAATCGATG TATCTGTTTA ACTAATGAAT 3360

CAATTGGCGC GCGTTGGGCA AAAAAATCCG TGTCTTTTTC TTCACAGATT GCTTCAATAT 3420

CATCAAACGT AGTTAAGACT TTTTGACAAG CTCTTTTCAA TTGAACAGAT TCTTTTTGCA 3480

ATCTTTTTTG AATTgCAAAA CCGCATTCCT AGAAATAGCG GCCGAATACA TCTCCCGTGA 3540

ACGATTAACC AAATTATGGG TTTCATCCAC TAAAAAAATA TTCTCTTCTG cACTTTGGGG 3600

TTCTTCAAAG AAGCGCCGTA AATAAACCGT TGGATCAAAT AGATAATTGT AATCGCCAAT 3660

AATTAAATCA CACCAAAGAC TTGCATCTAG CGACAACTCG AATGGACACA ATTGGTGCTT 3720

TTGTGCATAC CGTTCAATAA CCTCTCTCGT TAAATGATTT TCATGTTGCA ATAAATCCCA 3780

CAGTCCCTCA TTTAAACGAT CATAATAGCC ATCAGCAAAC GGGCAATGTT CTGGTGTACA 3840

ATTTCGTTCA GTTAAAAAAC AAATTTTATC TTTTGCTGTT AACGTAATAC TCTTAGTTTG 3900

CGCTTGTTTT TCGGACAGCG CTTTGACGGC ATCTTCAGCC ACTTGCCGCG TGATGGTCTT 3960

AGCTGTTAAA TAAAAAATCC GTTCACCTTC TTCTTCACCA ATTGCTTTCA TTGCTGGAAA 4020

AAGAGTGGAC ATTGTCTTGC CTGTTCCAGT TGGTGCTTCC ACGAAGAGCC GCTGTTTACT 4080

AAGAATTGTT TTATAGACGG CTACCGCCAA TTCTCGTTGC CCTTTGCGAT ACTCCCCGTA 4140

AGGAAAACTA AGATTTTTTA ATGATTTATT TCTCAAAATA CGCCACTCTT CTTTAAAAAT 4200

TACCCACTGT TCATAGCGCT TCATCAAATC AGTAAAAAAC TCCGCTAATG CAGCTTCTGT 4260

AAAATGGCGT TCTTGTCGTG TGATTTGTTT TGTGGTGGTC TGATAATAAG TGAGTTGAAG 4320

CGTTATTTCA GAAAGTTGTT CTTGTTGGCA ATACATGTGA CCATAACACA TCGCTTGATA 4380

CCAAAACAAT TCGACCTTTT CTGGCGGCAA TTCTTCAAAC GCTGGTTCGG ATGTTTTAAT 4440

TTCATCAATG ACCACTTGCC CAGTTTCTTT CACAAAAATG CCGTCTGCAC GACCTTCTAC 4500

ACAATAATTA TGGCCATTTA AAGGCACCTC AATTTTTAAA GATACTTCCT TTTGATAAAG 4560

GCTCCCCGCT TCCTTTTGCA ACATTCGATG AAGGCGACTA CCTTCTTGTG GTGTATGTTC 4620

ACTCACATGG CGAGAATCGA TACTTCCCCG CCTTAAAATA AATTCGACCA CTTGACGCAC 4680

CGCAATTTTT CTCGTTGATT TCACGTAATC CCACCTATCT GCCTCCCTAT TATAGAACAT 4740

ACATTCGCTA AACAAAAGCA AAAAATCCAA AACGAACACA ATCATGCGTT CATTCTGGAT 4800

TTCTCTTTAG ATTATCAACA ATTAAAAACA ACTTGTTTCT TCTAACTGTT TGGTTTCTTT 4860

TCCATTTGCT TTTGCCACGA CCAATTGCGC CATTTGTTTA AACGAATTAT TTAAACCAAT 4920

GCCGCCCTTT TTGGGCATCT TCACAATCGG TACTTGCTGA TAATAGTGAC TAACTGCTGG 4980

ATTTTCTTCT GCCATCGCTG CAATTGCAGG AATGCCTGCA GCATTAAACT TACAAGCTAG 5040

ACGAGCAGCC ATTTCTCCAA AATTAGGATA ATGCATGGCT GGTCCGCAAA GAACCGCATC 5100

GGCATGAAAC TTTTTTGCAA AGCCAATAAA TTTCTTTGCT ACCTCTTCCT CGTTTTCCAA 5160

ATAATATTGG TCACCACAAT AAAGTGTCGC AACAATTTCC GCACCTTGCT TCTGTAAAAA 5220

TGGTAGCAAT GTTACCCCTG GACCTAAGGG AGCTTTTTTA CCACTTGGTT TCAGCTGAGC 5280

TTGTTCATCA CTGCCAAATC CTGCTAAACT GTGGTTTAAG ACTAAAAGGA TTTTTTTCAT 5340

GCCTATTTCC TCCTTATTTA TAAATCATCG AAGTCAAAGG AATCAAAATC AATTTCTTCA 5400

CTCTCTtCTT CTTGTGCTTT CCATTCTTCA CGTAAATATT GGTCATCCAT AACTCGAATA 5460

AAAGGATAAT AAATGAACAT TCCTAGTACC AATAAGAGAG CTTGCAGAAT TGATCCTTGC 5520

CAACCGCTGA CTAAAAATCC TGAGATAATT GGCGGGGTCG TCCACGGCAA TTGAATACCA 5580

TTCGTCAATG GAACTAAGCC CCAATCCATA GCAAAATAGG CAATAATAAT ATTAATTGCG 5640

GGTACTAAGA CAAATGGAAT TAAAAGGATT GGATTTAAAA CAATTGGTAA CCCAAAAATG 5700

ATTGGTTCGT TAATCCCAAA AACACCAGGA ATTAGTGAAA GTTGCGAAAG CTTTTTAATT 5760

CGTTGaCTTT TACAGAATAA GACCATCACA ATGATTAATG AGAGTGTGGA aCCGGCGCCA 5820

CCAAATGTTG cAAcATATCT TGAAATTGTC CAGTGATGAT GTTGGGAATT TCATGACCGT 5880

CTTTAAATGC TTGTAAATTT TCAACAGACA ATGCCCGCAA AACAGGATTA AAGACTGCGC 5940

CAACCACACT ACCTCCGTTA ATGCCAAaGA ACCAAAAGAA GTGTAAAAAT AAATAAGCAA 6000

TGCTCATTGC ACCTAAGGTA TTACCAAGAC TCAATAAGGG CGTTTGTAGA AATTCAAAAA 6060

TAACATCTTG TAAATTACCA TAAGGAGTTA AAGCAGCCAG CCAATTGACT AAAAACATGA 6120

GAGTAACAAC TAAAAAACTC GGAATCAGGG CCTCAAAAGA CTTAACCACT GTTGGTGGCA 6180

CTCCTTCTGG CATTTTTATT GTCCACCCTT TACGAACAAT GGCCACAAAT AAACGAGTCG 6240

CAATTAATGC GACAATCATT CCTAAAAACA AGCCACTGGA ACCCATCCAT TTTAAAGGTA 6300

AACTAGTAAC TTCATAAACC TTTGAACTAT TTTCCGGTGT AAATAATGTT TGATAAGGCG 6360

TTACAATAAA AAATGAAACC AATGCGATAA TTCCTGCTTG CATTGCATCC ACATTGAATT 6420

GTTTTCCTAA ACTATAGGCw ATACCGACAA CGGCCAGAAT AGTCATTACA TCAAaACTAG 6480

CACTAGCTGG AACAGACATC ATTGTCGCCC AGTCTTTACC AAGCAAATGA GACATCCAAG 6540

TATCCCATTG TGGCAAAGGA AAATTAGCAA TTAATAAGAA GACCGAACCA ACAATCAACA 6600

AGGGCGTTGA TAAAAGGAAG CCATCACGAA TCGCTACTAA ATACTTATTk TCTCCTATCA 6660

CTTTTGCCAA AGGGGTCAAT AC 6682

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18724 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

AACTCCTAGA ACCAAGTAGG TACTGTTTAA CATCTATTTG CTTTGCAAAT AGTGTTTCTC 60

AGCAATAAGA AGGGATTTCT GAAAATTGTT CTATAATTTT TGGAGACCAA GTAGGTATTG 120

TTCATCATCG AATCACTACG TTCTTCGTGA TTTTCAACAA TTTCGGCTTA TTTCCGAAGA 180

CCTTCAACTC TTAGAGGTTT AGCTCTTAGA AGTTGTTGAG AACGAAGCAA CGCGTAGTGT 240

TGCTTCTGTT CCCGCCGTTT ATCAGTTTTT GAGCGTGAAG TAAAAATCCA AAGTGATTTT 300

TATCCCACGC CCATTTTCAT TTATGACGTT GGCTTCGTGT AAATTGAGAA GTCTCGTCCT 360

CTCTAGTGGT TGATTGCTGT GTTCGTTGTA CGGCTTGCTT TTCTGGCTCG CTCTTTTTTG 420

CAACATCAGT CACAACGTTT GGCTTCTGAG AAGGGGGAAC AGTTGTTTTT TTGTTTAAAG 480

TTGGCGTAGT TCTTTTTGTT TCTTTCTTTT GTTTCTTTTT ATTATTTTTA TCATTTCCAG 540

TGACTGTCAA ATAGATAATC AGAAAAATCA TTAGGATCGT TGCAACAATT AATAGAATAA 600

```
ACATGTTAAA TGAAATGAAC ATGCATGGCC CTCGTTTCTC GTTTTTTCTt AGTTTACCAA    660
AATTtCAAAA TTTCGTCACT CTTTTCTtCA GAAAAAACGA ATTTCAAAGG GGTTAAAAGA    720
AACTTTTATA AAAAAGCCTC CTTTTTTGTG TGAAAATGCA CAAAATTTCT GGaATCGTTT    780
GCTGTTAAAA CCAGATACTG GTAAAATAAT TATTGTATAA GaATCaTTTT AAAAAACGAT    840
TTTTGTTACA AATAACGTAT TTGTTGAAAA GGcACTGTGA AGTAGTTAGA GACGACAGCC    900
aAGTGACTTT TAATGACAGC CGCCCAATTA ATTTCATGGA GCCTTTTCGT AGAAAACGAG    960
GAGGTTATTT TGCGTGAAAG TGAGTATATT TTCGACATGC GTGGTGGATT TACTTTTTCC   1020
AAATGTAGGG CAAGCAATGG TAGAGGTTTT GGAACGTTAT GGCTGTGAAA CGGTTTTACC   1080
AACAAGCCAA ACGTGTTGTG GTCAGCCAAC GTATAATAGT GGCTATGTGA CAGAGAGCAC   1140
GACAACGCTT AAGAATCAAA TTGATAGTTT TGACGGTGCC GATTATGTGG TGGGACCTGC   1200
GGGGTCTTGT GTCGGCATGA TGAAAGAATA CCACAAATTT TTAGCAGATG ATCCTATTTA   1260
TGGACCAAAA GCACAACGTT TAGCTGAGCG AACTTACGAG TTTAGTCAGT TTCTTTATCG   1320
TGTCTTAGGG GTCAAAGATG TTGGCGCAAC GTTACATGGG AAAGCCACTT ATCATCGTTC   1380
TTGTCATATG ACGCGAATTT TAAATGAACG AGAATCACCC TTTGTTTTGT TGGATCATGT   1440
GAAAGGATTA GAAATGATTC CTTTAGGACA TCTTGAAAAT TGTTGCGGAT TTGGCGGAAC   1500
GTTTTCAGTG AAAATGCCAG CTATTTCTGA ACAAATGGTG ACGGAAAAAA TGAACGATGT   1560
GATTGATACA GGCGCAGAAA TTTTAATTAG TGCGGATATG GGTTGCCTGA TGAATATCGG   1620
GGGTAAATTC AATCGTGATG GGAAAAAGAT TAAGATTATG CATATTGCAG AAGTCCTTAA   1680
TCACGAAGTG GATGAAGCAC GCATGGACCA ACCACAAATC ATATCGGTAG GGTAAGGGGG   1740
AAGAACTGTG GGATTATCAA CGAGCAATAA GCCGTTAAGT GAACGAATTG AAGAAAGTAA   1800
AAAAGATGTG TTTATGCAAA AGGCTGTCGC TAAAGCACAA GATGCCCAAT GGGAGAAACG   1860
AGAAGGTGCG CGGGAAGCTT TAGGCAATTG GCCACAATGG CGTGAACTGG GAGAACAAAT   1920
TCGCCAACAT ACCATTCAGT ATCTGCCTGA CTATTTAGAA GAGTTCAGTG ATAACGTTGC   1980
CAAACGAGGT GGCAAAGTCT TTTTTGCTCA AACCGCCGAA GAGGCCAATG AATATGTGAA   2040
ACAAGTGGTT CTAGAAAAAA AGGCGAAAAA AATTGTGAAA TCTAAATCAA TGGTGACGAC   2100
CGAAGTCGAT ATTGATCCAA TGTTATTGGG ATTAGACGAT GTCAGTGTGA TGGAAACAGA   2160
CTTGGCGGAA TTCATTTTAC AAATGGATGA CTGGGATGAG CCGTCACATA TTGTTTTCCC   2220
AAGTATTCAT AAAAATCGCG AACAGATTCG TCAGGTTTTT GCGAAAAAAT TAGGGTATCA   2280
AGGAGATAAT GACCCGGTAA ATTTGGCACG CTGTGCACGA GAAGTGATGC GCAAATTCTT   2340
TTTGGAAGCG GAGATTGGCA TTACAGGTTG TAATTTCGCT ATTGCGGATA GCGGCTTAAT   2400
TAACTTAAAT ACAAATGAAG GAAATGCTGA TTTGACCATC AGTATTCCTA AGACACAAAT   2460
CGTGTTAATG GGGATGGAAA GAATCGTACC AACAATGCGT GAAGCAGAGG TTTTAGATAA   2520
CCTTTTAGCC AGAAGTGCCG TTGGTCAAAA TTTAACAACG TATGTGACGT TTGCCGGGCA   2580
```

AAAAAATGCG GATGAATCCG ATGGTCCAGA AGAATTTCAC GTGGTCATTT TGGATAACGG 2640

ACGTTCAAAA GCGTTAGGAA CAGCCTTTCA GCCAGTCTTA CAGTGTATTC GTTGTGGGTC 2700

TTGTCTGAAT GTTTGTCCGG TCTATCGCCA TATTGGAGGG CACGGTTATG GCTCTATTTA 2760

TCCAGGACCA ATTGGTGCTG TGCTTTCTCC AATTTTAGGT GGATATAAAC AGTTTGGGGA 2820

ATTGCCTTAT GCATCTAGTC TTTGTGGGGC ATGTACGGAA ACGTGTCCTG TGAAAATTCC 2880

ATTGCACGAA TTATTAATTG AACATCGTAA AGTGATGACC GATGATCTGA AAATGAAACA 2940

TGGATTTGAA GATTTCCAAA TGCGTATGGT AGGTAAAGCC ACTGGTTCAC CAGCAATGTT 3000

TAAAGCAGCC ATGAAAGTCG ATCATGCTGC AGCGGGCATT TTAAGTAAAC AAAAAGACAT 3060

TACTGTTGAA AACATGTACA ATCaTGGAGG CTACCTTGAT AAAGGACCAG GGTTAGTAAA 3120

AGGTTGGACC GATGTTCGTG ATTTGCCAAG ACCACCAAAA TCTTCAGAAA ACTTCAGAAG 3180

TTGGTTCAAA AAACATCAAG AAGGTGAAAA AAATGACTAA CGAAGCGATT CAAAATCGAG 3240

AACCTTTTTT ACAAAATCTA CGAGAAAAAT TAGACGTAGA AAAACAGCCT GTTTCGGCCC 3300

ATCCATTTGA GCCAGTTAAT CATTTACCAG AAGAACAGTT GGCAGATAAA ACACCAGCAG 3360

AGCTATTAAC GATTGTGAAA GAACGTGTGG AAACGATTCA CACGAACTTA GTGGAAACCA 3420

CACAAGAAAA TTTACTGACG ACGATTCAGC AAATTGTTGC TGAGTTTGGT GGGGACAATC 3480

TTTTATTGCC AACAGATGCA CGTTTTGAAG CCTATGGATT AGCTGATTTA GCTAAGTCAT 3540

TAGAGGCTGT TTCTGTTAAG CAATGGCAAC CTGGAAGTGA GCAACGAGAA GCGAATATTC 3600

AAACAGCTGC TCaGGCaAAT ATTGCGATTG CCTTTGCTGa GTTTTTAcTA GCTGAATCaG 3660

GAACGATTGT GGTGGAGTCC aATGCGGGAC aAGGCAGAGC CtTACATTTT TTACCGAAGC 3720

ATTACATTTC AATTATACCG TTTAGCAAAC TCGTTCCACG TTCGACTCAG CCAGCAGCTT 3780

TTTATACAGA AAAAATAGAA AAAGGCGAAA AAATCGGCTC AGCGATTCAC TTTATTTCAG 3840

GGCCTTCTAA TTCAGGAGAT ATTGAAATGC AACTGGTTGT CGGGTTGCAT GGTCCTTTAG 3900

AAGTTTGTTA CGTGGTTGTG ATGGATCGCT AAAACGCTAG AAGCATAATC AATTCAGCCC 3960

TTTTTAGAAT TGATTATGCT TTTTTTATGA GCAGGATCAA AAAAGTAAGA ATCAAATTTG 4020

AAAGAAATAC GTTTTTTTTC TGACCATTTT GTAAAAAAAG CGTACATAAA AATTAATTGG 4080

GTGTTAAGCA CTATAGGATT GATTTACTTT GTGGTAGAAT ATGAGACATA AGATAATAAG 4140

GTTTTTGGAG GAATGAGTTT GAAATCAAAA GAATTAATTA AAACAGTCGT CTTTTTTGCC 4200

TGTTTAGCTT TGGGTCTGTT TTTACTGAGA CAATTTGTAT TTACGCCTGT CGTAGTGAGA 4260

GGTCATTCAA TGGATCCAAC GTTAGCAGAT GGTGAACGGG TAATTACGTT AAAAAACACA 4320

GAAATTAATC GTTTCGATAT TATTACTTTC CCAGCGCCAG ATGAACCAGA TAAAAATTAT 4380

ATTAAACGTG TGATTGGTTT ACCTGGAGAT ACAATTGCGT ACAAGGATGA TACGTTGTAC 4440

ATCAATGGAA AAGAAGTTGA CGAACCCTAT TTAGATGAAT TTAAAAAAGC CTTAACAGAT 4500

GGTCAACCTT TGACAGGCGA TTTTTCATTA AAAGAAAAAG TACCAGCAGA TAGCTACTTT 4560

GTTTTAGGTG ATAATCGACG GAATTCAAAA GACGGTCGTG TCATTGGTTT TATTCATAAA 4620

AAAGATATTT TGGGTGAAGT GAAATTTGTG ATGTGGCCAT TCTCACGGTT TGGTCCAATA 4680

CCAGAAGTGT CAAAACAATA AAAGAGACAA TGAGTAAAAG AGAAGGACGG AAATGTGTAA 4740

AAACAAACAC ATGTTTCCGT CCTTCTTTTG TGGTAAAATA AAGGCAGTGC AAAGGAAAAG 4800

GATGTGAAAC CATGAGTGTG CAGTTTATTA GAGGAACCGC CGTGGCAGAC TTAGAAGCGC 4860

CACTTATACA GGCGACAAAA CAATGGCTAG AGGAAGACGC ACAGCATGAA GTATTCTATT 4920

TAGTCCCGAA CCATATTAAG TTTGAGCAAG AAATTCAAGT CTTGCAAAAG TTACGTCAAT 4980

TACAGACAAC TACGTCTGAT TCAATAACCA GTACCCGTCT GCAGGTTTTT AGTTTTTATC 5040

GTTTGGCTTG GTACTATTTA CAACATACAC CTTTTTACTC GGCAGATGTT TTATCTGATG 5100

CAGGTGCTGC GATGATTTTT CGCAAAATTT TAGTAGAAGC AGAAGAAGAG TTACAAATTT 5160

TTAGAGGTGA AATTAACAAG CCTGGCTTTA TTCaGCAACt ATTTCAGCTT TATCaAGAAA 5220

TGCGCGAAGG CAATATTGAG ATTGCGGAGC TTTATCCATT TTTAGAAAAA CAAACGGAGA 5280

ACCCTAAAGG ACAAGATTTA CAACTTAAGT TTCAAGATTT AACGTTAATT TTTACGCGAT 5340

TCCAATTGCA AATGAGTCAG TATGGCTATG AATCAGCGGA GATTATTCAA CATTTAAGCG 5400

AGTATTTACA AACGGTTGAT TTGTCGAATG TTCAATTCGT TATTTCAGGC TATCAGCAGT 5460

TTACTGCCAG AGAATTAAAG TTGATTGAAG TGCTAATGGC GCAAgcAGGT TCTGTGAAAG 5520

TAGCCCTTTT ATTGGACAAA CAATATCCGC ATGACTTACC TGATCCGCGG TCGTTATTTT 5580

ATGAAGCGGG ACAAACCTAT CATCAGCTTT ATCAACTAGC AAGACAGAAG CAAATCCCGA 5640

TTCTTTCAGA TTATGTGGAA AAGAAAGAGG TATTAATTAC AAATCCAGAT TTGCAAGGAT 5700

TAAACGATTA TTGGATTCAA TCGCAAGAAC ACCTGCCTCC ATTAAGCACA GCTGACTGGA 5760

GGGGCGATGG CTTATTTCTG TGGCGTGCTG AAAACGTTAA AGAAGAATTA ACACATGTTG 5820

CAACCGAAAT TCGCCGTTTA GTGGTGGAAG AAGGGTATCG GTACAAAGAA ATTCAGGTGT 5880

TGACTCGTGA TTTAGATTGT TATGAAAATT TATTAGAACC GATTTTTGCG GAACACGAGA 5940

TtCCTGTTTA TGTCGACCGA GATATGGCTA TGGATCGGCA TCCTTTAGTC GAATGGATTG 6000

AATCTTTATT TGCTATTCAT TCGTACAATT ATCGTTATCG CGATGTGCTC CGTTTTTTGA 6060

GAACGGAATT ATTTATGCCA ATGAATCAAC TCGCTACTTC TGAAGAAAGC TTGACTGATT 6120

GGTTAAACCA ACGGAACGCT TGGCGCAGAA AAGTTGATAT TACCGAAAAT GTAGTCCTGG 6180

CATACGGCTA TGAAGGTTAT TATTGGTCAC AAGAAAAAGA TTGGGAATTT ATTCGGTATG 6240

ATTTTGAAGC GGAAGAGCAA GAAGATGTCG CTACGATGGA AGAAGAGTCA AATGCGATTC 6300

GTCAATCGCT TCAACGGCTG TTACCTAGTT ACTTTCAAGC AATGATAAGT GCCAAAACTG 6360

GTTTGGAAGC AGCGACCGTC TTCTATCACT TTTTACTGCA AAGTGGTGTG GCAACACAAT 6420

TAAAAATGTG GCGCTTACAA GCTATCGAAG CAGGTCAACT AGAAACAGCT CGTAACCATG 6480

AACAGACTTG GGATGCATTG ATGTCATTAC TAGATGAATA CGTCACTGTT TACGGAGAAA 6540

GTTCGTTTGA TTTTACAACG TTTCAAGAAA TATTTGTTAG TGGTTTGGAA GGTTTGCACT 6600

ATAGTAAAGT GCCGACAGCG ATTGACCAAG TACAAGTTCG CGCAATGGAT TTAACTCGAC 6660

CTGGCGCTGC CAAGGTTACT TTTGCTATTG GGATGACGGA AGAAATTTTT CCACAAAAGA 6720

TTGAAAATAA AACACTATTG TCTGATGAAG AACGGCAGAC AATCAATGAT ACTTTAACGG 6780

AAAACCAATA TTTACGTGGG ACAACCGGAC GCAAAATTGC GCAAGAACCT TATGTTGCGT 6840

ATCTTGTCTT TTCTTCAGCG CGAGAACGGT TATATTTAAC CTATCCAAGT GTTAAAGATA 6900

CCGCTCAAGA AGTCAAACCG TCTCCTTATT TTAAAAATAT CCAAAAAGAT TTAAACTTGC 6960

CCGTTTTTGA AAAAAATGAA ACAACCATTT TCGATGATGA AACGACCAGT TTAGCCCATA 7020

TTAGTACGTA TCGGACGTTA ATCGGTGAAT TAACGCGTTT GAAAAGACAA CGGAAAGAAA 7080

CACAAGAAGG ACTGTTGCCT TTCTGGTTAA ATATGGAAAA AGCGCTAATG AACCAATCAA 7140

TTGCGCCGCT AGCCAAGCAA GTTTTCGAAA GTTTAACGCA TCAAAATATT CCTGAAAAGA 7200

TTGATGAGGT CTTAGCAGAA CCTTTATATG GTAAAGATAT TTATACGTCC GTTTCTCGGA 7260

TGGAGAGTTT TTATCGCTGT CAGTATCAAT ACTTTTCTCG CTATGGCTTA CGTTTGAAAG 7320

AAAGAGATGT CTTTGGGCTT TCGCCAGCAG CTACAGGGGA ATTTTTTCAT GAAGCATTGG 7380

ATCAGTTCTT TAAATTATTG ATTATGAATC AACGGAACTT ATCTGAATTA ACCGATCAGG 7440

AAGTCAATTT ATTGGCGGAG GAAGTACTAA ATAGTATTTT AGGGGATGCG CGTTTTTCTG 7500

TATTAACGAC CTCTAGTCGC ATGAACTATA TTCGCTATCA ATTGAGTCAA ACAATTAAAA 7560

AAGTCAGTTG GGCGCTGAAA CGTCAAAGCC AACGAAGTGs CATGACGACT GTTCAAACGG 7620

AAGTGTTATT TGGCCAAATT GCCGCTAAAA AAGGAATTTC TGGTTTGGaA TTACCATTAA 7680

AAAATCAAGG GAAGATTcAT GTTCGGGGAA AAATCGACCG GATTGATCAA TTGGTTACCC 7740

CCGAATCAAC CTATTTAGGT GTGATTGATT ATAAATCGAG TCACCGTAAA TTTAATATGA 7800

CAGAAGCGTA TTACGGCTTA GCCATGCAAA TGCTGACGTA CTTGGATGTG GCTTTAATGG 7860

ATGCTGTGCA ATTAGTTGGT CAAGAAGCTA AACCTGCTGG CTCACTTTAT CTGCATGTAC 7920

ACAATCCTAC ATTATCTTAT GAAGGCAAAG ACGACATTGA ACAACAAATG CTTAAAAAAT 7980

ATCAGTTTGA TGGTCTTTTG ATGAAAGACC CAGATTTATT GGATCATTTA GATACGAGTC 8040

TGCAAGCTAA ACAAAGTTCG TTGCTGTTTC CGATTGAAGA ATCCGCCAAA GAACAAATTA 8100

AACCTGGTCG TCGCCAAGAA GATAAATTTG TGACAGAACC AGAATTAGGC GCGTTGCTTT 8160

CCCATAACCG AAATAAATTT ATTGAAGCGG GGAATCAAAT TATTGGCGGC GAAGTTCAAC 8220

TTAATCCCGC TTATCAAGGC AAGGAACGGA TTGCTTGTCG CTATTGTCCG TTTCGAAGTG 8280

TCTGTGATTT TGATGTGATG CTAAAAGAAA ATAATTATCA TCGTATTGAG AATTTATCCA 8340

AAGAGGAAAT TATGGCGCGC TTGTTAAACA AAGATGAGGA AGGAGCAACC GAAGATGAGT 8400

AAAACAATTC CACTACGTCC AGCCAATGAG CAATTCACCG ACAGTCAATG GCAAGCGGTG 8460

TTTGATGGCG ATGAAAATAT CTTAGTTTCT GCTTCGGCTG GCTCAGGAAA AACCACCGTT 8520

TTAGTCCGCC GTGTCATTGA AAAAGTCAAA AGCGGCGTCG ATATCGATCG TCTCTTGATT 8580

GTGACGTACA CTGAAGCCGC CGCCCGTGAG ATGAAAGAAC GTATCCAAGT GGCTTTACAA 8640

AAGGCAATGA ATGAAGAACA AGATCCAGAA AGAAGACGGC ATTTTTCTCG TCAAATTGCT 8700

CTTTTGCCAA CAGCCAATAT CAGCACGTTA CATGCCTTTT GTTTAACTGT TATTCGTCGT 8760

TTTTATTATT TAATTGACAT AGATCCAGTT TTTCGGATGT TAACAGATGA AACAGAAACA 8820

TTGTTATTAA AAGAAGATGT TTGGGATGCA TTACGTGAAC AATTTTATGC GGAAAATCAA 8880

GAAGCGTTTT ACCAATTAAC AGCGAATTTT TCCAATGACC GTAGTGATGA CGGACTCACT 8940

AACTTAATTT TTTCTTTTTA TGAATTTGCC AAAGCCAATC CTGATCCAGA GGCGTGGATT 9000

AATGGCTTAA CACAAGCCTA CGAAGTAGGG GaTCAGCTAG GCGAATCAAA GTTATTTCAG 9060

ACGTATTTAA AACCGCTAGC CGTTGAGACG TTACAACGAA CTCTTCAGCG TTACGAAGAA 9120

ATGGTGACCT TAACAGAGGG CGAAGAGAAA CTTCAAAAGA TTTGGTATTT AGCTCmAAAT 9180

GAAAAGGAAC AAACAAAACA ATTTTTACAA TTTTTAGAGA GAAATGATTT AGAGAGCGCC 9240

TATAATTTGA CTGAATTGTT GAGCTTTGAC CGCTACCCAA CGGTTCGTGC CGAAGAGTTG 9300

AAACCAACCG CGGAACAAGC CAAACAGTTA CGAGAACAAA ATAAAAAAGC GTTAAATGAT 9360

TTGAAAAAGC AACTGTTTAC ACTTTCCCCA GATGCCATGA AACAAGTCTT GAAAGAGGCC 9420

ACGCCCATTG TTCAAGAAAT GGCTCACGTG GGGAAACAAT TTATGGAGGC CTATGGCGCA 9480

GAAAAACGGC TTAAAAATTT AGTGGATTTT AACGATTTAG AACATTATAC GTTAGCTATT 9540

TTAGCGAAAA ATCAGGCAGA TGGCTGGCAT GCATCAGAAG CCTCGGTTTA CTATCGCGAA 9600

AAATTTGATG AAGTGTTAGT CGATGAATAC CAGGATATCA ATCAATTACA AGAATCTATT 9660

TTGTATTGGT TGCGCCGTCC GCTTAGCACA GAAGGAAACC TTTTTATGGT AGGAGATGTG 9720

AAACAGTCTA TTTATTCTTT CCGCTTGGCA GATCCTACAT TGTTTATTGA AAAATACAAT 9780

CAGTATGGGC AAGGCAAAGA GGGCAAACGA ATTATTCTAG CTGAAAATTT TCGTTCTCGC 9840

AAAGACGTTT TAGACTTTAC GAATTTAGTA TTTAGTCAAT TAATGGATGA GCGAGTGGGC 9900

CAAATTGCTT ACGATGAGTC AGCTGCTTTG GTTCATGGTT TTGATCAATT TTCGGAGGCG 9960

GCTGATTATT CAACTGAATT GCTAATTTAT GAAAAAAAAG CAACAGAATC AGTTGAATTT 10020

CCAGAATTAC AATCCCCTGA GTTGTTGATT GAAGATAAGA CAGAAGGGGA ACTTTACGTT 10080

ACTGCTTTGA AAATTCGTGA ATTAATTGAC CAAAATTTCT TAATTTATGA TAAAAAGCTG 10140

AAAACGGATC GACCGATTAC TTATCAAGAT ATTGTCTTAT TAACGCCTAC GAAGAAAAAT 10200

AATTTAACTA TTTTAGATGT TTTTAAATCA TTAGAAATTC CAGTTCAAGT CAATGATGCT 10260

CAAAATTATT TCCAAGCAAC GGAAATTCGT ACAATGATTG CCTTGCTTCA ATTGATTGAT 10320

AATCCTTATC AAGACATTCC TTTGGCAGCT GTGTTACGTT CACCAATTGT TGGATTAAAA 10380

GAAAACGAAT TAGTGCTTAT TCGTTTGGCT AATAAGGAAA CCAGTTACTA CGAAGCATTT 10440

TTAACGTTTA ATCAAAAGAT GGAACCAACC ATGGAGGAAG CAGTTGTTCA GGAAAAAACT 10500

ATACGTTTTG CTGAATCATT AGAAAAATGG CGGGAACAGG CTGGACGCAA CCAGATTAGC 10560

AACTTGCTGT GGACGATTTA TCGGGAAACG GCCTATCTTG ATTATGTTGG CGGGTTGCCT 10620

GTTGGCAAAC AACGCCAAGC CAATTTATAT GCGTTGGTCG ATCGGGCCGC GGCGTATGAA 10680

AAAACAACCT TCCGCGGATT ATTTCAATTT GTTCGCTTCA TCGAAAAAAT GCAAGAAAAA 10740

GACAAAGATT TAGCAGAACC AGTCGTTTTA AGTGAAGAGA ATGCCGTACG CGTCATGACG 10800

ATTCACGCTA GTAAAGGGTT AGAATTTCCG GTGGTATTCG TTTTAGATAT GACCAAAGAA 10860

TTTAATGTCA GTGATTTGAA TGAACGCTAT ATCTTTGAAG AAAATTTAGG GGTAGGCATT 10920

CGTTATTTAC AACCTGAAGA ACGAGTTATG TACGATACTT TACCATTTTT GGCGATTAAA 10980

CAAGTACGAT TAAGAAAATT ACTATCCGAA GAAATGCGGA AGTTATATGT GGCGTTGACA 11040

CGTGCAGAAC AGAAACTTTT TCTGGTAGGC TCCTATAAAG ACCAAGCGGC GATGTGGAAA 11100

GAGTGGCTGA AAGTTGGCGA TGTAGAAACA CTGGTGCTAC CAGCTGAAAA TCGCTTGCAA 11160

AGTAAAAGTA GTTTAATGAA TTGGGTTGGG ATGACATTGG TACGTCATCA AAAAGCCGAT 11220

GAATATCAGC AAGAAGTAGT AGTAAGTAAC GTTCCGCAAG TCAAAAAGCA CCCAGCTAAT 11280

TTTCATATTC AGTGGTTCAA CGAAGAACAA CTACGAGCAG CCATTCAGCA ATTACAATTG 11340

CCAGAAAGAC AAGCAGAGGA TTTAGCCGAA AAAGCGCAAT TATCTGCTGA CAAAATCAAT 11400

CGTGGCCTTG CACGGTTAAG TTTCAATTAT CCGTTTGAAG TGGCTACTCG GACCACCAGT 11460

TATCAGTCTG TTTCTGAAAT CAAGCGAGTA TTTGATGACC CCGATAACAA AGAAATTGGT 11520

AAAATTGAGG TTCGGGAAGA CAATACTATT CAAGCGCAAC CGTTGATTGT CAATCGAATG 11580

ATTGAAGGAG ATTTAAGTAA GCCTAAATTT TTAGATACGG TTCAAGCGCC CTCAGCAGCT 11640

GAAATTGGGA CAGCAACCCA TTATTTATTA CAACTAATTG ATTTAAGTAA GCAACCTTCT 11700

TACGAGGAAG TGCGAGCTGT TCAAGAACGC TTAGTGGAAA ATAAATTGAT TTTACCAGCA 11760

ATTGCTGAAA AAATGAATCT GGAACAAATC GTGGCCTTCT TTGACACCGC CTTAGGAAAA 11820

CAATTGATTC AACACCATCA AACCGTTAGA AGGGAACAAC CTTTTTCTAT GCTGATTGAA 11880

GCAGAAGAGC TCATTCAAAA TTATCCAGAA ACGACACAAG ATGATTTATT AATTCATGGG 11940

ATTATTGATG GCTATATCGA ATTAGACAAT CAATGCATTT TGTATGACTA TAAAACCGAT 12000

CATGTAAAAG GCACCTCTCC CCAAGCAATT TCTGAAATTG TCGAACGGTA TCGCGGACAG 12060

ATGAATCTTT ATCGTAGAGC GCTACAAGAG GCAACGCATA AAGAGGTGTC GCACGTCTAT 12120

TTAATTTTAT TAAATGGAGG AGTAATTATT GACATGCAGA CmGGAAATGT TGTAGATTTT 12180

ATAAAGTAGA AAAAATTAGA AGCAGAAACA CTTACATTTT GTAAGGAATG TGCTATACTC 12240

ATTGAAGAGG TGAGGAAGAT GGAAAAAGTA AAAACAGAAT CGTTTGTACT TTGTsCCgaA 12300

ATTTGAAAAA TCATTTTCAA TACTTGGTAA AAAATGGAAT GGTTTAATTA TTGATGTTTT 12360

ATTAGAACGA GGTCCTCAAC GTTTTGGCGA ATTAAAAGAA rAAATTCCAA TGTTAAGTGA 12420

TCGTGTCTTA GTAGAGCGAT TGAAGGAACT AGAAGCTGAA GGAATTATCA CCAAAGCTGT 12480

TCGTTGCGGT GAAGGAAATC GCTTTGGAnT ATTTCCTGAC CGAAAAAGGC GAAGACTTGC 12540

AATTAGCAAT GGAACAAATT CAACATTGGG CAGAAAAATG GGTAAAAGAC GAAGAATGCA 12600

GTTGACCGCA AGCAAAACTT CTTGTAAACT AAACGTAATT AAATAACTAA AAAAACAGTG 12660

ATGGAAACAA GTAGTTTGAA TTGACGTGTT TTAGGGAGAG TCAGCCAGAG ACTGGAAGCT 12720

GATTTACACA GTCACCAAGc GAAATTCACT TCCTGAGTTT GTTTGTTTTC AAGAAGAAAG 12780

CAAACCGGTA ATTGCCGTTA AATGAAATGA GTGCATGACA ACAATTTGTT ATGAATTAGG 12840

ATGGTAACAC GATAGCTCGT TCCTTTGAAA GGAATGGGTT ATCGTTTTTT TTTTAGTTTA 12900

AATAAAAGAT AGAGGGGAAC GCATAATGAC ATTACAAGCG CAATTAGAAG CTCTTAGAGA 12960

CAATACGCTC AAAGAAATCG CACAAGTTGC TACTTTAAAA GAATTAAACC AAATTCGCGT 13020

AGAAACATTA GGGAAAAAAG GGCCAATCAC CGAAGTTTAA GAGGCATGAA AAACCTTTCA 13080

CCAGAAGAAC GACCGGTGGT GGGGGGCTTT GCAAATGAAA TTCGTGATTT ATTAACAGAA 13140

GCAATTGAAG CGCGCAAAGT TGTTTTAGAA AACGAAGCGT TAAATGCGGC ACTAAAAGAA 13200

GAAAGTTTAG ATGTAACTTT ACCTGGGAAA CAAATGCCTC AAGGCACACG TCATATTTTG 13260

ACACAAGTTA TGGAAGAGAT TGAAGATATT TTCTTAGGCA TGGGTTATCA AGTAGTGGAA 13320

GGATACGAAt AGAaTCCGAT CATTATAATT TCGAACGCAT GAATCTACCA AAAGATCATC 13380

CGGCACGCGA CATGCAAGAT ACGTTTTATA TTTCAGATGA GATGTTAATT CGTACGCATA 13440

CTTCACCAGT TCAAGCTCGA ACAaTGGaAA AACACGATTT TTCmAAAGGk GCATTACGAA 13500

TGATTTCTCC TGGGAAAGTA TTCCGTCGTG ATACAGATGA TGCGACACAT AGTCACCAAT 13560

TCCATCAGAT TGAAGGCCTT GTTGTCGATA AAAATGTCAC AATGGGTGAT TTAAAAGGGA 13620

CGTTAGAAGT CATGATGaAA AAAATGTTTG GtGaAGATCG TAAAATTCGT TTACGTCCTA 13680

GCTATTTCCC yTTTACAGAA CCTTCAGTCG AAGTAGATGT TAGCTGTTTT AAATGTGGCG 13740

GCGCAGGTTG TAACGTCTGC AAACATACAG GTTGGATTGA AATTTTAGGT GCGGGGATGG 13800

TTCATCCAGA TGTGTTACAA ATGTCAGGGA TTGATCCAAC AGAGTACTCA GGCTTTGCCT 13860

TTGGCTTAGG ACCAGATCGC GTTGCTATGT TACGTTACGG TGTAAATGAT ATCCGTAATT 13920

TTTATCAAAA TGATTTACGT TTCTTAAATC AATTCAAGGT AAAGGAGTAG TTGAAAAATG 13980

TTAGTTTCTT ATAAATGGTT AAATGAATAT GTTAATCTTT CAAACGTTAC TCCGCAAGAA 14040

TTAGCGGACA AAATGTCTGT GACAGGTATC GAAGTTGAGG GCGTTGCTGT ACCTGAAGAA 14100

GGATTAAAGA AAATTGTTGT CGGCGAAGTA AAAGAATGTG TTCCTCATCC AAACTCTGAC 14160

CACTTATCAA TTTGCCAAGT AGACATCGGT GAAGAAGAAT TGTCACAAAT TGTTTGTGGA 14220

GCACCAAATG TAAAAGCCGG AATTAAAGTC ATCGTAGCAT TACCTGGTTC AAGAATTGCA 14280

GGAAATCAAA AAATTAAAAA AGGTAAAATG CGTGGCGAAG TCTCTAACGG AATGATTTGT 14340

TCGTTAGAAG AGCTAGGATA TTCAGATAAT GTCGTACCAA AAGCCTATGC TGAAGGGATT 14400

TATTATTTAC CTCAAGAAGC AGTGAATGGG ACACCTGTTT TCCCTTATTT AGACATGGAT 14460

GATGCGATTA TTGAATTATC AATTACACCA AACCGTGCAG ATGCACTAAG TATGAGAGGG 14520

GTAGCCTATG AAGTTGGCGC AATTTATCGT CAAACACCTC AGTTTAATGA TCCCGAACTC 14580

AAAGAAGATG CTTCAGATAA CGTGGAAAAT TACGTAACAG TGACTGTTGA AGATTCACAA 14640

GATGCGCCAg cGTATCAAAT TCGTGTTATT AAAGATGTGA CGATTGCAGA AAGTCCTCAG 14700

TGGTTGCAAA ACCGATTGAT GAATGAAGGA ATCCGTCCGA TTAACAATGT GGTGGACGTG 14760

ACAAATTATA TTTTATTATT ATTTGGTCAA CCATTGCATG CATTTGATTA CCAAAAATTA 14820

GATAGCAAAG AAATCTTGGT TCGTCGAGCA ACAGCAACAG AAGAACTAAT TACATTAGAT 14880

GGTGAAACAC GTCAATTAAC AGAAGAAAAT ATCGTCATTA CGAATGGAAA AACgCCTGTC 14940

GGCTTAGCCG GAGTAATGGG TGGGGCTAAT TCTGAAATCA GTCAAGAAAC AACAACTGTG 15000

GCTTTAGAAG CGGCATTGTT CAATCCATTG TCTATTCGCA AAACGTCTAA ACAATTCAAT 15060

TTACGGAGTG AATCTTCAAG TAGATTTGAA AAAGGCATCA ACCAAGCGAC AGTTGGGCTA 15120

GCTTGTGATG TGGCAGCCGC GATGATTGCA GAGTTAGCCA ACGGAACAGT GGTCTCAGGC 15180

ACTGCAGTTG GTTCTGAAGT GGCCGTTAAA GAAGCGCAGG TAGCTGTGAC TTTAGAGCGA 15240

ATCAACCAAT ATTTAGGCAC TGCATTAGAT GAAGCAACGG TGAATGAAAT TTTTGAAGCA 15300

CTCGGTTTTG CTTATGAAGT AAATCAAGGC GCATATGAAA TTACGATTCC ACCAAGACGT 15360

TGGGATATTG CGATTGAAGC AGATATTATT GAAGAAGTGG CGCGCATTTA TGGATATGAT 15420

CATTTACCTT CAACATTGCC AAGTGGAGAA ACAGTTGCTG GAAGTCTAAC CAAAGCACAA 15480

CATGTTACGC GCCAATTAAA GAGCTTACTG GAAGGTCATG GTACGAGTGA AGCCATCAGC 15540

TATGCGTTGA CAACAGAAGA AAAATCCCGT CAATTTATGA TGAAAGAAAG TCAAACAACA 15600

CGTTTGCAAT GGCCAATGAG TGAAGAGCGT TCTGTGTTAC GTATGAACTT AATTTCTGGT 15660

TTATTAGATG ATGTCGCATA CAATGTGGCA CGTAAAAATA ATAATATCGC CTTCTACGAA 15720

GTAGGACGCG TTTTCTACCA AACAGAAGAT CCAACAAAAA ATTTACCTAC AGAAGAAAAT 15780

CACTTAGCAC TTGCTTTAAC TGGTAATACA ATGGTTAAAG ATTGGCAAAC AAAAGCAACA 15840

GCCGTTGATT TTTATACAGT AAAAGGCTTA GTGGAAAGTA TTGTAGCTGT TTTAGGATTA 15900

ACAGAAAAAA TCAGTTATCA AGCGACAACG GCGATTCCAG AAATGCATCC AGGCCGGACA 15960

GCATGGATTT ATTTAGAAGG TGAAGTGGTT GGTTTTGTCG GACAAGTTCA CCCAACGACA 16020

GCGAAAGCGT ACGATATTCC AGAAACATAT GTTGCTGAAT TAAACTTGCa ACAATTAGTA 16080

GCCATAGAAG CAGGCGGGGT TACTTATGAA GCAGTTTCTA AATTCCCAGC AGTTTCTCGA 16140

GACATTGCTT TATTAGTCGA TGAAACAGTT ACCAATCAAG AACTAGTTAA AACTATTTCA 16200

GATAACGCAG GTAAATATTT GAAAGAGATT CATTTATTTG ATGTATATCA AGGTGAAAAA 16260

CTAGGCGCTG GCAAAAAATC AATGGCGTAT AGTTTAACTT TCGTTAATGC AGAAGCGACA 16320

TTAGTGGATG AAGAAATTAA CCGTTCAATG GAAAAAGTTG AAAAAGCGCT AATTGAAAAA 16380

CATCAAGTAG AAGTAAGATA AAATAAAAAC AGGAAAGTGG GGCATAAGTC GAAATGACTT 16440

ATGCCCCACT TTCTAAATCT TAAGATAATC AGCTGAACAA AAGCAGCTCC AAGAAATAAG 16500

CCGAAATTCT CCAAAAATTA AAGAACAATT TTCGGAAATT CCTTCTTATT TCTTGGAGCT 16560

AAACACTTTT GTCTCAGCCT CCTATTTTTA ACTACATTAA TAACTTTTCG CCCAATTTCG 16620

TGAAAGCCAA CGTGAAAATT GTGAGATACT AAAGTTAATA ATGAAATAAA TTAGGGCGAC 16680

GATTCCGTAA ATAGTGAATA CTTGTCCTGT TTGTGCGTAg CGCCCCATTA AGATATAAGA 16740

TTTGCCAAAG AGTTCTT[illegible]A AAGCAATCAC GGAATACAAG AAACTTGTAT CCTTGATGAC 16800

AGTAACAAAC TGAGAAACAA TTGCTGGTAA GACATTGCGA ATGGCTTGAG GCAAAACAAT 16860

ATGATAAAGA ATTTGCCAAT TTGAAAAACC TTGAGAACGT GCCGCTTCAA TTTGGCCTTT 16920

GTCGACCCCA TTTAAACCAC CACGAATAAT TTCCGCTAAA GCAGCGGTGG TAAACACAGT 16980

AAAACTCACA ATTCCAGCAG GTGTGGATTT GATTTTGAAA ATCAAAAAAA TAACATAAAT 17040

CCACAATAAG TTTGGAATAT TACGCACCAC TTCTACATAA ATACTGGCAA GTAACTTCAG 17100

TGGTCCATGG GTTTGATTTC GCAAAACGGC TAGAATGGTG CCAAAAATCG TACTCAAAAT 17160

AATTGAGACG AAGGAAATAT AGAGCGTCAA TTTTAAGCCG TTAAATAGAA ATTGAAAATT 17220

ATTTGCTGTC AGTAATTGAC GCATTTCTTC TGAAAAAGGC ATCTGTTTTC CTCCTATCTG 17280

CTATATGCTT TCTTATTTTT TTCTTCCATT TTTCTAGCCA AATTCGCTAA TGGGAAGCAA 17340

AGAATGAAAT ATAAGAAACC AGCAAGCGCA AAGGCGGGAA TATAGTTCAA GTTGACAGAC 17400

GACCAACTAT TGGCAGCAAA CATTACGTCA GCGCCAGAAA TAATGGCAAC TGTTGACGTA 17460

TTTTTAATTA GGTTCACTAC TTGATTGGTT AAAGGCGGTA ACATAATCCG CCATGCTTGT 17520

GGTAAAATCA CATACCGCAT CGTTTGACCG TACGTTAATC CTTGGGAGTA AGCCGCTTCA 17580

AATTGCCCTC TTGGGACAGC TCCGATTCCT GAGCGAACCA CTTCAGAAAT GTAGGCGCCA 17640

TGATAGAGGC CTACACACAG AATGGCAATC GTAGTAATAG AAAATGAAAG CAGCGGGCTT 17700

AGCAGAGGAA AGCCATAATA GACTACGATA AATTGAATCA ACAAAGGTGT GTTTTGAAAA 17760

AATTCTACAT AGACACGACT AATTGCGTGT AAAAGTTTAC TTTTGGTCGC TGATAAACTG 17820

CCAAAGAAAA TGCCTAAAGC AAAGGCAACT AAGATCGCTC CTAAAGAAAG CAGAATAGTA 17880

TAGAGAAATG CCTGACCAAA AATTCGCCAA TCTTGAAATA AAGCCTCCCA ACGATATAAG 17940

GCAAATGGCC CTTCTTCAAG CCAGGAAATA TTGGCTAACA AAAACATAAA AATCCCTCCT 18000

AAAATTTTTG ATTCTTCAAG TCTTAGGAAT GATTAATTTT TTAAGCCCCA GTCTTGATAA 18060

ATTTTTTCTA GTGTGCCGTC TTTTTCCCAT TTTTCAATTA AACCATTTAG ATAATCATTT 18120

AATTCAGTAT TTGCTTTCTT TGTGGTAATT CCATATTCTT GTGGCGAAAA ACCATCTTTT 18180

AAAATTTCTG TCTTGCCATC CACATAGCCA GATAGGATTG ATTTATCCAC TGAAAATGCA 18240

TCAATTCGTT TAGAAGTTAA CGCTGTTTTT AATTCTGGAT AGGAACCTAA TTCTTGATAA 18300

GTAAATTTGA TTCCTTCCTC TTGCGCTTTT TTCTCTAAGT TTTCTTTTGT CGTAGCAGAT 18360

TGTGCAACGC CGATTGTTTT TCCGTCTAAA TCTTTCATGC TGGAGAAGTT ATCCTCTTTG 18420

CGCACAAGGA AACCAATCTC ATCTGTATAA TAAGGAGTTG TAAAATTATA AGTCTTTTTA 18480

CGTTCATCAG TAATTGTAAA TGTAGCAATT ACCATATCTA GTTCACCATT GTCTAAAAGG 18540

GGACCGCGCG TTTTCGCTGT CACGCCCACA AATTCCACAT TCTCTTCGCT GCCGGTAATT 18600

TCTTTCGCAA TCAAGCGTGC AATATCTgGT TCCATTCCTT CAtTTTTGTn TGATCAGGAT 18660

TCATATAGCC AAAATTAGGA ATGCTTCTTT GACACCTACT TnAATnTnCC CTGCTTTTTT 18720

TAAT 18724

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7254 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

AAAAATTACG TGATTCGCTT AGCCTGACCA GACTTTTGTC TAAGTCTTAA AACTTGATAA 60

GTAATAAGAG GTCGGGACAG AAGTGCTTAA CTCCGAGAAC AAAGTAGGTA CTGTTCAACA 120

TCTATTTGCT TTGCAAATAG TGTTTCTCAG CAATAAGAAG AAATTTCCGA AAATTGTCAT 180

AAAACACAAC GAGGAACGAG TTGATGTTGC ATGATACACT ACACTCCGTT TCGTTCGCAT 240

CAATTTCTAA GCATGGAACA TGCTAGAATT GAAGGACCCG GTGTTCTTTA ATTTTTGGAG 300

ACCAAGTCCT TCAATTCTTA GTGATTTATC ACTTAGAGAT TGATGAGATC GGAGCTCTGC 360

GAAGTAGGTA TTGTTCATCA TCGAATCACT ACGTTCTTCG TGATTCTCAA CAATTTCGGC 420

TTATTTCCGA AGGACCTTCA CCAAGTAGGT ACGATGATTT CATCTGTTCA TGCTTCACAG 480

TGAAACATCG CAACTCTTAG AGCTGTAGCT CTTTTGAAGG CTGTACTGTG ACGAGACGCA 540

TGTCGAGAAA ACACCTTAGA AGTTGTTGAG ATCGGAGCAA CGCGTAGTGT TGCTTCTGTC 600

CCCGCCGTTT ATCAATTTTT AAGCCTGGAA CAAAAATCCA AAGTGATTTT TGTTCCAGGC 660

tCTTACTTTC ATACTAATCT TAGGTCAGGT TGTCATCAAG TTTGCCAATA ATAGACTTTT 720

ACAAGCGCGC CTTTTTTGAG TAGTACACCA GCTGGTTGTT CTAGCAGGCA ATTTGTTTCT 780

AAAGTATTTG ATAAATTGGA TGAAAGATGA CTTTTTGTCA CTAAATTTAC TGTGCCATCT 840

ATTTCTTTCG CTCGGTAAAA TCGACGTAGT CTACTAGGGG TTAATTCATG CGCTAAAATT 900

ACTGTTCGTT CCTTGAGATT TAACTCGACA CAATTAAAAA AAGTAGCTAA TAATGTCCAG 960

TAAAAAAGAT GTAAATTAAC CATTGCCGCA AACGGATTGC CTGAAACGCT TAAAATGAGT 1020

GTTTCTTGAA ATTGACTTCC CATAACTGGG GTTCCAGGTT TCATATTTAC AAAATGAAAC 1080

AACTCTGTTC CTCCGATACG TTTAATTGCT AAAGGTAAAT TGTCCTTTTC ACCTACCGAA 1140

ACACCACCAC AAGTAATTAA TACATCAAGT TTTGGTGAAA TCTCAATCAG CCTCTTCGCT 1200

AATTCTTCAG GGTCATCTGA ACAATGGTCT AGGTGGATTA CTTCATTACC ACTCGCTTTG 1260

ACAAAAGCTG CTAACATGTA GAGATTACTA TTATAAATTT TCCCAACTGA CAGTGGTTCT 1320

TCCAAAGAAA CTAACTCACT ACCGGTCGCT AAAATCCCCA CCTTCATGCG CTTCAAAACA 1380

ACAACTTTAT CAATCCCTTG CATTGCCAAC AAACCAATCA TGCGACTATT AATCTGTTGA 1440

TGTTTTGCTA AAATTTTTTG CTGAAAAGCA ATATCTTCTC CAATTTCCGC ATAGTTTTGG 1500

CCCGCAGTGA CAGAGCGAAA AATTTTAACA TGGTCTTGCC CAAAATCCGT CCATTCTTGT 1560

GGCACTACAG CATCAAAATC GGTTGGAATA AAAGCACCTG TTGTTATTTT AATCGCACAG 1620

TTATTCGTCT GATTACAATA AAACATGGAA TCCCCTGCCA CTAAAGAACC GACGACTTTT 1680

AAACAACTTG GGCTTTCAGC AGTCGCACCT CGTGTTTCTA CTGCACGAAC GGCATACCCA 1740

TCCATCCCGG CACGAGGAAA ATGAGGAATC GCTGTCTCTG CAAAAATGGT TTCTCCGCTA 1800

ATCCGTCCCA AGGAATCAAG CAGCGCCACT TCTTCTGTTT CACATTGACT CTTCATATGT 1860

TGGGTCATTT TCTCTCGTGC TTCTTCCAAC TCGATCATGT TTTAACCATC CTTTTTTATT 1920

TATTCACTGG TTTAATCGCA ACGGCACAAA CTTTATATTC AGGAATTGTG GCGATATCAT 1980

CAAATACAGC ATTGGTGATT TCGTTAACAT TCCCATCTGG AAAGTGGAAT GTCATAAAGA 2040

CTTCTTGCGG AAATACTCGA TTTCCGACAG CCGCATAGGT TTCGATTTTG CCACGTCGAG 2100

AATGAACTTC AACTTTGTCA CCTTCTTGGA TTCCCAGAGC TTGTGCATCG ACTGCATTTA 2160

TCTCAATATA AGAGTGATTC GCTATTTGGT TAATCCCTTC CGTTCGTCCT GTCATTGCTC 2220

GGGTATTGTA ATGGTAAAGC ATTCGGCCTG TCGACATTAA ATACGGGTAC TCTTCATCTG 2280

GTAATTCTTG TGCTTGCTTA TACGGAATTG CTTTAAATAA ACCTTTCCCA CGAGCAAAAC 2340

TACCCACATG CATAATCGGC GTACCTGGAT CTGTCAAGCT CCGACAAGGC CATTGTAAGC 2400

TCTCCTTCTC TAATCGTTCA TAATTAATTC CACCAAACGA AGGCGTCaCA GCTGAAATTT 2460

CTTCCATAAT TTCTTTTGCA GATTCATACG CACAAGGATA GCCCATTCGT GTCATGACTT 2520

CACAAAAAAT TTCATAGTCT TCTCGCGCTT GTCCTCTTGG TTCAACAGCT TTACGGACTC 2580

TTTGCACGCG CCGTTCTGTA TTTGTAAATG TCCCGTCTTT TTCTGCATAA CTGATGCCTG 2640

GTAAAACGAC ATCTGcATAA GCAGCCGTTT CGGTCATAAA TAATTCTTGA ACAACTAAAA 2700

AGTCnTAAAC TTTCTAAAGC TTGACGAACA TGTCCTGTGT CTGGATCCGT AACAATTGGA 2760

TCTTCTCCAA AAATATAAAG GCCTTTGACG CTTCCTGCCG TTGCTGCTGG TAAGACTTTA 2820

GTAGAAGTTA AACCTGTGTT TCGATTTAAC GGCACATGCC AAGCTTTTTC AAATTTGTCG 2880

ATCACTTCTG GATTATTGAC TTTTTGATAG CCTGGGAAAT CATAAGGCAT ACAGCCCATA 2940

TCGCAAGCAC CTTGAACATT ATTTTGACCT CGCAAAGGAT TGACACCACA GCCTGGTTTT 3000

CCAACTTTAC CAACCAACAT TGCCAAATTA GACATACTCA TAACACCCTC AGTTCCTGTC 3060

GAATGTTCCG TTACACCTAA ACAATAAATG ATTGGTGCTT TCTCTGCTTT TGCATACATG 3120

CGTGCTGCTT GAATTAAATC TTCTGGATGA ATATGACAAA TTTCTGCGAC TTTTTCCGGC 3180

GTATAATCAG CCACCATTTT TTCTAAATCT AAAAATCCTT CTGTCCGTTC TTCAATAAAA 3240

TGGCGATCTG CTAAACCTTC TTTCAAAATC ACATGCATCA TGCCATTAGC AAATGCGACA 3300

```
TTCGTTCCCG CCTGAACTTG TAAATGCAGC GCACTATCCT TCACTAAGTT AATTTTCCGC     3360
GGATCAACAA CAACAACTTG CGTTCCTCTT TGGATGGCTT GGCGAATTTG CGCACCAATA     3420
ACTGGATGAG CTTCTTCTGG ATTGGAACCT ACTAATAAAA TCATGTCGAC ATCTTCAGTG     3480
ATATCAGCAA TAGGATTAGT CATCGCACCA GAGCCTAAGG TTTGGGCTAA ACCATGGACC     3540
GAAGCCGAAT GACAAACACG GGCACAGTTA TCAACATTAT TGGTACCAAA AGCAGCACGC     3600
ACCATTTTTT GAAAGACATA ATTATCTTCA TTTGTCGCTC TAGAGCACGA AAAACCAGCA     3660
AGAGCATCCG GACCGTTTTC TGCTTTAATT TCGTTAAATT TTGACGAAAC AAGCGTTAAG     3720
GCTcTTCCCA AGAAGCAGGT TCAAAAATGC CATTCCGTTa ATTAATGGCT CCGTTAAGCG     3780
ATCTCCTGAC CCAACAAATT TATATGAAGC AAATTTTCCT TTCACACAAA GTAAATTTTT     3840
ATTAGCTGGT CCATCCAGGG GTTCAACACC CACCAAACGA TTATTTTTAA CGAGTAAATT     3900
CATCTGACAG CCTGTCCCAC AATGCGGACA AGTTGTAGGA ATTTTCTGGG TTTCCCATTT     3960
GCGGTACTCT kTCGTATCTT TTGCTGTTAA CGCACCTGTA GGACAGGAAG AGACACAATT     4020
GCCACATGAT TCACAGATAG ATTGATCAAA CGCCTGCCCA TAACTCGGCA TCATTTTTGT     4080
TTCAAAGCCA CGATTAGCAA TACTAAGTAC ATCTCTTCCC TGCCGCAATT GACAGACACG     4140
GGCACAACGA CGACACATGA TACACTTTTC TGGATCATAA GAGAAAAATG GATTACTGGT     4200
ATCTTCTTGA TGGCAAGGCA TTCTTTTTCC CTCGGTAAAG CTAGTGACAT CAATCCCATA     4260
ATCTAGCGCA TACTGTTGTA ATTGACAATC CCCGTTTTTT CCACAAGAAA AACATTCTAA     4320
TTTATGATTA CTCAATAATA AATCAAGAAT AAAACGTCTG GAATCAGCAA CTTTTTCAGA     4380
ATGTGTTGCC ACAACCATGT CCTCTTGACA ATGTGCCGTA CAAGCAGTTG TTAAGCCACC     4440
ACGTCGGCCA CCTTCCACCT CAACCACGCA CATCCGACAA GAACCATCAG GTGCTAATTC     4500
CTTTAAATGA CAGAGCGTTG GGATTTCAAC ACCTAGCCCT TTCGCTGCTT CGAGAATGGT     4560
AGTCCCTTTA GGCACAGTTA CTTCTTGGTT ATCAATAGAC AACGTAACTG TTTGCGTATG     4620
CAACTTTGTT TTCATTTTTA CCAACCTCCT TTGCTGGTGG CATAGATTTC TGTTTCAGCA     4680
GTCGTTTCAA ATTCTTCTGG AAACAAAGCC AATGCACTCT TCATTGGATT tGCGACGGAT     4740
TGGCCCAAAC CACAAGCTGA TAATCGCGTG ACATGTgTGA GCATCTTTTc cAaTCGCGGC     4800
AAATCACTAG CTGGTTGCCT CGTTCCGATk GAATTTACTC AATAATTCCA AAATCCTAGT     4860
TGTTCCCAAT CGACAAGGGG TGCATTTACC ACAGGACTCA TGAGCAAAAA AGGCCGCCAC     4920
GTGTACTAAA TAGTCCACAA CATTAACTGA ATCATCCATT ACTACAATTG CACCAGAGCC     4980
GACAGATAAG TCATGGGCCC ATAAATCTTC GTAGGAATAC AAACAATCCC TTAACGTTGC     5040
GACAGCACCA ATTGGACCGG ATTGTCCACC GAAGTGAATA AATTTCAATG CTCGACCAGT     5100
TGCAGAGCCT CCGCCGTATT GCTCGCCATA AATTATTTCT TCTAATGGTG TTCCAAGGTT     5160
GACCTCGTAT AAGCCCCGAT TTTTAATATG TCCAGACAAA CAAATTAGCT TTGTTCCGCC     5220
ACCATCGGCA GTACCCATCT CTAAAAAGGC TTGTCCACCT TCTCTTAAAA TTACAGGAAC     5280
```

ACTTGCAAAT GATTCAACAT TATTTACTAA TGTTGGTTGC AAATATAAAC CAACATCTGC 5340

TAAGTGTGGT GGTTTGACAC GCGGCCGCCC CGTTTTTCCT TCAATTGAAT TTAAAAGTGC 5400

TGAATTTTCC CCACAAATAT AGGCGCCAGc ACCAGAAATA ATCGTAATAT CATAATTGAA 5460

TCCTTCAATT CCTAAAATGT TTTCCCCTAA AAAGCCAGCT TGTCGTGCAT TATCGAGTGC 5520

TTCTTGAAAT GTTTTTTGGA TTCTGCGATA TTCTCCTCGC ATATAAATAT AACCTGCTTT 5580

TGCTGAAAAT AAATAGCCAG CAATAATCAT TCCTTCAATC ACACTCAATG GATCTTCAGA 5640

TAAAAGTACT TTATCCTTAA AGGTCCCTGG TTCGCCTTCA TCAGCATTAC AAACAATGTA 5700

CTTTGTGGtT CCTTTTGcAT GATAAAGATG GcGCCATTTT TTTCCTAAAG GaTAGGcTGc 5760

TCCACCGsGa CCTCTTAAGT GcGcAATATC CAGTTCGTTT AAAATTTCTT CATCCGATAG 5820

TGAAATTGCT CTTTTTAAAC CATCAAATCC ATCGTATTTA CAATATTCAG CAACATCgGT 5880

CGCAGATTGC ATTTTCCCCA TTCTCTCTAA TAACATTGGT TGGTTTCTCT TAAGCATTGA 5940

CTCACCTCyT ACAGTCCTTC ATAGCAATCA TTTTGTAAAT GTTGAAwTAA TTGATAAATT 6000

TTCTCTTCTG TTAGTTGACT AAAAACGGTG TCCTTAATTT TAATTACTGG ACCTAAATCA 6060

CAAGCCCCGA TACAAGGAAT ACTATGATAC ATAAATAAGC CATCTGGTGT CGGCTGATTC 6120

TCTGGCACTT CTAAAATGGT TTCTAAAACT TCGGCAACCA TCGCCCCACC AGTATAATGA 6180

CACGGTGTGC TGTTACAAAT CTTCAAGACA TACTTCGCTT GCGGTTCCGT CTTTAAAATG 6240

GCATAAAAAC TGACGATTTC ATAGACTCGC GCTTCCGTTA AATGTAGATG CTCTGCAACT 6300

AATTGAGCTG TCTCCTGATC GATATACCCT TCCTCTGAAG CAAATTGTAA TTCAATTAAA 6360

ATATTTAAAA TTCGTTGTGG ATCAGCATCG TTTTCTAGAA TAATCGCTTC TTTTTCAATT 6420

AAGCTTAATG CACACATTCC TTTTTATTCC CCCTTTTTAG TGAAACTTGA AGTTGTTCTC 6480

CTTGATAATA GATACGCAAT TTTTTTGGTA TTTCAGCAGA TAGTAAACAA AAATCGGCTT 6540

TATCTATGAC TTTTTTTGCT GATTCTTTTT GGCGATTCTG GTTATTCATC ATAAGAAATA 6600

GTCCTGGCTG CACAACATTA CGTAATGAAT TTGATTCACA AATGATTAAT GCTTGGGCCG 6660

GCACAGCTTT CAGAAATTGC TGAATGGCTT CCGCCAAATT TTCTTGAAAA ACTTTCAATA 6720

GAAAAACTTT TTGACAACCT GCAGCTAATA ATTTCATGGT ATCTTTATTG CCAATCGTCT 6780

CTTGCTCTTC TACTAATTCA TAACCAGCTG TAATACTGGT ACAAATTCCG CATCCTTTTG 6840

TGCCTCGTTG ACACACCCCT CGTTGACCAC TAATCGTAAT AATCTTCAAG CCATAAATTG 6900

GATAGACTTC TTGGTACTGT TTAATAAGGG CTTCGGCTAA TGTCGTTTTG CCACTATTCC 6960

GACCAGTAGA ACCAATTTGA ATCATGTTTG GTATCTCAAT TGCCATAGAG CTCCTCCGTC 7020

TGAATCGTTT TCTCAACTCT TGGTTTGACA GTTGCTGGTT TTGTCATAAA GATCAATGGT 7080

ACGGCCATAA AAAGACTGCC ACCTACTGCA TTGCCTAAAA ATACCCATAA TGTATTCTGC 7140

AAATAATCCC ACCAACTGAT TGTGGTACTT TGTGAAAAAA TTGCCGCTGG AATAATAAAT 7200

GCATTAGCGA CCACGTGTTG AAACCCACTT GGGGATTACA ATTCAAAATG AGAT 7254

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7296 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

GGGGCCAATT TaGaAAATkG ACmATTGAAT AGAAAAGGaT GTGAACaAAT GGCaAAAGAa 60

ACAAATGTTA AgTTAGTCaC GGTGAGTGTT TTTGTGGCAA CATTTATGAC AGCCATTGAA 120

GGGACCATTG TGTCTACTGC GATGCCAACG ATTGTCGGCT CGTTACATGG CATGGAAATT 180

ATGAACTGGG TATTTTCAAT TTATTTATTA ACGAATGCGA TGTTAACACC GATTTATGGG 240

AAACTTGCGG ATAAAATTGG TCGTAAACCT GTCTTCATGA TTGGCATTAT CATTTTTATT 300

TTGGGCTCCT CGTTGTGTGG CTTTGCTCAA GATATGTTGA CTTTAATTAT TGCCCGCGCA 360

ATTCAAGGTG TGGGGGCAGG CGCAATTTTA CCAGTTGCGT TAACGATTAT TGCCGATATG 420

TATACATTGG ACAAGCGAGC GAAAATTTTA GGTTtAAACA GTGCCGCCTG GGGAATTGCT 480

AGTATTTTTG GTCCGTTAGC AGGTGGTTTT ATTGTAGATA CAGTCGGTTG GCATTGGATT 540

TTCTTCATTA ATGTTCCTAT TGGACTTGTT TTATTGGGCT TGATTAGTAT TTTCTTAGTT 600

GAACCAAAGC GGGAACGGAC CAAGATGCCA ATGGATATTT TGGGCAGCGT TACTTTGATG 660

GTAGTGCTGC TAACGTTATt GCTAGGTTTT CAAATGATTA GCGATAATGG TTTTACATTA 720

GTAACATTTG GTTGTTTAAG TTTGAGTGTG CTCTTTTTTG TAGCATTTGT GATGATAGAA 780

AAACGCGCGC AAGACCCAGT GATTGATTTG CATTTATTTA ATCAACCAAC GTTTGTTTTA 840

GTAAATCTTA TTGCAGCGCT TATTAGCGGT TTCTTAATGG GGATTGATGT CTACATTCCG 900

ATGTGGATGC AAGGTGTCTT AGGAAAAAGT GCAGGAATTG GTGGCTTAGT TTTAGCGCCC 960

TATGTCGTTA CTTTGGATGG CTGGATCATT TATAGCAAGT AGTTTTATGG AAAAATATGC 1020

TATGAAAAAA GTCTTAACGA TTGGGTTATC GATTCTATTA GTCGGGGCCA TCTTTTTAGT 1080

GGTAATGCCA ATGGCCGTTC CGTTTTGGCT TTTCTTTGTA GTGTCTTCTG TCTTAGGAGT 1140

TGGTTTTGGG ATTACTATCA CAACGACTAC GGTGACAGCA CAAAGTACAG TGGAGCCTGA 1200

AAAAATGGGG GTTGCAACAT CGTTTAATAC GTTGGTGCGT ACAATTGGGC aGACTGTGAT 1260

GGTGTCAATT TTTGGTGTGA TTTTAAATGc AGGAATGTTT GCGAAATTGG AAGCGAGCGC 1320

GTTAAACGTC GATGCAGATG TCATGAATCA ATTAGTGAAT CCACATACTG CaAATTaATT 1380

CCAGCTGCGT TGTTAAAACC ATTACGCGGT ATCCTCTATG CAGGTCTGCA TAATGTTTAC 1440

TTAGTCGGTG CGGGCTTAGT TGTTGTCGCT CyTTTGTTAA ATATTTTCGC AAAAGCGCAA 1500

CGAGCGAAGG TTTAGTGGTA CACTAATAGA AAGAGTATTG GAGGGAAAAG TAGATGGCCT 1560

TTTTACAAGC AAATATTTAT TCmAATGTAT TGGAAATGGA AGTGAATGTG AACGTAATTT 1620

TACCACAAGA AACGGTAAAA AAAGTGGGCA CAAGTACCCA AGGAGCATTA ACAGATATTC 1680

CCGTGTTGTA TTTATTACAT GGGATGGGAG GGAATCACAG TGTCTGGGCT CGCAGAACCT 1740

CTATTGAGCG GTATGTCGCT GATTTGGGGA TTGCTGTCAT TATGCCGTCC ACGGATTTAG 1800

CTTGGTATAC GGACACACAA TACGATATGA ACTATTGGAC GTTTATCTCA GAAGAATTAC 1860

CGAAAATTTG TCATCAACTC TTTCCACAGT TAAGTACAAA ACGAGAAAAA ACCTTTGCCG 1920

CTGGTTTGTC AATGGGTGGT TATGGCGCCT TGAAATTAGG ATTAGCGAAA CCAGAATCTT 1980

TTGCAGCCGT TGCTTCTTTG TCTGGGGCAG TCAGTTTATC ATCCACTAGT TTTGGGGAAT 2040

TATTAAAAGT CCGTAAGCGA AgcTATTGGG AAGGTATTTT TGGCCCGTTA GATCAAATTG 2100

AAGGTTCCAT TCACGATCCG CTGTATTTAT TACAACAATT AGTAGAAAGC CAAACAGAAA 2160

TGCCCCAATT TTACCTTTGT TGTGGGGAGC AAGATATGTT ACTTTCTGCT AATCAACAGA 2220

TGGCTCAAGC GTTGGAACAG GCACAGGCAA GTTACACTTT TGAAACTGGA CCAGGAGAAC 2280

ATGACTGGGT GTTCTGGGAT GAATGGATTC AAAAAGCCTT GGCTTGGCTA CCAATACCAA 2340

AATAGAGAAA GTAGTTTTCA ATATTTATTT TAAAAGTTAA ssCTTGGTCC GTCAGTCGAA 2400

ATGACTGACG GACCAAGGGc TTTTTTTATT TAGATGAAGA TGTGTTGTCT GCTTGCAACC 2460

CATCTAAGTA GGTTTTAACG GCTTGAATAA AGGCGGATTT TTGGTGGATG GAGTGTCGAA 2520

ATTGAATGCT TTCCTCCATG TACGTTAACT CATTGTCGGA TACATATTGT TTCATGTGCA 2580

ACACTTGTTC TTTTAAATAG TCATCTCCGT TCAACACTTG CGAAACGAGC TGTTTTTCCT 2640

CCGTTGCGTC CATAATCATC CCAAATAAAA GATGTTGTAA AGACGTTACT GCTAAATGAG 2700

TCATTTCTGG TGAAAAACCT GCTTCTCTTA AAATGCCCAT CATTTGATTC AAATGGCGTA 2760

ATCGTTGGGG ATAGGCAGGG ACGGTTTGGA TTTCGATAGC TACGGCACAG GGAAATTGTT 2820

GATAGAGATC ATAATAGTTT TCCATGAAGG CTAGTAAGTC AGAATACCAT TCACCACATA 2880

ATGCAGGTTC TTGAAAGTGC TCTTCAATCG CTTCAGCCAT ACTTTGAAGT AAGGCTTGCT 2940

TATTTTTAAA ATACCmATaA ATTGcGGGCG CCTGTACACC AAGCTGTTTG GCGACTTTAC 3000

GCaTAGAAAG TTGtTCTAAA GTGGGACTTT TTTCTAACAA AGAAAATGCG GCTGCAATAA 3060

TCGTGTCTTT TGATAATTTC GGTTCCATGA TTTTAGCTCC TATTAAACGT TATCTTCGTA 3120

AGTTTATCAA AAAGAAGGTA CACTTTATGA TAGGAAAAAC TTGTGAAAAA TTAACTTGTT 3180

CGCTTGTATT TTACCACTAA AAATGGTAAA TTATGAATCG GCCTTATTTT ACAGTGTTAA 3240

ATTGGTGAAG TCAATGAGCA AATTGATTTT TTTGTAATGG GGAAGAATTC GCAGACCAAC 3300

ACTGAAGAAG CAGGGAACTA TTTTAAGAAA GTAGGGATAA GATGGACCTT ATTATTCAAC 3360

ACGCCAAAAA ATACAAAGGC TCGGTAGTTA TTGCCTTGTT GGCGGTGATC GTAATGGTTG 3420

TCTCAGCACT TTGGCAACCA AAATTGTTAC AACAAGTGTT GGAAGCAATT ATGAATGACG 3480

ATAGCGACAA AATGAAAAAC TTAGGGATTC AATTAATTGC CATTGCAGGA TTAGGACTCG 3540

TTGCAGGGGT TATTAATACA ATTTTTTCTG CCAAAGTTGC CCAAGGTGTC AGTGCTGACA 3600

TTCGTGAAGC AACTTTCCGC AAAATCCAAA CCTTTTCATT TGGCAATATC GAAAAATTTT 3660

CAGCAGGTAA CTTAGTTGTT CGTTTAACAA ATGACGTTAC TCAAATTCAA AATGTCATTA 3720

TGATTGCGTT GCAGACACTT TTTAGAATTC CTTTTTTATT TATCGGGAGT TTTATTTTAG 3780

CGATGCTAAC CTTGCCACAA CTATGGTGGG TGATTGTTGC ACTTGTCATC GCAGTTATTT 3840

TGATTTCTAT GCTGTCCTTT TCGCAAATGG GCAAGCATTt CATGATTATT CAAAACTTGA 3900

TTGATAAGAT TAATGGTATT GCAAAAGAAA ATTTGTTAGG CATTCGTGTC GTCAAATCAT 3960

TTGTTCAAGA AAAAAATCAA TTGTCACGTT TTACAAAAGT GaGTGAAGAA TTaAcCACCC 4020

ATAACTTAAT CGTAGGTAGC TTATTTGCTG TGATGATTCC AGCGTTTATG TTGGTGGCGA 4080

ACTTAGCTGT TGTAGGCTCA ATTTTCTTTG TGAGTAACTT AGTCAAGGAC GATCCAACTT 4140

TAATCGGTGG TGTCGCTTCC TTTATGAACT ATTTAATGCA AATCATGATG GCGATTATTA 4200

TCGGCGGAAT GATGATGATG ATGACTTCTC GTGCCGCTGT TTCCATTAAG CGGATTAAAG 4260

AAGTGATGGA AACAGAACCA GATGTTACCT ACAAAAAAGT CCCAGAACAA GAGTTAATTG 4320

GTAGTGTTGA GTTTGACCAC GTATCTTTCC GTTATCCTGG GGATGAAGAA GATACCTTGA 4380

AAGACATCTC TTTCTCTATT CAGCCAGGTG AAATGATTGG GATTGTGGGC GCAACTGGCG 4440

CTGGTAAATC TACATTGGCA CAATTAATTC CACGTTTGTT TGATCCGACA GAAGGAAAGA 4500

TTGAAGTTGG TGGTGTCGAT TTACGAGAAG TGAACGAACA CAGTTTACGA AAAACAGTTT 4560

CCTTTGTGTT GCAAAAAGCA ATCCTGTTCT CTGGAACGAT TGCGCAAAAT CTACGCCATG 4620

GTAAACGTGA CGCTTCTGAA GCGGACATGG AACGGGCTTC AGGCATCGCC CAAGCCAAGG 4680

AATTTATTGA AAAATTAGCA GAAGGTTATG ATGCACCCGT CGAAGAACGG AGTAATAACT 4740

TTTCTGGTGG ACAAAAACAA CGGTTATCAA TTACAAGAGG TGTGATTGGT GAACCTAAAA 4800

TTTTGATTTT AGATGATAGC ACCAGCGCGT TGGaTGCTCG TTCGGAGCGG TTAGTCCGAG 4860

AAGCGTTGGA TAAAGAGTTA AAAGAGACTA CCACGATCGT TATTGCCCAA AAAATTTCTT 4920

CTGTTGTCCA CGCAGATCGT ATTTTAGTAT TGGATAATGG ACGTTTAGTC GGGGAAGGCA 4980

CGCATGAAGA ATTAGCAGCT ACGAACCCCG TGTATCAAGA AATTTACGAA ACTCAAAAAG 5040

GAAAGGAGGA AGCTTAATGA CTGATTTAAT AAAAGCTAGT AAATTTTTCT ATCATTATTT 5100

GAAACGTTAC AAAGTTTCGT TCCTCTTTAT TTTCTTAGCA ATTTTCGCAG CGACTTATTT 5160

ACAAGTCAAA GCGCCGCAAT TCGTTGGGGA AGCTATTCAG GAATTAGCGA AATATGCGGT 5220

TAATGTGATG CAAGGAAAAG ACGATAAAAG TGCGTTCGTT TCTGTCATTT GGAAACTACT 5280

CATTTTTTAT GTCTTAACTA GTGCCGCTAG TTTCATTTAT AGTATTCTCT TTACACAAGT 5340

CGTGGGGAAA TCGACGAACC GCATGCGGAT TGGTTTGTTT AACAAATTAG AAAAATTGAC 5400

GATTCGTTTC TTTGATTCTC ATCAAGATGG TGAAATTTTA AGTCGTTTTA CTAGTGACTT 5460

AGACAACATC CAAAATAGCT TAAACCAAGC GTTGCTACAA GTATTAACCA ATATTGCCTT 5520

ATTGGTTGGT GTCTTAATCA TGATGTTCCG TCAAAATGTG GAACTGGCAT GGGCCACAAT 5580

TGCTTCTACG CCGATTGCGA TTTTAATTGC GGTCTTTGTG ATTAGCAAGG CGCGCAAATA 5640

TGTCGATTTA CAGCAAGATG AAGTGGGTAA ATTAAATGGC TATATGGATG AAAAAATTAG 5700

TGGGCAACGT GTGATTATCA CTAATGGCTT ACAAGAAGAA ACCATTGACG GCTTTTTAGA 5760

GCAAAATGAA AAAGTTCGTG CCGCTACGTA TAAAGGTCAA GTGTATTCAG GATTACTTTT 5820

CCCAATGATG CAAGGAATGT CATTAGTCAA TACGGCGATT GTTATTTTCT TTGGTGGTTG 5880

GTtAGCAATC AATGGCTCTG TTGATCGTGC CGCTGCGCTA GGTTTAGTCG TTATGTTTGT 5940

CCAATATTCT CAACAATATT ATCAACCATT AATGCAAATT TCTTCAGGTT ATAGCATGAT 6000

TCAATTAGCT GTTACAGGCG CTCGTCGTTT AAATGAGATG TTTGATGAAC CAGATGAAAT 6060

TCGTCCCGAA AATGGCGAAA AATTAGAAGa GATTAACAAA GCTGTGGCTT TAAATCAtGT 6120

TGTTTTTGGC TATAATCCAG AAACGCCTGT GTTGAAAGAT GTTTCTATTC ATGTGGATAA 6180

AGGTGAAATG GTCGCTTTAG TTGGACCGAC AGGCTCAGGA AAAACAACGA TTATGAATTT 6240

AATGAATCGC TTTTATGATG TGAATGAAGG CGCCGTGACG TTTGATGGTG TGGACATTCG 6300

TGAAATGGAT TTAGATAGCT TGCGTTCACA TGTAGGCATT GTTTTGCAAG AGTCCGTTCT 6360

ATTTTCAGGA ACAATTCGTG AGAATATTGC TTTTGGTAAG CCAGAGGCAA CGGATGAAGA 6420

AATTGTTCAA GCAGCGAAAC AAGCGAATAT CCATGAATTT ATTGTAAATC TTGAGCAAGG 6480

CTACGATACT GAGATTACGG AAGAGAATAA CCTTTTCAGT ACGGGGCAAA AGCAATTAGT 6540

AAGTATTGCA CGAACGATTA TTACCAATCC AGAATTATTA ATTTTGGATG AAGCAACAAG 6600

TAATGTTGAT ACAGTAACAG AAGCTAAAAT CCAAAAAGCG ATGGACGAAG CAATTAAAGG 6660

GCGCACCAGT TTTGTGATTG CCCATCGTTT GAAAACGATT CTAAATGCGG ATCGGATTAT 6720

TGTTTTACGA GACGGTGAAG TCATTGAAgA AGGCAACCAC CATGAACTAG TTGAACAAGA 6780

TGGATTCTAT GCGGAACTTT ATAAAAtCAA TTTGTTTTTG AATAAGTCGA AAAAGAAGCA 6840

AAATGAGTCA TTCTCATTTG CTTCTTTTTT TGCTATGCTA AATAGAAAAA GGAGAAATGT 6900

CAATGGAAAA AATGAGGAAT ATACGGCAAT TACGGCGCAC GTTTTTGTTG CTAAATACTT 6960

TATTTGTACT GAGTGTATTC AGAGAGAATC GATGGACAAA ATCGGCACTG CTTTTTTTGG 7020

TTGTTGGCTA ATAGTTTTTT ATATGGACTG CAGTAGTTGA TTTATCTTGG GGGCGAAAAA 7080

GGCTGGCTGA TACATTTGAA GCGGTGGTAT TATTGCGCGG TTCTTCTCCA AATAGCAGGC 7140

ACATTGATTG TATATTTTTC GTTTTTTGTA TAAAGGTCTC TCGTTTCTTG CCGGAGAATA 7200

TGCTAAAATA AAAAGCAGCA GGGGGCGAGA AAACCATGGA AAAGCCnAAT nAACGCCnAG 7260

CTGGATTTAG AACGTACACG GGCCAATTAA TTTTAA 7296

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9927 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

GATTAAATGT CCTTGCCTCT GTTTTTGAAA ATCCTAAATC AAGCCTTTCA CACCTATATC 60

TTTTCGGTAG TACATTCCAG GAAATGAGTA ATTATCCAAT GTTTGATAAA CTTTTTTGCG 120

GGCTTCTTCA AGAGTCGAAG CAGTTGCTTC TGCTAAAAAG ATTCGTCCGC CAGCCGAAAC 180

CAGAGCCTGT TCTGAACGAG TTACCCCTGC ATAATAGATG GTTGCACCAG ACATTTCTGG 240

CAATGGTAGA TTTGTCTTAT ATTGATTAGG ATAGCCTTCT GCTGCTACAA CAACACCAAC 300

ACTCGCTAAC TCATTCCATG CAATCGTTGG TTGCTCTTTC ACCAGCAATG AAGAAATTAT 360

TTCTGAAAGG GGACTCTCTA AACGTGCCAA CACAACTTGT GTTTCTGGAT CCCCAAATCG 420

AGCATTAAAT TCAATAACCT TGGGGCCAGT TGCTGTTTGG aTTAAACCAG CATATAAAAT 480

CCCTTGAAAC GGACGTTTTT CAGCCACCAT TCCTTTAGCA GTCGGTTTCA GGACTTGGCT 540

AATGGCTTGC TGAACCGCTT CTTCTGGGAC CTGAGGCACT GGTGTATAAG CGCCCATCCC 600

CCCTGTATTC GGACCAGCAT CCCCATCGAA AATCCGCTTA TGATCTTGAG CAATCACCAT 660

CGGATATACC TTCTCATTAG AAACAAAAGC CATTAAGGAA AATTCTTCTC CTTGCAAAAA 720

TTCTTCGATT ACAACCGTTT GACCGCTTGC TCCAAATTGT TGGTCCTCCA TCATCGCAAC 780

TAGCGCCCGT TTTGCGGCTG CTTTTGTTTC GGCAACTACT ACTCCTTTTC CAGCTGCCAA 840

ACCATCTGCC TTAATAACAA TTGGTACTCC TTGCTGTTCA ATATATTCTA AGGCAGGTGC 900

CAAACGTCGA AAGACCCGAT ATTCCGCCGT TGGAATCTGG TATTTTTGCA TAAAATCTTT 960

CGCAAAACTT TTTGAACCTT CAATTATTGC AGCTGCTTTC GTCGGGCCAA AAATCAGTAA 1020

TCCCGCTGCT TGAAAGGCAT CTACAATTCC TTCAATTAAT GGCGTTTCCG GTCCGACGAA 1080

TGTCCAATCA ATCTGATTAT CCTTCGCAAA TTGAATTAGT TCCGTTTGCT GCTCTTCACC 1140

GATAGGTACA GTCCGAATAT TATCACCACT CATACCAGGA TTACCTGGTG CACAATAAAC 1200

AGTCTTGACT TTGTCTTCTT GCTGTAATTT TTTGGCGATT GCATGTTCAC GACCACCACG 1260

CCCAATCACT AATATATTCA TCGGTTTCTC CTTTAAAAAT TAATGTTTGA AATGACGAGT 1320

CCCAGTAAAA AtCATCGCAA TACCATAATG ATTCGCTAGA TCGATTGATG CTTGGTCTTT 1380

AATGCTCCCG CCTGGCTGAA TAATTGCTTG AATGCCaTGT TGTGCAGCAT ATTCCACGCT 1440

ATCTGCCATG GGGAAAAAGG CATCACTTGC TAAGACGGCC CCTTGTAATT TTCCAGCTGC 1500

TTGGGCTTGA TCAACTGCAA TTTTTACTGA ACCAATTCGA TTCATTTGAC CAGCGCCAAT 1560

CCCCACCGTT TGGCTTTCAT TGGCTAAAAC AATAGCGTTT GATTTTACAT GCTTGACTGC 1620

TTTCCAAGCA AAGTTCAATG CGTCTCTTTC GGCTGGTGTG GGTTGGCGCT CTGTAACAAC 1680

TTGCCATGCT TGGTCATTTT CAACAACTGT ATCCTGTTCT TGAATTAATA ACCCGCCCAT 1740

AACAGAAACT AATTCCACCG CTTTTTTTTC TTGAGCATGA AAATCTAACG TTAACAAACG 1800

AAGATTCTTT TTCTGAGATA ACAGTGCCAA AGCTTCTTCC TCAAAACTAG GCGCAATAAT 1860

TATCTCTAAA AAGATTTGAT GCATTTCAGC CGCAGTCGCT TGGTCCACTG GACGATTCAA 1920

AACAACAATG CCACCAAAAA TTGAAACTGG ATCTGCTTCA AATGCCTGAC GATAGGCTGC 1980

CAAAATCGTT TTTCCTGTTC CAATTCCACA TGGGTTCATA TGTTTAACAG CGACGACCGT 2040

AGGTTCTGTA AACTCGCTAG CAATCCGGAG TGCTGCATCA GCATCTCGAA TATTGTTATA 2100

TGACAGTTCT TTGCCATGAA GTTGTCGAGC GGAAGCAATG GATAATGATA CAGGTACAAC 2160

GGATTGATAA AAGGTGGCTT GTTGATGGCT ATTTTCGCCA TAACGCAATG TTTGCTTACG 2220

TTCATAAGTA AGCGTCTGTT TCTCTGGTTC AGTTTCACCC ACTTGCGTCG TTAAATAGTC 2280

AGCGATAAGC GCGTCATAAG CAGCGGTATG ACGAAAGACT TTTGCTGCTA ATCTTTTTCT 2340

AGTGGCTAAC TGGCTAGTTC CAGTGCTGCT AATTTCACTT AATACCTGTT CATAATCACT 2400

AGGATCAACC ACTGCTGTTA CGTCTTGATA ATTTTTTGCC GCTGAACGTA ACATGCTAGG 2460

ACCGCCAATA TCAATCATTT CAATCGCTTC TGCTTCAGAG ATCGCTGGTT TTAAAATTGT 2520

TTCTTTAAAT GGATAAAGAT TCACTACTAC AAGGTCaATA AAGTGAATGC CTTGTTCAGT 2580

CAGTGCTTTT TGATGCGCTT GGTTTCCCCT TTTTGCCAAA AGTCCGCCAT GGATATTCGG 2640

ATGTAACGTT TTGACACGGC CATCTAACAT TTCTGGAAAG CCAGTGATGT CGTCAATAGC 2700

AAGTGTCGGT ACCCCCGCTT CTTCCAAGAC AGTCCTTGTT CCCCCAGTGG AGATAATTTC 2760

AAAGCCATTT GCAACTAAGC CTGCTGCAAA TGTTGTCACT CCTGTTTTGT CTGAAACGCT 2820

AATCAATGCT CTTTTGGTCA TTTCATTGTC TCCTTTACAA TTTGTGAAAT TATTTTTGGG 2880

TACCATTCGT GCTCAAGAGC ATGAATTTTC TCTGCTAATG TGTCCACTGT ATCTTCGGTA 2940

TCAATCTTGG TGGTTGTTTG GAAAATAATC GGTCCTGTAT CAACCCCAGA GTCAACaTAA 3000

TGGATAGTGA TTCCTGTTAT TTTCACTCCA TAATGAAATG CTTCTTCAAT TCCgTGAAGA 3060

CCTGGAAAAC TAGGTAACAG AGAAGGATGA ATATTGACGA TTCTCTTTGG ATAAGCCTCC 3120

AGTAAAGTCT TGCCGATAAT TCTCAAATAA CCAGCTAACA CGATTAAATC GATCTGATGC 3180

TCTTTTAAAT GTTTCAATAC TTGTTCCTCA TATTGCCCAC GAGATGAAAA ATCAGATGGC 3240

GAAAAGCAAA CAACGGGAAT CTTGCGTTTT TGGGCACGGG TCAATACATA GGCTTCAGGC 3300

TGGTCACAAA AGACTAAGCT CAAATGCCCT GCGATTTTTT TTTGGGAAAA AGCTGCAGCA 3360

ATCGCCTCAA AATTGCTACC ATTGCCAGAT GCAAAAACAG CTATCTTCAT AAGGCTTCCT 3420

TAAAAATAAC GGCTTTCGTT TTCCTAGGTA CCAAACGACC AATTTCAAAA GCTTCTGGCA 3480

AAAGCTCCTG AACTTTTTCA ACATGCTCTG GACTCACCGC TAAAATCATC CCTAAGCCCA 3540

TATTGAAAAT TTCATACATT TCTTCAGGTG GAATTTCACC GTATTTTTGA ATCAATTTAA 3600

AAATTGGTAG CACTGGCCAA CTATTTAATT GAATCTCTGC AGCTAATGCG GAAGAAAACA 3660

TTCTTGGTAG ATTTTCTAAA AACCCGCCGC CAGTAATATG CGCAGCCCCA TGCACGAGTC 3720

CTGCTTTTAA CAATGGCAAT AATTCCTTCA CATAAATTTT TGTTGGTGTT AATAATTCTT 3780

CCCCTAACGT TTTGGGTGCC AATTCTGGAA GCTTGTCCGT TGTTTTAAAC GAATGTTTCT 3840

TAAAAAAAAT ATTTCTTACT AAAGAGTAAC CATTAGAATG AAGACCACTT GAAGGAAGAC 3900

CAATCAAGAA ATCGCCTTCT TTTACGTTAC TTGGAGTCAA GAGTTGACTT TTCTCAGCAA 3960

TACCCACGGC AAAACCAGCC ACGTCATAAG CATCTGCTTC ATACATATCA GGCATTTCAG 4020

CTGTTTCGCC CCCAATTAAG GCTGCATTCG CCTCACAACA GCCAGCCGCA ACACCCGCAA 4080

CAATTGCTTC AACTTTAGCT GGATTAACGG TCCCTAACGC TAAATAATCT AAAAAATATA 4140

ACGGTTCGGC GCCTTGTGCG ACGACATCAT TGACACACAT TGCTACACAA TCAATTCCGA 4200

TTGTCTCGTG TTTCTGTTCT TCAATTGCCA AGAGTAACTT TGTGCCTACG CCATCCGTTC 4260

CTGATACCAA AACTGGTTCT TTAAGCTTGT AACTACTTAA ATCAAAACAA CCACCAAAAC 4320

CACCAAGCAT GCCCAACGTT CCCGTCCGTT GTGTTTTTTG GCTATGCTTT TGAATTCGTT 4380

CTACTACTTC GTATCCCGCC TCAACATTTA CACCAGCTTT TGAATAGGCA TTTTCCATGT 4440

CAATTCCTCC TAGTTTCTTT TTTCAAGGAT GCTTGATATT GTTCTTCATA ATCATACAAA 4500

GGCGTTGGAT AGTCTCCATT AAAGTAAGCC ATACACAGCC CAGAATAAGG AGCATCATAA 4560

TTTAATTGAA TCCCATCAAT CAGCCCTGCT TCACTCAAAA AAGCCAGCGA ATCTGCGCCA 4620

ATCTGCTCTT TGATTTCTTC AATTGAATAG TTTGCCGCAA TTAGTTCGTC CCGCGTCTGT 4680

ATATCAATGC CATAAAAACA CGGATATTTT AATGGTGGTG AAGCAATTCT GACATGAACT 4740

TCTTGCGCTT CTGCTTCTTT TAATAAATGA ATAATGCGCC GAATAGTGGT CCCGCGAACA 4800

ATCGAATCAT CTACTAAAAT GACTTTTTTT CCTTTGACCA CGCCACGAAC TGCCGATAGT 4860

TTCATACGGA CACCTTGTTC TCTTAATTCT TGTGTCGGCT GGATGAATGT TCGAGCGACA 4920

TATTGATTTT TTACTAAACC TAGTTCGTAA GGTATCTGTG CTTCTTCTGC ATAGCCGCTA 4980

GCTGCAGACA ATGACGAGTT CGGTACACCA ATTACCATAT CAGCTGCGAC TGGACTTTCG 5040

ATTGCCAAGT TCTTCCCCAT TCTTTTACGA GCTGTATGCA CATTAACTCC GGCAATATTT 5100

GAATCTGGTC GTGCAAAATA AATATATTCC ATTGCACAAA TTGCCGGTTG GACTTCCGTA 5160

GTGAAAACCT CTATTTTCAA GCCTTCATCA GAAATGATGA CTACTTCACC AGGCGCCACA 5220

TCTTGAATAA AGCGAGCCCC AACTGTTTCT AATGCACACG TTTCGGATGC CACAACGTAC 5280

GCCCCATTGT TCATTTGTCC GATGGAGAGA GGGCGAAAGC CGTTGGGATC CAGTGCAGCA 5340

ATCATCATCG TTTCTGTTAA TAAAAGATAA GCAAAGCCAC CCTTAACCAA ATTCAAGCTC 5400

TCTTTTAATT TATCGAGAAA CGAGGCAGCA GTGCTTTTTC GAATCAAGTG CATTAAAATT 5460

TCAGTGTCTG AATTGGAATG AAAAATAGCG CCTTCTCTTT CTAGTTGTTT TCGTAAACGT 5520

TTGGCATTTG TCAAATTGCC ATTGTGTGCT AaCCCAaTtG ACTATCATAA AATTTAAATA 5580

AAAATGGCTG AATATTATCC ACGCTACCCG TACCAGCGGT TGCATAGCGT ACATGACCAA 5640

TAGCCGCCGT TCCTTCTAAC GCATTCAGTT GTTCTTGCTT TTGGAATACT TCTGCTAATA 5700

AACCTAAATC ACGGTGGCCA TCTAATTTCC CAGAATTGTT CGCCACAATT CCCGCGCCTT 5760

CTTGGCCTCG ATGCTGCAAG CTGTGTAAGC CAAAATAGGT CACTTGTGCC GCATCAGGAT 5820

GACCCCAAAT ACCGAAAATA CCACACTCTT CATTTAAGCT TTTTGCTTCA TAAGACATGG 5880

AATCGCATCC TCCCATAAGG CTTTGGCTTC TTTTGTTGAG CAATTGATCA CATCATCTAA 5940

CGCATGAATG ACCAAACCAG TTTCAGTCAC TTTCCCAATA TGTTGAGCTT TTCTTCCGAT 6000

AAGTGTTTCA AATGCTTCTT GATGCTGGGG ACTAACTGAA AGAATAAAAC GGGATTGCGT 6060

CTCCGCAAAT AGATATTCTT TTTTTAATGG CAGTGTCACC TGTAACCCCA GTTCGTTAGC 6120

AAAAGCCGAC TCGGCTAAAG CAACGGCAAC GCCACCTTCA GCGCAATCAT GCGCGCTAGC 6180

AACTAGGCCT ACTTGAATTG CTTTTAAAAC TAATTCTTGG TTTGCTTTTT CCTCTTTTAA 6240

ATCAAATGAG CGTAATTGCC CTTCAATTCG ACCCAGTTGT ATTTTTTGAA TTTCTGAGCC 6300

GTTAAAATCT GCGTGGGTTT CACCGATGAG ATAAATTAAG TCATCAACTT GTTTAAACGC 6360

TTGTGTGGTA ATTTGTGAAA CATCTTCAAT CACTCCCACC ATGCCAATCA TTGGCGTTGG 6420

ATAAATGGCT TGACCATCTG TTTCATTGTA AAGAGAGACG TTCCCTGAAA TCACGGGTGT 6480

TCCTAGCTGA CGGCAGGCGG CGGCGATTCC ATCCGCCGAC GTCCATAATT CCCAAAAACT 6540

TTCTGGTTTA TCTGGGGAAC CATAATTCAA GCAATCAGTG ATAGCCAAAG GTTGTGCTCC 6600

ACTCGCTACA ATATTACGGG CTGCTTCTGC AACCGCAATT TGCCCACCAA TTTCAGGATT 6660

TAAATACAAA TAACGGCTAT TACAATCTGT CGTCATCGCT AAGGCTTTAT TGGTGCCCCG 6720

CACTCTTAAA ACAGCTGCAT CACTTCCTGG TTGCACCACT GTATTCGTTT GTACTTGACT 6780

ATCATAGGTT TCGTAAACTG ATTTTTTCGA TGCAATCGTT GGCTGTTGCA ATAGGCGTTT 6840

TAATAATTCC GTACTATTCG TAATTTCTGG AATGAAGGGC GCTAATTTTT TAAAAGCACG 6900

CATCCGCTCT GGTTCTTGAT AGGCTTTCTG GTAGACAGGG GCTTCTTCTG CCAAAGCATC 6960

TACTGGTACA TCTGCGACTA TTTTGCCACC GTGAAAGACA CGATACTGTC CATCATCTGT 7020

TACGGAGCCA ATCGCTACCG CTTCTAAGTC ATACGTTTGA AACAATTCAC AAACTTGCTG 7080

TTCTGCTCCT TTTTCCACAC AGATTAACAT CCGTTCTTGG GACTCAGACA AGAGCATTTC 7140

ATAGGGGCTC ATCTGCGTCT CTCTCTGCGG AACTTTATCT AGCTCTAGAA TCAAGCCTGA 7200

AcCTGCTTTT GATGCCATTT CAGCACTTGA AGAAACCAAG CCAGCCGCCC CCATATCTTG 7260

GATGCCGATT AAAATATCTT GATGATTTTT GATTAAATCA AGACAAGCTT CCAATAATAA 7320

TTTTTCCATA AACGGATCAC CTACTTGTAC CGCAGATCGT TGCTGTTCTT CACCTTCACT 7380

AAATTCTTCT GAAGCAAAGG TCGCACCATG AATTCCATCT CGTCCTGTTT TCGCACCTAC 7440

GTAAAGAATG CTATTTCCGA CGCCTTTTGC TTGGCCTTTT TGAATCTCAT CATGACGAAT 7500

CAATCCCACA CACATAGCAT TAACAAGTGG GTTTCCTTGA TAACATGGCT CAAAAGCAAT 7560

TTCGCCGCCA ACGGTTGGAA TACCAATACA ATTTCCATAA CCGCTAATCC CCGCCACAAC 7620

TTCTTGAAAC AAATAACGTG TTCGTGCATC CGTCAGTTCG CCAAATCTTA GCGAATCCAG 7680

TAAAGCAATC GGCCGCGCCC CCATACTGAA AATATCGCGA ATAATTCCAC CAACGCCTGT 7740

AGCTGCTCCT TCGTAAGGTT CTACTGCAGA AGGATGATTA TGGCTCTCTG CTTTAAAAAC 7800

AACAGCTAAT CCGTCACCAA TATCGACAAT CCCAGCTCCT TCACCAGGGC CTTGAAGCAC 7860

TTGTGGCCCT TCCGTCGGAA ACTTTTTTAA AACGGGCTTG GAATTTTTGT AAGAGCAATG 7920

TTCACTCCAC ATCACGGAGA ATAAGCCAGT TTCTGTATAA TTAGGTtCTC TTtGCAAAAT 7980

GGTCTCAATC TGATGATATT CTTCATCTGT TAGTCCCCAT TCTGCATACA ACCGTGTTTC 8040

CTTGATTTCT TTTGGTGTTG GTTCATTCAT CATTGGCTTA ATTCCTTTCC AAAGTTTTTA 8100

ATGATTGAAG CAAAAAAGCT GCGTCCATCT GTTGAACCTA ACAATTCTTC CATCGCTCGT 8160

TCCGGATGAG GCATCATTCC TAGAACATTG CCCTGTTCAT TACAAATACC AGCAATATTC 8220

GCAAGACTAC CGTTCACCTT TTCTTGATAC GTAAAAACAA TTTGGTTCTT TTCTTTCAGC 8280

TCCTTCAACG TTTTTTCATC ACAATAGTAA TTCCCTTCGC CATGCGCTAC TGGTAAATGT 8340

ATCGAAACAT TTTCAGGATA CTCGCTTGTA AACTTTGTGT GCGTATTCAC TTTCAATGGG 8400

ACACTTTTAC AAATAAAACG AAGCGAATCA TTTCTTAACA ACGTGCCAGG AAGTAAGCCA 8460

ATTTCTGTTA AAATTTGAAA GCCATTACAA GTACCAAAAA CCATTTTGCC TTCTGTAGCA 8520

AAGCGAATCA CTTCATTGAT AATTGGCGAA AACCGAGCAA TTGCACCACA ACGTAAGTAA 8580

TCGCCATAAG AAAAACCACC AGGCAACAAT ACGCCATCAA ATTTTGCTAA AGAAGTGGCA 8640

TCGTGTCGGA CAAATTCTGC ATCGATGCCC ATCACyTCTT TGACAGCCCA CAATAAGTCT 8700

GCATCACAAT TGGATCCTGG AAAAACAATC ACTGCCAATT TCATTTAAAT TTCCTCCAAA 8760

GATAAAATTT CAAAACGATA GGATTCCATA TTTACATTGG CTAATAGTTG ATCACAAATT 8820

TGTTCAATTT TCTCTTCAAT TGCTTCTGTC TGCTGGACGT GCATTTCAAA ATACTTACCG 8880

ACACGGATTT CGGCTACTTC TGTAAAACCA AGCCGATGGA CTGCTTCCTT AACCGCTTCA 8940

CCTTGGGGAT CCAAAATTGA TTCTTTATAC GTAACATAGA CTTTGACAAA ATACaTACGC 9000

GCTTCCTCCT CAATTAGCTG TTAGTAAACG TTGTAAAACT TCTTCATATA CAGGAACTAT 9060

TTCTCCTAAT TCACGGCGAT AAACATCTTT ATCCAGATGT TCGTTTGTGT TTAAATCCCA 9120

TAGACGACAG GTGTCTGGAG AAATTTCATC TGCCAGTAAC AATTGATTGG CCTTCGTCCG 9180

ACCAATCTCG ATTTTAAAAT CGATCAAACA AATATTGAGG CGTTGAAAGA GTTGACTAAG 9240

CGCCTGATTA ATTTGTAAAG CCAATGCTTT AATCTCCACA ATTTCAGCAG GAGTGGCTAC 9300

TTTTAGAAAT TGAATATGGT CTTCGTTAAT AAAAGGATCA TCCAGACGAT CTTCTTTGTA 9360

ATAAAATTCA ATAATGGGCG TCACTAATTT AGTGCCCTCT TCAATCGCTA AACGCTTTGA 9420

AAAACTTCCT GCCGCATAAT TACGTACGAC AACTTCCAAA GGAATCaTTT CTACGACCTG 9480

AATTAATTGT TCGTGTTCAG ATACTTTTTT GATAAAATGA TTGGGAATCT TCTGTTGCTG 9540

TAAATGTTCA AAAATTAGAC TAGTGATTTG ATTATTTAAA GCGCCTTTCC CAAGAACTTT 9600

GTCTTTTTTT TGACCGTTTA AAGCTGTTGC TTGATCAAGA TATTCCACAA AAAGAACCGC 9660

AGCATTTTCT GTTTGATACA ACTTCTTCGC TTTTCCTGTA TACACAAGTG CTTTCTTTTC 9720

CATTCtTTCC CTCTTCTCGT TAGCTTCATG AACCCTATTC TTTCTAACTc AGCTGATTCA 9780

CATCaGACaA AAACTTACAT TAGTTACAAA AAAACaATAA TCGTTCGTGT TTGTCGAATG 9840

TAAAAATAAA AACTTGCTT GGAGCAAGCA TTTATTAAGT TACGTTCCAA ATACCAACAG 9900

TCTCAATTTC CTTTAACGTT TCTGATA 9927

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7678 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TGTCTGCmrT rCAGTAAGGG CGTTTTTGGC TGAAGCAAGA TATGAGGTGG ATGCCTGCAC 60

AGATGGAAAC GAAGCACACA CCAAGTTCTA TGAAAACACC TATCAACTGG TTATTCTTGA 120

TATTATGCTG CCCGGTATGA ATGGGCATGA ACTTCTACGT GAATTTCGGG CGCAAAATGA 180

TACCCCCATT CTGATGATGA CAGCCCTGTC GGATGACGAA AACCAAATCC GGGCGTTTGA 240

TGCAGAGGCA GACGACTATG TAACAAAGCC ATTCAAGATG CGGATTTTAC TAAAGCGGGT 300

GGAAGCCCTG TTACGGCGCA CGGTGCGCTG GCAAAGGAAT TTCGTGTGGG CAGGCTGACA 360

CTTCTGCCGG AGGATTTTAG GGTACTTTGT GACGGTACGG AGCTGCCCCT GACACGAAAA 420

GAATTTGAAA TCCTTTTGCT GCTGGTGCAG AACAAAgGCA GAACCTTAAC CCATGAAATC 480

ATTTTGTCCC GCATATGGGG ATATGACTTT GACGGTGATG GCAGCACAGT CCACACTCAT 540

ATCAAAAATC TGCGGGCGAA CTGCCGGAAA ATATCATCAA AACCATCCGC GGTGTAGGTT 600

ACCGATTGGA GGAATCATTA TAATGGAAAG AAAAGGGATT TTCATTAAGG TTTTTTCCTA 660

TACGATCATT GTCCTGTTAC TGCTTGTCGG TGTAACGGCA ACACTGTTTG CACAGCAATT 720

TGTGTCTTAT TTCAGAGCGA TGGAAGCACA GCAAACAGTA AAATCCTATC AGCCATTGGT 780

GGAACTGATT CAGAATAGCG ATAGGCTTGA TATGCAAGAG GTGGCAGGGC TGTTTCACTA 840

CAATAACCAA TCCTTTGAGT TTTATATTGA AGATAAAGAG GGAAGCGTAC TCTATGCCAC 900

ACCGAATGCC GATACATCAA ATAGTGTTAG GCCCGACTTT CTTTATGTGG TACATAGAGA 960

TGATAATATT TCGATTGTTG CTCAAAGCAA GGCAGGTGTG GGATTGCTTT ATCAAGGGCT 1020

GACAATTCGG GGAATTGTTA TGATTGCGAT AATGGTTGTA TTCAGCCTTT TATGCGCGTA 1080

TATCTTTGCG CGGCAAATGA CAACGCCGAT CAAAGCCTTA GCGGACAGTG CGAATAAAAT 1140

GGCAAACCTG AAAGAAGTAC CGCCGCCGCT GGAGCGAAAG GATGAGCTTG GCGCACTGGC 1200

TCACGACATG CATTCCATGT ATATCAGGCT GAAAGAAACC ATCGCAAGGC TGGAGGATGA 1260

AATCGCAAGG GAACATGAGT TGGAGGAAAC ACAGCGATAT TTCTTTGCGG CAgCCTcTCA 1320

TGAGTTAAAA ACGCCCATCG CGGCTGTAAg CGTTcTGTTG GAGGGAATGC TTGAAAATAT 1380

CGGTGACTAC AAAGACCATT CTAAGTATCT GCGCGAATGC ATcAAAATGA TGGACAGGCA 1440

GGGCAAAACC ATTTCCGAAA TACTGGAGCT TGTCAGCCTG AACGATGGGA GAATCGTACC 1500

CATAGCCGAA CCGCTGGACA TAGGGCGCAC GGTTGCCGAG CTGCTACCCG ATTTTCAAAC 1560

CTTGGCAGAG GCAAACAACC AGCGGTTCGT CACAGATATT CCAGCCGGAC AAATTGTCCT 1620

GTCCGATCCG AAGCTGATCC AAAAGGCGCT ATCCAATGTC ATATTGAATG CGGTTCAGAA 1680

CACGCCCCAG GGAGGTGAGG TACGGATATG GAGTGAGCCT GGGGCTGAAA AATACCGTCT 1740

TTCCGTTTTG AACATGGGCG TTCACATTGA TGATACTGCA CTTTCAAAGC TGTTCATCCC 1800

ATTCTATCGC ATTGATCAGG CGCGAAgcAG caAAAAGTGG GCGAAgcGGT TTGGGGCTTG 1860

CCATCGTACA AAAAACGCTG GATGCCATGA GCCTCCAATA TGCGCTGGAA AACACCTCAG 1920

ATGGCGTTTT GTTCTGGCTG GATTTACCGC CCACATCAAC ACTATAAATA TTTAAAACTT 1980

AAATGATTTT GACCGACAGG TATAACCCTG CCGGTCTTTT TGTTTTTCGC CGCTACAGGA 2040

AAACTACAGA TTGACTACAG GGAAAGTACA GATACGCTTG CCATAATAAC AATCGTACCA 2100

GCCACAAATC GTAGTTTTAT TGCAAAGGAG GCATTCAATC AAATGGAAAA AAGCAACTAT 2160

CATTCCAATG TGAATCATCA CAAACGGCAT ATGAAACAAT CTGGGGAAAA ACGGGCTTTT 2220

CTATGGGCGT TCATTATCTC GTTCACAGTC TGCACGCTGT TTTTGGGGTG GAGATTGGTT 2280

TCCGTATTGG AGGCAACACA GCTACCGCCC ATCCCTGCAA CTCATACAGG CAGCGGGACT 2340

GGTGTAGCGG AGAATCCAGA GGAAAACACT CTTGCCACCG CCAAAGAACA GGGAGATGAA 2400

CAGGAATGGA GCCTGATTTT AGTGAACAGG CAGAACCCCA TCCCCGCCCA GTACGATGTG 2460

GAACTTGAGC AGCTGTCAAA TGGTGAGCGG ATAGACATTC GGATTTCTCC CTACCTCCAG 2520

GATTTGTTTG ATGCCGCAAG AGCTGATGGA GTTTACCCGA TTGTCGCATC CGGATACCGG 2580

ACAACAGAAA AACAGCAAGA AATCATGGAT GAAAAAGTCG CCGAATACAA GGCGAAAGGC 2640

TACACCTCTG CACAGGCTAA AGCGGAAGCA GAAACTTGGG TGGCCGTGCC GGGAACAAGC 2700

GAGCATCAGC TTGGTCTTGC TGTGGATATC AATGCGGATG GAATTCATTC AACCGGCAAC 2760

GAGGTTTACA GATGGCTGGA TGAAAACAgC TATCGCTTTG GTTTTATTCG CCGCTACCCG 2820

CCAGACAAGA CAGAGATAAC CGGTGTGAGC AACGAGCCGT GGCATTACCG ATATGtCGgC 2880

ATCGAAGCTG CCACAAAGAT ATACCACCAA GGGCTTTGCC TTGAGGAATA TTTAAACACA 2940

GAAAAATGAG AAAAGGATAT AATGCTATGA ACAGAAAAAG ATTGACACAG CGCTTCCCGT 3000

TCCTGCTTCC AATGAGACAA GCGCAGAGAA AAATATGCTT TTATGCGGGA ATGAGATTTG 3060

ACGGCTGTTG CTATGCACAG ACGATAGGAG AAAAAACGCT TCCCTATTTG CTCTTTGAAA 3120

CGGATTGTGC GTTATACAAC CACAATACCG GATTTGACAT GATATACCAA GAAAACAAGG 3180

TGTTCAACTT AAAGCTGGCG GCAAAGACCT TAAACGGCCT ATTGATAAAA CCGGGGGAAA 3240

CCTTTTCTTT CTGGCGGCTG GTACGCCATG CGGACAAAGA TACCCCCTAT AAAGACGGCC 3300

TTACGGTGGC CAATGGTAAG CTCACCACCA TGTCGGGCGG CGGTATGTGC CAGATGAGCA 3360

ATTTACTATT TTGGGTGTTC CTGCATACGC CATTGACAAT TATCCAGCGC AgcGGTcACG 3420

TAGTAAAGGA GTTTCCAGAG CCAAACAGTG ACGAGATCAA AGGGGTGGAT GCAACCATCT 3480

CAGAGGGCTG GATTGATTTA AAAGTGCGAA ACGATACCGA CTGCACCTAC CAAATATGGG 3540

TGACCCTAGA TGATGAGAAA ATCATCGGTC AGGTGTTCGC CGACAAACAG CCTCAAGCAT 3600

TATACAAAAT TGCAAACGGC AGTATTCAGT ATGTCCGTGA AAGTGGCGGG ATTTATGAAT 3660

ATGCCAAGGT TGAACGGATG CAAGTTGCCT TAGGTACCGG GGAAATAATA GATTGCAAGC 3720

TGCTTTATAC AAACAAATGC AAAATCTGCT ATCCCCTCCC GGAAAGTGTG GATATTCAGG 3780

AGGCGAACCA ATGAGAAAAA GTATGGGCAT TACTGTTTTT GGATGCGAGC AGGATGAGGC 3840

AAATGCTTTC CGCACCTTAT CACCAGATTT TCATATTATC CCTACGCTGA TCAGTGATGC 3900

GATATCGGCA GACAACGCAA AATTGGCCGC TGGCAATCAA TGCATTAGCG TAGGCCATAA 3960

GTCCGAGGTT TCCGAGGCGA CAATTCTTGC GCTGAGAAAG GTCGGGGTAA AATACATTTC 4020

TACCCGCAGC ATCGGCTGCA ATCACATTGA TACGACTGCC GCCGAGAGAA TGGGGATCTC 4080

GGTTGGCACA GTTGCGTATT CGCCGGACAG CGTTGCGGAT TATGCTTTGA TGCTGATGCT 4140

GATGGCCATA CGGGGTGCAA AGTCCACCAT ACACGCCGTG GCGCAACAAA ATTTCAGACT 4200

GGATTGTGTC CGGGGGAAAG AGCTGCGGGA TATGACTGTG GGAGTTATTG GAACCGGCCA 4260

TATAGGGCAA GCGGTCGTCA AAAGGCTGCG GGGATTTGGA TGCCGTGTGC TAGCCTATGA 4320

TAACAGCCGA AAAATTGAGG CAGATTATGT CCAGCTTGAT GAGCTTCTAA AAAACAGCGA 4380

TATTGTTACG CTCCATGTGC CGCTTTGTGC GGATACCCGC CATCTGATCG GCCAGAGGCA 4440

AATCGGAGAG ATGAAGCAAG GCGCATTTTT AATCAACACT GGGCGCGGGG CGCTTGTCGA 4500

TACCGGGTCG CTGGTGGAGG CACTGGGAAG CGGAAAGCTG GGCGGTGCGG CACTGGATGT 4560

GTTGGAGGGC GAGGATCAGT TTGTTTATAC CGACTGCTCG CAGAAAGTGC TTGACCATCC 4620

CTTTTTGTCG CAGCTCCTAA GGATGCCAAA TGTGATCATC ACACCCCATA CGGCGTACTA 4680

CACCGAGCGT GTGCTGCGAG ATACCACAGA AAAAACAATC AGGAATTGTC TTAACTTTGA 4740

AAGGAGTTTA CAGCATGAAT AAAATAAAAG TCGCAATTAT CTTCGGCGGT TGCTCGGAGG 4800

AACATGATGT GTCGGTAAAA TCCGCAATAG AAATTGCTGC GAACATTAAT ACTGAAAAAT 4860

TCGATCCGCA CTACATCGGA ATTACAAAAA ACGGCGTATG GAAGCTATGC AAGAAGCCAT 4920

GTACGGAATG GGAAGCCGAT AGTCTCCCCG CCATATTCTC CCCGGATAGG AAAACGCATG 4980

GTCTGCTTGT CATGAAAGAA AGAGAATACG AAACTCGGCG TATTGACGTG GCTTTCCCGG 5040

TTTTGCATGG CAAATGCGGG GAGGATGGTG CGATACAGGG TCTGTTTGAA TTGTCTGGTA 5100

TCCCCTATGT AGGCTGCGAT ATTCAAAGCT CCGCAGCTTG CATGGACAAA TCACTGGCCT 5160

ACATTCTTAC AAAAAATGCG GGCATCGCCG TCCCCGAATT TCAAATGATT GAAAAAGGTG 5220

ACAAACCGGA GGCGAGGACG CTTACCTACC CTGTCTTTGT GAAGCCGGCA CGgTCAGGTT 5280

CGTCCTTTGG CGTAACCAAA GTAAACAGTA CGGAAGAACT AAACGCTGCG ATAGAAGCAG 5340

CAGGACAATA TGATGGAAAA ATCTTAATTG AGCAAGCGAT TTCGGGCTGT GAGGTcGGCT 5400

GCGCGGTCAT GGGAAACGAG GATGATTTGA TTGTCGGCGA AGTGGATCAA ATCCGGTTGA 5460

GCCACGGTAT CTTCCGCATC CATCAGGAAA ACGAGCCGGA AAAAGGCTCA GAGAATGCGA 5520

TGATTATCGT TCCAGCAGAC ATTCCGGTCG AGGAACGAAA TCGGGTGCAA GAAACGGCAA 5580

AGAAAGTATA TCGGGTGCTT GGATGCAGAG GGCTTGCTCG TGTTGATCTT TTTTTGCAGG 5640

AGGATGGCGG CATCGTTCTA AACGAGGTCA ATACCCTGCC CGGTTTTACA TCGTACAGCC 5700

GCTATCCACG CATGGCGGCT GCCGCAGGAA TCACGCTTCC CGCACTAATT GACAGCCTGA 5760

TTACATTGGC GATAGAGAGG TGACCCGTAT GGAAAATGGT TTTTTGTTTT TAGATGAAAT 5820

GTTGCATGGT GTTCGTTGGG ATGCCAAGTA CGCTACATGG GATAACTTCA CGGGAAAACC 5880

AGTGGATGGG TATGAGGTGA ATCGCATCAT CGGCACAAAG GCCGTGGCGC TTGCTCTGCG 5940

CGAAgCACAA ATCCATGCGG CAGCCCTTGG CTACGGCTTG CTTTTATGGG ATGGATATCG 6000

GCCAAAATCT GCGGTGGACT GTTTCCTGCG TTGGGCGGCG CAgcCGGAgG ACAACCTCAC 6060

AAAAGAAAAA TATTACCCCA ATATTGAGCG AGCCGAGTTG ATTACAAAGG GCTATGTGGC 6120

CTCACAATCC AGCCATAGCC GTGGAAGCAC AATTGATCTT ACGCTCTACC ACTTGGATAC 6180

AGGGGAACTT GTTTCAATGG GAAGCAACTT CGATTTTATG GACGAACGGT CGCACCATAC 6240

AGCAAAAGGG ATAGGGAATG CAGAGGCACA AAATCGAAGA TGCTTGCGTA AAATCATGGA 6300

AAGCAGCGGA TTTCAGTCCT ATCGCTTTGA ATGGTGGCAC TATAAGTTGA TTGATGAGCC 6360

ATACCCCGAT ACCTATTTTA ATTTTGCTGT TTCATAATGA AAGTATTTGA TTTTCTAATT 6420

ATGTATAAGT TGGCTACAAA TTACTTAGTA TTTCATCAGA CCAATTACTC TCTTGTTTAC 6480

AGAAAAATTC TGCGCTGATG GAATCTGCTT TATTATGCGG GCGAAAAATG AAATTGACCA 6540

TATTTTTTCa GAACTTTACT CTGTACCGaA TTGCCTGCAA AAGCCTTATT TTAAGCTGAA 6600

AGTTCAGGAA TTGCTTTTGT TTTTGTGTAT GCCCCTCGTG ATTTGTACAC CTATCTTAAT 6660

TGGCTTTGCA ATTCTCATTC CGTATCTCTG CTTTAAGAAT TTGGaAAAAc GAAGCATTGT 6720

GAATCGGCTG CGGGCAGAGC AAAAAGAGAA CCAGCAGAAA CAAGTCGTTC TTGCTCTGCT 6780

GATTCACTCG GAACTGTTTG ATTCGGGTTT TCGTTGAAGG TCAAGTAGCT GCTCTGTCAG 6840

GAAGTCCAGT GTGTTCAGCA GAATCTGCTG ATTGTCACGG TTGCATGACT GAAATTTTCC 6900

CATGAAACGC TGGAGTTCTT CATCCTCAAT AGAGTTTGAA GCTTGAAACA GGACATTCGG 6960

AGAAATTCCC AGTCGTTCAA TTAACCGGGA AAGTGTTTCA TAAGACGGAT TTTTTCTTCC 7020

CTTTTCGATG TCCTGAACCG TTTTGACCGC CAAGCCAGAT TGATCCGcCA GTTCCTGTTG 7080

GGTCAAGTTA CATCGAACCC TTTCGGCTCG GATGTGCCTA CCCAGAAATT TCAGGCTGTC 7140

TTCCGGCATG ATtGTTCACC TCATATATAT TCTATCGTAC CAATTTGTAT TTTGGTACTC 7200

CCTTAGAGTT GTGTTCCAAT CAGTATTATC ATGCCGGTTT TTGAGTTTAA TATGCAATAA 7260

AGCCCTGTTG GAAGCAATAA AAAAAACAGC TTCTCACGGC aGGGCCTAAT TGCCaCGCCC 7320

CCGTGGTTTG TCTGTAAAAT CATGGGGGTA TTTTATGTTC AAAAGAATAA TGGGAGGTAA 7380

TGCGTTAAAA AAAGTGTAGC TCCGCCGTGG GCATTATCCG CCCCGCTAAG CGCACTATCG 7440

TTCCTATACA ATACATATCA TCATTCCTAC TACCTTATTC GGTCGCATTG CGCCAGGAAG 7500

CCTAAGAAGC TTCCTGGCCy TTTTGTTTGT CGAATTTTGT TGAATGATAA CGAGAATTGT 7560

CGATGTCGCA GCCCGCCTTC ATCGGTTCGC AGCCCATGAA CTGATGAAAG GAGTAAACGG 7620

CTATGACCAA AGAATATTGG GTAGCGGTCT TAGAACAAGA AGATCGTnTG ATAAGTAA 7678

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4732 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

ATTTATTATA AGGGAGCGAn ATTnCATGGC ACGCGTAGAA AGTTtTGAAT TaGATCACAA 60

CACAGTAAAA GCACCATATG TTCGCCTTGC TGGCACAGAA CAAAATGGTG ATGCGTTAGT 120

CGAAAAATAT GACTTACGTT TCTTACAACC AAACAAAGAT GCCCTACCAA CAGGCGCATT 180

ACACACGTTG GAACATTTAT TAGCAGTTAA CATGCGTGAT GAATTAAAAG GAATCATTGA 240

CATTTCGCCA ATGGGTTGCC GCACTGGTTT TTATATGATT ATGTGGGATC AACATTCACC 300

ACAAGAAATC CGTGATGCAT TAGTCAACGT TTTAAACAAA GTAATCAATA CAGAAGTTGT 360

TCCAGCAGTC TCTGCAAAAG AGTGCGGAAA CTACAAAGAT CATTCTTTAT TTGCAGCGAA 420

AGAATACGCA AAAATCGTCT TAGACCAAGG AATTAGTTTA GATCCATTTG AACGTATTCT 480

GTAATCTTTA GCAATTAACG ACAGAGAAAC CCACTATTAA GCTGTGGGTT TTTCTGTTCT 540

CTTTCTTAAT TTAAACGGAG GTTTTTTATG AATCAAACAA TCGAACAATT ACTAAGTCAC 600

CGTTCCGTTC GTCATTTTAA AAAGCAAGCG TTAACGGACA AGCAGAAACA ACAGTTAATT 660

ACTGCAGCTC AAGCAGGATC AAGCTCCAAT TTTTTACAAG CTTATACAAT TATAGAAATC 720

AAAGATCCTG AGCTACGCAG AGAATTGGGA CGTTTAGCGA ATTGTGAAGA CTATGTCGTT 780

AACACTGGTG TGTTCTATGT TTTTGTTGCT GATTTGTATC GTCATGCAAC GATTTTGTCA 840

AAAGAGGGGC AATCATTGGA ACCCTTAAAG ACACCAGAAT CATTATTAGT AGCTGCCGTT 900

GATACAACGA TTGCTGCGCA AAACATGGCC ATTGCGGCTG AATCTATGGA TTTAGGGATT 960

TGTTATATTG GCGGCATTCG AAATGACTTA GACACAGTAG CCAAACGCTT ATCTTTGCCT 1020

GAATTAACTG TACCACTTTT CGGATTAACA ATCGGTGTAC CAGAAACGTT AAATGGTGTA 1080

AAACCACGGA TGCCTTTTGA AAATATTTTA AGTGAAAATC ACTACCAATC AGACAAATTG 1140

ACAGATATGC ACACATATGA TGAATTATTA AAAGATTATT ATGCCAGCCG CTCAAGTAAC 1200

GCCCAAACAG CGGATTGGTC ACAGAAATCT TTGTCCTATT TTTCTTATAA TCGTCGCCCA 1260

GAAGTAAAAA TTTTCTTGCA AAAACAAGGA TTtGATGTCT AAACAAAAGA GCGTGGGCTT 1320

ACGTTATTTC AACGTAAGCC CACGsTsTTT TGTTAATTTT TATTTATCTT TACTAAATTT 1380

GTCTTTTAAG TCTTCAAAGC CTTCTTTAAC TTTATCAGTT AATTTTTCAG CGCCTTCTTT 1440

AAGGTCGTCG CCGGCATCAC GTGCTTTGTC TTTTACGTCA GCAAAGGTTG ATTGAGCTTT 1500

TCCTTCAAGT TCTTTGCCTT TATCATCCGT AACTTTGCCT TGCGCTTCTT TTGCAGTACC 1560

TTCTACTTTA TCTTTTGCAT CGTCAAAACG TCCTTTTAAA TCTGCCATGC TAATCCCTCC 1620

TAAATAATTA ATTACATAAT AAGAGTATCA TTCTTAGCCA TTTCCTTCAA ATATAATGCA 1680

TGAACTACGA ATAAATGGTT ATTTTTGAGC CATTAAAGCT TTTTTTGCCA AGGCAAAGCA 1740

ACCGATTGTT CCTGGATTAT CTTCTAAAAG AGGCGTCACC AAGTATTTTT CTAAAGGCGG 1800

TGTTTCCACA TATCCATTGA CTAATTCAGC AAATTTTTCA CGAACTTTCG GCATCAAATG 1860

ACGTTGTTTC ATGACGCCGC CACCTAAAAT GATCACTTCT GGCGCTAATA ATAAAGTCGT 1920

ATTGTACGnC GtTGCGCTAA ATAATAAGCT TCTAATTCCC AAGTTTTATG ATCCTCTTCC 1980

AATAAATGTC CTTTTTTACC AGAACGACCT TCAACTGCTG GTCCTGCTGC AATCCCTTCT 2040

AAACAATCTC CATGATAAGG ACAATTTCCT GCATACGTAT CTTCAGGATG ACGACGAACT 2100

AACGCATGCC CCATTTCTGG GTGGCTAAAG CCTTCAATAA ACTCACCGTT TTGAATCGCT 2160

CCAGCGCCAA CACCAGTTCC AATTGTATAA TATACACAAC TAGATGTATG TTGCCCATTT 2220

CCAGCAACAT ACTCACCATA TGCCGCAGCA TTCACATCCG TTGTCCAAGC CATTGGCACA 2280

TCAAAATGTT GTTTCATAGT TCCTAACAAG TCAAAGTTAC GCCAAGCTAA TTTTGGTGTA 2340

GAAGTGATAT AACCATACGT AGGAGAATCA ACGTGAATAT CAATCGGACC AAATGAACCA 2400

ATCCCAATCG CTTTTAAAGG ATATTGTTGG AAAAATTCTA TTACTTTTTT CATTGTTTCT 2460

TCTGGGGTTG TTGTGGGAAA ACTGACACGT TCTACGATGG TCAAATCATC TGTCCCAACG 2520

CCACATACAA ATTTTGTGCC ACCGGCTTCG ATACTTCCTA AAAGTTTTTC TGTCATTGCA 2580

ATATGCTCCT TCGTTTCCAT TTTACAAGTA AGCGGATACA TTACTTTTGA TTAAATAACT 2640

AACATACACG TTTAGTTTAC CACTTTTTTG TCGTTTCGAC AGAACAATTT AACAATTTCA 2700

GTCAGGTTGC CCAATGAAAT CTGGCAAGTT GCTAAAGAGC ATTTTCCTAA AACAAAAAAA 2760

CGAGCATTCA CGGTTATCTA ACAATGAATG CTCGTTTTTG TTTATTGATT CTTTAGCGAA 2820

AAATATAACA ATAGTCCGAT TGCAATCCAG CCACCACTAA GTAAAAGCAT TAACCAACTT 2880

AAAAGCGTTT CTTTTAGAAT ACTTCTGGGC TCTTTTGCAT GCTGAATAAA ACTCATACTA 2940

GTCACCAAAG AAAAGAGGAA AAACATAAAA CAAATGACCG CTAAACCGCC AGACAACCAA 3000

AACCAACTAG CCATTATTCT CACTCTCCTT TTCGCTTGTT TTTTGATAAT TATTGCGCCA 3060

AATGACACTA TAAATTGCCC CAATTGTGAT TACTAACAAC GTGTAGAAAG TCGGTAACCA 3120

CCATGTGGGT AGGACCCATT TGACCGTAGC TTCTTGATTC AATTTTTTTG CTTCACGAGT 3180

GTTAATCGTA AATTCTTGTT GAAAAGGCCA AGACTGTTGT TCATCTTGTT GTAAAATAGC 3240

ACGTCCTGTA AATATATAAG TCCCAGCATC TAGCGGTTGT TTACCCAAAT CAAGATGAAA 3300

ATTCATGGAA CTATTTGGGG CAAAGCGAAC ATTTTTTAGT GTCCGTTGTG CAATTTTTTT 3360

AGAGGATCCT TTTTTTATCA CTTGACCTTC TAAACGAACT TCTGGAAAAA TAGCAGCATA 3420

TGGATTTTGA ATGACCGCCT TTAGCGATTT GGCTCCTAAA GCAATTGTAG GTGTTACTTT 3480

CGCTAATTTC AGTTGGTTCA CCTCTGTTAA AGGAACATTC GCAGCACTCC TCATGGCAAT 3540

CGCAACATTA AACCCGTAGC GACTGGTTAA TCCCGTTGTT TGGTTCTCTT CCTCTGTTTT 3600

TTCATCTTTG AGATGCGCAT AAACCCCGCC CAAAACAACT CCTTTGATAG GCTGATCAAG 3660

TCCCTTCACC TTCATCTTTG CTGATACTGT TTGATGAGGC GGCACACTCA CTTCCTCATT 3720

TTCTAAAGTA ACCAAGTCTG TCAGCGGAAC CTTTAACGAT GAATCAAGTG GTTTATCAAA 3780

ATCATTGTAC ACAATCAAAC CATTATCATT TGTTGaTGCA TTTGCTAGAG TCAATACGAT 3840

GGaCAATTTT TCGTCTGTCT TATTTTTTAT GTCTAAGACA ATAGTGGAAG ATTGCTTTGG 3900

TGCCAGAATT AAATCATAAT AACTAGATGT TCCTGACACT TGATTTTCTG GTAAAACAGG 3960

CGTTACTGTA AATTTAGTCT TCATTTCTTC TGTAGCTTCT ACTTTTGGAA CAATAATAAA 4020

AAATGTGAAA AACAGAAATA GACAGAwAAC ACCGACTTTT TTTCTACTCA TTTCTTCAAG 4080

ATCTCCTTTA TAAAGCTGAA TGTACCTACT AGTCTAACAA ATCAGCGAAA CAATGCTGAT 4140

CAATCTTTGA CTTTTGGGAA GGATAATCTT GAAAAAAGTC TCCTTGAAAA ATACCAAGGA 4200

GACTTTTTAT TaTTCACTAT AATATAATTC TTTTTTTATC TGTGTGGTTG ATACTTCTGG 4260

CGTTCGTGGC AAATAAAGAA CCTTTGCGGT TGTTTCAACT TCTAGAAAAT CAAAAGCCCC 4320

TTGCCAATCA TCGCCCATGA CTAAAGTATC AATTTTAAAT TCAGCAATAT CGGAAACTTT 4380

TTGTTTCCAA CTAGTTTCAG GAATAACTAA ATCAACATAA CGAATTGCCT CTAATAATTG 4440

TTTACGTTTT TCATATGAAA AATAACTTTG TTTCTTTTTT TCTAAATTAA AAGCATCTGT 4500

TGAAAGTCCA ACAATCAGGT AGTCCCCTTG TTGTTTCGCT TTTTTCAACA AATTAATATG 4560

ACCATAGTGT AGGAGGTCAA ATGTGCCGTA AGTAAGTATT TTTTTCATAA AAACCTCCAT 4620

TTGTACTTTG ATATTTCTTC GTTAATTATc AATTTTTaTG AATAGAAAAA AGGCTACTCT 4680

TGaTTGaCTT TGaTGGTTTT TCGATGCAAA TTTTTTATTT TnCGTAGCCG Tn 4732

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7294 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

GGGaaAATwt TTTTAtgkGc tATGGcGTAt TACtAGTTAC gTTGTTGCCC TAGTGTCCcT 60

TATTTAAAAt ATTTGGtGAA AAATCaGCGT TTATGGACcA TTGgATACGG CCcATTTCcA 120

GTGGGGAATG TTTTkGcCTG TGCTTTTTGG CTTTTTCATG CTGATTGGCT TTGATCAATT 180

TTTTATTGCG TTTCATGGAT TATTTTTTAA CAACGATGCT TGGTTATTTA ATCCGCTGAC 240

TGATCCGATT ATCAATGCCT TACCTGAAGC CTACTTTATG CATTGTTTTA TCTTGTTTTT 300

```
TATTTTATTA GAAGGCCTGT TTTTGATAGG TATTTTAATC GGAAAACGTT CTTTAAAAAA    360
ATAAAGAGAG AAAAGTGAAC AATTCCAAAA TTGCTCGCTT TyCTTTTTTA TTCAaTCtAG    420
TTTTAAAAAC TAGATTGAAT AATTGTCAAA AGTATGATAT TCTTTTTTAG AAATGGTTAG    480
GAGGGTTTTT ATGGACGAGA TAGAAAAATC ACTACAAAAT TGGGGCCAAC AATTGGAAAA    540
CGTTCATTTG CCAAGATGGC ATGAGTTACC AGATATTGAA TTATATATGG ATCAAGTGAT    600
TACTCTGATA GAAAAATACT TATCCCCATT GATTACACTG GAAAAACATA CTTTGTTAAC    660
GTCCTCAATG GTTAATAATT ATGTCAAACA CGGGTTAATC CCGGCCCCAG TGAAAAAGCG    720
CTACAATCAA AAGCATTTAG CGTTTTTAAT TGCGATTACG TTGTTGAAAC AAGTGTTAAC    780
ATTGCCAGAA ATTAAGCAAG GCATTTTATT CCAAGGAGCT ACGGTTGGTA TTCGTGAAGC    840
gTACAATTTA TTTTGTCAAG AACAGGAACG AGCGATTTAT GTGATTGCTG CTCAAGCGCA    900
AGAAAAAGAA GTTCAGGCCA AATCACAAGA ACCTATGGGG ATTGAATATT TGGCTGTAAA    960
AGCAGCGACA ATGTCTTTTG CTACTAAAAT GTTTAGTGAA AAAGTGATTG AATTGGAACA   1020
AGAATACTTA AAAGAAATGG ATGAGATGAC ACATGAATAA AGAAAAAATC GCACTACTAG   1080
TTGATTCTGG GACAGATGTT CCAGAGGCAT TAGTCAAACA ATACGGCATG TATGTCTTGC   1140
CCTTACAGAT TATTTATCCT GAAAAAACCT ATACAGATAA GGTAGATATT ACACCAGAAG   1200
AAGTGTATCA ACGTTTAGAG AAAGAAATCC CTAGTACTTC ATTACCTGaT GGTGCAACAA   1260
TCCAAGCTAT TTTTGACaAA ATTAAAGAAG CGGGTTACGA AAAAGTCTTA GCGGTGACCA   1320
TTTCTAGTGG TTTAAGTGGA ACGTATAATG TAGTTCGCTT GCTTGGAGAA CAAACCGAAG   1380
GACTAGATGT ATTTGTTTTA GATACAAAGA ACATTGGTAT CGGGGCTGGC ATTCAGGCGA   1440
TACGAGCAGC AGAATTAATT GAAACAGGAT TAGGATGGCA AGAATTACAA CAAAAATTAA   1500
CAGAAGAAGT AGCCAACGCA AAAGTATTTT TTAATGTAGC CACTTTAGAA TATCTACAAA   1560
AAGGTGGACG AATTGGTTTA GTCACTTCTA TTTTAGGGAA TGCGCTAAAA TTAAACCCCA   1620
TTATTTCTTG TAATGAAGAA GGCATTTACT ATACGGTGGC TAAGTCTCGC GGCCGTAAAA   1680
AAAGTCTTGA TAAAACGTTT GAGTTGGTGA CGAACTTTAT AGGTGAGGCA CCCCGTTTCC   1740
GTTTGGCAGT CGCTCACGGT GCGGCAGAGG AAGAAGCmAA AGCmATGATG GAACGCTTGA   1800
AAGCAGCTTT TCCACAAGCA GAAGAAATTT ACTTTGGAAC AATTAGTCCG GCATTAGTTG   1860
TTCATACAGG ACCAGGACTT TTAGGTGTTG GCATACAGTT ATTAAATGtT AtTAAATGAA   1920
AAATCCCCTC AAATTTTTTT GaGGGGrTTT TTCaTTTAAG ATTyCyCTTT TyCATAGrAA   1980
TATCaTTtAG CyGGaAAAAC aACGrAAATT wTwTTGaAAA TyTCACGAAT GTGTTACAAT   2040
CCATCTACAA AGGGGGAtTT TTTGTGaAAA AAGCAATAGC AGAATGTTTA GGAACATTTA   2100
TTCTTGTCTT TTTTGGGACA GGAACAGCCG TTCTAGGAAA TGGTATGGAA GGGATTGGAA   2160
CAACGGGGAT TGCCTTAGCA TTTGGGTTAA CAATCGTTGC TGCTGCCTAT AGTATTGGCA   2220
CAATTTCAGG TGCACATCTA AACCCAGCTG TTTCTATAGG AATGTGGCTT AATAAACGAA   2280
```

TGACAACAAT GGAACTAATC TACTATGTAG TGGGACAAAT TGTTGGTGGC TTGATTGCCT 2340

CATTTGCTTT ATTAAGCATT CTAAAAGGTG CTGGTAAAAG TATTGAGAAC TTAGGTCAAA 2400

ATGGCGTAGG GGAATTAAGT GTTGCTGGTG CATTGACAGT CGAAATTATT TTAACATTTA 2460

TTTTTGTTTT AGTTGTCATG ACAGTTACAA GTGCTAAGAA AGGCAATGCT TCATTAGCAG 2520

GAATCGTAAT TGGGTTAACA CTAACAATGA TTCACTTAGT CGGCATTCCT TTGACAGGAA 2580

CTTCTGTCAA CCCAGCTCGT AGCATCGCGC CAGCCGTTTT TGCGGGTGGT TCAGCGTTAT 2640

CTGAGTTGTG GATTTTTATC GTTGCGCCAT TAATTGGTGG ATTATTAGCC GCTCTTGTAG 2700

CAAAATTCGT CTTAGATACG GAAGAATAGA TTAGCGAAAG AAGAATCTTT GATAACAAGC 2760

GTCAATTGTT GTCAGAGATT CTTTTTTTGT TATACTAGGA GAACCTAAAG GCGCATTTTT 2820

AACCAAAGAT GAAAACATCA TCAAAATAAG GTATAATAGG AAAGTATCAA ATAAGATTAG 2880

AAGGAAAAAG AGGTGTCTCT TTGTGAAGAA AATTTTAGTA GTTGATGACG AGAAGCCAAT 2940

TTCAGAGATC GTTAAATATA ATTTGGTTAA AGAAGGATAT GAAGTATTTA CTGCTTATGA 3000

TGGAGAAGAA GCACTTGAAA AAGTGGAAGA AGTGGAACCA GACTTAATTA TTTTAGACTT 3060

AATGCTCCCT AAAATGGATG GCTTAGAAGT CGCGCGAGAA GTGCGCAAAA CACATGATAT 3120

GCCAATCATT ATGGTGACTG CCAAAGATTC TGAAATTGAT AAGGTTTTAG GATTGGAATT 3180

AGGAGCCGAT GACTATGTAA CGAAACCATT TTCAAATCGT GAATTAGTTG CTCGTGTAAA 3240

AGCCAATTTA CGGCGAGGTG CAACCAATGC GAAAGAAGCC GAGGTGACAA CACAATCTGA 3300

ATTAACGATT GGTGATTTAA CCATTCATCC TGATGCATAC ATGGTCTCAA AACGGGGTGA 3360

AAAAATTGAA TTAACCCACC GTGAATTTGA GTTACTTTAT TACTTAGCAA AACATATCGG 3420

ACAAGTGATG ACTCGTGAAC ATTTATTACA AACCGTTTGG GGTTATGATT ATTTTGGGGA 3480

TGTGCGGACA GTGGACGTAA CCGTACGTCG TTTAAGAGAA AAAATTGAAG ATAGTCCAAG 3540

TCATCCAACG TATTTAGTTA CTCGTCGTGG GGTTGGTTAT TATCTAAGAA ATCCTGAACA 3600

GGAGTAATCA GGTATGAAGA AAAAAGTTCA CTTTTTTCAA TCTGTTAATT TTAAGATTGC 3660

TTTATCATTT ATTCTGCTGT TACTGATTGC GATTCAAATC ATTGGTGGCT ATTTCATTCG 3720

TGAATTGGAA GCCACTACAA TTTCCGATTT TAAAAAAAAT ATGGATTCCC AAGTTGTCCA 3780

ATTGTCAAAC ACGTTAAGTA CGCAGATGAG CAACAAAGAT CTCGAACGTA GTGACGTTGA 3840

TGCAAATTTA AAAAAAGCGT TATCTGATTT TTCAAATGCA GATATTTCTG AAGCGAGAAT 3900

TGTCGATGAT AAAGGGATTA TTCGGGCAAC CAATGATTTA AATCAACAAA ATATTATTGG 3960

GAAAAAGAAT GATTATCGTG ATTTAAATGA CTTTACGAGT AAAAAATATC AAGCTTTAGA 4020

TAATGATAAA CGCGTGTATG TGAATGTCCA GCCGATTCAA TCGCCTACTG GAGAAACAGT 4080

GATTGGCGTC CTTTATGTGA AAAGTAATTT AGAAAATAAA TACCAAGAAA TTACCAACAC 4140

AGCAAGTATC TTTTTCACTG CTTCTATTAT TGCCGCAGCA ATCTCGATTA TTGTGACTTT 4200

ACTGATTGCA CGATCAATCA CGAAGCCGAT TGGTGAAATG CGCGAGCAAG CCATTCGAAT 4260

CGCTCGTGGT GATTACGCTG GAAAAGTAGA AGTCCATGGA AAAGATGAAT TAGGCCAATT 4320

AGCAGAAACA TTTAATCAAT TATCAGAACG GATTGAAGAA GCACAAGAAA CAATGGAAGC 4380

AGAAAGAAAT CGTTTAGATA GTGTCTTAAC GCATATGACA GATGGTGTCA TTGCGACGGA 4440

TCGCCGCGGA AAGGTGATTA CGATTAATGA AATGGCCCTT TCATTATTAA ATGTAAAAAA 4500

TGAAAATGTG ATTGGGACCT CGTTATTAGA GTTGTTAGAT ATTGAAGAAG ATTACACATT 4560

GCGGAAGCTG TTAGAAGAGC CAGATGAACT GCTGATTGAT CGCTCAACGT CTGATCGTGA 4620

AGAAGACCAA ATGATTATCC GGGTAGACTT TACAATGATT CGTCGGGAAT CAGGATTTAT 4680

TACTGGCTTA GTTTGCGTAC TTCATGACGT CACAGAACAG GAAAAAAACG AACGGGAAAG 4740

ACGGGAATTT GTTTCCAATG TTTCTCATGA GTTGCGAACG CCTTTAACAA GTATGCGTAG 4800

TTATATAGAG GCTTTGAGTG AAGGAGCTTG GGAAAACCCT GAGATTGCGC CGAATTTCTT 4860

AAAAGTCACG TTAGAAGAAA CCGACCGGAT GATTCGTATG ATTAATGATT TGTTAAATTT 4920

ATCTCGGATG GACTCTGGGa ATACACATCT TCAATTAGAG TATGTGAATT TTAACGAATT 4980

GATTAATTTT GTCTTGGATC GCTTTGATAT GATGATTGAA AATGAGCAAA AAAATTACAA 5040

AATTTGCCGT GAATTTACTA AACGCGATTT ATGGGTAGAA TTAGATACAG ACAAAGTAAT 5100

TCAGGTTTTA GACAACATTT TGAACAATGC GATTAAGTAT TCGCCAGATG GCGGCGTCAT 5160

TACCTGCCGA CTAGTTGAAA CACATAATAA TGTCGTCTTT AGTATCTCGG ACCAAGGTTT 5220

GGGCATCCCT AAAAAAGATC TCGGGAAAGT CTTCGAGCGT TTTTATCGTG TGGATAAAGC 5280

ACGTGCGCGA GCACAAGGTG GGACTGGTTT AGGTTTAGCA ATTTCTAAAG AAGTAATTCG 5340

GGCCCATAAC GGGAGTATTT GGGTGGAAAG TACAGAAGGT GAAGGATCGA CTTTCTATAT 5400

TTCACTACCA TATGAACCTT ATGAAGAGGA TTGGTGGGAA TGATGAAATT ATCAGAATGG 5460

ATTACAAGAA TTGGCTTGAT TTTGATGGTC ATTTTAAGTA TATATTTTTC AGTCAATATC 5520

TGGCTGAATT CTGCCAAAAA AATACCAGAA ATGAAGTCGG GAAGCCAAGT CACAACAGCT 5580

GTCAATGAAA AAGCCATTGG CGATGTCTAT TTACCTTTGC AATTGATTCG AATAACTGAT 5640

GGAAAAGCGA TGCAAAGTAA TCGTGAAACA TTAATTAGTA ATGTTCAAAA TGATATTAAA 5700

ATGGCTACGT TTGGTAAATT GACACAAGTT GTGACAAAAA ATGCAGAGCA ACTTAAGCGC 5760

TACAACCAAA TGGAACAAGG CATTGAACTT CTTTATCAAG GtCCCTTTTT AATCtCgGAC 5820

TATGCTTCGA TTTATAATCT ATCCATTAAT TTTACTAACT TtAATGAGTT GACGGACCAG 5880

TATTTTACGA AAATTCAATT GGgAtTTTAA CGAAAATAAG ATACGTTTTT TAGATTATGA 5940

TCAAGCCAAC GTCTATGAAG CGCCCATGAC TGTCAATAAG GCGCGCTTAA TGGGAATTAT 6000

CAATAAAGAG GGATTACAAT ATCAAGACGT TTCCGAAAAT ACGCTAACCA AACAAGGGCA 6060

ATGTTATTTA ACCAATGATA TGAAGTTGAA AAAGTACAGT TATATCTTAG CTTCGCAACC 6120

AGTTACTCGT TTTAGGAATG CTTTTTTCAA TGAAACGGAA GATATCCAAA CCAATGAAGA 6180

CAGTCAAGAC TTAACCTACA CGAGTAAAGA AGAACGATTG TTTGCAGAAG AAAAACTGGG 6240

AAAAATTGAT TTTAAAGGGA CCTTGCCAGA AGAGAATAAA CGGGACTCAA TCTATAATCA 6300

AAGCTTTTCT TATGTAAAAC GTTTAGGAAC CAATATGGGG AATTTGCGTT ACTTTGATCG 6360

AACGAAAGAT AGTGTCAATT ATCGGACTTT TGTGGAAGGT TTCCCAGTGT TCAGTAATGA 6420

TTTAAAAGGC CAAGTGGATA TTCGCATCAC GAACAACGAT GGTGCTGCAC CAAGCGTAAC 6480

CATTAACACA AGTGTGAATA CGATCCAAGT GCCGATTCCT TCAGAAGAAG AAGTGACGCT 6540

GGAAAGCACG GAAAAATTGA TTAAGCGTTT AGAAACGGCT GGTGCTAAAA AGGAAAAAAT 6600

TCAATCGGCT GTTATCgGTT ATACGTGGCA GACAATTGAA GAAGTCAAAC AAGTCGTTGA 6660

TTTAAGTCCG GAATGGTATG TCCTTTATAA CAACAATTGG TACACTGCAA CGGACTTGGT 6720

AAAACAATTA CCAAGCTTGG AGGTGGGCTA ATGGATTTTA AACGAATTGA GTGGATTTTC 6780

TTCTTAGCAT TTTTAGGATT AAATATCTTT TTATTTGGAA TTTATCGGGA AGGTCAACAA 6840

GAGGAGAGCA ACGTCTCTTC TTCCAGCCAA ACAGAATCGA TTGAAAGTCG GTTGGAAAAA 6900

GATAACATCT CGTATAAAGG GACACTTTCT TCAGAACGAT TGGAAGGTTA TTATTTAAGT 6960

GGCGAACAAA CCAATTTTTC TGCTGCTTTA AAAATCCAAC GTGAAAAGAA TAAAAATTTT 7020

TTGAGAAATG GGCTGCAAAT TGCGGATAAT ACTTTAACGA GTGTACCTAG TAAAAACTAT 7080

TTTATTGATC CTAAGAAAAT TGATAAAGAT TTAAGTACCT TTTTAAATGA AAAAAATGCT 7140

TTATTATTCG GAGACGAATA TCAATACTTA CCAGAATTTT CTCATTTAAA AGAGCCGACG 7200

GCAGAAATTG TGGCTGCACA ATCGTATAAA GGAATTCCTT TTAGAGACGA CACGGCAAAA 7260

TTAAGTATTT TAGCAGATAC GTCAGGTGAA TTAT 7294

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3422 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

nATTTCATCT AACTTTTTGA CGCCnTAACA CCCTATTATC ACTTTTTACA TCTTATTTCG 60

ATTTAAAAAT AACGATAAAA AGCCCGTTCC TAACAGAATT ACTGCTGAAA CATAATAAGG 120

GAAAAAGTGA TTGATATCAA ATAAAGCTCC TGCTGCCATT GGTCCTAAAA TATTACCAAA 180

ACTTGTAAAT GTCGAATTTA GTCCGTTGaT GGTTCCTTGT TGATCTCCAG CATGTTTGGA 240

TAAATAAGTA GTTACTGCTG GTCTAAACAA GTCAAACGCT AAAAAGACAA TAAACGTTGa 300

AAATACAACA ACTAAATTAT TTTTTGTAAA GGCAATCACG GCAATAAAAA TGGCACTTGC 360

AAAAAAGGTT AATTGGATTA AACCCATTTC ACCTATTTTT TGGACGATTG CATCAAAGAA 420

AAACAACTGA CAAATCAACG CTAAAATACC ACTAACCGTA ATCACGATTG CTATTTCACT 480

TGTGGTAAAG CCAAAATTAA TAGTCGCCAT AATACTATAA ATAGATTCGA ACGCTTGCAG 540

GCCAAAAGAG GAAATTAAGA TAATCACAAA TAATGAGGTA AACATTGGAT TTCTTAAAAT 600

ATCCATAAAT GAACCTTTTT TCGCTTCAAC AGCGGCTAAA ATTCGTTTCT CTGGCTCCTT 660

CAAAACAGTT AATGTCAAAA TAAAACCAAT AAACGCTAAA AATGCGGCTG CAAAAAACGG 720

GGCGCGAATA CCTAAATAAG CaATAAAACC ACCAACTCCT GGTCCGATAA TAAAACCACC 780

ACTAATTGCG GCTGACACAA GCCCCATCGC TTTCGGACGT TCAGAAATCG TGGTCATATC 840

TGCCACAAAG GCTGTCACTG ACGGCATTAA TAAAGCGGCG GCAATCCCGC CTAAACCACG 900

AGAAATATAA AAACCGCTTT TCGCTTGGGC TAAACCAAAT AATAACTCAG AAATTGAAAA 960

CACCAACATG CCCGTTGCAA TCATTTTTTT CCGACCAATT TTATCCGAAA GGACACCTGC 1020

GATTGGCGAT GTGATTAATT GGGCAATCGC AAAAATAGAA ATCATCATTC CCATTGTGGT 1080

ACCAGAAAAA TGCATTTCTT CTTTTAATTG GGGAATTACG GGAATGACTA GGCCTACGCC 1140

TAGAAAAACA AGAAATAAGT TAGAAATTGC TAAGTACATC ATACTATTTT TTTTCGTCAT 1200

AATTTGCCTC CTCCATAAAA AAACAGCCAT CAAAAATGGC TATTTCTGTT TCTTTGTATG 1260

TACTTTCTTT GACAGTGTTT CGGCTGAAAC ACTCGTTGAG TCATTGAGAT ACATTTTACA 1320

CTAATTCACT GAAAATTTCA CTAGTTTTTA CTTATTTTTT TAAAGGAAAT TGTTATTTTT 1380

TAAAGACATA ACTTAAGTAT CCAAGTATTT AATGAATCGC TTTTGTTTTT TATCAGAAAC 1440

ATAGCCAATG TGTTCATAAA ATTGATGTGC TGCTGTACGC TGTTCCCCTG AATTTAAGCG 1500

AATCGCAGAC AAGTTACACT TTTTAgCTcC GTTcTAGTTT CGTCATTAAt TGtTTACCAA 1560

TTCCACGGCC TTCTGCTTCT TTTGCCACTG CTAATCCCaG AACGTtCAAT AACGTTTGGG 1620

AATAAATTGT TTCATACAGT TGCGCATGCA CATAGCCTAA AATCTTGCCt GTCTTCGTTT 1680

CTTCATAAAC ATATAAAAAC TCTCTGTCAC TGGCTAGTAG TAATTTCTCA ATCTGTTTAC 1740

AGGTCACGTC AACGGGAACT TCATAGCCCA ATTGTTCTTT ATTAATGATA GTTAACGCCG 1800

CTGCATCTGT TGGCTTCATT TTTCTAATCA AGGGTAACAC CCTCCTTTAC ATTCTTACTT 1860

AAATAATAAC GAAAGCATAA TATCCGCAAA ATCATTTAAT TCTGAATCAT TGTATTGCTC 1920

TAACAAAACT GTTTCGCCTC GATACTCTTT TAAAGCTAGT TGTATCAGCG AATGATGATT 1980

AGTAGGAAAA TGTACTAGGC CCCATTCCCC GCCAGCCCTT TTTGATAAAA TTTTATTCTC 2040

TTGCTTAAAT GCCATTACTC TACACATATT TAAAATAACG TACATAGGTT GTTGAACAAT 2100

TTCTTTTTTT GCCTCTTTCA CATCAGCTAC AATACTTTCA AAATATTTCT TTTGATCTAT 2160

CCAGCCAAAA ACGGAATCAA TCGGTTTTCC TACGAGTACT TTTCCATATA ATTTTGTAAT 2220

CATAAGATGT GCCACTAGAT CCTCATCGTA TCCCATCATA TCTTCAATAT ATTTATTTTT 2280

ATTAGTCAGA TATCTGGAAC GATGTATCGG AGAAAAATGT AAATCGAAGG GAATCGGATA 2340

ACTCAAATTT ATTGTATCTT tTAATTGtAG CACAkGGAAT TCTAtCCCTT TTTTTGGkGc 2400

CCACTTATCT AAAATATTAA TTGTATAATC TAATAATTGT TTCTTCTCAT CAATAGTAAG 2460

CTTTTCATAA ACCACAATCA GTACATCCAA ATCACTTACT GAATGACAAT AGGTTCCAAA 2520

AGCATAAGAA CCATGAAGAT ATATTCCAAC TAAATTTTCT TTTATTATTT TCATATATCC 2580

ATTTTCCATT TGAGCTAATA ATTGATGAAC CACGATTATT CCTCCATTAT TGATTGACAA 2640

TTGCTATAGT TTTTTCGATA AATAAAGAAC TAACTCATCA TCATTAACAC ACCGTTCGTT 2700

TGGTTTCAGC TGCTTATTTT GAAACCAGAC TCCTGAACCA TCTGGCACAT ATCCTTGTTT 2760

GATATATAAT CTTTGCGCTG CACCATAACC AGGATGCAGG CCTACTCCCA AAGAAACAGT 2820

ATCTGCATAA CTTTTGGCAA CATTTTCTGC ACGATTTAAT AATAACCGCC CCACTCCTTG 2880

TTTTTGAAAT GGTAAGAATA CATTGAAATC AGCTATTTCA GGATATAAAT TTTTAAAAGG 2940

TCCTTCCTTA GCTAATGGCA GTAACGTCAC ATAACCTAAC AGCGTTGCCT TTTTCTCCGC 3000

TACAAAAACC GTTCGTTGCT TGACTAATTG TTCTTCTAAA TAGTTCATTA AAATCTCTTG 3060

GCGACTTGGC CAACCTTGTT GGrcAAATTG AGTATCCAGC GCTAAGATAT CTTTCTCTTG 3120

CATTTCTCTA ATAACCACTA GCTCAGACAT CAGCATTTCC TCCCGAAAAG CGTTTTCATT 3180

TTATAAAACG AAGCGAATAA TTCTTCGCTC CCTTATTCAT TTTCAAAACC ATCGAGTTGA 3240

TAATATTTGG GCAAAACAGC AACACATGTT TCTTCTGAAT TAACAGCAAT AACGAACCCT 3300

TGCCATAAGG nCCACGTCAC CGCTTCACCT AATTTCTCTT TCACTCGTTG nATCATTTCA 3360

GCTTGCGTTA AGTCTCGTGC AGGrAAATGT TGAACTTTAT CATTACTGAT TAAAAAAATA 3420

AA 3422

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9172 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

CCGGACCATC AAAATCACTT TTTTCCTGCT CTTTTnAAGA AACGAAnGGA CCCATGTCTT 60

TTCAGAAAAG TTTGTTACAA TnACAAGCGT AGAGAGGATG TTAGGGTATG GATAAAATGA 120

AAGCTATTAC TTTTTTTGAT TTnGATGGAA CTTTGTTAGA TGGAACATCG CAAATAACTC 180

CTGAGATTAC CGCAGCAGTG GCGGCCTTGA AAGACAATCA AATTTTACCT TTAATTGCGA 240

CAGGACGGAC CTTGTGTGAG ATTCAGCCAA TTATGAAAGC CAGTGGCATT GATTCTGCGA 300

TTGTAATGAA CGGGCAATTT ATTCATTATG AAGGAAAAAC GATTTATTCA GATGAATTTA 360

CAACAGAGGA ATGTGTTTCA CTGCATGAAC ATGTCAAACA GCGTGGTCAT GAATTGGCTT 420

TTTACAATGA GCGCCGTATT TTTTGTACAG GGCATACTGG CACAGTGAAA CAAGCCTATG 480

ATTATATTCA TTCAGCTGTT CCAGAAATTG ATCCCACAGG TTATGAGAAT GATGCCGTCA 540

ACATGATGTT GGTCTTAAGT CAACATGGCG ATGATGATGA ATATTACTAT GAGCGGTTTC 600

CTGAATTGAC CTTTTATCGC AATGGTCCTT TTTCAATAGA TATTGTTCGT AAAGGGGTCT 660

CTAAAGGCTC TGGTGTCAAA AATCTTTTTA ACACCTTGGG ATTAAACGGC ATTCCGACTT 720

```
ATGCCTTTGG TGATGGCATC AATGATTTGG CGCTGTTTGA GGCCTGTGAT TATGGAATTG    780
CCATGGGTAA TGCGCGAGAA GAACTAAAAG AAAAAGCCAC CTTTATTTCA ACTAAAAATA    840
CTGAGAATGG CATTGTTAAT GGCTTGAAAA AATTTGATTT ACTATAAGAG AAAACTACCT    900
TCTTCTATTG ATGTACTTTA GAAGAAGGTA GTTCTTTTGT ATGTTCATTT TTCTTTTTCT    960
GCTGTGCACA GAGAGCTACT CCCACTACTT GAATCAGCAA AATCGCTCCC AACAAACAAC   1020
TTAATACGAT GTTTCCTGAT AGGATTGTtC TTCTGATTAA AATAATTAAT ACAACACTGC   1080
TAATTCCGTA GAGTAACCAT TGGTTGTCCA TTTTTATTGA CCCCTTTGTT TTTCTATGTT   1140
TAGGATAGGA TGTTtCCTAC TAAAATGCAA GGnAAGATCC TAGATTCTTA CAACTAACCC   1200
ATAATTTTTA CTAATTCCTT ATTTGTTTTC ATTTATTCTT AAGATTCATT TCATTTTTGg   1260
TAAAAAATCA CACTTTTTTT CTATCTCTGT TTTAAATTGC GTACAGTTCT TAAAAAAAAk   1320
GATACACTAA AaGTGaACTC GaAAkGAArG GrGCCATtCC TTTGGtTAAT AGCGAAAAAA   1380
GAAAACAATT AATTATTGAT ACATGCATTT TAGCTGGAAA AATTATGCTG GAAAGCGGAT   1440
CTGAAGTATA TAGAACCGAA GATACAATTA CTCGGATTGC AGCAAATGCT GGCGAGCCAG   1500
AGAGCGTTTG TTACACGACG GCTACAGGGA TTTTCGTGGG CTTTCGTTCC AGTAATTATA   1560
CGCAATTAGA AAACATCCCC CAGCGCTCGA TTAATTTGGA AAAAGTTAGC CTTGTCAATC   1620
AGTTATCAAG AGAATTTGCG CAAAAAGAGA TTACTTTACC AGAGCTTTAT CAGCGCTTAA   1680
CTCTATTAGA AACAGATACC CCCACTTTTT CAATTTCGTT ACGCTTGTTA GCAGCCGGGA   1740
TTGTTAGTTG TACGTTAATG TACATTTTTG GTGGCACTTG GcAAGATTTT ATTGCAACTT   1800
TTTTTGTTGG CGTGATTGGG TATGCCAGTT ATCTTTTTAC CCAAAAGTTA TTCCAAGTTC   1860
CTTACCTCGA TTCATTCGCA GCCGCTTTTG TCATTGGTTT GTTGGCCTAT TTGGCTGTGC   1920
ATTTTCACTT AGCGGTCAAC ATCGATAATA TCATTATTGG TGCGGTAATG CCGTTAGTTC   1980
CAGGTGTGGC GATTACGAAC TCTTTTCGCG ACATTTTAGC GGGCCACCTC ATCAGCGGGA   2040
CGGCGCGAGG GACAGAAGCT ATTTTTATTG CTGGGTCTGT TGGCCTGGGA ATTGCCTTGA   2100
TTTTTAAATT ATTTATGTAA GGAGTGTTCA AATGGAACTA ATTTTTCATC TTCTCTTTAG   2160
TTTTTTAGCG ACTGTGACGT TTGGTATTAT TACGAACATT CCTCGTAAAG CACTGGTGGC   2220
TTGTGGCATT ACTGGCATGA CAGGCTGGAT GATTTATTAT GTCTTAACGC AAACTTTTGA   2280
TGCTAGTCAA ACCTTTGCAA ACTTTTTAGG AACGGTCGGG TTAGGAATCG CTAGTATCTT   2340
TTTTTCCCGT TACAAAAAAA TGCCGATGAT TATTTTTTAT ATTCCGAGTT TGGTTCCTTT   2400
AGTGCCAGGA GGCCCTGCCT ATCAAGCAGT TCGGAGCATT TTGTTAGGAA ATATTGATGA   2460
CGGTTTGCAG TTGATTTTAA AAGTTGTCTT TACTGCTGCA GCCATCGCAG CGGGCTTTAT   2520
GGTGACCTCG CTACTGGAAC GAATTGTCAA ACGTTTTTTC CCCAAGCGAC TCTTCCGATT   2580
GGTCAAAAAA TAAAAAAACA AGCTCATTTT CTGAGCTTGT TTTTTTATTC TACTTCTTCA   2640
AAGAACAGCA ACAAATCTTG TAGCGTTAAT TGTGCTTTTT CAGCAGCGTT TAAATCTTGA   2700
```

```
ACAATTTTAC CTTTTTGTAA AACAATTAAA CGATTGCCAT ATCGTAAAGC ATCTTCCATT     2760
CGATGGGTAA TCATTAAACA AGTTAAGTTC CCTTCTTCAA TCCGTTGTTC CGTTAAGTGC     2820
ATCAACTGCT TGGATGTTTT GGGATCTAAG GCAGCAGTAT GCTCATCTAA AAGTAATAAT     2880
TCTGGTTTTG TGATAGTGGC CATTAATAAA CTTAAGGCTT GGCGTTGACC GCCAGAAAGG     2940
TTACCTGTCG GTGTATCCAG ATGATTTTCC AAGCCATTGC CAACTTCTTG GCAGATTTTT     3000
GTAAACAAGG CGCGTTGCTC ATTTATCTTC CGTAATGTTA AGGGACGTTT TTGTCCGCGT     3060
TTCATGGCAA GCAATAAATT TTCTGCCACA GTCATTCTTG GTGCTGTCCC CATTTTAGGA     3120
TCTTGAAAAA CACGCGCTAG AGATTTGGCC CGTTTGACTT CAGAAAGATT GGTGCGATTG     3180
ACTCCTTCAA TTAAAAATTG TCCTTCTGTT AGTGTCAAAG TTCCTGCTAA ACTATTGAAT     3240
AGCGTACTCT TGCCTGCGCC ATTTCCACCT AAAACAGTAA TGAAGTCCCC AGGATAAATG     3300
GCTAAGTTCA CATGATCCAA AATCACTTTT TTTTCATTCA TCCCATTATC AATGATTTTT     3360
GTCGCATTTT TTAATTCAAC AACTGGTTTC ATTTCGTTTC CCCCTTATCA GAAAAGCCGT     3420
GTAGTCGTAA TGCTTTTTTC AATTGTGGGA TCATTAAACA GATTGCTAAA ATCACTGCTG     3480
AGAAGATTTT CAAATACGTC GTATCAAAGC CTAATTTAAT CACTAATAAA ATCAATAGTT     3540
GATAAATAAT ACTCCCTACA ACAATCGCCA TCAGCCGTTC ACCAAACGTT AATTCTCCAA     3600
ACAGTACTTC ACCAATAAtG AGTGAAGCCA AGCCAATCAC AATGACACCG GTTCCTTTAC     3660
TTACATCCGC ATATCCATCA tTTTGGGCAA TCAAAGCGCC aGATaAGCCA TCAAGCCATT     3720
CGaTaAAACC AaTCCTAAAA TTTTCATGCG ATCCGTTTGA ATGCCAATCG AGCGAGCCAT     3780
TTCTTCATTG TCACCTGTCG CAATATAGGC TTGTCCTAAA CTAGTGTTAA AGAAAAATAA     3840
CAGTAAAATC ATTACCACAG CCAAAACACT AATCCCTAAA AAAATAGTAT CATAATAATC     3900
GGGTAAGGCC ATTTTTTGAA AAACATCTTG AATCTTCGGT TGGTTTAATA AAGACTTATT     3960
TGGCGACTTC ATCACAAATA AAATCACGGA ATTTAAACCA GACATCACTA AAATTCCTGC     4020
TAAAATAATT GGAATTTTTC CTTTTGTAAA CAGTAAACcA GTCACTAAAC CGGCAAGCAT     4080
TCCTGCTCCG ACCCCTAAAA GAGTCGCTAT AAATGGTGAA ACCCCTGTTG TAATGGCAGT     4140
TACACAAACC GCCCCCCCTA ACGGAAATGA ACCTTCTGTG GTCATATCTG GAAAATTTAA     4200
AATACGATAA GTCATAAAAA TACCTAGTCC TAACAAGGCC CATAACATTC CTTGTCCAAT     4260
TGCTGAAACA ATCATCGCTT CTCCTACTTT CTTTGATAAT TAGTTAATCA ATTTCGATTT     4320
TTTTGCTAAC TCTGCGGGAA TCGTAATTCC TAACTTTTGA GCTTGTTTTT CATTGATAAT     4380
TGTGTCGCCT GTATTAAAGG TATAAATGGG TGTAGTTGCT GGTTGACTTT TTCCAGACAG     4440
AACCTCTGCT GCCATTTTGC CGGCTTGTAC CCCCAACGCA TGTTGATTAA TCCCTACAGT     4500
TGCTAAACCG CCTTGTTCTA CCATTGTATC CACCGAAGGA ATAATCGGCG TCTTGGTTCG     4560
ATTTGCTTCT CCGACTACCG TTTGCATGGC ATTAGCAATC GTATTATCAG TTGGTATATA     4620
AATCACATCT GTTTCCTTAG TCATCACTTG AACCGTTTGT GCAATTTCAT TACTTGATGG     4680
```

CACAGCATAT GTTTTCACTG TCAATCCTTT GGCTTCAGCC GCTTTCTGCG CTTCAGCTAC 4740

TTGGTATCGA GAATTTTCTT CTGAGGAAGC ATAGAGAATA CCTATTTTTT TGGCAGCTGG 4800

TATCAATTCT TTTAGTAAAT CCATCTGGGC TGACAAAGGA GATTGATCAC TAACACCCGT 4860

GATATTACCG CCTGGCTTTT GATTGTCTTT CACCAATCCT GCACTAACAG GGTCCGTAAT 4920

GGCGCCTAAA ATGATTGGGA TTTCTTGGGT AGTGTTGGCT AAcGCTTGCG CAcTGGCGTG 4980

GCGATGCCAA TTAAAACGTC TGCCTTTTTT TGTACTAATT GTTGACTCAT GATTGCTAAC 5040

TTACTCTGGT CGGCTTGACC ATTTTGAAAA ATAATTTCTA AGTTTTTGCC TTTGTCATAG 5100

CCATTTTCCT TTAATCCTTC TTCGACACCT TTATAAATTT CGTCCAATGC GGGGTGACTG 5160

ACAAACTGCA AAACCCCTAC AGTAACTGTT TCTTTCTTTG AATTCGTCAC TTTTTGATTC 5220

ACAAAAAAAG CACCGATTAA ATAAATAGCT AATAAAGACA CGACAAACGT TAAACGTTTA 5280

TTTTTCATTT TGCATTCTCC TAACTGCCCT TTTGTTTGGC AACTTTTTCT TTTTAATAAT 5340

AAAACCATTA TCTATAGGGC GTTTCTTCCG TTGAAAGTGA AACTGAGATT CGTAATCCTC 5400

TGCTTAGTCA AAATCCAAGG AACACTTTTT TACAAATGTC AGGATTTAGC TGCGACGAGG 5460

ATTGCAACTT TGGTGAGAAT CGCAGGAGTA CTTTTTCTAC GAATGGGACA ACGTTCTCTT 5520

CGTTCACCTT GTACTACTCA TCACTCTTGA AGGCCAAACT GCGGCGAGTA TACGCAGCCT 5580

GCATCTACTC TGTCnCTTCG TGCCAATCGT AGTGCTTGTC AGGATTTAAC TGCGACGAGG 5640

ATCGCAACTT TGGTGAGAAT CGCAGGAGCA CTTTTCTACG AATGGGACAA CGTTCTCTTC 5700

GCTCACCTTG TACTACTCAT CATCTACTCT GTCACTTCGT GCCAATCGTA GTGATTCGTA 5760

GCAAAAAAAC AACGACTTAT CCTAAGTGTT TAGGAGTGGT CGTTGTTTTT ATGCAAAACA 5820

AAAACCACAT AGATGGGTCC TCTATGTGGT CGTTTTTTCT TCTTGAATAT AACCAGCATA 5880

GATACACTTT TTAGAATAAG TCTTCTAAAG TTGTATCTAT GAAAGAATCA TACATACAAC 5940

TATCTACGCC AGTTTTGCCA TGATTGGTGC ATTTCTTTTA CGTTAGATAA TTGTGTCATG 6000

TATTTTCCCT CCTAGTTATG TTGTTAAAGG TAGTCTACAC TATATTGATT TTGTTCGTCA 6060

AGAAAAAGTT TTTATTTTCT TGCATTTCTT TTACTTCTAA AAAACGAGCC ACTCTTATGA 6120

GAGGCTCGTT TTTCTTTATT TCAGCATATT TACCACTTAA TTGTTTCGGT TTCTTTCAAT 6180

TTTCACGGGA ACTTTAACTG CTTGTTTTTG CGATTGAATC AGCCATAAAC TTACTAAAAT 6240

CGTCATTGAA ATTACTAATG CCATCGAAAA TCCCCTCTTT CTAATTCCTT AAAACTTTCT 6300

ACAATCTAAG TTTAACATAG CTAGCTATTA TTCACCAGTT TCTGGTACTG TATCAGTGCT 6360

TTCTGTCGTC GTTTCAGTAG ATTGAGCAAC GCCAGTCGTG CTGGTCGTCG TCTGTGTTTC 6420

CGCGGTCTTT GCTGCTTCTT TTTCAGCTTC TTGCGCCGCT TTTTGCTGGT CTTTCAAAAC 6480

GCCAGCCATT GGTAATCGAT TTAATTCTGG TATTAATTGT GTTTTTAAAT CCATAGATTT 6540

TTCGACAACC GTCCGTGTCT TTTCTTTGCC GTTTTCAAAG AAAACAAGCG TAATCTCATC 6600

GCCGCTAACA TTGTATTGCT CTTGTGCTTT GCCAGCATTA TAAATTAACG GATAGATATA 6660

CAGCGAACGA TGAAATTCAT TCATTTTATC TGAATTAAAC AGCTTCACCA TTTGATCGTA 6720

TTGCTTGCCA GATGGTTTAG CAAACATGAC TGACATGGCT CCTGTTTCTT GGATTTTCTT 6780

GTCTGCTGTT TTGTAATCTA AGACACTGAC TTTCTCTGAT TTGATGGGTC CTGCGACTAA 6840

CGATTCAACT TGAGCAGCTC GAATATTTTT TTGCATCGTT GGAATAAAAA TGATTGAGCC 6900

AATTATTAAT ACTAATGCTA CAATAGCGGT AATCATTAAC AGTTTTCTTT GTCTTAATTC 6960

ATAAAGTGCC TCTTTTAGTT CTTCCATAAA ATTATTCAAC TTCTAATCGC TCCTCGCTTT 7020

TAGTTCTTAT CTATTATACT AATCATTTTC TTTCTTGCCT ACCCTTGCGC TTTAATTACT 7080

AAAAAAACCA GCAAACTTCC CCAGCTTTCT AAGGAAAACC CAACTCCTTT TTAACAAGCT 7140

TCTTTCTTAA ATTCAAAAAC TCCGCTATAA TAAAGCTATT AGTCGGAAAG AAACGAGGGA 7200

TTTATTATGT CATTTGaTGG CGTATTTACT CATAGCATGG tTCATGAGTT GaCAGAAACA 7260

CTCGTAAGTG GCCGGATTTC AAAAATCCAT CAACCTTATG AAAACGAAGt CGTTTTAGTC 7320

ATTCGTGCGA AAGGTAAAAA CCATAAATTA TTATTATCAG CACATCCTAG TTATGCGCGC 7380

ATACAACTAT CTACAATTAC CTATAGCAAT CCGGAAACAC CTCCTAATTT TGTGATGATG 7440

TTACGAAAAT TTCTAGACGG TGCCATTTTA GAAACGATTC AACAAATAGA AAATGATCGC 7500

GTGATTCACT TCACCTTTTC AAAAAGAGAT GAACTAGGTG ACTTACAAAA TATTGTTTTG 7560

ATTGTTGAAT TAATGGGACG TCACAGTACT ATCGTTTTAG TTAACCAATC TTCTGGAAAA 7620

ATCTTAGATG CGATAAAACA CATTGGAATG AGCCAAAATA GTTATCGTTC GTTGCTTCCT 7680

GGAGCAACTT ATATTGAACC GCCAAAACCA ATGGGACTTA ATCCGCTCAC GGCCTCAAAA 7740

GAAGAAGTCT TTGCGCTTCT TTCTACAGCT CCTGAATTAA ACGGACGCTA TTTACAACAA 7800

CATTTCCAAG GTCTAGGCAA AGATACAGCA GACGAATTAA GTGCCCGTTT ACTTGCACAA 7860

CCAAATGAAA AAATGGCCAT CTGGACGGAA TTTTGGTCCT CAGTCACTGA AGCTGTTGTA 7920

CCGACATTAA CAGTTACTGA AAAAAAAGAA TATTTTACGC CCATTCCTTA TCAATCATTA 7980

GTTGGTGAAC AAACAACGTA TGCGACACTC AGTGAATTAC TAGATGCCTT TTATTCTGGT 8040

AAAGCAGAGA AAGACCGTGT CAAACAACAA GGCGGCGAAC TAATTCGCAA AATTGAAAAT 8100

GAACTAAAAA GAAATCAAAA GAAACTTAAA AAATTACAGC AAACCTTAGC CGATACAGAA 8160

AACGCTGAAA ATTACCGTCG CGACGGAGAA CTATTAACAA CCTTCATGGC ACAAGTCCCA 8220

AAAGGAGCCA CCGAGGTGGA ATTACCAAAT TATTATGAAG AAAATGCACC GTTGCGTATC 8280

TCGCTAAATC CTGCACTTTC ACCTAATCAA AATGCACAAA AATATTTTCA AAAATACCAA 8340

AAATTGAAAA ATGCTGTCCG AGTTGTTAAA ACACAAATTC AGCAAACACA ACAAGAAATT 8400

TCTTATTTAG AGTCAGTTGT GGCACAACTT GAAATTGCGA CACCAATGGA CATTGAAGTG 8460

ATTCGTGAAG AACTGATTGA ACAAGGCTAT TTAAAACGTA AAAAAAATAA AAAACAAAAA 8520

CAGCCGAAAA AAAGCCAACC CGATCTTTTT TATGCCACAG ATGGCACACC TATTCTCGTT 8580

GGAAAAAACA ATTTGCAAAA TGATCAATTG ACTTTACGAA CCGCTAAAAA AACTGATTAT 8640

TGGCTCCATG CCAAAGATAT TCCTGGATCT CATGTAATTA TTCGTGACGC ACATCCTAGC 8700

GAAGAAACCC TTACTGAAGC CGCACTTCTA GCAGCTTATT TTTCAAAATA TCGACTTTCT 8760

TCACAAGTGC CTGTTGACTA TGTCCAAGTA AAACATGTGC ATAAACCCAA TGGAGCAAAA 8820

CCAGGTTATG TGATTTATGA AAACCAACGA ACGTTATATG TGACGCCAAC AGAAGAATCG 8880

ATCAAAAAAA TCCAGcAAAA TAAGGCTTCT GCTTCCTAAT TGTATTAAAC TAAAAAACAA 8940

CGTACTGcCG ATTGATCGTA CGTTGTTTTT TTGCATTGGC AAAGACATTT ATCAAGTGAT 9000

AAATAACACA AGTAAGCTCA ATTAGTTTAC AAACTAGCTA TTTCTCtCTA TACTGACTGT 9060

TGGAGGTGAA ATTCATGGTC TTTTTACACA TTATTTTTGG GATGTTTGCC TTTATTGGCG 9120

CTTTAGGTTT AGGTGTCGCC TTCCGCGATA GCTTTACAGA ACTTTCAACA CC 9172

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 788 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

ATTATTGATC GAGAAAATGA AATTCGAGAA TTTAAACCTG AAGAATATTG GAGCATTGAT 60

GGGAATTTCC AAAAAGGTCG CAAGAAATTC AAAGCAAATT TCTGGGGCGT AGATGGTAAA 120

AAGAAAAAAT TACCAAATGC AGAAGCGGTG AAAGAAATCA CTTCTCGGAT TGACGGTAAA 180

GATTACGATG TTACAAAAGT TGAGAAAAAA GAACGGAAAC GTAATCCTGC ATTACCTTTT 240

ACAACCAGTA GTTTACAACA AGAAGCGGCA CGTAAATTAA ATTTCCGTAC CCGTAAAACC 300

ATGATGGTTG CGCAACAATT GTATGAGGGA ATTGCTTTAG GGAAACAAGG CACTGTGGGG 360

CTGATTACCT ACATGCGTAC AGACTCAACC AGAATTGCTG ATTCCGCCAA AGCAGAAGCT 420

GCTGAATTCA TTGAAAAGAC cTACGGAGAT GAGTTTTCAG CACATGGCGG CCGCAAAGCA 480

AAAAATACGC AAGGTGCACA AGATGCCCAT GAAGCGGTTC GTCCGTCAAG CGTTTTAAGA 540

ACACCTGATG AAATGAAGAA GTATTTAGAT AAAGATCAAT TAAAATTATA TACTCTTATT 600

TGGTCGCGCT TTGTTGCTAG TCAAATGACA CCAGCGATTT TAGATACCAT GAaGGTCACG 660

CTTCAACAAA ACGGCGTAAC GTTTATTGCC AATGGCTCCA AAGTGAAATT TAAAGGGTTT 720

ATGCAAGTAT ATGTTGAAGG ACGCGATGAT GGTAAAGAnG ACAAGGAAAA ATATTTTACC 780

TGAAnTAG 788

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4313 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

TTATGATTAA GGCAACAACG CCCCAAGCAC GATATGGAAA TCCCACGCCA CGTGTAGCTG 60

AAACACCTAG TGGTATGCTA AATGCCaTTG GATTGCAGAA TCCAGGTTTA GAAGTGGTTA 120

TGCAAGAAAA ATTACCGAAA TTAGAAAAAT ACCCTAACTT ACCGATTATC GCAAATGTTG 180

CAGGTGCTTG CGAAGAGGAC TATGTGGCTG TTTGCGCAAA AATTGGGCAA GCGCCTAACG 240

TCAAAGCAAT TGAACTGAAT ATTTCCTGTC CCAATGTGAA ACATGGCGGC ATAGCTTTTG 300

GTACAGATCC AGAGGTTGCT TTTCAATTAA CACAAGCAGT CAAAAAAGTC GCCTCTGTTC 360

CGATTTATGT TAAACTATCA CCAAACGTAA CGGACATTGT GCCGATCGCC CAAGCAATTG 420

AAGCAGGGGG CGCAGACGGT TTTTCTATGA TTAATACATT GCTAGGGATG CGCATTGATT 480

TAAAAACGAG AAAACCTATT TTAGCCAATC AAACTGGCGG ACTATCTgGC CCAGCCATTA 540

AACCTGTTGC TATTCGCTTA ATCCGACAAG TCGCCAGTGT TTCTCAGTTG CCAATTATTG 600

GAATGGGCGG TGTTCAAACA GTGGATGATG TCTTGGAAAT GTTCATGGCA GGTGCTAGCG 660

CTGTCGGTGT TGGTACGGCC AACTTTACAG ATCCATATAT TTGTCCGAAG TTGATTGACG 720

GATTACCTAA GCGAATGGAA GAGTTAGGTA TTGAATCGTT GGAACAATTA ATTAAAGAAG 780

TGAGAGAGGG GCAGCAAAAT GCACGATAGA CCAATTATTG CGTTAGATTT TCCCACGCAA 840

AAAGAAGTGG CTGTTTTTTT AGAAAAATTT CCTAAGGAAG AAGCCTTGTT TGTTAAAGTG 900

GGGATGGAAC TTTTTTACGC AGAAGGACCA GCGATTGTTC GCTGGCTTAA AGAACAAGGA 960

CACGATGTTT TCTTGGATTT AAAATTACAT GATATTCCGA ATACTGTGGA AAAAGCGATG 1020

ACAAATTTAG CAAAATTAGG CGTTGCCATT ACTAATGTCC ATGCAGCAGG GGGCGTTCGA 1080

ATGATGCAAG CAGCTAAGGA AGGTCTAATC AAAGGGACTC AGCCAGGCGC TAAGGTTCCT 1140

GAGTTAATTG CTGTAACCCA GTTGACTTCT ACTAGTGAAG AAGAGATGCA CCATGATCAA 1200

TTAATCAACG TTCCTTTAGA GACAAGTGTC ATTCATTATG CTAAATGTGC GGAAAAAGCG 1260

GGTCTGGATG GTGTGGTTTG TTCAGCATTA GAAGCACGTG GTATTCAAGA AGCAACCAAG 1320

CAGACGTTTA TTTGCTTAAC ACCTGGTATT CGTCCAGCAG GAAGTGCAGT AGGTGACCAA 1380

CAACGGGTCG TTACGCCGCA ACATGCTCGC GAAATTGGGT CCACATATAT CGTTGTTGGC 1440

CGACCTATTA CCCAAGCAGA AAATCCTTAT GAAGCCTATC AAGAAATCAA AAAAGACTGG 1500

AGCGAAAAGT AACATGACAA AAGTAGCAAA AAAGATTGCC AAAGATTTAT TAGATATTGA 1560

AGCGGTATTT TTAAACCCTA ATGAGCCGTT TACCTGGGCA AGTGGCATAA AAAGCCCGAT 1620

TTATTGTGAT AACCGTATCA CGATGAGTTA TCCTGCTGTG CGTAAAGAAA TTGCAGAAGG 1680

CTTAGCGGCA AAAATTAAAG AGACTTTCCC AGAAGTCGAA GTGATTGCTG GGACAGCTAC 1740

CGCAGGGATT CCACATGCGG CATGGGTAGC CGATATTTTA GGATTACCAA TGGTTTATAT 1800

CCGCAGTAAA GCCAAAGATC ACGGCAAAGG GAATCAAATT GAAGGCCGAA TTTCTGAAGG 1860

ACAAAAAATG GTAGTTATCG AGGATTTAAT TTCAACAGGC GGTAGCGTGT TAGAAGCGGC 1920

AGAAGCGGCG GAACGTGAAG GAGCAACGGT TTTAGGCGTT GCGGCAATCT TCACTTATGA 1980

GTTACCCAAA GGAACGGCAA ATTTTGCAGA TAAACAGATG ACGTTACTAA CGTTAACCAA 2040

TTATTCAACT TTAATTGATG CTGCATTAGA AGCAAATTAT ATCGAAGAAA AAGATGTTAC 2100

CTTATTACAA GAATGGAAAA AAGATCCTGA AAaCTGGGGC AAaTAACCGT TGAACTAAaG 2160

AAGGTTCTTG TCGTCTGATT TTTAGTGTCA TTTTCAGTAA CCAGATTTAA CGGTAGGAGA 2220

ACTTAGGTCA TTTGCCGAGG AGCGTAATGA TCTAAGTTCn TTTGTTTGAC GTAGAAAAGA 2280

TGTTAGGATA GAGTTTTAAC ATGAGTTTTA CCTGATTACT ATGTGGAAAT GrAGGaAGAG 2340

TGTATGAAAA AAATAAAAAA TATGGACGTk GAATGGAGCT ATACAGGCCA ACGATGGACC 2400

CAGAACATTG GCATACGCTA TGTGACTGGT TTGCAGAAGG TGCTAAGTTT GCTTATCAAT 2460

CGCCAaTTGC TTTAGAAAAA GAAaGTGCTG AGACAGTAAA TTCTCAGATT ACGTTTCATT 2520

ATAAAAAAGA AGAATTCACT GAAAAAGAAT TCAAAAATAC GTTTCATTTT GTACCACCAA 2580

ATACGGAAAG TTACGTAATG TTTGAAAATG TTGCGTATCA TTTAACCGAT ATCCATTTTC 2640

ATATGCCTAG CGAACATTTG CTATCGGGAA AGCAGTATCC GCTAGAATTC CATTTGGTTC 2700

ATATGAACGA TGCTGGCGAA AATTTAGTTG TCGGTTGCCT GTTTACCATC ACAGAGGAAG 2760

AAAATCGTTT TTCAGAAGCG AATCATCCGA TGGACTGGGA AAACGGCACG CACCAACAAT 2820

GGTTTAATCC GTCAATCTTT TTACCAGAGG AGCGTCTACA CTACCATTAT GTAGGCTCGT 2880

TAACAACACC GCCAACTAAG GGTCCAGTGA AATGGTTTGT ATTTGATACC ATTCAAAAAA 2940

TGGATCAAGC CTTTTTAAAT AAGATTAAGG AAGGTATGTT GGCCTTTAAC AATCGGCCCC 3000

TTCAGCCGTT GAATGGGCGT AAGATTTATT TTTCAAATGA TTAGTGTGTG GAGGAAAAAG 3060

GGTTGAATAT TCAACAAATG AAGTATGTTG CTGCCATAGC TAACAATGGG AGTTTTCGGG 3120

AAGCAGCCAA AAAACTATTT ATTACACAAC CTAGCTTGTC AAATAGTATT CGAGAATTAG 3180

AAGAAGAACT AGGGATTTCC CTCTTTTTAC GTACCAATAA GGGCGCTTTT TTAACAGAAG 3240

AAGGAATGGT TTTTTTAGAA CAAGCAGAGA AAGTTTTAGT TCAAATGGAG TTATTGGAAA 3300

ATCGCTATCG CGAGACAGTG ACCAGCGAAC GATTCTCCAT CTCTTCGCAA CATTACGATT 3360

TTTTAGGTGA GGTAATTGCC AAAGTTCTTA AAAAGTATGG GGATCAATAT AAAGATTTTC 3420

GTGTCTTTGA AACCACGACC TTAAAAGTAA TTGAAGATGT CAAAGGATTT CACAGTGAGC 3480

TGGGGATTAT TTATTTAAAT GAACAAAATA GTGTCAGTAT TGAACGGTAT TTAGAACAAG 3540

CTAATCTAGC CTACGAGGTA ATTAGTACTT TTAATACACA CATTTTTTTA GGAAATCATC 3600

ATCCTTTAGC CAAGCAAAAA GAAATTCATT TAGAGGAATT AGTTCCCTAT CCACAAGTTC 3660

GCTTTAATCA AGAAGGAAGC AATTTTTCCT ATTTTTCTGA AGATTTAGTT GAAATTCCTG 3720

AACAGGAATC GGTCATTCAT ACTACGGATC GTGGAACCTT GATGAATTTG TTGGTAGAAA 3780

CCAATGCTTA TGCTTCTGGT TCAGGAGTCG TAACGGGTTT TACGAAGAAA GAAATTCGGT 3840

TAGTGCCATT GGCGCCAcTT TGGAAAATCG TATTTGTTTA CTCTTTCCCA AAAATCGGGA 3900

AATCAGTCCA ATTGGTCGCT ACTTTATCAA AGAnTTAAAA GcGTTGTTtA AAAAAgAGGT 3960

AGATACCAAA ACATTTmGCa CTAAGGaTaA AAAGAAGAGg TCGGGACAGA AGTGTTTAAC 4020

TCCGAGAAAT AAGAAGGAAT TTCCGAAAAT TGTTCTTTAA TTTTCGGAGA ATTTCGGCTT 4080

ATTTCCGAAG GATTTGCTTC TGTTCCCGCC GTTTATTAGT TTTTGAGCGT GGAGCAAAAA 4140

TCCAAAGTGA TTTTTGTCCC ACGCTCTTCT TTTTTGGCTC ATTTTGAAAC AGTAGAAATG 4200

GTTkCTCTTT CAATTAATGT GACAGGCACT TGAATGTCTT CTATGATTTG AAGGTCTTGT 4260

TGCATCGTTT GTAATAACAC TTCACCAGCT AGTTTTCCCA TTTCTTTAAA GTC 4313

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8921 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

ATTnTTnCAT GCGATAAATG GTATTCCTGG TTTATCAGCC GTGAAGCCTA AAGCAGCTTT 60

CTATATTTTT CCGAAATTGA TACGGCCAAA TTTGACTTTT ATGATGATGA AAAATTTGTT 120

TTaGATTTCT TACACAAACA TCATATTTTA TTAGTCCATG GTGGTGGGTT CAACTGGCAA 180

CAACCCGATC ACTTCCGTAT TGTTTATTTA CCGAAGATGG AAGACTTAAA AACTACAGCA 240

GATAAAATGC GTGAATTTTT AAGTACCTAC AAACAAAAAT AAAAGAACTA CCAAAAAATA 300

GCCTCTAACG AAGCTATTTT TTGGTAGTTC TTGTTTTATC AACAATACTA TTCTGCGGTT 360

TCACATGCTT TTTTACTAAT AATCATTCGG TTGGTATATT TTTCTTCCAT ATCAGGACTA 420

GGATTTGCTA CTACGATGAG ATAAGAATTT TCATACTCCT TCTCCACAAA GCCTGAATAA 480

TTTGTTTTTT CCCAAATGAA GTGAACCTTT TTCTCTTCGG CCATACAATA ACACTCCTTT 540

TTAATAAAAA TGAACCTATT TATTTTCTAA GAAAACCTTA GATAATGTTC TTGCAGTATA 600

TTATGCAATA AAGTACAATA AAAGCATAAT GCGAATTCGC ATATTTTGAC AAAATCCCCA 660

TTTTGCAAGG AAAATGGCCG ATATAAATAG GAGCTATTGT GATGAATTCT TTAGGTGCAG 720

TTATTAAAGA AATTAGAAAA AATAGAAAAT TGACACAGAA AATGTTATCA GAAGACATCT 780

GTTCGCAAAG CGTTTTAAGT CGCATTGAAA ACAATGAAGA GTTGCCAAAC GTTCTCGTAA 840

TGCAACAACT TTGTGATCGT TTAGGTGTGA CTGTTGATCA GATTATGCGT TATAAATCTG 900

GTGATGTGCA CGTTGTTACG TATTCATTTG AAAAAATGGC TGAATATTTT CGCCATAAAA 960

AATATCAATT ATTATTAAAT TATTTAAAGG AAAATCGTAT TGAAGAACAA TTGTATTTAG 1020

ATACGGACTG GCAAAAATAT TATTATTATT TAGGtAGTTG TGAATTGTTT GTTCTAAATG 1080

ATTATGAAAA AGCTATTGCT AGCTTGCGTA AAGGGTTATC TTTCACCTAT AAAGCCGATA 1140

AGTTAAATGT TTCTGATTTA GAGATTCAAT TAATTAGTTG TCTAGGTGCC ACTTATGGTT 1200

ACATGGGCAA TCGAGTAGAA GCAGAACGAT TTTTAAGTTT GAGTATTCAT TATTTTAATC 1260

AATTGCCAAA TGAACGCAGT AATGCAGAGT TAACGAAAAT ATTTTTTAAT TATGCTGATT 1320

TCTTATTTAA AAATTATACA GAAAAAGATG CCGAAATTTA TGTGGATCAA GGAATTACTT 1380

GGGCCCGCAA AAAAAATAGT TATTATTATT TAAGTGAATT ATTAAATCTA AAATATTTAG 1440

TTTTAATGAG AAAGAATAAA CACGAAGAAG CAGAACGTTA TCTGAACTTA GCACAACAAA 1500

TGAAAAATGT GGAAAGTGGC AATCTTTAAA AAATCCTCAG AATTGCTTAA AAGGTCTACA 1560

CCATTTCTTG ACTTTTGTAC TGATTAAGAG TACTTTTGAA GTAAGTAAAA TAAAGAAAGA 1620

AGGATGAAGA TGAAAACATT TATCTTTGCA GACAAGTTCT TTTTAAAAAG CGATGTTAAA 1680

GGACCTGGTT ATTTAGAAAT CACGGATGGT ATCTTTGGTA ATTATACAAA AGATGAACCG 1740

CAAGGGgACG TAAAAATTAT TCGTGAGGAA GGCAAATGGA TTGCCCCAGG TTTAGTCGAT 1800

ACTCATATTC ATGGTTACAT GAATCATGAC GTAATGGATA ACGATGCAGA AGGCATTAAA 1860

GTGATGTCTG AAGGCTTATT ATCTTGTGGT GTAACTTCAT TTTTACCAAC AACGTTAACT 1920

TCAAGTAAAG AACGTTTAAC AGACGTTGCA CGTACGATTG GTCAAGTCTA CCAAGAAGTA 1980

CCTGGTGCAA AAATTCAAGG GATTTACTTT GAAGGACCTT TCTTTACTGA AGAACATAAA 2040

GGCGCACAAA ATCCAAGTTA TTTTGGAGAT CCAGACCTTG ATACATTCCA CGAATGGCAA 2100

GAAGCTTCAG GCGGAATTAT TAAGAAAATT GCGTTGGCAC CAGAACGTAA TGGTGTGAAA 2160

GAATTTGTTG AAACCGTTAC AGATGAAGGT GTCGTTGTTG CTTTAGGCCA TAGTAATGCG 2220

ACCTTAGAAG AAGCTGATGT GGCAGTTGAA GCAGGCGCTA gCGTTTTTGT TCATGCATAT 2280

AATGGGATGC GCGGCTTAAA TCACCGTGAA CCAGGGATGG TTGGTGCATT GTTAACGTTA 2340

CAACACGTTT TCTCTGAATT AATTTGTGAT GGACACCATG TACATCCGCA AGCGGCTGAA 2400

GTATTGATGG AAAAAGCGGG CCATGATCAC GTTGCGTTAA TTACAGACTG CATGATGGCT 2460

GGCGGTATGC CAGATGGCAA CTATAATTTA GGTGAATTTC CTGTTGTGGT AGCAGAAGGA 2520

ACGGCTCGTT TGGATACTGG GAACTTAGCA GGTAGTATTT TAAAACTAAA aGAaGCAaTT 2580

AAAAATGTAG TTGATTGGGG TATTGCAACA CCTGCACAAG CAATTATGAT GGCTTCATTA 2640

GTTCCtGCGA TTAGCTGTAA AATTGATGAT CAATGTGGCA TGATTGCAAA TGGTCGCGAT 2700

GCTGACTTTA TCGTATTAGA ACCAACCATG GAATTAGCAG CCACTTATTT AGATGGCGTA 2760

GAACGTTATC GAGCATAAGA CACAAGAAAA AGCTTGCGGA AATTTATCCG CAAGCTTTTT 2820

TTATAATTCT TTTGGTAAAA TCGCAGAAGT GATTTCTTTG TTCTCGTAGG CAATAAAAAA 2880

ATCGATTAAT TGCTTGTCAA ATTGCTCAAT ATCCTCTTTG TAAAGGGAGC TAGGAATCAC 2940

AAAGGTTTGC TCTAGGGTGC CTGCTAGATA ATCGCCATCT AAACTTTGGG CGACAACTAA 3000

AAATTCTTGT TCGCGCCAAT GTTCCACATC TGTTTCAGAA ATGGTTTCTG GAGAGGCTAA 3060

AAATTTACAA TTTTCGACAA ACTGTTCATC AATTTCTTCA CTGAAAGTTA ATTCAGCATC 3120

CTCTTCTTGA GGTTGAGAGA GGATTGCTTG CAAAAATGTT TGTGGTTTTG CATCGGGTAA 3180

TGTTTCTTTT AGCTGTTCAA AATCCAATGA AAGTTCCTCC TTATTGTTTG TTTAAAATAG 3240

TTACTGAAAA ATCAATGGGT GAATTTGTTT TAAGGCTTTT GGCTTCTAAG ATGATTTTTT 3300

CAGCCAACAA ATCGTTTCCG TTATATACAG GTGTCACGCG ATAGCGAACG GTTTCGCCGC 3360

GATTAAGAGC TGCCCGAATC TGATTTTCAT ATTTCGTCAT ATAAGGAGTG TTGACAGGAT 3420

TCTGATACAA GGTGGTTAAG TTACGTGGAT CATCACCACT GCCTCCCATT TGACGGCCGA 3480

TTAGGTGGCC ACGCGAATGA TTAGCTTTTC CTGAAATGAA ACCAGCTGGT CGAATATCTT 3540

TATTAGCGGA AGTTCCTGTA TTGACCATTG CGGGTTTTAA TAGCGCATCA GCGCCTGTCG 3600

CTCGTTTTAG ACGATCTAGA GAATGATAAG TAATCCAGCC GTTTGAGTTA TCCTTTAATT 3660

CCTCTGCTGA AAATGTGGCT TTACCATTTT CGATGGGTnC GGGATTATCT CCTTTTGGTG 3720

CTGGTGTGCT TGAAaGGGgT GGTTGTTTGT GGTTGAACAT TAAATAAATC TTGTAAAGCA 3780

TCAGGAACTT TAACGCCAAC CGCGCCTAAA ATTAAAATGA TTAATGCGCC AATCAGCATG 3840

ATGGTTGAAT TAAATGGTGG TTTCTTCTTT CTGTTTGCCA TAACGTGCTC CTCTAATTTT 3900

ATACTTGCAA TAGTCCAAAT TTAAGTGGAC CATCTTTTCT ATTATACGTA AATTTGCTAA 3960

AAAATCAAAA AAGGCTTTTC TTTGAAAGAC TTGTTTTCTC AGATATGTTA CTATTTTAAT 4020

GTAGAAGGGa AgGGATAGTG GATGGCCTAT ATTGAAGTCA AAAATGAATA CAAGCGTTAC 4080

CAAATGGGAG AGACAACTAT TACTGCGAAT GACGGCATTT CTTTTGAAGT GGAAAAaGGA 4140

GAAGTAGCGG TTATTTTAGG CCCCAGCGGT GCtGGAAAAT CAACTGTTTT AAATATTCTT 4200

GGTGGAATGG ACAGCTGTGA TGAAGGGGAA ATTATTATTG ATGGCACAGA TATCGCTCAG 4260

TTTTCAGAGA AACAATTAAC CACTTATCGA CGAAATGATG TCGGGTTTGT TTTCCAATTT 4320

TACAATTTAG TCCCAAATTT AACGGCAAAA GAAAATGTCG AATTGGCTTC GCAAATCGTT 4380

GCGGATGCGT TGGATTCGAC CAATGTTTTA CAATCGGTGG GTTTAGGGGA ACGTTTAGAT 4440

AATTTTCCAG CGCAACTTTC TGGTGGTGAA CAACAACGAG TAACCATCGC CCGTGCGATT 4500

GCTAAAAAGC CCAAGTTACT TTTATGTGAT GAACCGACAG GAGCGATGGA TTATGAAACA 4560

GGGAAACAAA TTTTAACCAT TTTGCAAAAC ACTGCTCGTG AAACTGGAAC GACCGTCTTA 4620

ATTATTACCC ACAATTCAGC CATTGCAGAA ATGGCCGATC GGGTAATTCG AATTAATGAT 4680

GCGAAAGTTC GCGAAATGAC TGTGAATGAT CAGCCTAAAT TGGTTGCTGA AATTGAATGG 4740

TAGGTGCCTA AAATGAAGAA AACAGCGCTA ATTAAAACAA GCTTACGTGA AATTAGACAA 4800

TCTACTACAC GATTTTTATC GATTATGGGG ATTATTTTCC TTGGTGTAAT GGTCTTTGTT 4860

GGTTTGAAAG CTACTGGACC AGATATGATT AAAACCGCGA ATAATTATTA CCAGAAAGAG 4920

CAGCTTCCAG ATGCGCGAAT AATTTCTACA ATGGGACTGG AAAAGAAAGA TTTAACAACA 4980

CTACAGTCTT TAAAAGATGT CGAAACAGTT GTGCCGCGGT ATACAATAGA TGCCACGATT 5040

GGGCCGCAAA ATAATGCAGT AAAATTGTTT GGCTATCGAA AAAATCAAGC AGGATCGGTT 5100

AACTATCAAG TAGTTGACGG TCGGTTACCT AAACAAACAA ATGAAATTGC ATTAGATACA 5160

TTAGCGAAAA CTAGATATGA CTATAAAATG GGTGATAAAA TCACTTTAAA CGATGCAGCG 5220

ATTAAAGAAA AAGGGCTGAA ACAAACTCAT TTTACTGTTG TTGGATTTAT CAATTCACCA 5280
GAATATATTG ATAATACATC CAGGGGAACC ACAACGGTCG GGTCTGGAAC GTTAAACTTT 5340
TTTGGCGTGG TTTCGGAAAA AGCCTTTGAC AGTCAACGGT ATCCAGAATT GCTAATCTCA 5400
TTTCGTTCAT TAAAACATCA AAATACGTAT TCCTCTGATT ATGAAAAGAA AAGAGAGCAG 5460
GCCTTAAATC AAGTAAAAGA GGCACTGAAA AATCGGCCAG AAGAACAGGT TGCGGCTCTG 5520
CGTGATTCGG CTCAGCCGGA CATTAACCAA GGTCGGCAAG CGCTGGAAAC GGGAAAACAA 5580
GCACTTGCGC AGTTGGaGCA ACAACCAGGA ATTCCAGCTG AGATGCTTGA AAAGCAAAAA 5640
GACGAATTGG CTAAACAAGA ACAACTGTTG GCTGAAAAAG AAACAGAGCT AGCCAATTTG 5700
AAAGCCCCaA CGTATTATTA TTTTACTAGG GAAGACAATC CAGGATTTTC TGAATATCaA 5760
GATAACGCTG ATCGGATTTC TTCTTTAGCC ACGCTGTTCC CGTTGTTCTT TTTCCTAATT 5820
GCGGCATTAG TCAGCTTAAC TACCATGACA CGTATGGTTG AAGAAAAAAG AATGGAAATT 5880
GGTAGTTTGA AAGCTTTAGG TTATCGTAAT CGTGAAATTG CTTCGATTTT TATTACCTAT 5940
GCAACGGTTG CCAGCTTAAC AGGTGCATTA TTAGGATTGG CAGTTGGCTA TTATCTATTC 6000
CCTAAAATAA TTTTTGATGc CTACGGTCAA ATGTACAATA TTCCTGATTT AGTTACACCG 6060
TGGTACTTAA ATTACAGTTT ATGGGGCATC ATTGTAGCCT TAGCTTGTAC TGTGGGAGCG 6120
GCCTTAGTGA CATTAAGAAT CGATTTATTA AGTACCCCAG CTACCTTGCT AAGACCCAAA 6180
GCGCCGAAAG CAGGCCAACG GATTTTACTA GAACGAATCC GCCCTTTATG GCAACGCATG 6240
AGCTTTATTC AAAAAGTGAC CGCTAGAAAT CTGTTCCGCT ATAAACGACG GATGCTGATG 6300
ACCGTTATCG GAATAGCTGG TTGTATGGGG CTGTTAATTG TTGGTTTTGG TTTACGCGAT 6360
TCAATTGTGG ATGTCGCTAC AATTCAATTT AATAAAATCT GGCATTACCA AGCAGTCGTG 6420
ACTTTTAAAG AGCAAACAAC ACCTGAAGAG ACAAAAGAGT ATCAAGAAAC GTTGCGCAAA 6480
ATGGATGGTT TGAATAAAAC AATTCCGTTA TATTCAGAAA TCTTTAAAAC GAAGGGCAAA 6540
GGCGCACCAA CTCAAAATAT TACTCTTTAT GTGCCACAAG ATCCTTCGAA AATGGCTGAT 6600
TTTCAATTAT TTAATGATCG AGTCACTGGA GAAAAATATT CCCTCAACGA TGACGGTGTA 6660
ATCATTAATG AAAAATTAGC TAAATTATTT GGCTATAAAG TCGGAGATCA ATTAAATTTA 6720
GAAAATAGTG ATAATCAAAC CTATCACGTG AAAATTGCCG CGATTGCAGA GAACTACACG 6780
GGTCATTTTG TCTATATGAC GCCGAAACTC TACCAGACGA TGACGAAACA GAAGCCGGAG 6840
TATAATACAG AATTCCTACT TTTTGACAAA AAGTTATCCT CAAAACAAGA AACCTCAATC 6900
GGAGAGGCGT TGACAAAACA GCCGAAAGTG TTGAACATTA CGTTTTTAAC TGCGATGAAA 6960
GGTTCGTTTG ATGATATGCT GAAAAGCCTT GATATTGTCA TTTGGGTACT AATTGCTGTT 7020
TCTGGTTCTT TAGCGTTAAT TGTGTTATAT AATTTAACCA ACATTAATGT TTCTGAACGG 7080
ATTCGAGAAC TTTCCACGAT TAAGGTTTTA GGGTTTTACG ATCGTGAAGT CACCACCTAT 7140
GTTTACCGTG AAAACATTAT TTTAACGTTC ATTGGTATTA TTGTTGGCTG CTTCTTTGGG 7200

AAAATTTTGC ATCAGTATAT TTTAGCAACC GTTGAAGTTG ATTTAATTAT GTTTTCGCCA 7260

ATTATTCATT GGCCAAGTTA TCTGTATTCC GCAGTCATCA CCATGTGTTT CACGTTGTTT 7320

GTGATGGTCA TTATGCATCG GAAATTGAAG AAAATCAATA TGATTGAAGC GTTAAAATCA 7380

AATGAATAGC ACATTTTAAA AAaCGAGGAG TCCGTAGAAG TTCGATAATT CTAGGGCTGC 7440

TCGTTTATTG CTGCTTCTTT TATTAAGAAA AAATAGGTCT TAAGATGTAT CTTTTGAGAT 7500

TTATTTGGAA ATTACTACTC TTTTTATTTA AAAATTGGTA CACTATTACT AGGTTATAGC 7560

ACAGAAAGGA GTGTGGGAAA AGTGTTAGAA GGTCTACCAT TGGAAACGGT TTATTTTTAT 7620

GCATTAGTGG GGAGCsCgGT CCTCGCTTTT TTACTAATTA TTTTTGGGGA TGTTTTTAAT 7680

TTTGATGGAC CGGTTGaCCC AATGCTGATT ATTCCATGGA TTGCTTTTAC CTCGCTGTTT 7740

GGGTACCTTG GGGAAGAATT AACAGCGGTG AATAGTTGGC TTATTTTAAT TGTCAGTGGT 7800

ATTCTTTCAA CCATTATTGT TTTCTTCCTA AATTTTTATG TGTTGGTCCC TTTGAAAAAT 7860

TCGGAGGCCA CAATTTCTAT TTCGGAAAAA GACATGGAGG GTCGTGTTGC TACTGTGATT 7920

ACCCCAATTC CTGTCCGGGG AATGGGCGAG ATTCAATTGA AGAGTGTCAC AGGTTCGTTA 7980

AGTCGACCAG CAGCATTTTA TGTGCCACAA GAAGTGGCTG CGCCAAGAGG TAGTGAGGTA 8040

TTGATTATTG AATTGAAAGA ACGGGTTTGT TATGTCATTC CTTACGAAGG CAGTCTAAAA 8100

ATATAGAGTA GGAGAGAAGA AAATGGGGAT TTTGTTACCg ATTATTATTG CaGTTTTAGT 8160

ATTATTAATG TTATTAATTG TGTTTGTTTC TAAATATCCA GACAGCCcAA ACCTGATGAA 8220

GCGTTAATTA TCcAGCGGGA GCTATCTAGG CTCTAAAAAT GTTCATGTAG ACGAAGGTGG 8280

CAACAAAATT AAAATCGTTC GTGGCGGTGG TGCGTTTGTC TTACCAGTGT TCCaACGTTC 8340

AAATCGAATT AGTTTGCTTT CAAGTAAATT AGATGTTTCC ACACCAGAAG TGTACACAGA 8400

ACAAGGGGTT CCTGTGATGT GTGATGGGAC ATCAATCATT AAAATTGGTT CGTCAGTTGA 8460

AGAAATTGCG ACAGCAGCGG AACAATTTTT AGGAAAAACA ACGGAAGAAT TAGAAAATGA 8520

AGCACGTGAA GTATTAGAAG GACATTTACG TTCGATTTTA GGTTCAATGA CAGTGGAAGA 8580

AATTTACCAA AATCGTGATT AATTTAGCCA AAGTGtACAA GAAGTTGCCa GTGTTGACTT 8640

AGCTAAAATG GGCTTAGTTA TTGTGTCGTT CACAATTAAA GAAGTTCGTG ATAAAAATGG 8700

ATACTTGGAT TCATTAGGGA AACCAAGAAT CGCTCAAGTT AAACGTGATG CAGATATTGC 8760

AGAAGCAGAA GCCTTGAAAG AAACTCGCAT CAAAAAAGCA GAAGCAGAAA AAGAATCACA 8820

ACAAGCGGAA TTGCAACGTC AAACAGAAAT TGCAGAAGCT TCAAAAGAAA AAGAATTGAA 8880

ATTAGCGTTA TACAAACAAG AACAAGATAT TGCCAAAGCC A 8921

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5523 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

TGCAAATGCA CGAGGAGGAA CAAAAGAATG GAACArGGAA CAGTAAAATG GTTTAACGCa 60

GAAAAAGGTT TTGGGTTTAT CTCACGCGAA GACGGAAGCG ACGTATTCGT ACACTTCTCA 120

GCTATCCaAG GTGATGGTTT CAAAACTTTA GAAGAAGGTC AAGCAGTGAC TTTTGATGTT 180

GAAGATTCAG ACCGTGGCCC TCAAGCAGTT AACGTAGAAA AAAACTAATA AAAAATAATC 240

CAAAAACACC TAACATGCGT TAGGTGTTTT TTTGTTAGAA AAAATAGAGA AGAAAAATGG 300

ACAAGCATTT TTCTTCTCTA TTAAATGAGA GCTTTATTGA TCACTAACTG TGGCAGTCAT 360

AATCATTGGC AGAATCATTG GATGACGTTC TGTTTGTTCA AACAAGAAAG GCTGTAAAGC 420

GGTCGTCATG GCTTCGCCTA ATTTTGCTTC GGTACAGTTT TTATCTTTCA TTGCTTCACG 480

TAGAGCATTG AATAAAAGAC GTTGGCCTTC ATGAATCATA TCACCAGACT CACGCATGTA 540

GACAAAACCA CGGGAAAGAA TGTCTGGACC AGCCATAATT TCTTTATTTT TAATATCAAC 600

TGTTGCTACC GCTAAAACTA AACCTTCTTC AGATAAGATA CGACGATCCC GTAAAACAAC 660

ATTCCCGATA TCACCGACAC CATTTCCATC AACGTACACG TCATTAGCGT TGAAGTGACC 720

AGCAGGGCGT GCGCTATCAG CCGTTAAAGC TAAGACATCG CCGTTGCCCA TGATAAAGCA 780

ATTTTCTTCA GGAACCCCTG TATCTTGAGC GAGCGAAcAT GGATTTTTAA CATTCTAAAT 840

TCACCATGAA CAGGCATGAA ATATTTAGGC TTCATCAAAC GAAGCATTAA TTTTTGCTCT 900

TCTTGTCCAC CGTGACCAGA AGTATGGATA TTATTGATTT TTCCGTGGAT AACTTCTGCA 960

CCGGCTTCAG ATAATAAGTT GATTAAACGG TTAACGCTTG TCGTATTCCC AGGAATTGGT 1020

GAACTTGAGA AAACAACCGT ATCGCCTGGT TGAATACTAA TTTGGCGGTG TGTACCGTTA 1080

GCAATCCGAC TAAGCGCGGC CATAGGTTCT CCTTGAGAGC CAGTACATAA AATCATTGTT 1140

TCATTTGCTG GCAATTGATT CAGTTCCGCC GCATCGACAA ACGTTCCTTT TGGCACTTTG 1200

ATGTAGCCTA AGCGTTCGCC ATTGACAATG GCATTTTCCA TGCTACGACC AAAAACAGCA 1260

ATTTTGCGAC CTGTTTTCAC GGCTGCATCA GCAGCTTGTT GTAAACGGAA AATGTTTGAA 1320

GCAAAACTGG CAAAGATAAT CCGACCATCA ATTTTTTCAA AGATTTTTAA AATAGAGGTT 1380

CCAATTGTTT TTTCGGATTT TGTAAAGGTG GGTATTTCAG CATTGGTACT GTCTGAAAGA 1440

AGACAAAGGA CGCCTTCTTC ACCTAATTTC GCCATTCTGT GTAAGTTGGC CGGCTCACCA 1500

ACGGGTGTAA AGTCAAACTT AAAGTCACCT GTTGCGACAA TGTTACCAGA AGGCGTTTTA 1560

ACCACAACGC CCAAGGCATC AGGAATACTG TGAGTCGTCC GGAAAAAGCT AATCGCTGTT 1620

TTTCTAAAAC GAATCACCGT ATCTTCATTA ATTTCGTGTA ATTCCGCTTC TCGCAATAAA 1680

CCATGTTCAT CTAGTTTGTT GGTAATTAAA GCCAAGGCTA GTGGTCCCGC ATAGATTGGG 1740

ATATTCGCTT GACGTAAAAG ATAAGGAACG CCACCAATAT GGTCTTCGTG ACCGTGGGTA 1800

ATAACGAGTG CTTTGACTTT ATGCAAATTT TGAACGATGT AGCTATAATC TGGAATCACG 1860

TAGTCAATCC CAAGTAAATC ATCTTCTGGG AATTTGATCC CAGCATCAAT AATGATAATT 1920

TCATCTTGGA ATTGTACCCC ATATGTGTTT TTTCCAATTT CGCCTAAACC ACCGATTGCA 1980

AAAACGCCAG TTTCGTTATT TTTTATGTTT ACTTTCATGT TAAAACTCCG TTAGAGTGAA 2040

GTCCGCATGT TCTTGCTCGT AAGCAAGATG TTTTTCGTCT AATAATTGAA CGTATTCAAT 2100

ATTGTATGGG GTATTTTCTT CAACTAATTG GCGCACGATT ACATCGCTAT CGGCTTCCAT 2160

GTAAAGAGAT TTTGTGTCTT CTCTTTTTGG GTTTCTGATT TTTGTTTCTT GATAGTATAC 2220

TTTGTAAATC ATCGATTCAT TTCTCCTTCA GCTACTAATA GTAGCAATTT ATAGTTTTTT 2280

TCATTTTAAG ACAGAAAATT AAACACAACA AAACGAGGTG CTGATCACCT AAAAAAGAAT 2340

GCTCATCAAA AACATCTGTT AATGATGAAA TGAAGCCTCG ACTTTGCTTC ATTCTGGGTA 2400

TTTGATTTTA TGTGTTCTTT TTCAACCGTC CCGTGTGTTG TGAATAACAA CAATTAAGCT 2460

TATTTTATCA TAGAATGATA AGGAAGTATA GAAACACTCA CTGGACTTTT AAAAAAAACA 2520

AACTTCTTGT TTCGCCTTAT TCTTGATCAA GTTCGAGTGC CTTTTTTTTA GGACGCCAAT 2580

ACCAACTTAA GCCAATCACG TGAGCTAAAT AAACAATCGC TAAAGTCGTA GTCATCAACG 2640

TATTTCGGAC AATTAAACAC GGAATGGCAA ATAAAAGAAA CATGGAAATA AATAGTTTCC 2700

ATTTACCTTT ACGAGTAATT TTTTTCTCAA AGAAGGCTTC TTGGATATAT TTTTGATAAT 2760

AAGAAGAAGC TAAAAATTTG TTATAAAGTT TATCTGAACT ACGCATCCAA AAAAAGGCCG 2820

TTAATAAATA AAAACCAGTG GTGGGTAAAA GCGGCAAAAA GATTCCTAAC GTGCCTAAAG 2880

CAAACGTACA GGAACCGAGA AAAATGAAGA ATAATTTTTT CATTCAAAGG GTCCTTTCAG 2940

TTTTTCCCAA TTATGGCACA ATCAAAAGGA AGTAGCAATT TTAATCGTAT AAGTAAATCA 3000

TTAAAATTAT CTATTTTTTA GAAAATCTAA AAGAGAAAGG ATAGAAAGAG AAAGGCTGTT 3060

TTTTCTAGCC AGACAGCTTG TATTACCGAT TTCTTTTTGA TACAGTGGGA GTACTAGAAA 3120

ATTTAAAGTA AGATAAGAAG GAGATGATTA ATTGAAGTTA ATGTGGCGTT ACACAaTGCG 3180

TTATAAGAAA TTACTTTTCG CTGAtTTTAT TTGTGTGTTT GGTTTTATTT TGATTGAATT 3240

AGGCTTGCCG ACCATTTTAG CACGAATGAT TGACAAAGGA ATTATTCCTA GAGATATGGA 3300

CTATATTTAT CAGCAAGGGA TTTGGATGGT TGTAATTACG ATTAGTGGCG TTGCAATGAA 3360

TATTTTGCTA GGGTACTTTG GTGCCAGAAT TACAACAAAC ATTGTTCGTG ATATTCGTGA 3420

TGATTTATTT GAAAAGATTC AAACTTTCTC ACATAGTGAA TATGaAAGTA TTGGGGTTTC 3480

TTCCTTAATT ACGCGAACTA CCAATGATGC ATACCAAATT ATGCTCTTCA TGGGAAATAT 3540

TTTACGTCTT GGCTTTATGA CGCCAGTGAT GTTTATTGCC AGTCTTTACA TGGTGATGCG 3600

AACGAGTCCG TCGTTAGGAA TGTACGTTTT AGGTGCCTTG CCTTTTCTGC TGCTAGCAGT 3660

TGTCGGGATT GCTCGTTTGT CAGAACCGTT ATCTAAAAAG CAACAAAAGA ACTTAGATGG 3720

AATCAATGGG ATTTTAAGAG AAAATCTTTC TGGATTGCGG GTAATACGCG CATTTGTTAA 3780

TGAAAAATTT GAAGAATCTC GTTTTAATAA AGTCAATGAA ACTTACACTA AAAGTTCAAA 3840

AAGTCTGTTT CGTTTAATGG CAGCAGCCCA ACCAGGGTTC TTTTTCTTAT TTAATATTGT 3900

GATGGTCTTA ATCATTTGGA GCGGGACTGT TCAAATCAGC CATGGGGATT TAGAAGTTGG 3960

GAATTTAATT GCTTTTATTG AATATATTTT TCATGCGCTG TTCTCGTTTA TGTTATTTGC 4020

CAGTGTCTTT ATGATGTATC CGCGGGCTGC GGTTTCGGCT TCACGGATTC AAGAAGCCTT 4080

AGACATGGAA CCAGCTATTC GTGAAGAAGA AGGCGTAACA GAAACAGCTA CTAAAGGCTA 4140

TCTAGAGTTT AAAAATGTAA CCTTTGCCTA TCCTGGACAT GCGGAAAGCC CAGTTATTCG 4200

CAATGTAAGT TTTAAGGCGT CACCTGGTGA AACAGTGGCC TTTATTGGGA GCACAGGTAG 4260

TGGGAAATCA ACATTAATTC AATTGATTCC ACGGTTTTAT GATGTATCAG AAGGTGAAAT 4320

TTTAATCGAT GGAGTGAACG TAAAAGAGTA CAAACTTAGT GCATTACGCA ATAAGATTGG 4380

CTATATTCCA CAAAAAGCGT TACTTTTTAC TGGTACAATT GCTGATAACC TTCGCTACGG 4440

TAAGGAAGAT GCGACGTTAG AAGAAATGGA ACGAGCAATT GACATTGCTC AAGCCACAGA 4500

ATTTGTTTCG CAAAAACCGC AAGGCTATGA TGAACCTCTT TCAGAAGGTG GCACGAACTT 4560

TTCTGGTGGT CAAAAACAAC GTTTaGCTAT CGCACGGGCG ATTATTCGCA ATCCGGAAAT 4620

TTATATATTT GATGATAGTT TTTCTGCGTT AGATTATCAA ACAGATGCAA ATTTACGAGC 4680

GCGTCTGAAA AAAGAAACAA CAGAATCTAC TGTTTTAATT GTGGCACAAC GTGTTGGAAC 4740

AATTATGCAT GCGGACCGCA TTGTTGTTTT AAACGAAGGC GACGTGGTCG GAATTGGCAC 4800

ACATCGTGAA TTACTTGAGA CTTGTCCAAT TTACTATGAT ATTGCGGCTT CTCAATTGTC 4860

AGAGGAGGAA TTAGCATGAA ACACGCCTTT TCTTCTATGA AACGAATCGG TCGCTATATT 4920

AAACCGTACC GAGTGACGTT TTATTTAGTT ATTTTATTTA CAATATTAAC CGTTGCCTTT 4980

AATGCAGCGT TGCCTTATTT GACTGGATTA CCGACGACAG AAATTAGCCG TAATATTGCG 5040

GCCGGCGAAT CCATTAATTT TGATTATGTA ATCCAATGTT TAATTTGGAT TTTAGTTGTG 5100

GGAACAGGTT ATTGTGTGGC ACAATTTTTG TCAGGCTTTT TAATGACGAA TGTCGTTCAA 5160

CAATCCATGC GCGATTTGCG TCGCGATATT GAAGAAAAAA TCAATCGTTT GCCAGTTTCT 5220

TATTTTGATA AGAACCAACA AgGAAATATT TTGTCACGGG TGACGAACGA TGTGGATGCT 5280

GTCAGCAATG CGATGCAACA AAGTTTTATC AATATTGTTT CAGCAGTCTT AGGTATTGTG 5340

ATGGCGGTAG TGATGATGTT CTTAATCAAT CCGCTGATGG CGATTTTTTC AGTGATTATG 5400

ATTCCGTTGT CTCTGATTAT TTCCAGAACA ATTGTTAAAA TCTCCCAGAA ATATTTCCAA 5460

GGAATGCAAA ATTCTTTAGG AGAATTAAAT GTTATGTCCG GGGCCGAGnT CGAATTTTAT 5520

AAT 5523

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5794 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

ACAGTCGATT TCGAGCACCG GCAACTTTTG CTACATCAAT GATAGACGTC ATTTTTTACC 60

TCCAAAGTGG AATGTATTCT TATAAGAAGT ATAGCATACA TTTTTTTGAA AATAGTGCTT 120

GACATGGGAA CATTCCCATA ATTTATCCTA AAGATGTAAA CGATAACAAT TTAATCAAGG 180

AGGAAGAAGA CGGATGACAA GAGGAGCATT AAAAATTGCA ACAATTGGTG GCGGTTCTAG 240

TTATACACCT GAATTAATTG AAGGATACAT TAAAAGAAAA GATGAACTGC CAATTAAGGA 300

AATTTGGTTA GTCGATATTG AAGCTGGTAA AGAAAAATTG GAAATCGTTG GGGCAATGGC 360

GAAgcGCATG ATTAAAGCTG CTGGTTTAGA TTGGGAAGTT CATTTAACAT TGGACCGTGA 420

AGCAGCCTTA AAAGGGGCGG ATTTTGTTTC AACACAGTTT CGTGTAGGAT TATTAGATGC 480

TCGCATTAAA GATGAACGGA TTCCTTTGTC CCATGGAGTA TTAGGCCAAG AAACAAATGG 540

CGCGGGAGGA ATGTTTAAAG CATTTCGAAC AATTCCTGTC ATTTTAGCTA TTATTGACGA 600

TATGAAACGC CTTTGCCCAG ATGCTTGGTT AGTGAATTTT ACTAATCCAG CTGGAATGGT 660

TACGGAAGCT GCAATTAAAC ATGGTGGTTG GAAAAAAACA GTTGGCTTAT GTAATGTGCC 720

GATTGGTCAT AGAAAACAAG CGGCCGAAAT GCTGGGAATT CCAGAAGAAG ACTTATTCTT 780

TAAATTTGCT GGAATTAATC ATTTCCATTG GCACCGCGTT TGGGACAAAG AAGGACAAGA 840

GCGAACACAA GAACTCATTG ATTTAATTTA CGGaCCTAAG CAAGAGCAAG AAAGTCATTt 900

GAAAAATATT TTTGATGCGC CGTTCCATTA TGAACAGTTA AAAGATTTAG GAATGTTACC 960

ATGTGGTTAT CATCGCTATT ATTATATTGA AGATGAGATG TTGAAGCATT CCATCGAAGA 1020

GTATGAACGG GGCGAAACAA GAGCACAAGT GGTTAAAGAA ACGGAAGGTC GTCTGTTTGA 1080

ATTGTATAAA GATCCCAAAC TAGATTACAA ACCAAAAGAA TTAGAAGAAC GGGGCGGCAC 1140

ACACTATAGC GATGCAGCTT GTGAAATGAT TGCATCAATT TATAATGATA AACGAACGGA 1200

TATGGTTGTT TCTACTGAAA ATAATGGGAC AATTACTGAT TTACCGTATG ATTGTGTTGT 1260

AGAAGTATCT GGACCTGTGA CAGCTCATGG CCATGAACCT TACAACTGGG GGGCCTTTCC 1320

GCCAGCAGCA CGAGGAATTA TTCAAAATAT GAAAGCTATG GAAGAAACGG TTATTCGAGC 1380

AGCCATCAAT GGTGATTACG GCGCAgCATT gCATGCTTTC ACTATTAATC CATTAGTTCC 1440

TGGCGGCGTC ATGGCGAAGA CCTTGTTGGA CGAATTATTA ATTGCCCACA AAGCGCATTT 1500

ACCGAATTTT GCAGATGCGA TTAATAAAAT TGAAGAAACA CAACCAGAAA CAGTCACATA 1560

TGTGGCCGAA TTAATGAAAA GTAATTAAAA AATATCGCAC GCTAAAGATT CGTTTAGTGT 1620

GCGATACTTT TTTTAGAATA GGAAGCAAAG GATTCCGTTT TTTTCTTGAC AGTAAAATGA 1680

TAGGTTTGTT GCCGGAACAG CCACTGAATA CCAGAGTAAA TTAAAATTTC AGCAATAATA 1740

AACGTCGAAA GTTTCTGTGC GGAAAGGAAT AATAGCGTTA AAATTAAAAA TAAGACGACA 1800

CAGGTATAAA AAACATGTTG CAATTTTTTA GAATTGATAA TCGTGATTTC TTTGGTATCT 1860

TCAGTAACTT CAATGCGCCG ATAAAATTGG AACCAGCCTT CATAGAACGT TAAAACATCT 1920

CCAGGATCAG CCTCAATTTC AATAGGTACA TTAGTAAAAA GTTTCCCTTG ATGCACACCA 1980

TTTTTATAAA ATGtAAAATA GCGTAATGCG CCCCAAGCTG ATGTTTTTCG TGTCAATGTA 2040

ATCTTCATGG TTTTCACCTC CGtCTTTATT TTGCCmTwGg TtTCGGTAAA ATTGgTATCC 2100

TACTAAAGTC CGATTCTCTC ATTTTAGGGT CAGAATTGAC AAAAGAAAGC AGAGGTGTTA 2160

ATGTGGGAGT AATTAAACGA GGAGGTATAA TTGTATGTTG GCATCTATCA GAATTTGGAT 2220

TGACTTTTTG ACAAAATAAA TTGAATAAAT CTTATCAAAA AGTCTGAAAT TAAAATTCAT 2280

GGGTGGCGTG TGTACTTTTA TACGAAACTA CAAGGATAGC GATTAGGCTG TCCTTTTTTG 2340

CGTACCCAAG ATTCTTTCGT TTCAGAAATG AAAGAAGAGG GCAAAATGAA AAAATTAAGT 2400

ATTCATCTGA AAGACGTAAG TATTCACTTT TCGGGAAAAC CAATTCTAGA AATTGATGAA 2460

TTATTTGTAT ATGAAAATGA AAAAATTGGG ATTATTGGTA AAAACGGCGC AGGAAAAATCA 2520

ACACTGCTCA ATTTAATTAT GGGTAAGATT CAATCAGATA AAGGAAAAGT TCAAAGATTG 2580

AACGACTTTC ATTACTTGGC ACAAGTAGCT GAAGAAATAA CGAATGAATC AGAGAAAACT 2640

GACAAAAATT GTTTGCTAAA TCAGAAAAAT CAAAAACTAA GTGGTGGCGA AAAAGTCCAA 2700

AAACGCTTAG CAACATTATT TTCAGAGTAT CCAACCGGGG TTATTTTAGA TGAACCAACC 2760

ACACATTTAG ATAAAGAACA TCGTCAGTTG TTAGTGGCAG ATTTAACGTA TTATTATGGG 2820

ACCGTTTTGT TTGTTAGTCA TGATCGCTTT TTCTTGAATC AATTAGCAGA GAAAATTTGG 2880

GAAGTTTCGG ATGGACATGT CAAAGAATAT TTAGGGAACT ACGATGCGTA TTGTCGTCAA 2940

AAAGAATTGG AACAGCAAAC ACAATATAAT GTCTATCATC AGTATCAAAA GGAAAAGAAA 3000

AAATTACAGG AATCTTACGC AAAGAAACAA GCACAAGCGC AAAAATCTAG TCATGTTTCA 3060

AAAAAACAAA AACAAAAGCA AATTAAACCT AGTCGTTTAG CTGGTTCTAA ACAAAAAGAT 3120

ACCGTACAAA AAGCACTCCA AAAACAAGCG AAAGsGwTTA ATGCCAGAAT TGACAGATTA 3180

CCCGATGTTG CGCAAGCAAA ACAAGAGCGA AAAATTATTT TTCCTACTAA TAATCAGTTC 3240

TCTTTATACA ATCCATATCC AATTAGAATT GAAAATTTAA CTTTCGCTTA TGAGAATCGA 3300

ACAATTCTTA ACCAAGTGAA CGTTCAAATT CCTTTAAACG AAAAAATAGC ACTTTGTGGa 3360

AAAAATGGTG CTGGGAAGAG TACCTTCCTT CAACAAATTG AAGCCCGTCA CCCGGCCATT 3420

TACTTTTCTC CCAAAGTCCG GTTAGGAACG TaTCATCAAT TAGATTATCG ATTAAAAAAT 3480

GATGAACCAC TGTTGACTTA TTTATTGAAG CGGACAAATT ATTCTGAAAA GATTGTGCGT 3540

TCACTTTTAT ATCGACTAGG TTTTCAACAA GAAAATTTAC AGACAAAGAT ATCTTCTCTA 3600

AGTGGTGGCG AACATGAAAA ATTACTTTGG CGCAATTATT TATAGAACCA AATAATATCA 3660

TCTTATTGGA TGAACCCACC AATTTTCTTG ATTTAGACAC AATCCAAGCA CTTGAAGAAT 3720

TTATAAGTGC CTATCAAGGA ACAGTAATTT TCACTTCGCA TGATGAAACG TTTGTGGAAA 3780

AAGTAGCCAC CCGCACTATT TATTTAGAAA ACGGGAAAAT TATTGATAAA TAaAAAAAGT 3840

rAACAAAAAG TCGCTATTTT AGCGACTTTT TGTTTAGACT TCTTCAATTG GTAAATAGGT 3900

GTATTCTGCA ATTTCTGCGG CATGGGAGCC AGCAACATGT CCAGTGACAA ATGCAGCTGT 3960

AACATTGTAG CCTCCAGTAT AGCCATTAAT ATCTAAAAGT TCACCAGCAA AAAATAAACC 4020

ATTGACTAAT TTGCTCTCCA TTGTTTTAGG GGTCACTTCT TTGAGGGAAA TCCCACCGCC 4080

TGTGACAAAA GATTTTTCCA AAGGCAATGT TTTCGTAACA GTGAACTGAA AGTCTTTCAG 4140

TAGTTCCACA AAGGAAAGmC GTTGCTTTTC TGTTAGTTGT TTGGCAGGCA CTTCTTCAAA 4200

AGATTTTGTC GGAAACACAT CCAAGGCTAC CGTGACAGGT TGATTACCGT TGCGAGTTAA 4260

TTCTTGGTTA ATAAAACTAG AACAGCGGAG CGCGGCAGGT CCTGAAATGC CAAAATGTGT 4320

AAACAGCATA TCCaTTTGAT GATTAACTAA AGGTTTTCCT TTTTGGTTCA AAACAGTTAA 4380

ATTAACATCT TGTAAAGAGA GACCTTGCAA CGTTTTATCC AGGATAAAAG GTTCTTCAGA 4440

AATAATAGGT GATTCGGTAG GGTAGAGCGG GCTGATGGTA TGCCCCATTT TTTTGGCTAG 4500

TTTATAGCCA TCACCAGTTG CTCCTGTGGA AGGATAAGTG CGGCCGCCAG TTGTTAATAC 4560

AACACACGGT GCATAAATTT TTTCCAGTTC GGTTTCAACG CCAATTATTT GATCGTCTTT 4620

TCGTAGTAAT TTTGTGACCT GTGTTTTTGT AAAAACAGTG ACTCCTAATT CGTTAATGCG 4680

GTTAAATAGC GCATCAACAA TTGACTTCGA TTTATCTGTA ACAGGGAACA TGCGTCCGTG 4740

ATCTTCTTCT TTTAAGTGAA TACCATTGGA TTCAAAAAAG TTCATGATAT CATAGTTATC 4800

AAATTGTGAA AATGCGCTGT ATAAAAATTT TCCATTCCCA GGAATAAATG AAATGATTTC 4860

TTCTGCGGGC CGATTATTGG TTACATTACA GCGGCCGCCA CCAGTCATTA ATAATTTTTT 4920

CCCAACACGG CGATTTTTTT CAATCAATAA TACTTGAGCG CCTGCTTCGG CGGCCGCAAT 4980

CGTGGCCATC ATACCGCTCG TCCCAGCACC GACAATAATT ACATCAAATT TCTTCATAGT 5040

TCTCCTCCAA TTCAATAGGT AgTGTACCAT AATTACGGGA TTTTTTCTGG ATTTAAGGAA 5100

AAAATAAACG TTTGACTATG GTAATTTTTG AATTTTCAGT TAAAATAAAC CTATAACAAT 5160

GAGAGGATGA TTAAATGGAC GCAAAACAAT ATGTTAAAAA TATTCAAGAA AAAATTCACC 5220

AATTAGACCA AGGTCAAACA GAATATTTAC AAGCAGTCGA TGAGTTTTTA CCAACCGTAG 5280

AAGGTTTTTT AGAAAAAAAT CCTCAATACA TTGAAGCAAA TGTTTTAGGT GTCTTGATTG 5340

AACCTGAACG TATTTTTCAA TTTCGTGTGC CGTGGCAAGA TGACCAGGGC AATTGGCATG 5400

TGAACAGAGG CTATCGTGTC CAATATAATT CAGCGATTGG ACCATACAAA GGTGGTTTGC 5460

GGTTTCATCC AAGTGTGAAT TTAAGTGTGA TGAAGTTTTT GGCCTTTGAA CAAATTTTTA 5520

AAAATAGTTT AACTGGTTTG CCAATCGGTG GCGGCAAAGG CGGTAGTGAT TTTGATCCAA 5580

AaGGTAAGTC AGATGCTGAA GTGATGCGCT TTTGTCAAAG TTTCATGACC GAACTTyAAA 5640

AACATATTGG TCCAAGCACA GATGTTCCAG CAGGCGATAT TGGCGTAGGC GCTCGTGAAA 5700

TTGGTTATTT GTTTGGCATG TATAAAyGTT GCGCAACTAT GATGCAGGAG TGTTGACTGG 5760

CAAACCTTTA GGATACTGGG GAAGCCAAGC AAGA 5794

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4232 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

```
ATTCGTATGC CaATTATrGT gGCtaTTTGg CgAAAAAgCc AnCnGGAGAA gTCaGrGnAA     60

AATGTGGAAA ATTATAGGAA GCTACTGGGT TTTATGCTAT GATTTTTAAT AAaACAAAAT    120

GAAGGGGCGG GAAAACATGT TATTTGGTAC AGCAAAAATG AATCGGGAAA ATCATTTAGA    180

AATCGGTGGT TGTGACACCG TAAAGTTGGC GCAAAAATTC GGGACGCCGT TATTTGTCTA    240

TGATGTCGCT CATATTCGCG CACAAGCACG TGGTTTTAAA CAAACGTTAA ATCAATTAGG    300

TATCAAAAAT AAAGTGGTTT ATGCAAGTAA AGCATTTAGT TGTTTAGCTA TTTATCAAGT    360

GTTAAAGGAA GAAGATATTG CTTGTGATGT TGTCTCTGGT GGAGAACTAT TCACAGCTTT    420

AAAAGGGGGC ATGGAGCCAG CGGAAATTGA GTTCCATGGA AATAATAAAA CCCCTGAAGA    480

ATTACGTTAT GCTTTGGATA ACAAAATAGG TACAATTGTC ATTGaTAATT TTTACGAAAT    540

TGACTTGTTG GAAGAATTAT TAGCTACACG AAACCAAACA CAAAAAGTTC TTTTTCGAGT    600

GAGTCCTGGC GTGGATGCAG AAACACATGA TTACATTTTG ACAGGCCAAG AAGATTCAAA    660

ATTTGGTTTT GATGTTGCAA GTGGCCAGGC AACAGAAGCA CTCGTTCGTC TTTTATCAAA    720

CCCCTCTTTT GATGTACAAG GGGTCCATTG CCATATTGGC TCGCAAATTT TTGCTGTGGA    780

AGGCTTTTTA GCTGCTGTTG AAAAAATGTT CACTATTTTA GAAGACTGGC GGCAAGTCCA    840

TCAATTTACA GCACGTGTTT TAAATATGGG TGGTGGTTTT GGTGTGCAGT ACACGCAGCA    900

AGATGAGCCA TTAGCACCAG CAACATTTGT TGAAAAAATT GTTTATTCCT TGAAAGGTCA    960

TTGTGAACAA tTAGGGTATC CaCTGCCTGA GCTTTGGATT GAACCAGGCC GCAGTTTAAT   1020

TGCCGAAGCG GGCACAACGA TATATACAGT TGGCGCGCAA AAAGAGGTTC CTGGTGTTCG   1080

TCATTTTGTT TCCGTAGATG GTGGCATGGG TGATAATATA CGGCCTGCTT TGTATCAAGC   1140

AGTTTATGAT GGTTTTCTAG CGAATCGTGA TGGGCATGAT TCAGTCAAGG AAGTCACAGT   1200

GGTTGGTAAA TATTGCGAAT CAGGAGATGT CTTACTTCGA GATATTTTAT TGCCAGAAGT   1260

TAAAGCAGGG GATTTATTGG CCATTAGCAG TACGGGGGCT TACGGTTATT CAATGGCAAG   1320

TAATTACAAC CGAAATCCTC GACCAGCGgT GGTTTTTGTC GAAGATGGAC AAGCAAAATT   1380

AGTTGCTCGC CGTGAGACAT ATGAAGATAT GACAACGTTG GaTTGTTGAT AGATTaAAAG   1440

TAGCCGCCTC TTTTTGAGAC gGCTGTTTTT TTTCcAAAAA ATAACmAAAT GTAAGGGTGA   1500

TTTCTTTCAG AAaTGaAAAA TCCaTAGTAT ATTAAATGAA GAAGGAAATA CTTTAGAAGA   1560

GGGTGTGATT TGCAGATGAA ATGGCTCCTC GGGTTACTGA TCATTTTTGC AATAATTGGC   1620

GTTGTTTTTT CTTTTTACAT TGCCAATTAT TTTTTGAAAA TTGCGTTGCT TAAAGACAAC   1680
```

CCTTGGTATC ATAAAAAAGG GCATCGGTTA TTAAATCCTG ATAATTTTCA AGAACGGGAA 1740

ACAAGATATA CAAAAATAGA AGAACAACAA AAACAAGAAG GCGAAGCTTT TTGGACAGAG 1800

TCTTTTGCTG AAGATCGTTG GTTGAAAATT AAAGATGAAA CATTGTATGC ACGGTGCTTT 1860

ATTCCTTATC CAGATAACCA TCGTTGGGCG ATTTGTGTGC ATGGTTATCG TTCAAACGGA 1920

AAACGCGATA TGGCATATAC TGCATTACGT TTTGCAGAAG AAGGCTACAA TGTTTTAGTT 1980

CCCGATTTGC GTGCGCACGG GAAAAGTTCA GGGAATAAAA TTGGGATGGG ATGGCTGGAT 2040

CGGTTAGATT TATTAAGTTG GCTTAGTGAA GTTCTTGCCA TTGATATGGA AGCAGAAATT 2100

ATTTTAGTGG GCGGTTCCAT GGGAGCGGCT ACTGTCATGA TGGCAAGTGG TGAAAAGTTA 2160

CCAACGAATG TCCGCGGATT AATAGTGGAT TGTGGCTACA CTTCTGTTTA TGATGAATTT 2220

AAATACGTCC TACATGAATC TTTTCATTTA CCCGCTTTTC CTATTTTAAC AATTGCTAAT 2280

CAATTAGCAT TAAATAACTA TGGCTTTCAA TTGAAAACCG CTTCTTCCGT GAGACAATTA 2340

CACAAAAATA CGTTGCCCAC ATTCTTTATT CACGGTACAG GTGATCGCTT TGTTCCAATG 2400

ACAATGTTTG AAGAAAATTT AGCGGCAACT CAGGGGATAA AAAAAGGCCT CATTGTTGCT 2460

AAAGCACCAC ATTTATCTTC AAGTGTCTAC GAACCTGAAA ATTATTACAG TAGTATTTTT 2520

GAATTTTTGG AAGAAAACTG tCCTGCGGTA AAAACAATAT CCGACTAAAA AAGACATGGA 2580

TTTTCATGAT TCCATGTCTT TTTTTTGTCT GACTCGTTTA GTTAAATTTA AAGTGCTTAA 2640

AAAGAAAAAT AACGGCGCCA AACTACCAAG GCAAAGAAAC CATTTCATTG AATAAGTTAA 2700

CTCTTGCTGT AAAAGAGATT CTCGTATATA GCTGAAACCT TCGCCATGGA CCCGCAAATA 2760

AATATAAATG TAAGCAAAAA GAATCACTAA AGGCAAAATA CAGACTTCTA ATATTTCCAA 2820

AAGATAAGGA AATGTTTTAT TTAATAATTT GAGGCGTGTC TTATTCAATA GATAAGCAAA 2880

AATGAAAATT GACCATAAAA AACCAAAAAA ACTAATCAAA TAAAAAATCA GATTCGCACT 2940

TCCTGCATTC ATAAAAGCTT TCGGGAAAAG TAGTCTGCCG ATGAATGTAT TTTTAATTGT 3000

AAAAGGCACA TCAAAAATGG AAATATAAAG ACAAAAAAGT ACTAATGTTA ACAAGGCATG 3060

CCAAATCTTT TTCTCCATTT CATTTAAAAA ATATCTAATT ACTTGGTACA TTTCCATAAA 3120

CTGTCGCTCC GTGTTTGAAA TAATCGTAGT TTTGACTATA CCATAGCTGT TCTCAATTTT 3180

CTTGAAAAAA ATACACGCAA CCAGCCTCGA CTATACGAAA CACACAGAGT GTCTAAAAAA 3240

TAAGTTGCCG ATCGCATTTT TTCGCATGCT AGTGTTGCAC TTCTGTTTCG TTGGTGTTTT 3300

CTAATTGCTG ATTTGTTGAT TAAAAAAGGG AAGGAGGGAG CGTTTTTCTA GTATAATAAT 3360

ATAAAAGAAT ATAGAGAAAA TCGTAGGATG GGAGTTATTC AAAGTGGCGG AAATTACAGT 3420

TAAGACAATT TCAGAGAAAG ATGATTTAAA GAAGTTGTTT TTGGGTCAAG AGTATGACCA 3480

ACAGCAGCCA GTTAAAAATA TTTTAAAAGC TAGAACATCG AATGTCCATA TTGAAAAGTA 3540

CGAGGTTGTG GAGGATGTTT TTTTAAAGTT AACTGGGAAG GCGACCGCGC ACAAAGATAT 3600

TATGAACTCA TTTTGGACAA CGTATAAAAT TATGTTACAA CTTGTTTATC CCGATTTCTT 3660

TAGACCAGCG GAGGTAATTG GGGAAAATCA AGAAAGCTAT TTAGAAAAAC CTTCTGAACA 3720

GAATTTTCTT GCAGTAAATA GCAAGTATCC GCCTCATGAT TCCGGACCGA TATTTAACTT 3780

ATTTTGACCA AGTTTTTCCT GACTATTTAC CAAATCCTGT ACCAAAAAAA TATACATGGA 3840

ATGAATTTCT TCTCGACAAT TTTACTAAaT TTGAGCGAGT TCATCmGGAC CCACAaTTAA 3900

aGCGTTTTGC CgAACTTACC CaTTCAATTG GAAACaTTAC AGTAGTTCCA CTAGGATTTA 3960

ACAGCGGAAG AAGTTTATCG TTTAAAGATT ATTGGGATTA TTCTTTAGAA CAACTTTCGA 4020

TATTTTTAGC TTCATTTCAT TCATGGGAAA GCTACGTTCA TACGTATGAG ATGCAACCAT 4080

TTTTAAATGA ACAGTATCAA CCAGTAGCGC TATGGAAAAA TCATTTGAAA AAAGACTCGT 4140

TTATCTTACC TCAGAATATA GAGGAGATAA ATGAATATTT AGTTCAAGTA AATCAACGAA 4200

TAGAAAAAAG AGGGCAGCGA ATAGTGAATC GG 4232

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24004 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

TGGAnTTATG GTTGGGTTTT nAGGTTGGCC GnAGGTTGGG GCCAAAGGTT ACCCGTTATT 60

TTCCTTGGAC CAAnCCACTT TATTTCGGAC CAAAAACCCA TTTTAAGAAA CCATTTGTAC 120

CCAAATGGTA AAGCAGAAGA TTTAGATGCT GCTTCCGCAG AATATGAAAA AATTATCGAT 180

GCACATCCAG TAGATATCCA AATCTTAGGT ATTGGTCAAA ATGGACACAT TGGTTTCAAC 240

GAACCAGGCA CACCTCTAGA TAGCTTAACA CATGTCGTTG AGTTGACTGA ATCAACAATC 300

AACGCAAACA AACGTTATTT TGACAAAGTA GAAGATGTCC CAACACGTGC TGTTTCTATG 360

GGGATTGGCT CAATCATGAA AGGTAAAAAA ATGATTTTAA TGGCTTATGG CGAAgCAAAA 420

GCAGAGGCCA TTAAAGGAAT GATTGACGGA CCAGTAACAA CAGACATGCC AGCAAGTGCA 480

TTACAAAACC ATCAAGATGT TGTTGTAATC ATTGATGATG CAGCTGCAAG TAAATTATAA 540

TCGTTGATGA AAAAGAACTT GAAACATAAG AATTTCCTTA TGTTTCAAGT TCTTTTTTTG 600

TTATACTGGA AGTAATAGAA GAAGAGGTGA AAAGGATGTT ACAGCTAACA ATTCCAGAAA 660

TTATTTTACG GTTAGTCCTA GCGATGGTGA TGGGTGGCGC AATTGGTTTT GAAAGACAAT 720

ATAAGAATCG CCCAGCAGGG ATGCGCACTC ACATTTTAGT TTGTATGGGC GCGGCTATTA 780

TTGCGTTAAT TCAGTCGCAA ATTGCAATCA ATGCATTGAA AGATGCGTTA gcTAATCCGG 840

CTTTGACGGG GGTCATTCGT TCAGATCAAG CAAGGTTGAT CGCACAAGTT GTAAGTGGAA 900

TCGGCTTTTT AGGTGCTGGA ACCATCATTG TAACGAAACA AAGTGTCACC GGATTGACGA 960

CAGCAGCGTC ACTATGGACC GTTGCTGGTT TAGGGATCGC GGTGGGCATG GGCTACTATA 1020

CGATTGCCTT AACAAGTTTT ATCGGTGTAT TATTTGCCTT AACCATTGTC AAACGAATTA 1080

```
TTCACGTACC AACAATCAAA AAATTAGAAA TTCGCTATCA ACATAAACAA GAAACCAAAG   1140
AATTTATTAA TCAGTATTTT GAAGAAAaAC ATATTGAAAT TTATGATGTG AACTTCAATG   1200
TTTCTTGGGT GGATGATACA AAAATCTATA CGAATATCTA TACTATTGAT TTGCCTAAAG   1260
GCCTAACGTA TGCCGATGTT ATCGAAGATC TATCTGTTTC AAATAATGTA ACAAAATTAC   1320
GATgATTAAT GTGTAGAAAC ATCTATTaTA ATTAATGACA ACCGCTATAA GTTAAAATGC   1380
TTATGGCGGT TTTCTTCTAG GACCTAGGAC TGATTTTTTA AGTTTAATAT TTCAAAAAGA   1440
TTACAAAAGA GTTATGATTG TATTACGAGT TAGAAAGAAA GGTGATAAGT AAGATGAATA   1500
AAAAAAATAA ATTTCTTAGT ATTATTGTTG TTGTCATTGT GATTATTGGT GCATTTGTGA   1560
TGAAAGCTTA TAATCAATTG ATTACTTTAG ATAATAAAGT CGAAGCAGAA TGGTCACAAG   1620
TAGAGAATGT AATGCAAAGA CGTGCTGATT TAATTCCTAA TCTTGTCAGT AGTGTTCAAG   1680
GCAGTATGAC TCAAGAAAAA GAAGTTTTAA AAGAAATAAC GGAAGCACGT AAAGCATACG   1740
CAGGAGCAAA ATCTTCTTCT GAAAAAGGGC AAGCAAATGA GCAAATTGAA AAAGGATTAG   1800
GTAATTTTGT TACAGTATTG AATGAAGATT ACCCAAAGTT AGCTTCTTCT GAAAATGTCA   1860
AAACCTTAAT GACTCAATTA GAAGGATCTG AAAATCGAAT TTCTGTTGAA CGCCGTAACT   1920
ATATACAAGC CGTTGAAACA TATAACCAAC AGATTTCAAA ATTCCCTGAT AAAATAATTG   1980
CAAAACTTTT AGGATTTGAG AAGAAAACTA ACTATGTTGC GACAGAGCAG GGAAAAGAAG   2040
TGCCGCAAGT AAAATTTAAA TAAAAATTTA TAGAAAATTT GGGTTAAAAA GAAAGAAGTG   2100
AGGAAGGTTG GATAAAAAAG TCGAAAAAAT TCAAGAAAAA GTTCGTTTGT TTTATAAAAA   2160
GCATCAGCGG AATACAACGA TTGTTGTGGT AATTAGTTTA AGTATGATTT TAATATTAGG   2220
AATGGCTGCT TGGTATGAAG GAAAGGGAAT CTCTTATGAA GCAAAACAAA CTGCACTGAG   2280
AGAGAAGCAA CATGCCTTAC AAAAAAAGAG AGACGATATC TTAGCAGGGA AAATGACTGA   2340
AGCGTCAACC CAGACAACAA TAAATAACTT TTCTAAACAA GAAATAACTG GTGAAAATGA   2400
ATTAGTGACA ATTACTGGAA CGAATGGGCA GTTGAACATT AATGATGCTA ATATCTATGT   2460
TTCAGATAAT GCAGGTATTG TCTCAAATCA ACTTAAACAG AAAATGTTTC AATTGAATCA   2520
ACAATTGTTG GAAAATGCGA ATGGCGCACA ATTTATGCTT ATTACCGTTC CTGCGTTACC   2580
TAGTGGGGAA TCTGTTGAAT CTTACAGCAA TAAGATTGCT AATCAATTAG GTGTAGGGGA   2640
TCGTGAAAAA AATAATGGTG TGGTTTTTCT AATGGCAATT GAGGATCGCG AGTCTCGCTT   2700
GGAAGTTGGA TATGGCTTAG AGAGTATTTT AACGGATAGC TATGCGGATG ATATTATTAA   2760
TAATGAAGAA GTAAAAGAGG CTTTTAGAGA CGAGGATTAT AATACAGGTC TCAATAAAAT   2820
TATTGATCAA GTATCGGCTG CAATCAATTC AAAAACCGCA CAAGTGGATA ATGAATTAAC   2880
AAATATACAA ACAGAGTTAA ACACTAACAC AACGAAAAGA AATGTTCTTT TAATGGCAAG   2940
TATTGCTGGG ATGCTCATTT GTGCTGTTTA TATTTTACAA ATTTTAAGAA CAAGAAAACT   3000
TGTTAAAAAA ATGTATCAAG ATTATTTGAA TTGTTTACCA ACAAAGGCTG TCCTAAATAA   3060
```

CTCAGAACAG ACAAAGAAAG TATTAAATAA GATGAAAAAG ACTTCTTTTT ATTGTTTATA 3120

TCTTAATGGA GCAACGTTAT TTGCCACGAA AGGGAAAATA AGACGTGCCA CTAAGCGTGG 3180

TTCTTTACTA AGTATATATC CAAATGCAAA AAAACAATCT TTTGGGCGTT TATTAGTAGG 3240

AGATACTTTA TATAGTTATG ATGGTAGTGT TTTGACGTAT GCTTATTTAA ACTCACAATA 3300

CAATCCGTCT AATCATAGTA GTTCTGGTAG CGGCAGTGGT GGTTCATTCG GTGGCGGCTC 3360

ATTCGGCGGT GGCGGCGCCT CAGGAAGTTG GTAAAATAAA AAGCTAGAAA CGTTGATCGT 3420

TTCTAGCTTT TTGCTATTTA TCTAAACCGT AAAGATAGTC ATCATCTTTC ATGGCTTCAA 3480

CTGTTCCTAG TAAATAGCCA TTTCCTACTT GGGAGAAGAA GTCATGGTTG CTAGTTCCTG 3540

TCGAAATACC ATTCATTACA ATCGGGTTCA CATCATTGGC TGTATCCGCA AATAAGGGAT 3600

CCATACCTAA GTTCATTAAA GCTTTGTTGG CGTTGTAACG TAAAAAGGTT TTTACTTCTT 3660

CGGTCCAACC TAAGTCATCA TATAATTCTT CTGTATAACG TTCTTCGTTT TCGTACAATT 3720

CGTAAAGTAA ATTATACATC CAGTCTTTTA GTTTATCTTG CTCATCTTCT GGTAATTCAT 3780

TAAATCCTAA TTGGAATTTA TAGCCAATAT ACGTGCCGTG CACTGATTCA TCACGGATAA 3840

TCAATTTAAT AATTTCAGCG ACGTTCGCTA GTTTATTGTT TCCTAAATAG TAAAGAGGAG 3900

TGTAAAACCC AGAATAGAAT AAAAACGTTT CTAAAAATAC ACTAGCAATT TTTTTCTCTA 3960

ACGGAGTACC GTTTTTATAA ATTTCGTTAA TCCGTTCTGC TTTATATTGC AGGTGTTTGT 4020

TGGTATTTGT CCACTCAAAA ATATCATCAA TCTCTTTTTT TGTATTTAAT GTACTGAAGA 4080

TAGAAGAATA ACTTTTGGCA TGGACAGATT CCATGAATTG GATGTTATTT AAGACGGCTT 4140

CTTCATGAGG CGTCCGCACA TCATTTCTCA GTTGATCCAT GCCGCTTTCT GATTGAACAG 4200

TATCTAATAA AGTTAAACCG CCGAAAACGT AGCCGACAGT AGTTTTTTCT AAATCGGAAA 4260

GAGTCCGCCA ATCATCTAAG TCATTGGATA ACGGAATCCG TGTATCTAAC CAAAATTGTT 4320

CGGTTAATTT TTCCCAAGTG GATTTATCAA TGACGTCTTC AATGGCGTTC CAGTTAATTG 4380

CTTCATAATA CGTTGCCATG GTTATTCCTC CTGTATGCTG GGACAGCAGT CTGTTTGACT 4440

AGTTGTCTCG TTTCTTCGCT TAAAACGTTC TACAGAGTCC AAATAGGAGT GTGCAGCCGT 4500

TCTTCTTTGT CCAGTATATC TGAAAATGTC CTGTTTAAAA AGAGAACATT CCTAAAAAAT 4560

AAGACGCTCA AGCAGAACGC CTTATTTAAG TGCTTTTCAT TAAATCACGC AAGACTCACA 4620

TTGATTGCTG CCAATTTCTT CAGCATCATC TGTAAATGTA CGAATATAGT AAATTGATTT 4680

AATCCCTTTG TTGAACGCAT AATGACGTAA AATGTTCAAG TCACGGGTTG TTTGTTTTGG 4740

TGTTTCTTTC CATTCATACA AGCCTTCAGG GATTTCTGAA CGCATAAACA AGGTTAAACT 4800

CATTCCTTGA TCAACGTGTT GTTGCGCTGT TGCGTAGACA TCAATCACTT TGCGCATATC 4860

CATGTCGTAA GCAGAAGTAT AGTAAGGGAT TGTATCATTA GCTAAATAAG GCGCTGGATA 4920

GTAAATCTTA CCAATTTTCT TCTCTTGACG TTCTTCAATC ATACGAGTAA TTGGATGAAT 4980

GCTGGCACTA GTATCATTAA TATATGAAAT TGAACCATTT GGCGCTACGG CTAAACGATT 5040

TTGATGATAA AGACCATCTT TTTTCACAGC ATCGCGTAAC GCAGCCCAAT CTTCAGCCGT 5100

TGGAATCTGA ATGTCTTTGA AAATTTCTTT CACTTTATCA AATTTTGGTG TAAAGTTGCC 5160

TTCAATATAT TTATCGAAGT AAGCACCAGA AGCATAATCT GATTTTTCAA AATTATGGAA 5220

GACTTGGTTA CGTTCTTTCG CAATTTGGTT ACTTTCCATT AAAGTCCAGT AATTTAATAA 5280

CATGAAATAA ATGTCTGTAA AGTCTAATGA TTCTTCAGAG CCGTATTCCA TATGATTTTT 5340

AGCGAAGAAG GTGTGTAAGC CCATTGCGCC TAGTCCGATT GTATGGCTTA AACGATTTCC 5400

ATTTTGGATA GTTGGCACGA CATCAATTTC AGAAGCATCT GTTACAAAAG TTAACGCACG 5460

CGTCATAGCG CGGACAGATT TTCCAAAGTC AGGACTTTCC ATTAAATTAA CAATGTTAGT 5520

TGAACCTAAG TTACAACTAA TATCTGTTCC TAAAACTTCA TATTCTTGTT TCCCATTAAT 5580

AACAGAAGGC GTTTGAACTT GAAGAATTTC AGAACATAAG TTACTCATGA TAATCTTGCC 5640

ATCAATCGGA TTGTTTTTGT TAGCTGTATC AATATTGATG ATATAAGGAT AGCCAGATTC 5700

TTGTTGTAAT TTAGAAATTT CATTTTCTAA GTCACGGGCT TTAATTTTAC GTTTACGTAT 5760

GTTTGGATTG GCCACTAGGT TATCATATTC TTTGGTAATA TCTACATAGG AATAAGGCAC 5820

ACCATATTCA CGTTCCACAC TATAAGGACT AAATAAATAC ATATCTTCGT TGTTACGTGT 5880

TAATTCGTAA AATTTATCTG GAACAATAAC ACCAAGAGAC AACGTTTTTA CACGGATTTT 5940

TTCATCAGCG TTTTCTTTTT TAGCAGATAA GAACATTTCA ATATCTGGAT GGAAGACATT 6000

TAGATAAACC ACACCAGCAC CTTGACGTTG ACCTAGTTGG TTAGAGTAGC TAAAGCTATC 6060

TTCAAATAAT TTCATAACAG GGACAACACC ACTTGCTGCG CCATCATAAC CTTTGATTGG 6120

TGCGCCCGCT TCACGTAAGT TGGAAAGAGT AATTCCTACG CCGCCGCCGA TTCGTGATAA 6180

TTGAAGGGCA GAATTAATTG AACGCCCAAT GCTGTTCATA TCATCTGTTA CTTGAACTAA 6240

GAAACAAGAA ACCAATTCAC CTCGACGTTT ACGCCCAGCG TTTAAGAAAG AAGGTGTGGC 6300

TGGTTGATAC CGTTGATGGA TCATTTCATC CGCCAATGTT AAAGCCAACT CTTCATCGCC 6360

ATCCGCAAAA TACAAGGCAT TAAAGGCAAC CCGATCTTCA TATGTTTCTA AATACTCTGT 6420

TCCGGCGTTG TTCTTTAAAG CGTATTGAGA ATAAAATTTA TAAGCCGCCA TGAAAGACTT 6480

GAATTGGAAA TGTTGTTCAT CTAAGAAAGC ATATAATTTT TCAATAAAGT CCATTGAATA 6540

TTTTTCGATG AATTCAGTTT CAATAAAATC TTCTTTCATT AAAAAGTTGA TTTTTTCTTC 6600

AGCAGAGCTG AATTTACGTG TATTCGGTTC TACATTTTCT TTAAAGAAGG CTGCGAGTGC 6660

TTCTTTATCT TTATTTAAAG GGATTTGACC ATCAACAGGA CGGTTAATCT CGTTATTCAG 6720

TTTGAAATAG CTGACATCTT TAATTTCTTT TAGACTCAAT TTCTTCCACT ACTTTCTTAA 6780

AGGTCTCCAC GTCTTCGTCT GTTCCGCTAA ACTCAAAGGA GTATAGAAGC GGAACGTGAT 6840

AATCGCGGGC AATATCTTTG GCGGTATAGA CAAATAATTC AGCAAAATTA CGGTTGCCCC 6900

CACCAGCTAC ACCAACTAAC AGTTCTTGGT TACTTTTATA ATCAAGAAAG TCATTCACTA 6960

CCTCTGTAAT TTCTGCATCA TAGGTAGGAA TGACAAGAAT AAAAGGCTCA TTTATTTCAA 7020

```
AAAAAGGATT CGCTGGCTCA AGTTCATAAG CAGGCAAATC CAATTTTTTA ATAAAGCGCC    7080
TTGTTTGACC GGTGACTGAA AAATAAACAA GTTTCATCTT AGCTAGCCAA TTGTTTTAAT    7140
TGGTCTGGAC GGAAGCCAAC AACGGTTGTT GCTTCAGATG TGATTACCGG TACGCTTTGG    7200
AACCCTTGTT CTTTTAACCA GTCGATAGCA TCTGGTTGTG CGTCGATATT GATTTCTTCA    7260
AAAGCGATGT TATTTTCACT TAGGaAACGT TTCGCCATTT TACATTGGAT ACAGTTATTT    7320
TTTGAAAAGA TTTTTACGTT CATTGaAAGT TCCTCCTCTT TTTTTCTACA TAGACTAGTA    7380
TAAATCTATG TAGAAAAAAG TCAACACTTT AACACTATAT ATTGTGGTGC GTTTTTCTCA    7440
TAATACTAGT TTTTGTGTTT TTGAAAGAAA AATGAAAACA CGTGAATCCT CGTTTTTAGG    7500
AGAAAAAACA CATTGGTCAG TGTTTGACAA ACAAAAAAAT TAAAATTTTG AAAGAGAACA    7560
TTTAGTGAAA AAACACTAAA TGTTGTGTTG ACATTTGAGT TAGGCACAAG AAGAAAAAAT    7620
AAGCGGATTT GTTATTGAAT GAGTGGGATT TTACAGTAGC TAGTAACATA AAAAACAAGT    7680
GCTGACTTTT TTTCAAAAAT AGTAACCTGA AAAAAGTGTC TCTATATAGA AGGAAAATAG    7740
AAGAAAACAG GAGATACATC CAGAAATTCA TTGACTTTCC GTGTAAGCAA GCGTATTCTT    7800
TTAATTGAGA ACGATTCTCA TTTTCGTTTG AAACTTTTTC ACATGGAGGA AAGTAAATGA    7860
TTACGTTAGC GCAGGCAGAA GTTGGGAAAG TCTATACAGT CGAAAGCGTC CAAGCAGATA    7920
TTCAGACAAA GAAACATCTG AACAATTTAG GTGTCGTTGC AGGCCAAGCA GTTGTCCTTG    7980
TCAATTATCA GAATCAAAAT GGTATTGTTT TATTACATAA TAGTCGAATT GCGCTGACAG    8040
ATACGATTTT ACAAGCGATT CATGTGGAGG AGCGCTCGGC GAATGAGAAA GTTTGGGTCT    8100
CCTTGGATAC CTTAAAAGTT GGAGAACGCG CTACAATAGT AGGGATCCAC GGACAAGGAG    8160
CGGTGAAACG TCGGTTGATG GACATGGGCT TAACAAAAGG AACAGTAATT TTTATTCGCA    8220
AAGTAGCTCC TTTAGCTGAT CCAATTGAAA TCAATGTACG TGGcTACGAA CTAACGTTAC    8280
GAAAAAGTGA AGCGGaACTA ATCTTGGTTG AAAAGGAGGA AAGCGAATGA CTCAGCAAAT    8340
TGCACTAGCC GGCAACCCTA ATAGCGGTAA AACGACAACG TTCAATATGT TAACAGGAGC    8400
CAACCAATAT GTTGGAAACT GGCCAGGGGT GACTGTGGAA CGCAAAGAGG GGGTGGCAAA    8460
GAAAGATAAG ACACTCATTA TTCAAGATTT GCCAGGGATT TACTCTCTTT CTCCTTATAC    8520
GCCTGAAGAA ATTGTGGCAC GAGATTATCT TTTGGAGGAC CAACCTTCAG TGATTTTGAA    8580
TATTCTCGAT GTGACCAATT TGGAACGGAA TCTTTATTTA ACAACACAAT TAATTGAAAC    8640
TGGATTACCG GTTGTTTGTG CGTTAAATAT GATGGATTTG TTGGAAAAAA ACGGCCAAAC    8700
ATTGAATAGT GAAAAGTTAA GTTATGGCTT AGGCGTGCCT GTTGTTGAGA TTAGTGCTTT    8760
GAAGAATCGT GGGTTGGATC ATGCTTTAAA ACAAAGTAAG CAGCGAGCGA ATGCCATAGA    8820
GACAGAGGTT ATTTATCCTT CTTATGATAA TCGCTTAGAA GCAGCACTTG CTGAAATTGT    8880
TGATATTTTA GGGAATACTG TTCCTGAAAC GCAACAGCGT TGGTATAGTT TAAAACTTTT    8940
TGAACGAGAT GTTCGAACCA AAGAGCAATT ATTATTGAGT TCTTTTCAAG AAAAAGAAAT    9000
```

TGAAGAAGTC ATTCAAATTA CTGAAAAAAT CTTTCAAGAT GAAAGTGAAG CGATTATCAT 9060

TAATGAACGC TATGCCTTTA TTGCTCGATT AATTGCTTTG TGTGCGACGA AAAAAACAGA 9120

AATGACATTT ACTCACAGTG ATAAGATTGA CCGAGTGGTA ACCAATCGGT GGTTGGcGTT 9180

GCCAATCTTT GCGTTTGTGA TGTGGTTGGT CTATTACTTA TCTATTCAAA CGGTTGGTAC 9240

AATGGGGACA GACTGGTTAA ATGATGTCTT TTTTGGGGAA TGGGTTCCTC AATTTGTGGG 9300

AAATTGGTTG GCGCAATGgC AAGTAGCGCC ATGGATGCAA AGTTTAATTT TAGATGGGAT 9360

CATTGCAGGT GTTGGGGCGG TGTTAGGATT TTTACCACAA CTGGCCGTCC TGTTTTTATG 9420

TTTAGGTTTt TAGAAGATTG TGGCTATATG GCCCGGATTG CCTTTGkCAT GGATCGCTTG 9480

TTTCGGAAAT TCGGCCTATC TGGTAAGTCG TTTATTCCAA TGTTAATTGC GACAGGTTGT 9540

GGGGTCCCAG GCGTGATGGC TAGTCGAACG ATTGAAAACG AACGCGATCG GCGAATGACT 9600

ATTATGGTCA CGACTTTTAT GCCATGTTCA GCGAAGTTGC CAATTATCGC TCTGATTGCA 9660

GGTGCATTTT TTCCTAACCA AAGTTGGGTC TCGCCCTCTG CTTATTTCTT AGGGGTTGCT 9720

TCCATTGTGT TATCAGGAAT TGCTTTAAAA AAGACCAAGC ATTTTTCGGG AGATCCAGCA 9780

CCATTCATTA TGGAACTTCC CGCCTATCAT TTGCCACAGC TTAGAAGTGT GGTTCGGCAT 9840

GCCTATGAAC GATGTCGCTC ATTTGTAAAA AAAGCAGGTA CGATTATTTT TGTTTCCAGT 9900

ATTCTCATTT GGTTCATGTC ACATTATTCT GTGACGTTTC AACCCGTTCC AGAAAGTCAA 9960

AGTATTCTAG CATTTCTAGG TAAAGGGTTG GCGGTGCTAT TTATCCCATT AGGCTGGGGT 10020

AATTGGCAAG GGGCAGTTGC CACAATTACA GGATTGATAG CGAAAGAAAA TATTATTGGT 10080

ACCTTAGGTA TTTTATTTGG CAATGTGAAA GATGTTTCTG AAAACGGGGT AGAAGTCTGG 10140

GGCGCACTGC AACACACCTT TACGCCAGTC GCGGCCTATT CGTTCTTAAC TTTTAACTTA 10200

TTATGTGCAC CTTGTTTTGC TGCAATTGGC GCAATTCGTC GTGAAATGGG AGACCTCAAA 10260

TGGACATTAG GAGCCATTGG TTATCAATGC GGTTTGGCTT ATATGGTCAG TTTTGTGATT 10320

TATCAATTAG GCCATGTGTT AGTCGAAAAA GGCACACTAA CTCTGGGGAC TTTTTTAGCG 10380

ATGGGTGTTG TTTTAGCTGG ATTTTATTTC CTTATAAGAA AACCAAAGCC AGACAAAGAA 10440

AGTGTCCAAG CAATCACTTC TTTAGAAAGA GGGTAATAAC GGATGGCAAC AATTATTTTA 10500

TCCATTTTAA TTTTTGGAAC AGCAGGAATC ATCACCTATC GACGAATCAA AAAAGGGGAA 10560

AGTTGTGAAG ATTGTCAAAC AGCTTGTCCT GTAAAAAAAG AACAATCTTC ACAaGAATAA 10620

CAACTGAAAA AAATTCCCTC TTGTATAAAT AATAGCGGAG TGCTACAATA TTGAGGTAGC 10680

TGAAAAGCCA TTTTATACAA CGGGGGCATA GCTCAGTTGG TTAGAGCATC TGACTCTTAA 10740

CCAAATGCTC TGCTTGTGAA AAAAGCTTTA TAAAACTTTA TAAATAGGTT TTTAGTTTGT 10800

GAATTTTTTT ATTGCCAACA TACTGGCAAC AAATTGCCAA CGAAAAAAAT TCTAAAAACA 10860

AGAAACGATA CCACGAAGAT ATTGTTTTTT TGTTTTTTAT TCTTTAAACC CAAATGCGTT 10920

TTCTAATGCT TCTGCTGCAC GCTCTTTTAC TCCTGCAGTT AAATGTGTAT ATAAATTAGC 10980

AGTTGTCGAA TACGATGCAT GTCCCAGCCA TTCTTGAATG GATTTCATAT TCTGTCCACT 11040

ATCTAACATT ATTGAAGCAT TAGAATGACG CAAGTCATGA AAGCGAATTT TTTTCAATCC 11100

ATTTTTCTTA ATCAACTTAC TCAAGCGATG TGTTACATAT TGAGGCTTCA TCAACCGTCC 11160

ATCTTCCATA GTAAATACGT ACTCACTATC TTTTAAAATA TAATCCTCTC CTAAAAGAAT 11220

ACGATTATCT TGTTGCCATT TAGCTTGATT TACTAATGCT CGTTCAATCG TTTCAACCAA 11280

AGGCAACGTA CGATAGCTGG AAGAGTTCTT TGTTTGATCT TTCTTTATTA ACTGGGACAT 11340

ATTCTCTGGA TTTTCTATGA TTACATGGTT TATCATGATT TTTTTGTTTA AAAAATCAAT 11400

AGCAGACCAT TTTAATCCTA AAACTTCACT TCTTCGCAAT CCATAAAATG CTGTTAAAAG 11460

AACTACCAAA TAGAGCTTTT CTTCTGACAA AATATCAAGT AATTTCAAGA TCTCGTGTTG 11520

ATTATAAACT TGAGAAACAA ATTGTTTCTT TTTTGGCTTT CTTACTTTAT CTGCTACATT 11580

TGTACTTATC AATTCTGTAA TAACAGCCTC TTGAAGAGCT TTGCGTATAA ATGCGTGATA 11640

ATGAATGACC GTATTAGCTG TACAGTCATC ATTTAACATA TATTGGTAAA AATCTTGTAA 11700

ATGATAAGGC TTTAACCTAT TTAGTTTAAT TTTTTTAGGA TAGAAAAACT TGTAAATTCG 11760

ATTATGGACA TAATGCTTAT ACCCAACATA TGTGTTAAAT GCTACTTGTT TTTTTGATAC 11820

TTCTAAATAA TTAAGCAAAT AGTCATAAAA CAATAAATCT GGATTAGCAC ATTCACATTC 11880

CATAGCTAAT TGACTTTCCA ATtCTAGTAA AAATTCATCT CTAATTTTTT CtGCTCGTTT 11940

TTTATTTCCT TTCTTTTTTA ATTTTGTATC TTTCCACTTT TGTCTAATTT TTCCGTTAGA 12000

GTCTTTATAA GATACAATTA CTTGTAATGT TCCTCTTTTC TCAATTATGT TTGCTCTTAA 12060

TTTCAATTCA ATCTCTCCTC TTAAACTGAA ATATCGGCAA AAATTTTGAC TATTCTATTA 12120

TAGCATGATT TTTACCTTCA TCTTGTATTT TTAATATATA CTCAATGACA GCTTTTTTAG 12180

GTATTTTGTA CCTTGCTCCA ATTTTTATAT TTTCTATTGA TTTATCTTTC AGCAAATTAT 12240

AAGCTTTATT TTTTCCAATA TTTAACATTT TAGTTAAATC AGTAATTTTC AATATCGAGG 12300

GATAACTTTC AAATATTTTT TCATAATCTT GAAAATCCAT AAAATTCACT CCTACATTCT 12360

GATAATAAGT ATTGCACTTT TTTATTTTAG GTGTTAGAAT AACAAAGTAT TAATTTGGGC 12420

TGTTAGCTCA GTTGGTAGAG CAGATGACTC TTAATCATCG GGTCGCGGGT TCGAGCCCCT 12480

CACAGCCCAT TGGGTGCCAA ACCCTTAAAT ACAGTGTAAA CGTTGGTGCT TAGCATTTAG 12540

AAAACGTATA TACTGTATTT TTTTGTTTTA AAAAAAGACT GTAGTACCCC GCTTTTTTAC 12600

TAAAAGCCCA CTGTAGTCAT TGATATTTTT GATAAGAAAG GAACATAATT ATGAGTAATG 12660

AAATTGCTAA TATAACAATA ATCCGAGAAA CTTTACAAAA CTATACTGCA AATGAAATCT 12720

CTAAACATAC AGGATTAAAT TTGAGTACTA TAAAAAACTA AAAAGTGGAG AACGTTTAAT 12780

AGAAAAACTA AATTTACATG ATGCTATTTG TCTCACTGAA TTTGGTTTTA AAAATAATAG 12840

AAAAAGTGTT GAAATCAACA TATGGAAGTA ATTCTTAATC AGCTGCTTTT GTCAAAAGCA 12900

CATAGGAATT TCACCTCTCT CCAAGTTTAC GGAGAGCcTT ACGGAAGTAT CATTATCCCA 12960

AATTCAAATC ATCGCTTtTT ATgACTATAA AAATtaTTTT AGAATGTTtA TCGCTCCTTC 13020

TTTACGATAA GAAATCGTGG CACAACTAAA ATATGCTCCT TGGAGTGTTA ATGTACTGAT 13080

AGAATGTATT TAATTTTCAA TGATCAATAG GTTTTCTACA TAAAAGAATT TACTAATCTT 13140

TTAAATATTA GAAGATTATT TTAAACAGAA AAAGTGGAAA GGGGCTTTTT TTAGTTAGTT 13200

AATTTTAAAT TTTAAGATAG CTTCTATCAA TTTAATTTCT AGCTGTCTTA GCATATACTC 13260

ATCTACATAG GTAATAAATT GCTCCGCCTC ATTAAATCCA TTTGTCAAAC ATAACTTAAC 13320

TATATAACTA TGATAATGTT TGAGTATCAT CTTGATCGAA ATTGGATCTT CTTCTAGAGC 13380

AGATTGAATA ACGGATAGCG GAACTAATGG ATACTGCCGT TTTTTTTTGT TGTCTACCAT 13440

AGGTCATACT CCGTTTTTAA TTGTTCAACT GCCTTTTTTC GTTGATACCA GATACCACTT 13500

GTACTCATGT TGAGAATTTT ACCGATTTCA GTATTATTAA AACCTGCAAA ATAATAAAGT 13560

AGGATGACCT TTCTTCTTAT CTCTGATAAA TGATCAATGG CTGTTGATAA CTTTTGATCT 13620

GAAATTAATA TTTCCATTCC AAACAAGAAA TAGACATCTT GTTCATCAAA ATGAGTTTGA 13680

GAGTACTTGC CtTCAGTTGA ATCTGTCAAT GGTTCCTCAT TATACTTTAT GTTAGkGTTT 13740

CGTTTTCGAA TAGACGTAAT TTCGTTTCGT AATACTTTTT TGCAGTAGGC ATCAAATTGG 13800

TGTTGTATAG ATTTCTCATT ATAATTTTGT TGCATGATAA TATGCTCCTT TTTCCCCTTT 13860

CCACTAATAC AACAATCAGA GAGCCCTTTT TTTTCGAAAA AACTAGAAAA AAACAAAAGA 13920

CACGTAGTTT TTATCTACGT GTCTTCCATT TTTTTCAGAA CAACTGCAAA GAAAAGAAAT 13980

ACATATTATA CAAAACAAAA AAATATTCTA ATTTTTTTCA TATAGAAATA ACTCCTATTT 14040

TAAAATTATA AATCGTAACG AATGCGATTT ATCCAATGTA ATTTTACATA CTTTATTTTA 14100

TTTGTCAACA AAAAATGAAC GATCATAATA ACGTTTTCCA AACTTATTTC GCTAGATACT 14160

ATTTTAATTA TTCTTCACTA TGACATCATA ATAAAGTTTT AATCAGAAAA TTTAGACGAT 14220

ATGAAACGGA ATAATCAATT AAAAGCTAAA TATGCTTTGA GAAAAAATAA TTTGTTTTTA 14280

GCGGAAGACA TATAGATAAA AACTATGTGT CTTTTATTTT TCTAGCTTTT TCGAAAAAAA 14340

AGGGCTCTCT GATTGTTGTA TTAGTGGAAA GGGGAAAAAG GAGCATATTA TCATGCAACA 14400

AAATTATAAT GAGAAATCTA TACAACACCA ATTTGATGCC TACTGCAAAA AAGTATTACG 14460

AAACGmAATT ACGTCTATTC GAAAACGAAA CmCTAGAATA AAAGAAAATG AGGAACTATT 14520

GAACGTCTCA AATGAAGCTA AGTACCCTTA CTCATTTTGA TGAACAAGAT ATCTATTTCT 14580

TGTTTGGAAT GGAGATATTA ATTTCAAATA AAAAGTTATC AATAGCCAAA CCAGTTATTG 14640

GACATAAAAA ATTTTTTATT TTGCAAGTTT AATGATACTG AAATCGATAA AATTCTCAAT 14700

ATAAGTATGA ATGGATTTAA TGAAGTGGTC AATTTATTAC CTATGTAGAA GAGTATACAT 14760

TGAGGGAGTT ATAAATTTTC AGAATCTATA TTGAAATTTA AAGTTAAATA ATTAAAAAGT 14820

CCCCTTTCAA TTATTTTGGT AGAAAATTTA TAGaTCATTG AAAATTAAAT ACATTCTAGC 14880

AGTACGTTAA CACTTCAAGG AGCATATTTT AGTTACGCCA TGATCTCTTA TCGTAAAGAA 14940

GGAGCGATAA ATATTCTAAA ATAATTTTTA TAGTCAATAA AAAGTGATGA TGTGAAGTTG 15000

GGATAATGAT ACTTCCGTAA GGCTCGCCGT AAACTTGGGG AGAGGTGAAA TTCCTATGTG 15060

CTTTTGACAA AAGCAGCTGC TTAAGAATTA GTTTTAAGAG TAAATTAAGT AAGACAAAAA 15120

GCATCATGTA AGTTTAGTTT TTAAGCACTT AATTTTGTAT ATTTAGAGAT TTCATTTGTA 15180

GTATATTTTT TAGTTTCTTG aATTTTAtTA TATTAACtAT TTCATTACTC ATAATCATGT 15240

TCtTTTTTTT AAATCtAAaT GATTATAGTT TAGTAAAAAA GAGAGGATGG AGTCCCTTTT 15300

GAGAATATAA AATACGGTAT ATACGTTTTT TATGCTAAGC ACCAATGTTT ACTCTGTATT 15360

TAATAAACTA ACATAAGAGA CTATAACCTT TGACCCGATA ATTATGATTA TAGTTATTAG 15420

TATAAATCTT TAAGTAATAG AAAAGATGTG TTGAATGAAG TATACATTCT GTAATTAACA 15480

ATGAATCATC AGTCATTGGA TAGGAGGCTT TATTTTAAAA GTGTTAAACG ATTCATCTTT 15540

AAGGTGTAGA CAAGAATAAG TAAAAATAAA AAGCTAACTA TCTCTCGAAA TTAAATATTC 15600

TACAGAATGT TGTTTAGAAA AAAGGACTAG TAACAAAAGC TACCTACATT AGGTAGCTTA 15660

AAAGAGTGAA AAGAATATAT GTTAAATTAC TAATTTTATA GCATAAAGTA AAATATTTGT 15720

CAAAAAAAAG GGGGAGGGGA GAAGGATGTA TTGTGCACGT GATCCATGCA GAAAGAATCA 15780

TTGCCTATAA AATAmAAAAA TTTACGAGGA GAATGATTCT TGAAGCAACA AATGAAAGCA 15840

AAAAAACAAT ATAAGACATA TAAAGCTAAG AATCACTGGG TAACTGTCCC TATTCTTTTT 15900

CTAAGTGTGT TAGGAGCCGT AGGATTAGCT ACTGATAATG TACAAGCCGC GGAATTAGAT 15960

ACGCAACCAG AAACAACGAC GGTTCAACCC AATAACCCCG ACCTGCAGTC AGAAAAGGAA 16020

ACACCTAAAA CGGCAGTATC TGAAGAAGCA ACAGTACAAA AAGACACTAC TTCTCAACCG 16080

ACCAAAGTAG AAGAAGTAGC GCCAGAAAAT AAAGGTACTG AACAAAGTTC AGCTACCCCA 16140

AATGATACCA CAAACGCGCA ACAACCAACA GTAGGAGCTG AAAAATCAGC ACAAGAACAA 16200

CCAGTAGTAA GCCCTGAAAC AACCAATGAA CCTCTAGGGC AGCCAACAGA AGTTGCACCA 16260

GCTGAAAATG AAGTGAATAA ATCAACGTCC ATTCCTAAAG AATTTGAAAC ACCAGACGTT 16320

GATAAAGCAG TTGATGAAGT AAAAAAAGAT CCAAACATTA CCGTTGTTGA AAAACCAGCA 16380

GAAGACTTAG GCAACGTTTC TTCTAAAGAT TTAGCTGCAA AAGAAAAAGA AGTAGACCAA 16440

CTACAAAAAG AACAAGCGAA AAAGATTGCC CAACAAGCAG CTGAATTAAA AGCCAAAAAT 16500

GAAAAAATTG CCAAAGAAAA TGCAGAAATT GCGGCAAAAA ACAAAGCrGA AAAAGAGCGy 16560

TAyGArAAAG AAGTCGCkGA ATACAACAAG CATAAGAACG AAAACAGCTA TGTCAATGAA 16620

GCGATTAGTA AAAACCTAGT GTTCGATCAA TCTGTCGTGA CGAAAGACAC TAAAATTTCG 16680

TCGATTAAAG GCGGAAAATT TATCAAAGCA ACTGATTTTA ATAAAGTAAA TGCAGGGGAT 16740

TCAAAAGATA TCTTTACAAA ATTACGGAAA GATATGGGyG GGAAAGyTAC TGGCAACTTC 16800

CAGAATTCCT TTGTAAAAGA GGCAAATCTT GGGTCTAATG GTGGGTATGC GGTTCTTTTA 16860

GAAAAAAATA AACCAGTGAC AGTGACCTAT ACAGGACTAA ACGCTAGTTA TTTAGGACGT 16920

```
AAAATTACAA AAGCAGAATT TGTTTATGAA CTACAATCCT CACCAAGCCA AAGTGGAACG    16980
TTAAATGCAG TATTTTCAAA CGATCCGATT ATCACrGCTT TTATTGGTAC AAACAGAGTC    17040
AATGGTAAGG ATGTTAAAAC ACGCTTAACG ATTAAGTTCT TTGATGCGTC AGGTAAAGAA    17100
GTACTACCAG ATAAAGATAG TCCATTTGCG TATGCGCTGT CTTCTTTAAA TTCAAGTTTA    17160
ACGAATAAAG GTGGCCATGC GGAATTTGTT TCTGATTTTG GGGCsAACAA TGCGTTCAAA    17220
TACATTAATG GyTCrTATGT GAAAAAACAA GCGGATGGAA AATTTTACTC ACCGGAAGAT    17280
ATTGACTATG GCACAGGACC TTCTGGATTG AAAAATAGTG ATTGGGACGC TGTAGGTCAC    17340
AAGAATGCCT ACTTTGGTTC AGGTGTAGGT CTAGCyAATG GrCGTATTTC CTTTTCTTTT    17400
GGTATGACAA CAAAAGGAAA AAGTAATGTG CCTGTATCTA GTGCGCAATG GTTTGCCTTT    17460
AGyACTAACT TAAATGCGCA ATCAGTGAAG CCTATTTTCA ATTATGGGAA TCCAAAAGAA    17520
CCAGAAAAAG CAACGATTGA ATTCAATsGA TACAAAGCCA ATGTCGTTCC TGTwCTTGTG    17580
CCrAAtAAAG AAGTCACTGA TGGyCAGAAA AATrTCAATG ATTTAAATGT GAArCGTGGC    17640
GATTCTTTAC AATACATTGT GACAGGGGAT ACGACAGAAC TTGCCAAAGT AGATCCAAAA    17700
ACAGTAACmA AACAAGGGAT TCGAGATACm TTTGATGCAG AAAAAGTGAC GATTGATTTA    17760
TCCAAAGTGA AAGTTTATCA AGCAGACGCA AGTCTrAACG ArAAAGACTy AAAAGCTGTT    17820
GCTGCAGCrA TTAATTCAGG AArAGCTAAA GACGTGACTG CTTCTTATGA yCTcAaTTTA    17880
GATCAAAACA CCGTCACAGC AATGATGAAA ACCAACGCrG ACGGyTCyGT TGTTTTAGCA    17940
ATGGGGTATA AATATTTACT TGTCTTGCCG TTTGTAGTGA AAAATGTAGA AGGCGATTTT    18000
GAAAATACAG CTGTTCAGCT GACAAAyGAT GGnGAAACGG TAACAAATAC AGTGATTAAC    18060
CATGTGCCaG GTaGTAATCC TTCCAAAGAT GTAAAAGCAG ATAAAAACGG TACAGTTGGC    18120
AGTGTTTCTC TACATGATAA AGATATTCCG TTACAAACAA AAATTTATTA TGAAGTGAAA    18180
TCTTCCGAAC GTCCAGCyAA CTATGGCGGA ATyACmGAAG AATGGGGCAT GAATGATGTC    18240
TTGGACACGA CCCATGATCG TTTCACAGGk AAATGGCACG CTATTACrAA yTATGACCTT    18300
AAAGTAGGGG AyAAAACGTT AAAAGCAGGA ACAGATATTT CTGCCTACAT TCTTTTAGAA    18360
AACAAAGACA ATAAAGACTT GACGTTTACr ATGAATCAAG CATTATTGGC mGCsTTAAAT    18420
GAAGGAAGCA ATAAAGTAGG CAAACAAGCT TGGTCTGTGT ATCTGGAAGT CGAACGGATy    18480
AAAACAGGTG ACGTAGAAAA CACGCAAACA GAAAACTACA ACAAAGAGCT TGTkCGTTCT    18540
AATACsGTGG TGACGCATAC rCCTGATGAT CCAAAACCAA CCAAAGCCGT TCATAACAAG    18600
AAAGGGGAAG AyATTAAyCA TGGAAAAGTk GCTCGTGGTG ATGTTCTTTC TTATGAAATG    18660
ACnTGGGACT TAAAAGGGTA CGATAAAGAC TTTGCCTTTG ATACAGTCGA TCTTGCGACA    18720
GGCGTTTCTT TCTTCGATGA TTACGATGAA ACGAArGTGA CACCAATCAA AGACTTACTT    18780
CGTGTCAAAG ATTCTAAAGG GGnAGACATT ACGAACCAGT TCACGATCTC wTGGGACGAT    18840
GCCAAAGGCA CGGTGACrAT yTCTGCCAAA GACCCACAAG CCTTTATTCT AGCGyATGGT    18900
```

```
GGGCAAGAAT TGCGTGTAAC rCTCCCTACA AAAGTCAAAG CcgATGTTTC TGGkGATGTT    18960
TATAATTCAG CGGAACAAAA TACATTTGGy CAACGAATTA AAACCAATAC yGTTGTCAAC    19020
CATATTCCAA AAGTGAAyCC TAAAAAAGAC GTGGTTATTA AAGTyGGTGA CAAACAAAGt    19080
CAAAATGGyG CCACAATCAA ATTAGGGGAG AArTTCTTCT ATGAATTTAC AAGTAGTGAC    19140
ATTCCTGCAG AATACGCTGG wGTTGTGGAA GAATGGTCGA TTAGCGATAA ACTAGACGTC    19200
AAACATGACA AATTTAGTGG cCAATGGTCT GTGTTTGCCA ATTCTAATTT TGTTTTAGCA    19260
GACGGAACCA AAGTGAATAA AGGGGACGAC ATTTCGAAAC TATTCACGAT GACCTTTGAa    19320
CAAgGGGTAG TGAAAATCAC GGCCAGTCAA GCCTTTTTrG ATGCGATGAA TCTAAAAGAA    19380
AACAAAAACG TTGCACACTC ATGGAAAGCG TTCATTGGTG TAGAACGAAT TGCGGCAGGA    19440
GACGTTTACA ACACAATCGA AGAATCTTTC AACAATGAGA AGATTAAAAC kAATACGGTA    19500
GTGACrCATA CGCCAGAAAA ACCACAAACr CCACCAGAAA AAACAGTGAT TGTACCACCA    19560
ACACCAAAAA CACCGCAAGC ACCAGTAGAG CCATTAGTGG TAGAAAAGGC AAGTGTrGtG    19620
CCAGAATTGC CGCAAACAGG CGAAAAACAA AATGTCTTAT TAACGGTAGC TGGTAGTTTA    19680
GCCGCAATGC TTGGCTTAGC AGGCTTAGGC TTTAAACGTA GAAAAGAAAC AAAATAATTG    19740
AAGACGAAAA GAAGGACGAA yTGmTTGTCC TTCTTTTTTT AGTTATGGAG GGGAAAmATT    19800
GGAAGCAGTA GTAGtAGAAA GAGAGGCAAA GGGGATGAAA GAAATTGCCA TCCAAGAAAA    19860
AGATTTGACC TTACAGTGGA GAGGAAACAC AGGTAAGTTA GTTAAAGTTC GATTrAArAA    19920
TACACGTGCA ATGGAAATGT GGTACAACAA ACAAATTACC GAAGAAAACA TTCAAGAGAT    19980
CACCACGTTG AATATyATTA AAAAyGGAAA ATCTyTGGCA TTAGAAGTAT ATCCAGAAAA    20040
AAGTATCTAT GTGAAACCAA ATTTAGGCAG AATCAATGTG CCTGTCTTTT TTATCAAAAC    20100
ACCTATTAAT AGAGGGATTT TTGAAGAGAT CTTCGGCGAA ACTTTAAAAT CATAAATAAA    20160
AAATTATTTG GAGGAAATTA CAATGAAAAA AATTATTTTA TCAAGCTTGT TTAGTGCAGT    20220
ACTAGTATTC GGTGGCGGAA GTATAACAGC ATTCGCTGAC GATTTAGGAC CAACAGATCC    20280
AGCAACTCCA CCAATTACCG AACCAACTGA TTCTAGTGAA CCTACGAATC CTACTGAGCC    20340
GGTGGATCCT GCAGAACCGC CAGTAATACC AACTGATCCA ACAGAACCAA GCAAGCCAAC    20400
CGAGCCTACA ACACCGAGTG AGCCAGAAAA GCCAACAGAA CCAACAACGC CAATTGATCC    20460
TGGAACGCCG GTTGAACCGA CTGAACCAAG CGAGCCAACA GAACCTAGTC AACCAACCGA    20520
GCCTACAACA CCAAGCGAAC CAGAAAAACC TGTTACTCCA GAACAACCGA AAGAACCAAC    20580
TCAACCAGTG ATTCCAGAAA AACCAGCAGA ACCAGAAACA CCAAAAACTC CTGAACAGCC    20640
CACTAAACCA ATAGACGTAG TCGTTACACC TAGTGGAGAA ATTGATAAAA CGAATCAATC    20700
GGCAGGAACA CAACCAAGTA TTCCTATTGA AACAAGCAAC TTAGCGGAGG TAACACATGT    20760
ACCAAGTGAA ACTACTCCAA TTACAACAGA AGCTGGGGAA GAAATTGTAG CAGTAGATAA    20820
AGGTGTTCCG TTAACCAAAA CACCAGAAGG ATTAAAACCA ATTAGCAGCT CGTATAAGGT    20880
```

TTTACCTAGC GGAAACGTTG AGGTAAAAGC AAGTGATGGA AAAATGAAAG TATTGCCACA 20940

TACAGGAGAG AAATTCACAC TCCTTTTCTC TGTATTGGGA AGCTTCTTTG TATTAATTTC 21000

AGGATTCTTT TTCTTTAAAA AGAATAAGAA AAAAGCTTAA TCAGAAAGAG TAATTTCGAT 21060

AAATTTCTGT GAGGAAAGGA GGTGCTTTGC TTGTTTATTT TATTTACTAC mAAGCnAAGT 21120

CATGTATTAC TAAATGAATT AGGGTTTAAT AGAAAAATTG ACCTGGTACA AGGTATTGAA 21180

CAGTTAACAA AAATCAATGC GGATTTAAAA AAATCGACGG ATCGAATTCT AATTACGTAT 21240

GAGATTTATT ACAACGTGAC AGACGAGAAA GCTTTTTTAA GAGAAACAAT GACTTTGCCG 21300

GTAGTTGAGG AAACTGTCCT CCAAAACATG GTACAGACTG TAGAAACCAA TGAAGAAGTA 21360

AGACAAGAAG ATTTAGAAGC GTTTCTTTAC CTGTTCAAAA GAAATGTGAG TAATAAGTCG 21420

AAGAGAATAG GCCGTTCAAC AGGTTCTTTA AAAAAGAAAG CAGAAAAGCA ACAAAAATCA 21480

ACTGTGTTTA CAACAATTTC TCGCTATAGC TATTTCTCAG TGGCTTTACT TATATTTTTA 21540

GCAGTAGGTG TTTTTACAGG TGCGTCATTT GGTTCTCAAA GTAAACCAAC GAGCCAAAAT 21600

GATTCTAAGT TAGAGGAACG ATTAACACAT TTACAAGAAC AAGTGCATGA GCAACCACAA 21660

GTTGATGTTG TTGCTCGGTT TTTCTTATCC aGTCTGTACT CTGGTGAAGA GAGTGACAAA 21720

GCCACGCAAA AGCAAGTTAA AAAGTATGTT TCTGAAGGTG TACTCAAAGA GATTAAGGGG 21780

TCTAAAGAAC AAATAAGAAC GTTATTTCTT TGGGAATCAA AAAAAGAAGG CAACCAATGG 21840

CTCCTTACGT ACGTTGTTAC CTTGAAAGTG GCTAAAGAAG AAACTGAAAT TCGTCAAATG 21900

AAAGTAGCTA TGGAAAAGAA AAACAATAAT TATCAAGTAA CAGAACTTCC AGAAGAAAAA 21960

GAATTTACTA TTAATGAATA ATAATTGAGG AGGCAATTCT ATGAAATATG AACGTCCATT 22020

AAAAAGAGAG TCACAAATCA AGGAGTTTGA GTTAGGCACG CACGCGGCAG TAATTGAAAA 22080

AGTTCAAAAG AAACGGTCTC AAAAAGGGAA TGATATGTTT TTACTCTCTC TTTTGGGAAA 22140

AAGTAACGAA AAAGGGGTTT ATTTTCTTAC cTTCGGGAAT GATTATACTG AAGATAATTT 22200

ACGCTATATT TTAGCCAGTA TCCAAGACAA CGGTGTAGAG ATTCCAGACG TTGACTTTGG 22260

CTATAATCGA GAAACTTTTG AATTTCTGAA AGGTAAAGAT GTCTATATTC AGGTTGAAGA 22320

mCAAGAGTAT AAAGGGAAAG TTAAACaTGC GGTAACaAAT TTTTTAACTC aGGATGAATT 22380

TGAAGAAAGC GAAGAAATGG AGTTTTCAGA AAGCAATACT GAAGAAGATT GGTAAAATAA 22440

GCCATTTCTT AGAATATAAA AACTGAAGGA GAGACAAAAC TATGGATATA GTCACAAAAA 22500

TTATTACACT AATTGGAGGA ATCATTGGAC TTGTTTCAGC AGTTAGTATT ATGTTTGGCG 22560

TGAAAGAAAT TCGTTCAGGT ATGTCGAATG ATGATCCACG TACGCTTGAT AAAGGAATAG 22620

AAAAAGTAGT TGTTGGTGGG GCTGTTATTT TAGCAATTGG TGGGGTAGTC GCCTATGTTA 22680

TTACACAGGT AGGaGCTATT CGTTTTTAAA AGGAGGAAAA AAATGAAGCT ATTAGTAGAA 22740

ATGATAGTCA ACGGTCAAAC TGAATGGGAA GTTGTAGAAG AAGAAAATGC GCCACAGGCA 22800

ATTATTCAGA GTCGAGGAGG CTTTTCATTT GATGAGAATG GAGAGCTCAT TGTCAATGAT 22860

GATGAAATAA GTTATACAGG CGTGTTTGAG GTTTGCGAGA CTAACTTATT GGATTTCACT 22920

GTTAAGGAAG CAGAAATCCA TAGATTTTAT CACAAAAAAT TAGAAAAGCT AGGGaTAGAT 22980

CCTCTAACGT TTGAAAACTC GCmAGAAATA GCTAACTGAG CGGAGGGGaA GCnATGAGAA 23040

ATCCAAAAGG AACGACCGAT CAGCTATTTA AAAGTTTAGC TGAATACmAT CCmACmACgA 23100

aTGrAGCCAT GTCTAAAAtA GcGAAAGTCC tAGTCCCtTT AGGAATCgCT ATTTTAGGCA 23160

TTCTTTTTAT GATTGAATTG GCGAATACCC AAAAGAAATT TCAATCAGAA GACGGTGGTT 23220

TGACAATCGA GCTATTAACT AACATTGCAT TGAAGTATGT AATCGCGTAC GTCTGTATTA 23280

TGGGTTCTGG CTACATTATT GATGGCATTG TTTGGTTTAC AATTCAAGCA GCCAAATGGA 23340

TTAATTCCAT TGTAACAGCG ACAGGAACAA GTGAAGCGAT CCCACAAATG GAAAAAGTTT 23400

CTTGGTGGGC AAAACCAATT GTATTTTTCT TTCAAATATT TGCATATATT GCATTAATAT 23460

TATCTAGTAT TATTGCAAAT ATTCTTATTT TCTTACGAGG GATACAATTG TATATGATTA 23520

AAGCTCTTGC ACCGTTGTTT GTGGCGTTTT TTGTTCATGA TGAATTACGT TCGATTGCCA 23580

TGGGGTACAT GAAACAAATC ATGGCCTATG CGCTTCAAGG CGTTCTGTTG GTTCTATTAC 23640

TTGGGTTAAT TCCGATATTA ACAGCAAACG ATTATCTATC CTTTGAATCT CTTGAAGGTG 23700

GCATTTTGgC AGGTGCTGGA GCGATCATCA TTAATGTTAT GACGTACTTT AATTTAATTT 23760

TAAAATATGT GGCAGTGATT ATTCTTCTAG TTGGTTCACA AGGATTTGCC AAACGATTAG 23820

TAGGAGCGAT GTAAAATGGC GATTAGTTCA GAATTCTATA rAGATTTGTC AAAAGTAGAG 23880

AAAAAAATCT GGGGAATTAC AGTCAGAGAA TTTAAAGCGT ATGTTTGCTT TGTTTTTATA 23940

GGGATATTTC TTCTGTTAGA AGTATTTTTT TTGCCTGATC TTTTATTTAT GGTTTGTGCG 24000

TTAG 24004

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2707 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

ATTTGAAACC TGTTGTTGAA TCCATGAATA CTACTGGTTG GTAATTTGGA nnGATATCTT 60

GTTTCATACG TTTTTTCGCT CCTTTGCCCT GATCCATTTG CTGTATCAGA GTTATTTTGT 120

CGGTATACAA CAATGCTATC TTACCATGTT TTCAGTTGGT ATGCAATTAT CTTTTTGGAT 180

TTCTTTTAAT TTGTTTGCCA AACGAAACAT CATGGAAAAC TTCAAATAAT TGTTGATTAT 240

TTTGGGTCCG TTTCAACATT TTTATGAATT GATCCGTATA TTCCAGCGAG TCTCCTGACA 300

TATTATTTCT CAATTTCCAC GTTTCTTCTA GTTGTTCTGG TGTCATTAGC AACTCTTCTT 360

TTCGCGTACT TGATTTTTTA ATATCAATTG CTGGGAAAAT ACGGCGTTCT GCTAATTCAC 420

```
GAGATAAATG GAGCTCCATG TTGCCTGTAC CTTTGAATTC TTCATAGATG ACATCATCCA    480
TTCGGCTACC TGTATCCACC AATGCCGTGG CTAAAATAGT CAGACTGCCC CCTTCTTCGA    540
TATTTCTGGc AGCACCAAAA AATCGTTTAG GTTTGAAAAA CGCAGCTGGa TCTATTCCGC    600
CACTTAAGGT TCGACCACTA GGCGGGACCA CTAAATTATA GGCCCTTGCT AAACGAGTGA    660
TGCTGTCCAT TAAAATAACT ACGTCCCGTT TGTCTTCAAC TAAGCGCATT GCTCGATCCA    720
GAACCAATTC AGCAACTCGC GTATGATTTT GTGGTTGTTG ATCAAAAGTT GAAGAAACCA    780
CATCGCCTTT AACGCTTCGC TCTAAATCAG TCACTTCTTC AGGGCGCTCG TCAATTAACA    840
ACAGAATCAA TTCCACATCA GGATGATTTT CTGTAATCCC ATTAGCAATT TCCTTTAAAA    900
CACTTGTTTT TCCCGCTTTT GGCGGCGCTA CAATTAAACC ACGTTGCCCA AAACCAATTG    960
GGGCAAATAC ATCAATCATT CGTGTTGATA ATCTGCCAGC CGTTGTTTCT AACGTTAATT   1020
GTTTTTCAGG ATACAACGGT GTTAGCGCTG GAAAATGTGG CCGCTCTTTC GCTTCTTCAG   1080
GATCTTTCCC ATTGACACTT TCAACATGCA TGAGTCCATA GTAACGTTCT GATTCCTTTG   1140
GCGGACGCGC CTTGCCAGCT ACTTTGTCCC CATTTCTCAA TCCGAAACGA CGAATTTGAg   1200
AAGAAGAAAT ATAAAtGTCT TCTGCACTTG GTCCATAATT AATTGGTCGC AAAAAGCCAT   1260
AACCATCCTG AGAAACAATA TCTAAAATTC CTTCCATGAA GAAAAACCCT TGCTTCTCAG   1320
CTTGTGCACG AATCACAGCT AAAGACAATT CTTTTTTGGT CATTTGGCAT AATAAGGATT   1380
TTAAcTCTTt CGCGTAgCAT AAATATCTTt TAACGTACTG tTTTCCaATt CAGCCATgTa   1440
AATAGTCGCT CATTGATCCA CCTCAAnCGG ACCTTCTTGT TCTTCGGTTT CTAGCATTTG   1500
AATATCAGCG CCTAAAGCAG TTAACTTTTC GATAATGTGG TCGTACCCAC GTAAAATATA   1560
TTCTACATTA TAAATAGTGG TCGTACCTTC AGCCATCAAG CCGGCAGTGA CTAGGCAAGC   1620
GCCTGCCCGT AAATCAGAGG CGACTACTTC TGCACCGTGC AATTGATTTG GGCCATTTAA   1680
GATAATCATA TTGCCTTCAA CCGAAGCATC AGCCCCCATC CGCACCAACT CAGGGATATG   1740
CTTATTTCGT TGAGCATAAA TCGTATCAAT AATTTCACCA GTTCCTTGTG CTTTTAAAAG   1800
TAATGGCGTT AACGGTTGTT GCAAGTCTGT GGCAAAACCA GGATACGGAT AGGTTTTCAC   1860
CGTCGTCATT TTCAGATCAT GTGATGGATG AACTTCAATC ATGTCTTCTT CAATTGTCAT   1920
CTTAACGCCC ATTTCTTGCA ATTTCGCAAT GAAGCTTTCT AAATGCTCAT AAATAACATT   1980
TCGAACCTTG ATTCCTTCAC CCATAGCTGC TGCCATAGCT AAATAAGTCC CAGCTTCAAT   2040
TCGATCTGGA ATAATCGAGT GACGACAGCC ATGTAATTCT TCGACTCCTT CAATCCGAAT   2100
AACATCCGTT CCAGCTCCTC GAACTTTTGC ACCCATGTTA TTTAATAATG TAGCAATATC   2160
AATAATTTCT GGcTCACGCG CAGCATTTTC AATAATCGTT TTGCCTTTGG CTTTAACTGC   2220
GGCTAACATG ACATTGATTG TAGCGCCAAT AGATACCATA TCCATAAAAA TTCGGGTTCC   2280
TTGTAAGCCA GCTTCGTCTG TTCTCAAGTA CATCGCCCCA TGTTCGTTGG TTACATGAGC   2340
GCCTAATGCT TCAAATCCTT TGATATGTAA ATCAATTGGA CGAGGGCCTA AGTAACAACC   2400
```

GCCAGGTAAG CCGACAACGC CTTCCCCAAA TCGACCAAGA AGTGAGCCCA TAAAATAATA 2460

GGATGCACGT AAGCTATTGA TTTTTCCTTT TGGCATTGGC ACAGAAACAA CTTCTGTTGG 2520

GTCAATAATT AATGTATTAT TTTCAAAAGT AATTTTAGCA CCCATAATTT CTAAAATTTC 2580

AATTAATGAG TGAACATCCT GAATGTCAGG AACGCCATCT AAAGTCACGG GCGAATCTGC 2640

TAAAATAGCA GCTGGAATTA AAGCGACCAC ACTATTTTTG GCGCCGCTAA TGGTTACTTC 2700

GCCTTTT 2707

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10636 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

TCACCTTATT GATAGCGCTT CCTTATATTG nAATAATGCG ATTTTCTAAA CAAAAAATTA 60

TAGGTGGTAT TGATAATGAA AAATCCGTAT GTCCAAATAA CCGACCAAGA ACAATATAAA 120

AAGTCAATTC GTCGAATTCC TCCAAAAGAT CCTAATATTT TTTTAATTGA GAGACCTACA 180

GCAGGTGAGA TTATTGCTCT ATGGTGGACA ACCTCTTCCC GTGGCAAATC AACTGATAAA 240

TATATACCAA AATATTTCAA GCGTATCTAT CGAGTAGATA TGATTAAACA ACGTAAACTA 300

TTCATCAAAG CTAACGTTAT CTCCGAAGAA AATAATACTT ATGTACTGAC TGAACTAGGA 360

GAAATGTTAG TACATAAATA CTTTTATATT ATTGAAGAGC ATAATAATTC TTCCAATGAT 420

TCTTTCAATA CTGAACGAAG AAAGTTAGCT GATAAAATTC TATATTCGCT AGGCGGTGAA 480

ATATAATGCA AACGGAAGTT AAAGAGTTTA TTGCTTTTGA TATTGAAACA ACAGGTTTAG 540

ACTTCCATCA GGATGCTGTT ATTCAGATAT CAGCTGCAAG ATACGTTGAT GGTATTGAAG 600

TTGATTATTT TGATACATTG GTAAATAGTG ACTACATTCC AGATGAAATT ACCAAATTAA 660

CAGGAATCAC TAATGACCAA GTACTAAATG CGCCAACTTT AGATGAAGTT ATGCCTTATC 720

TTTTTGATTA TTTAGGAGAT ACGATTCTAG TTGGACATAA TATAAAAAGT TTTGACTTTC 780

CTTTTTTAAA AGCAAAAGGA TATAATATTG CAGAAGGACA TGAAATTTAT GATACTCGCT 840

ATTTTGCAGC CACACGAAAG CATGGGGCTG TAAACAATCA ATTAACTACA TTGAAATTAC 900

TTTTTGGTAT CGATGCGATT TCACATAATG CATTGAATGA TGTTAGGATA TCTGCAATTG 960

TATTTATGGA ATTATTAAAG ATAGAACCTC AAGAAAAAAG CTCTACTCTT AAGAAAAGTA 1020

GCTTGGACTC TTTTGAATTA GATGATGATA CTACACCTTT TTTTGAAGGA ATGTCTTTTG 1080

TAGTAACGGG GGCATTTGAT GATTCTAAAT ATAACCGTAA ACAAATTGAA ACACTGATAA 1140

AACAACATGG CGGTAGGGTA TCGTCAAGTT TATCTGCCAA GACTGATTAC TTTATCCAAG 1200

GTATTCAAAT TTCTTCCCAA CTTAAAGACG GAAAGCACAG CAGTAAAGAA CTAAAATATA 1260

TAGAGTTACG TGAAAAAGGG GTTGATATTT ATAAATTTAA TGGTAATCAG TTTTATGAAC 1320

TAATTACAAA CTATAGAAAA TTGAGGAAAA ATTAATGACA GAACTAAAGA AAATAACTTC 1380

AAAAAATGGA CATGTTTACC TCTTTAAAGA TATACAACAT AAACAACGAA ATGAGTTAGG 1440

CATAGCCTCT GGACAATCAG TACTGATTTC AATTTTTGAA CCGGATCCAT TCTATTTTAC 1500

AGAAGATATT CAACTTATTC ATTATAAAGA TATCGTTAAA ACCGAATCCG TTTCTTCAAA 1560

AGAAAATGAA TTTTTTGAAT ATGTAAAACA AAATTCACCA GTTGCTTATA AAAAGAAAAT 1620

CTCGGATAGT TATAGTATCT CTGAGTATAT TAATTATGCT CCACAACTTA AATTTAAGTA 1680

TATTATAAAA GATAGTATAC TAACTATTAC GGGGGAAATT AAAGACAGAA ATGTTAACTT 1740

TCCAATTTCT CAGCCTTACG TATATGTCGA TAAACCAGAA AATGAAATTC CGTTAGATAA 1800

TAATTTAAAA TTCCAATATG TAGTCAACGA TTTCGATATA CATGCTCCCG AATACAAAAT 1860

ATCGTTTATT TTAGGTACTT CAAAAGCCGA GTATGGTCGT TCTTTCAATA TGAAAAATTT 1920

CGAAGGTGTT TTACGCACCC CATCTGAAAT TAGTTCTATG AAAGCAAAAA AAAAGAAATA 1980

GTTACTTAAA TAAAGCAAAA AAGTTCTACG CCTAAAATTC GAATTTAGAC GTAGAACTTA 2040

TTTATTTCAA TGTCCGAACC AAATAAACTT CTTCACAAAA GCCTTTAAGT CCATTATAAA 2100

TCCGTTGTGC TCGTGTTTTC TTTTCACATA AAGCAAAGAC TGTCGGTCCA CTGCCGCTCA 2160

TCAACGCTGC ATCAGCGCCG TACTTCAGCA TGCGATCCTT GATTTGTTGC ACGATTGGAT 2220

GTCTGGCAAT CGTGACACTT TCCAACGCAT TGCCAACCGT TTGGGTCATC CCAGTATAAT 2280

CCCCATTTTC AATAGCTATT CTTAGTCCAG CAATATCTGG ATGGTGCAAT TCATCTACCG 2340

CTAAATCATT AAAAACGGTA CTAGTGGATA CACTAATTCT TGGTTTAACT AACACAATCC 2400

AACATTGAGG CATGGTCGGT AATGCTTCAA TTTTCTCCCC ACGTCCGTTT GCAAAGGCTG 2460

TTCCGCCCCT TAAACAATAA GGAACATCCA TACCAATTTG GCTGCCAATT TCACAGAGTT 2520

CGTCCATCGT CAACCCCAGA TTCCACAGTT TGTTCAAGCC TCGTAATGCT GCGGCACAGT 2580

CACTGCTCCC CCCTGCTAAA CCAGCTGCCA CGGGGATACG TTTTTCAATA TATATTTTAA 2640

TTCCTTTATG AATATTATAG GTACGCTTTA ATAAGTCAAC TGCTTGATAG ACATGATTCC 2700

GTCGATCCAC GGGTAAAAAA GAGCTGTCCG TTTCAATACG AATGATGTCT TCTTCTAGAT 2760

TTTCAAAAGT TAAATGATCG GCTAAATCTA CACTTGCAAA AATAGATTCA ACTTCATGAT 2820

AGCCATCCAC ACGTTTATGC AGAACATCTA ACCCAAGATT AATTTTCGCA GGAGCTCGTT 2880

CGATTATTTC CATCGGTTCC CTGATTCACC TTCCTTAATC TTTTGTCTAC CAACATTTTA 2940

ATCTTTTAGA TTCTATCATA TGTCCATGAA GAAGTACACC GAAGATTGCT CTTTCTCTGA 3000

CGACAGATGG CTGAAATTAA AAAGAGCTTA GGAACGAAGG GCTCTTCCTT CACTCCTAAG 3060

CGGACTTTCT AAATTATACG GCTTGACCGA TGAATTCGAT TTCAACGGTG CGTGTTAAAA 3120

CATCTGAATA GCTATAAGAT ACACGTTCAA ATGAGTTCTC GTCTGGATCC AAGTCAACAA 3180

CAAAAACAGA TGGATAAGTC TCCGTTAAGA TTCCTTTACG TTCTGTTTGG CGCTTTCTTC 3240

CAGTCTGAGC AACTAACATA ATTTCACTGC CGATACGGCC TTCCAAATCT TTCTTAATTG 3300

ACGATAACGT TGTTGGCATT GCATTCACCT CGAGAAACAT TATATCACTT TTTTTAAAAA 3360

ATTTCAAGTT TATCATACTA GAAGTCTGTT TGCAACTACT TTTGCTCGAC AAACCGGCTG 3420

TTTTGTCGAA TTTTTATTCA TTTTCTAATG TATATTTCAT TTTTGCTTTG GCACGATCAT 3480

AGGCACTTCG GACAGCTGAA AAAGTTCCCC CACGTACTTT AGCAATTTCT TCTAACGTGT 3540

AGCCTTTTAA GTATGCGTTT AAAATTTGTT GTTCCAACGA GGAAAGACGT GTCGGAAAAG 3600

CTGCCAATTT TTCTTGAATC ATTAATTGAC CTAATGGGTC GGTAGCCGCT GTCCCTAAAT 3660

AGTCCACTAA CTCTTCACCA GCGCTTTCTA ATCGTTGTTC CAGAGAAATA GACAAGGTAT 3720

CTGTTCGCCG CTTATAGGCG CATTGATGGC GAATTAAGCT ATGTACGTGA TTTTCAAAAT 3780

TTCGTTTAAA AAACTTTCCT AATGATGCTT GTTTTTCTGA ATCAAATCGT TCTAACGTCC 3840

GAAAGAAGAC AATTCGCCCT TCTTGAAGCC AGTCCTCAGT ATCATAATCT TTTAAATAGT 3900

ATTTTTTCAT GCACTTATAA ATCACTGGCT GATATTGCCC ATATAAGGCT TGAAATTGTT 3960

CTTGGTCTCC TGCTAATGCT TGTTGAATCT GTTCTAACAT AAAAAGCCCT CCTTTTCGAT 4020

AAACGTATCG AATGATTTAG GAGAGCATAG CATGTTAGAA ATTTATGACT TAAGCGAACA 4080

TTTGTTTAAT CTTTTTTTTG AGACAATTCT TCTAGCTTCT CAGAAAGTTT TGATAACTGC 4140

TCCACGTTCC ATGGAGAATT TCGACGAAAT TGCTGGAAAT GGATATCCGT TGCATGAATG 4200

GCAATCGTCT TTTCTGTTTT CTGGACAGCT TTGTAAAGTT CATTGGCCGA AGTCCGCAAT 4260

GCACCTTGTG AAAAGACCAC CCATTGTTCA GCCAAATCAC TCGTTGCGAC TGTTACTTGC 4320

GTTAAACGAT CATTTTTTTC ACCAGCTAGA CGTTCAATAT AACTATCTGC TGTTTCATCT 4380

TCTTTGGTGA AAACCACGGT CAAAAGGTAC TTTTTGTACG TTTGCTGAAT ACCTGGGACT 4440

AATTGGGCAT CAAAGACCAC AATGATTTCC AACCCTTCAT ATTTGGCGTA ATTGGACAGA 4500

CGACTAAGAA GCAGTTCACG CGCATCCTCT AAACGCTCTT GCTTTTTCAA CTGGACCAAT 4560

TCTGGCCAAG CACCAATCAT ATTATAGCCA TCAACGATTA ACAATTGTTT TTTCATACAA 4620

TCGCCTTCTT CTATAATGGG TTGCGGCCAC GATAGACTTC ATACATTAGC AAGCCTGCAG 4680

CCACACCAGC GTTTAAGCTT TGAACGTGTC CAGTCATTGG AATCGTTAAT AATTCATCCA 4740

CTTCTTTATG AAGACCTTGA CTCATGcCAC GACCTTCATT TCCGATGATT AAGGCAATTG 4800

CGCCTTGTGT ATTCCAGCGA CGATAATCCG TCCCAGACAT ATCTGTACCA AAAATCCAAA 4860

ATTGATTCTC TTTTAGCGTT GCAATACTTT GCGCTAGATT CGTTACACGA GCCACTGGAA 4920

TATGTTCCAC GGCCCCTGTT GACGCTTTGG TTACGGTTGG TGTAATTCCC ACTGCACGAT 4980

GTTTCGGAAT GATAATACCG TCCACGCCTG TCGCATCAGC TGTTCGCAAA ATTGAGCCAA 5040

AGTTATGTGG GTCCTCTAGA CTGTCTAAAA TTAGGAAAAA TGGTGTTGCT GTCTTTTCTT 5100

TTGTTTGTTC CAATAATTCT TCTAAAGTTA AATATTGATA CGCCGTGATT GCTAAAACCA 5160

TGCCTTGATG CACGCCGTGA TCACTCATCG TATCTAATTT TGCTTTAGGC ACCCATTTAA 5220

CTGGAACAGC CTGTTCTTTC GCTGCTTGTT TTAATTGCTC AATTTTCTCG CCACGAGCAT 5280

CTTCTTGTAA AAACAATTTA TTCCCGCGTC CTTGCTGCAA CGCCTCAACA GTGGCATGGA 5340

AACCAAAGAC AAAGTTATCT GCTAGCTCTT CTGAGGTTCC ATCAGCACCT GCTGTCGCTG 5400

GTCGTGGGTT CCGTTCTTTT GAACGCTTAT CTGAACGTTT CTTACCTTTT TTAAAAGGAC 5460

GAGATTTTTC ATTTTTCATT GGTTTCACCT ACTTTCTTGA TACACCAGTT AATTAATTCT 5520

TCTAAACGTT CTTTTTGTTG TGTTAAATGT AAATAACCCA TCAAGGATTC GAAACCGGTC 5580

GCTACCCGAT AGGTCGTCAC ATCAGCATTT TTGGCAAAGG TGTGACTTTT AGCATTCCGA 5640

CCGCGTTTGT ACATCACTTC TTCTTCCTCC GTTAGCATTT CTCCCGCTAA CATCGCTTGC 5700

ATTAAGCTTG CTTGGGCTTT TGCTGACACG TAATGTGTCG CCATGCGATG CAAAACATTC 5760

GGTTTCGTTT GTCCCAAGGA AACTAAATAA TCACGAATGT AAATTTCATA AATCGCATCG 5820

CCAACATAGG CCAAAGCTAA ACCGTTTAAT TGTTTGTAAT CTCTCATTCA CTTCTTCTCC 5880

ATCTCGTTCC TTGCGGAGTA TCTTCTAAAA CAATGCCTTG TTCTTTTAAT AAATCGCGGA 5940

TTTCATCACT TCGGGCAAAA TCACGGTCTT TCCGTGCTTG GTTCCGTTCT TCAATTAACG 6000

CATCGACTTG TGCATCCACT AATTCTTCGT TTTTAAAGAA TAAACCAAAA ATAGCTAACC 6060

AGCCACTGAA AAGCTTATCC ATCGCCACTA GAACTGGTTC CGAAACTGTC GCTTGTTCAC 6120

TGTACTGATT CATCATCTTC GCTAATTCAT AAACGACCGT AATCCCGTTA GCTGCATTGA 6180

AATCATCGTC CATTTCCGTA ATAAAACGTT GTTCTAATTC GTTTAATTGA GCCAAATAGT 6240

GTTCATCCTC TGGCAAACTT GCGACCGCAT CTGCTTGTCT AAAACGTAAA TTTTCAAACG 6300

CATTTTTCAA TTTTTGATAA TTGGCTGCTG CCTCTTTTAG AGTGCTTTCG CTATAACGAA 6360

TCGGGCGACG GTACTGGGTG GTGGACATAA AGAAACGTAA AATTTGCGGA TCCACTTTTT 6420

GAATCATTTC ATGAACAGTA ATGAAGTTAC CTAGGGATTT ACTCATTTTT TCATCGTCTT 6480

CACCAATTGT GACATAGCCG TTATGCATCC AGTAATTGGC AAAGGTATGG CCTGTTTTGG 6540

cTTCGCTCTG TGCAATTTCA TTCTCATGAT GTGGAAATTC TAAATCTTGT CCACCACCGT 6600

GAATATCAAT CGTCTCTTCC AAGTGTTTTG TCGCCATGAC AGAGCATTCA ATATGCCAGC 6660

CGGGACGTCC TTTCCCCCAA GGAGAATCCC AAGAAATTTC ATCTTCTTTG GCACTTTTCC 6720

AAAGAGCAAA ATCCAATGGG TCTTCTTTTA ATTGTTGTTC AACGCCAGTT CGTTGACTTG 6780

CCCCCACTTC TAATTCATCA ATCGATTGAT GGCTTAATTT ACCATAATTA GGGAATTTAC 6840

GGGTGCGATA ATACACATCA CCTGCGACCT CATAAGCAAA ACCTTTTTCA ATTAAAACTT 6900

CGATGAAAGC TAAGATATCA GGCATGTGGT CCATCACCCG AGGATGTAAC GTTGCTGGTT 6960

GGACATTCAA TGCTTGCGTG TCTTCTTCAA AGGCTTTGAT AAAACGCTCC GCCACTTCCG 7020

GCGCAGTAAT TTTCAATTCT TTAGCTGCCT TGATAATTTT ATCATCGACA TCTGTAAAGT 7080

TCGAGACATA ATTCACTTCA TACCCACGAT ATTCAAAATA ACGACGAATC GTATCAAAGG 7140

CGATCGCACT GCGCGCATTA CCGATATGGA TATAGTTATA CACGGTTGGT CCGCAGACAT 7200

ACATCCGAAC TTTACGCGCC TCAATTGGCG TAAATACTTC TTTTTCTCTG GTCAATGTAT 7260

TATAAATTTT AATCATGCCC TTTTTCCACC TTTCTTCCAT TTAGACGAAC GACTTTCGCA 7320

GGAACACCCA CAGCTGTGGC ATCTGCTGGG ATGTCTGAAA TAACAACAGC ACCCGCACCA 7380

ATTTTTGAAC GTTCACCAAT CACCACTGGC CCTAAAATTT GAGCTCGTGC CGAAATCATC 7440

GCTCCTTTTT TCACTGTTGG ATGGCGTTTA CCCGTATCAC GACCGGTCCC ACCTAACGTC 7500

ACTCCGTGAA ACAAAACAAC ATCTTCTTCT ATTTCCGCGG TTTCCCCAAT CACAATGCCC 7560

ATGCCGTGAT CGATAAAAAC TCCCGCTCCA ATTGTGGCTC CGGGATGAAT TTCTACGCCC 7620

GTTATGAAAC GCCAAAATTG TGCATGAATC TTCGCCAACA AAAACAACTG ATGCCTATAT 7680

AAGAAATGAG ACAAGCGATG CCAAAACAGG GCGTGCAAGC CAGGATACGT TAAAAGAACT 7740

TCTAGTGTGG TCCGCGCTGC AGGATCGTTT CTTTTAACAG CTGCGACGCT ACGTTTCAAC 7800

CAACTCATAT GGTGAACTCC TTTCTATTAT ACAAAAATAA AAAACGTCTT TTAAGCAACA 7860

AAAAAACGTT GCTTAAAAGA CnCCCaTGCG TGGTTCCACT TTTCTTCACA AGAACTAGCT 7920

CTTGCCTCGA TACAGTTAAC GCCTGCCGAC GTCTTCGGTT ACTTTTTTCA CCAAAGCCAC 7980

TCATAGATGC ATTTCGACAA AAGATGCTCG CTGTTCTCAC CACCCACAGC GTCTCTGAAA 8040

TCCATCTGAT TGTCTACTTC TTCTATTCTT TGTGTTTTTA TAACACTTTA TTTAAATGAT 8100

CTAATGCTTT TTCTTTGCCT AATAATTCGA TGGTTTCGCC TAATTCAGGG CCATGCATTT 8160

GTCCAGAAAC AGCAATTCGA ATTGGCATAA ACAAGTTTTT GCCTTTAACG CCGGTTTCTT 8220

TTTGGACGGC TTTAATCGCT GCTTTAATAC TTGGGACATC GAAAACGTCC AAGGCTTCTA 8280

ATTGTTCTTT AAACGCATGT AACACAGTTG GCACTGTTTC TCCTTGCAAG AATTCTTTCG 8340

CAGCATCATC TAAAACTGGA TGTTCATTAA AGAACAAGTT TGATAGTTCC ACAATTTCAG 8400

CTGCATAACT CATTTGTGGT TGATACAAGC TAACTAATTG TTTTAACCAT TCGATTTTTT 8460

CAGGAGATGG ATTTTCTTCA ACACGACCGT CAGCGACTAA ATAAGGAATA CACATATCTG 8520

TTAGTTCACT TAAGTCCATT TGTTTCATGT ATTGATTGTT AATCCATTCT AATTTTTTCG 8580

CATCAAAGGC CGCTGGTGAC TTGCTTAAGC GTTGTTCATC AAACAATTTA ATCAAGTCTT 8640

CTTGTGAGAA GATTTCTTCT TCGCCAACAG GAGACCATCC TAGTAAAGCA ATAAAGTTGA 8700

ACATGGCTTC TGGTAAGTAA CCTAATTCAC GGTATTGTTC AATAAATTGT AAAATTGTTT 8760

CATCACGTTT GCTCAATTTT TTGCCAGTTT CTGAATTAAT GATTAAGGTC ATATGGCCAA 8820

AGACAGGTGG TGTCCATTCA AACGCTTCGT AAATCATTAA TTGTTTTGGT GTGTTCGCAA 8880

TGTGATCATC GCCACGTAAA ACATGAGAAA TTTTCATTAA GTGATCATCT ACTGCAACGG 8940

CAAAGTTATA CGTTGGCATC CCATCGCGTT TTTGGATGAC AAAATCGCCA CCAATGTTAT 9000

CTGACTCAAA AGTAATTTCA CCTTTCACGA TGTCATCAAA TTTGTATTCT GTATTACGTG 9060

GCACACGGAA ACGAATCACT GGTTCTAGAC CAGCCGCTTC TTTGGCTGCT TGTTCAGAAG 9120

GTGTTAAATT CGCACAAGTG CCTGCGTAGT GAGGCATTTC CCCACGAGCT TGTTGTGCTT 9180

CACGTTCTGC TTCTAATTCT TCTGGCGTAC AATAACACTT ATACGCACGA TTGCTACTTA 9240

ATAATTGATC AATCAATGGT TGGTAAATTT CTTTACGTTC TGATTGACGA TATGGACCAT 9300

ATTCTCCCGG ATGAGCAGGT GATTCATCCC ATTCCATCCC TAACCATGCT AAGTTTTCCA 9360

ACTGACTTTT CTCGCCGTCT TCGATGTTCC GTTTTTGGTC GGTATCCTCG ATACGAATGA 9420

TAAATTCTCC GTCATTGTGG CGGGCAAATA AGTAGTTAAA TAATGCTGTT CTTGCATTCC 9480

CAATGTGCAA GTGTCCAGTT GGACTTGGTG CGTAACGCAC ACGAACTTTC GTCATGTTCA 9540

TTCCTCTTTC CATTTTTTAA AAACAGACGG TTTAACGTCG TTTTCAGTTA TTTTTCCAAT 9600

AAAGCAACGG CAATTGCGCC GATTCCTTCT TTTTTACCAA CAAAGCCCAT CGTTTCCATG 9660

GTCGTTGCCT TAATGTTCAC TTGTGTGACC GCTAAGTGGC AACTCGCGGC TAGATTCTTT 9720

TTCATCTCTG CCAAGTATGG GCTCATCTTT GGCTCTTCCG CTAAAATTGT ACAATCAATA 9780

TTGCCAATAG TAAAGCCCGA CCGCCCAACT TTTTCATTTA CTTTTTTTAA TAAGTTTACA 9840

GAATTCGCAT TTTTGAACTG CGGATCGGTT TCGGGAAACA GTTGCCCAAT ATCACCCAGA 9900

CCAGCGGCAC CAAGGATTGC ATCAATAATT GCATGTGTTA AAACATCGGC ATCTGAGTGA 9960

CCTAACAGAC CTTTTTCATA AGGAAGAGTA ACCCCACCAA TAATTAAAGG TCGTTGCGGA 10020

ACTAATTGAT GAACATCAAA ACCTTGACCG ATTCGAATCA TACTAAGTCC TCACTTTGCT 10080

TAATTTTTAT CGTTTTTAGG ATGGTTATCC TTGCGATCAT TAATCCCGCG ACCCGCGTGC 10140

GCCGGTTTGG CAAAAATCAT ACGCCCAGCT GCGGTTTGTA AAGCACTTGT GACGACCACT 10200

TCGATTTGCT CATTCATGTA ATGCTGGCCA TCTTCTACCA CAACCATTGT GCCATCATCT 10260

AAGTAAGCGA CCCCTTGTTG ACGCTCTGTT CCGGCTTTTA CAACCATCAC ATTCATCGTT 10320

TCACCTGGAA TCACCACTGG TTTCACAGCA TTAGCTAAAG CATTAATGTT TAAAACAGGC 10380

ACATTTTGAA ATTCAGAGAC TTTATTTAAG TTGTAGTCAT TGGTCACAAC CACGCCATCT 10440

AATAATTTTG CTAATTTGAT TAATTTACTG TCTACTTCAG AGATATCCTC GAAATCTCCG 10500

TCATACATCT CAACGGAGAT ACCTTCTTCT TTTTGTAAAG CATTTCAACG CTGGTTCCCT 10560

GCCCTTTAAA ACTAATAATT TnAACTGTAC CGCCCATGCC GACAACACGT CTCGGAATAT 10620

TAATTAAGCC ATAACT 10636

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6531 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

GAAnGGGTCT TTTTTCGAGT AGACGTTTGA TnTTAGCGTA TCCAAGAAAG GAAAACCCtA 60

ACATAAAaAT TTTTTATTTA CGAaGGAGAC CGATTTATAT GGAAAAGAAA GAATTTCACA 120

TTGTAGCAGA AACAGGAATT CACGCACGTC CAGCTACTTT ATTAGTACAA ACTGCAAGCA 180

AATTTAACTC AGATATTAAC TTAGAATACA AAGGTAAATC TGTTAACTTA AAATCAATCA 240

TGGGCGTTAT GTCTTTAGGC GTTGGTCAAG GTTCTGACGT AACAATCACT GTTGATGGTG 300

CTGACGAAGC TGAAGGAATG GCAGCAATCG TTGAAACATT ACAAAAAGAA GGATTGGCTG 360

AATAATGTCT GAGATGCTAA AAGGGATTGC CGCAAGTGAT GGCGTTGCTG TTGCTAAAGC 420

TTACCTGCTA GTTCAACCTG ACTTGTCTTT CAACAAAACT TCCGTAGAAG ATACTGATGC 480

AGAAGCAACT CGCTTAGACG ATGCTTTAGC AAAATCTACT GAAGAGTTGC AAGCAATTCG 540

CGACAAAGCA GCTCAAAGCC TTGGTGAAGC AGAAGCGCAA GTTTTCGATG CTCATTTAAT 600

GGTTTTATCT GACCCAGAAA TGGTTGGTCA AATCAAACAA AATATCCAAG ACAACAAAGT 660

CAATGCAGAA GCTGCCTTGA AAGAAGTTAC GGATATGTAT ATTGGTATGT TTGAAGCAAT 720

GGaCGACAAT GCTTACATGC aAGAGCGTGC aGCAGATATT CGTGACGTTG CCAAACGTAT 780

CTTAGCACAT TTATTAGGTG TGACTCTTCC AAATCCTTCA ATGATTAACG AAGAAGTAAT 840

CGTGGTTGCC CATGACTTAA CACCTAGTGA TACTGCACAA TTAGACCGTA CGTATGTTAA 900

AGCCTTTGTG ACAGACATCG GCGGACGTAC TTCTCACTCA GCCATCATGG CTCGTTCGTT 960

AGAAATCCCT GCAATCGTTG GTACAAAAGA AATTACTGAC AAAGTAAAAG CAGGCGATAT 1020

TTTAGCAGTG AACGGAATCA TTGGGGATGT TATTATTGAC CCAACAGATG CTGAAAAATC 1080

TGAATTTGAA GCAGAAGCAA AAGCTTATGC AGATCAAAAA GCAGAATGGG ATAAACTAAA 1140

AAATGCAGAA ACAGTGACAG CTGACGGCAA ACATGTTGAG TTAGCTGCAA ACATTGGTAC 1200

ACCAAAAGAT TTAGAAGGCG TACACAAAAA CGGCGGCGAA CTGTTGGTTT ATATCGTACA 1260

GAATTCTTAT ACATGGATTC ATCTGACTTT CCAACTGAAG AGGACCAATA CCAAGCGTAT 1320

AAAGCAGTTC TTGAAGGAAT GGAAGGAAAA CCAGTTGTGG TTCGTACAAT GGATATTGGT 1380

GGGGACAAAG AGCTTCCTTA CCTAACATTA CCACACGAAA TGAACCCATT CTTAGGTTAC 1440

CGTGCATTAC GTATTAGCTT ATCAGAACTA GGTGACGGCA TGTTCCGTAC ACAAATGCGT 1500

GCATTATTAC GTGCGTCTGT TCATGGTAAC TTACGTATCA TGTTCCCAAT GGTTGCTACT 1560

TTGAAAGAAT TTAGAGCAGC GAAAGCAATC TTTGAAGACG AAAAACAAAA ATTAGTCAAC 1620

GAAGGTGTTG AAGTTTCAAA CGATATCCAA GTAGGTATTA TGATTGAAAT TCCAGCAGCT 1680

GCCGTATTAG CTGATAAATT TGCCAAAGAA GTTGACTTCT TTAGTGTTGG AACTAACGAC 1740

TTAATCCAAT ACACAATGGC GGCAGACCGT ATGAACGAAC GCGTTTCTTA CTTATACCAA 1800

CCATACAACC CATCAATCTT ACGTTTAATT AAAAATGTAA TTGATGCAGC ACACGCTGAA 1860

GGTAAATGGG CTGGTATGTG TGGTGAAATG GCGGGCGATC AAACAGCTGT TCCATTACTA 1920

TTGGGTATGG GCTTAGATGA GTTCTCAATG AGTGCTACAT CTATCCTTAA AACACGTAGC 1980

TTGATGAAAC GTCTAGACAC AACTAAAATG GCTGAACTTG CTGACCGCGC ATTAAAAGAA 2040

TGCGACACGA TGGAAGAAGT CTTTGCATTA GTTGAAGAAT ATACAAAATA ATTAAGTGAT 2100

TATGAAAGCA CATCGACAAT TGTCGATGTG CTTTTTTATT GTTGCGATTC ACGTATTTTT 2160

TTCTTGAAAT AATAAATCGT GCTTCCCAAA ATGAATAGAA AAACCAGCGT GATGCTACTC 2220

ACCAGTTTCC AAGGGAAATT GGTGTGTTTT TCAGGTGGGA GGTTGACGGC GGTAGCATTG 2280

TATTTTTTAG CTTCTTTTTC TTGGATGACT AAGTGTTTTT CCCACGAATA GTCTTGATAG 2340

CCATCGTTTG CTTTGATTTT tACAGTATAA TTTCCAGGAA CAAAAGGTTG TTCCTTTAAG 2400

TCAATCCGAT AATTAAAATT TGTATTAGGA GCCATTGTTA ACGACTCTTG GTGATTTTCG 2460

TAGCGAGGTT TGGTTTCTTT TTCATAATAC AAAGAGGCAT CGACTTGTAA TTTTTTAATC 2520

ATAGCCGCTT GTTTGTTTTG GAGATTCATT AATACGCGGT TACGACCGTT AGCTTGATCG 2580

GTTTTTACTT CATTTAAAGA TAATTCTGGC TGAACAGGTT CATCTGTTTC ACTAAGTAAG 2640

ACACCGACTA CATAAGAAAA ACGATTATTG ATGGCCACAC CGTTTTCTGA GTGAGCAGGT 2700

TGTTCATCTG ACTTTTGTTC AAAGTAGAGT CCTCCTAAAA CGTACCCATT AAATGGTTTT 2760

TCCGGTAAAG TAAGCTGACA TTCAACAGTC TTGGTTGAGT GTTTAGGTAG GTGGATTTCT 2820

TTTGGCAATT GCGCAATCTC ATTAAAAGTA AAAGGTGCAG ACGGATCTTT TTTTGTTTCA 2880

GCATGAGAAT AATCTGCCAA GCCATTATCA TTAGTAATCG CTGCATTGGC TGAAGCTAGC 2940

ACGGTAATGT CGTTGGCTGT TTGATTCGTT AGTTCGATTT GTAGCTTTTG TGTTTGGTTT 3000

GGAGCGACAC GTAAATCAAA ATAACTTGTC TCTTTGGTTC GTTGATTGTC GGGTAAAATG 3060

GCTTTTACAG AAAAAGTCAT TTCGCTTGCT TCTACAGAAT AAGAGCCAAT AAAAAATAAA 3120

AAAATGATGA CGCTAATAGC ATGGAGCCAT TTAAAAAGCT GCTGATTCAT AGTGTGCACT 3180

CCTTTTTGAC AATTTTTTAT AACAAGGTAT TGCATGAAGT TGTTTCATGC AATACCTTGT 3240

TTGTTTTCTT TTAAAGTGGT GTATCATCCA AAATCCAAGT AAGCGTTGTT TTATATTGTT 3300

TAGCTGCAAC TTTTTTCGTT GTCGCAGGAA CAGTTAATTG AATGCCTTGA GCAGCGGTcg 3360

sTCCGCTACC AAAAGCTAAT GTCCAAGTGC CCATTCCTGT TGAAAGAGCG GCAGTCGCCA 3420

CAGGAGAAGT GGCGGCACCA GTTGGATCCA AGGTTACTGG GACCGTTAAT GGCGCCAAAG 3480

AAAGTAACTG TGTTGAATTG GGTGTTGCTG CTAAAAAGGT CAAGGATGCA TTATCCAAAA 3540

CAGCGGGTGT TGAATCACTT GTAGCAAATT GCGCACTCTG TTTAACTGAT AATTTCCAGC 3600

CAAGATTTAG ACCACGTTTG TCAGTTACTT GAACATAGTT TGGCACGCTA ATTAAATCGC 3660

CAGTACTATT TTGCACTTGA TCCAGTTGTG CCGAATAGAT CGCTGTTCCG GCTTGAATTT 3720

GTTTTGATCC AAAATGGATA TTTGAAACAT AGTCAATACT AAGGGGTCCG GCTGTACCAG 3780

GTTGATGAGG ATCAGCAGGA TTAGGTGTCA CAGGCTGAGA AGGGTTCGTC GGGTTGACAG 3840

GTGGCGTGAC AGTATTATCA AGAGCAAACG TAATATCAGC GTTGGATTGA ATAGAAGCGA 3900

CTTGAGCTGC TTCAACTTTT GAACTATTTA AAAGCATAAG AGGCGTAgcA ACGAACAAAA 3960

TGCTTGAAAT GCGCACGATC TTTTTCATTT TTTATTCCTT CTTTCTTGTT TATTTTTAAC 4020

TTTTAAAAAA GCTAAATCCT AATTAGAGGG GTGTATCAGA AAGAATCCAT TGTAATTCGG 4080

CAGTGTAAGA TTCTTTGACC GTTAAATTAA GGTCAGGAGC AACGCTTAAA AAAAGTCCAT 4140

```
TTTGTTGGTT CCAAGGAATT TCTACATTGT TATCGTCCGT TGCGGTGTGT TGATAAACTT     4200
CGGTTGCTAA AGAGGAGAGA GGAAAATCAC TGCCATTTTT TCGATAAATG AAAGCTTGCT     4260
GGATCGTTTT ATTTGTAGAA GTTGAATGCA TGGGCTGACT CTCTCTTACC AACAAATGCC     4320
ATGGTGTTTT TGCAACTCGA CCATCGGCAA TGAGGAAAGG AGCGCTGATG GTCGATGGTG     4380
AGTATATGGT TCGTTTGCTG GGGAAAGGTA AATTTTTAAA GGAAACTGTA GTTGGCATCT     4440
GCACAAAACG GAGGGTACCG TTTTTGATAG GTAAAGATTG TTCTTTAGTG ATTGGCGTAA     4500
TTGcGTTCCG CCACCATCAT TGCCAGTAAC CGTGCTAGAA AATAAAAGAA ATTGGTTTAT     4560
CCATTCTTCT GTAACCGTTG CTGTATATTG TAAGGTGAAT TGTGTATTGG CTTCCGTAAA     4620
ATTATCTGCT GGCAGCGAAT TATGATAACT CCAATACGGC AAACCAGTGG AACTATCAGT     4680
TAAAAGTTGA AGCCCAGCTA TTGGCGTACT TAAACCTGCT GCTGATTGTT TTGTTAATGT     4740
AGCAGCTGAA AGTTGGGACA GTCCTTTAGG CACCGTTGCT CGTAATTGCT GATTCATCCa     4800
CGTTGCTGGT TGATTGGTTA AACGTGTGGT TAACGTAAAA CGAATGATTT GACCACTTAC     4860
GACAGAGTTT GCAGCTTCCC AGAGCCCTTG TGCATTTTTC CGCTCAATGA TTTGCGTAAC     4920
CGTTAGTTGA GGCACAGGAT AAGCAACTGT TTGAGAAACA GTAGAACTGG TTAACAAAGG     4980
CAGGCCAGTG CGATTGCTGA GACCTTGGAC CCACAAGCTG ACCGTTTCTC CCCCTTGGAA     5040
AGTGGGTAAA TTAGCAAATT GAAAATGGCC ATTGGAATCT AAGGACGTTG TGGAAATAGA     5100
ACCACCATTT ATTTGTACTT TAACTTGATA GGTTCCGTTA GATACTTGAG AATCTTTACC     5160
TGAAAGTTGA GTTGAGCCGG CATTGGCGAC TACGTTTGTG ACTGGTTGGA TTGACCAATT     5220
GACACTTTGG TCAGCTGAGG GGCTGCTTTG AATATAGTCG GTTGACGTTT TTGAACGAGC     5280
AAAGACAATT GTAGAGAGCG TGTCGCCATA TTCTAATGGT GAAGCTAACG ATGTCGAGAA     5340
ATTTCCGGCT GCATCGACAG AAATCCCTTT CACCACAATG GGGGTAGTGG TGCCATCATG     5400
GCGGTTCAAG GTTAATTGAA CAGAGTAGTT ATAGTTTGGG TCTTGTACTT GTTGAATGGT     5460
CCCTGTTACA GTAGTGTCCA GATGATTTGG TTGAGTTAGT TTGGCTGGTG TGACTTGAGG     5520
AACGGGATAA TTACCTAAAC TAAGTTGAGC TTTGCTAATT ATGCCTGTTT TTGTATTATT     5580
CCCAATGATA ATATTAATTG CGCCTAATCC aGTGCCGACA TCTGTAAAGG TTCCTCGAAT     5640
CGGATAGTCT TTGCCGTTAA GAGCAGGAGG GATTGAAATT CCTCGTTTAT AAAGTGCGCC     5700
CACATCAAAC GAAATGGCGG CAGACCATTT GGTTCCGGGA AGTAACACCA AGGTGCTACT     5760
TGGAAAAGTC AGATAAACAG AATGATAAGT GCTACTATAA GAGAGcTGAC GCCGTTAGTT     5820
GAACTGTGAA TATCTTGAGT GGCTGTTCCT GGATAACTGA CAGTCCCTGT TAAAAGAGTC     5880
AGAAAATCTT GTTGCTTTAG AGTGTCGTTA TCTAACTGGC TTGAAATTTC TGTAGGAAGT     5940
TGAATAGCAA TAATCGGTCG ATCATTAATT GCTAAGCCGA TAAGTGTCAC ACTGTATTCA     6000
TACGAAAGTG TGATAATATC TTGGTTATAC GTTCCCGTGC TTAAGGTACT TTTTAAGGTG     6060
CTATTACCAA CAGAAACACT AGCTGGCGTA GGTGCCATAG GTTCATCGTT TTCAGAGGAA     6120
```

GCATGGGTTT GCTCGGTTTC TTGTTCCACG ATATCGGTTG CTTGAGTTGT TTGTTGTTCT 6180

GTTGAAACTG AAGTAAATGT TTCACTAAAA GTAGAGGCGT GAGTTGTAGG GTTCAGTGTT 6240

TCAGAATGAG CTATCAAAGG TTGAGTCAGC GATAAAAAAA GCCATGATAT GATGCAAGAA 6300

AAGATTCCCC ATACGAACTT ATTCAAAACC ATCCTCTCCT TTTTTGaTAT TATATGtATA 6360

ATGACGAAAA TATATAATGT GTAAATATTA AGTACATACA TTCTATAATG GGCCATnCAT 6420

AAAGTAAAGC GATTACATCC GATGGTGnAC AGGATATTTT GTTCACAATC TTTTAGTGTT 6480

TTAAAGAGGG AAATTTCTCG AAATTCTATG CGTTTTnAAT CCGGGAAAGA T 6531

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6399 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

TTTTCGGGAA AGAAACACCT ACCATTTTTG ATGTTTTGGG CAGGGAATAG GAGGTAATAC 60

ATGGAATGGA GTTTTCCTTG GCTATCAATA GATGGCGACA GAATGTACGA TGATAGCGAC 120

TTCTCGAAAT TTTTTGAGGG TCTTTTTTCT TATGGCGTTT CTCTAACAAC AGCTAACGCA 180

TTAAAGGTTA CTGCCAGTCC TAATGGCGGG ATGAAAGTAC AAGTAGATTC AGGCTATGCA 240

TTTGTCGGTA AAGTCTTCTT GAATAGTACT ACAAAAGCTT TAAGTATTGA TGTAGCAAGC 300

AGTATGCAAG ACAGGACGGA TAGTATTGTG GTTCGCGTGG ATAAAAGTGT GCGAGATGTA 360

TTTTTAGCTG TTAAAAAGAA CGATACGACA GTAACCCGTA CATCGGACGT TTACGAGCTA 420

CAGTTAGCGA CAATAAGAGT GCCCCGAAAT GTTTCTAGTA TTACAGGCGA TTTAATCACA 480

GATAAACGGG CAGATACGAA AGTATGCGGG TATTCTTCGC CGTTTCAAAA GGTTAACGTT 540

TCAGGGTTAG AAGAGCAATA CGGTACAATA CTAAAAAAAA TTGCGGAAGC CAACAAGACG 600

AGTTATGAAA AAATCCTAAA TGATTTTAAA AACTATGTTG CAAAAGCGCA AACCGATATG 660

GATTCTAATA TTGAAGAAAT TATCCAATCA GGTAATGGAA AGGTAAGTGC TTTTGATGTT 720

TTAATTCACG AATGGTTTGC AGCTTTAAAA AATGAATTAG ATACAAATCA AGCATCAAAT 780

TTACAGAATC AAATCAATGA AATGAAAGCT ACTGAAGAGT TACCAGCTAT AGAACATAAT 840

TTACTCGGTT ATCCTAATGT ACAAGTTTTG TATTGGGAAT ACGGTATTGG CCTATCAGGA 900

TTAGCTAATG AGCCAACAGG TCTAGGCGGT AGCAATGTGA AAAAGATTCC TCACAGTGTA 960

GAATATCTTG ATTTATTCAG TTTCAAAGTT AAAGTGCCAA TGAACTTTAA ATTGGTAAAT 1020

CCAACAGTAA CAAAAATAGA TAGTCGAACT ATTCGCTTTA TTGAAGCATT TAAAGTTATA 1080

GAAATTAAAT TTTAGGAGGA AAAGAATGTA TACATTTAAA AAAGGTGATG CAGACTACCA 1140

AGTCATGCTA AATGAAAATT TTAAAGAGAT AACTGATTCT TTCGAAGATG GTTTATTTGT 1200

CAGAAAAAAT TCTAAAATAT TAGATTTAAA CGATGCTATA TTGCCAGGCA TTTATTCAAT 1260

CCCAGCTACA GGAGTTGATA ATAAACCTTT ACCTAATTCT GGAAGTTTAT TCGTTTGTAA 1320

AGACCCAGGA GGGGTTAGAC AATTATTTCA GACAGAAAGA ACAATATTTA TTCGACAATT 1380

AGGGGGAGTA CCTTCATCTT GGaCGGATTG GaAAAAAGTA GCTTTTGAAA AAGAACAACC 1440

TTTTGAAGCT TGGTATTCAC CAGGAACTAA CCATGCTGGA TTCAAAAATA AGGCAAGATA 1500

TAATTTAGGG CCAGAATTTA GCAACATAGG CCAGCGACTT GGGTTGCCTA TGAAAAGTGA 1560

ACCGTTGGAG TGGAATAGTG GGCGATGGCA AGCAAAAGTT CTTAGAGACT GCAAGCTAAA 1620

TATAAGCGGA ACTGTAAAAT ACCATGTTGG TAGTTCGAGA GGTGTTCTCT ATGCTTATAC 1680

TCATATAGAC AAAGGCCTTG ATGAAGGCGT AGGTGACTTA GGTATTGGAT CAGCAGTCGG 1740

AGCTGTTGGC GGCTTGGATT ATAAAAATGT CGCAGCTTTT GATTTAAACG TTACACTAAA 1800

AAAAGGTGAG TATCTTGCGT TTCGCTTAGA ATTAGCAGCA GATAAGCAAC TTGATTATAC 1860

TCAATTATCT TCTATGCATA TCACAGAATT AGTATAGGAA CTGAATTTTA AAATAAAAAC 1920

CGTTTAGCAA AAAAGCTAAG CGGTTTTTAT TTATAGAGAG AAGGAACTTT ATGTGAAGGA 1980

TGAACTGATT CAAGATGTTG TAGAGCGTTT AGTGCGTATT GAGACAAAGC TCGATAACTA 2040

CGAAGTGTTA AGAGAAAAGA CAGAAGATGC TAAGGAAAAG GCTGATTATG CTTTCTCTAT 2100

TGCTAAAAAT AATGAAGAAG ATATCAAAGA AATAAAGGAA AATCAGAAAT GGTCTTGGCG 2160

GACAATTGCG GGCATAGGGG TGTCAGTAGT TGTCTATTTA TTTACAAAAT ATTTAGGAGG 2220

TGTATAAGAA TGATTCTACC AGATAAGTAC TACAAAATTA TCAAATGGGG CGTGCTAACA 2280

GTGCTACCTG CAAGTTCTGT TTTGGTTGCC ACACTAGGTA AAGCTTATGG ATGGCAGCAA 2340

ACAGATATGA CTGTTTTAAC TATTAATGCC ATAGCAACTT TTTTAGGAGT GGTAACAGGA 2400

GTATCAGCAT ATAATTTAAA AGACAAGGAG AAATAAAAAT GAAAAAGAAA ATTTTAGCAG 2460

GAGCGCTTGT CGCTCTGTTT TTTATGCCTA CAGCTATGTT TGCCGCAAAA GGAGACCAAG 2520

GTGTGGATTG GGCGATTTAT CAAGGTGAAC AAGGTCGCTT TGGCTATGCA CATGATAAAT 2580

TCGCTATTGC CCAGATTGGA GGCTACAATG CTAGCGGTAT TTATGAACAA TACACATATA 2640

AAACGCAAGT GGCAAGTGCT ATTGCCCAAG GTAAACGTGC GCATACCTAT ATTTGGTATG 2700

ACACTTGGGG AAACATGGAC ATTGCGAAAA CAACAATGGA TTACTTTTTG CCACGTATTC 2760

AAACGCCTAA AAATTCCATC GTTGCATTAG ATTTTGAACA TGGAGCGTTG GCTAGTGTTC 2820

CAGATGGATA TGGAGGATAT GTAAGTTCAG ATGCCGAAAA AGCAGCAAAT ACAGAGACAA 2880

TTTTGTACGG TATGCGCAGA ATCAAACAGG CTGGCTATAC TCCAATGTAT TACAGCTATA 2940

AGCCATTTAC ACTAAATCAT GTAAACTATC AACAAATCAT CAAAGAGTTT CCTAACTCTT 3000

TATGGATTGC TGCGTATCCT ATCGATGGTG TGTCACCATA TCCATTGTAT GCTTATTTCC 3060

CAAGCATGGA TGGTATTGGT ATTTGGCAAT TCACATCCGC TTATATTGCA GGTGGTTTAG 3120

ATGGTAACGT AGATTTAACA GGAATTACGG ATAGTGGTTA TACAGATACC AATAAACCAG 3180

AAACGGATAC GCCAGCAACA GATGCAGGCG AAGAAATTGA AAAAATACCT AATTCTGATG 3240

TTAAAGTTGG CGATACCGTC AAAGTGAAAT TTAATGTAGA TGCTTGGGCA ACTGGGGAAG 3300

CTATTCCGCA ATGGGTAAAA GGAAACAGCT ACAAAGTGCA AGAAGTAACT GGAAGCAGAG 3360

TATTGCTTGA AGGTATCTTG TCATGGATTA GCAAAGGTGA TATTGAATTA TTGCCAGACG 3420

CAACAGTCGT CCCTGATAAG CAACCAGAAG CGACTCATGT GGTACAATAC GGAGAAACAT 3480

TATCAAGTAT TGCTTATCAA TATGGAACAG ACTATCAAAC GTTGGCGGCA TTAAATGGAT 3540

TGGCTAATCC AAATCTTATT TATCCTGGTC AAGTTTTGAA AGTCAATGGA TCGGCAACAA 3600

GTAATGTCTA CACGGTTAAA TACGGCGATA ATTTATCTAG TATTGCAGCA AAACTTGGCA 3660

CTACTTATCA AGCTTTAGCT GCATTAAACG GATTAGCAAA TCCTAACTTG ATTTATCCAG 3720

GTCAAACATT GAATTATTAA TAGCTTTAAT ATAAAATAAG GATACACTTA TTTAAATTTC 3780

TCTCGAGTCG CCGTCCCCAA GGCGGCTCTT TTCAGGACCA TTAGCTCAGT TGGTTAGAGC 3840

AAACGGCTCA TAACCGTTCG GTCACAGGTT CGAGCCCTGT ATGGTCCATA GCTAAAAAGG 3900

GCACGAAAAG GGCAAAAGTA GTAATACTTT GTCAGATTAT GTTGTTTTTT ATAAATTTTA 3960

TTAGTATTAT ATAGCATTAT AAATGCCTGT TTTACGGGCT TTTACAATTT TTGTGAAAAT 4020

TGGACCAAAT CCTTGGGATC CAATGGGCGC ATTAGCATAA AACAGAAAAA GAAGGCCCCC 4080

GCCGTCAGGG CGGGGGCCTT CTTTTTACGT TGGGCGGCGA ACTTCACGGA TTTCGCCCAA 4140

TGTTGTATAA TCATTGCCTG CTAAACGAGC AATCGGTTGG AGTTTTTCAG GTAAAATGTA 4200

CTGATGTTCT TCATCGAAAA CGGCTGAATC AAAATAGAAA TCGGTAATTT CTACAATGAA 4260

TAAATCCGTA ATAATTTGAC CTTCATGGTT AGCAATAGGT ACATATTGAT GTAAGCGAGC 4320

TTCCATCCGA ATCGGCGCGG CTTTGATTGC TGGAACAGCA ACAGTCTGGC TGGCAATGGT 4380

TTCAATACCA AATGTGTCAA TTTCACTGAT CTCTGCTGCT AAAGACGCCG AAGTCTGATT 4440

CATTTGAGTA AGAACGTCTT CGTTAACGAA ATGGATGACA AGTTCTTTAG TAGCCAAGAT 4500

GTTTCGTGCC GTATCTTTAG GCTCTTCGTC TGGTCGTAAA ATGGACAACG TTGCTAAAGG 4560

AATTTCAGCT GCCGCCGCAT TAAAGAAACT AAAGGGAGCA GCGTTGACGA CACCTGTAGC 4620

TGTATCTTGT GTTGTAACCC AAGCGATTGG CCGAGGAATA ATACTTCCTG ATAAAAATTT 4680

ATACGCTGTT TTTGCAGATA ACTCGTTACT ATTGTAATGA ATCATAGGAG CCTCCTAGAA 4740

AAAATACTAA TTTCTTTGAA CATTCGCATC TGAAGTATCG AAAGGACGCA CAAATTCTTC 4800

AATCCCTGCC CGTTTTGGTT CTAAAAATGG CGGTAATGAT AATTTTTCGC CAGCGTGTTC 4860

ATAGGTTTCA TCCTCTAAAA ATCCAGGACC ATCGGTTGCC AGTTCAAATA AAACATGTGT 4920

TGCTGCCAAG AAaTATTCAG ATTTAAAGTA AAAaCGTtCC ACAAAGCCAG AATTAGGAAA 4980

TTCCAAACGA TTAATTTTTT CAATCCAGAA ACGTAGTGCT TCCTCATCGG CAACGCGCAA 5040

GGCTAAGTGA TGGACATTGC CGTAGCCTTC TTGAGCCGCT GGCAAATCTT CCCGATGCTC 5100

TACGATGATG CTGGCACCAT TCCCACCTTC GCCAACTTCA AATAAATGAA AACTACCTTC 5160

AGCATCAATC AATTTGAACC CTAAAACTTC AGTTAAAACT AGTTGCAGAT GTTCGTAATT 5220

TTCGACAGTT ACAAAGGCTG GTCCTAAGCC GATTAAAGCA TGTTCACTAG GGACGTTAGA 5280

TTTTTTCCAA GGCGTGCCGG CTGCAACACA TTGGTTCTTT TCATCAGAAA TAAGTTGGTA 5340

ACGTTGATTG TCAAAGTCTT CAAATTCTAA ATACTTTTTA CCAAAACGTT CTTGAATTTC 5400

TCCATGATCA ATTGCATATT CATTAAAGCG TGATACCCAG TAATCCAAAG TAGCATCATC 5460

TTTTACACGA AAGGAAGTAC GAGAAATTGT ATTGGTTCCT TTGGTTCCtT TrGGTAAGCC 5520

TGGGAAGTCA AAGAAAGTCA TATCTGTTCC GGGAGAGCCG GCATCaTCAG CAAAAAwTAA 5580

aTGATAkGTT TCAATATCGT CTTGGTTGAC AGTCTTCTTA ATTAAACGTA GACCTAAAAT 5640

ATCGGTGAAA AAGTGATAAA TTTTTTCGGC ACTTGAAGTC ATGGCAGTGA CGTGGTGTAA 5700

ACCTAAAATT TGTGTATTCA TAATAAAAAT CCTCCTAAAA GAAAAATAAT TTTTTGATTA 5760

ATTAACTTAC TAAAAGTAAG TTTAAATGAT TAGTCCTAAA AATGCAAAAG AAATGTCTCG 5820

AATTCGAACT ATTATTTTTT AGTTGATTTT TTAAAAAAAT AAGAATATAA TAGCCTGAGA 5880

TATTTTAATA CATAAACTTT CGTAAACATC GCCGGATGCT TGCGACTAAA GTGTTAATGA 5940

ACAGTTCTTA GGCCATAGCT GAACTGTTTG TTGGCACTTT TTTTGTTTGG AAGGGAAGGG 6000

AAACAAGATG AGTAAAGCTT GGTTAAAAGT TATTTTTGGC GCTTTTTGTG AAGTGATTTG 6060

GGTCATTGGC ATGAAGCATA GCACAACATG GTGGGAGATT TTAGGGACAG TAATCGCAAT 6120

CTTTATCAGT TTCTATGCGC TGATTAAAGC AGGAGAAGAA TTGCCAGTCG GTACAGCCTA 6180

TGCTGTATTT GTTGGCTTAG GTAGTGCTGG AACTATTGTG ACAGATCATT TTTTATTCCA 6240

CACACCATTG GGCATAGGAA AAGTTCTTTT TCTGTTGCTG TTATTAATAG GTaTAATTGG 6300

CTTAAAAATG GTAACTGGrA ATAAAGAGGA GGAAACAAGA TGAGTTGGTT ATACTTAGTT 6360

ATTGCAGGCT GTTTTGAAAT TTTTGGCGTA GGTTCAATT 6399

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3956 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

AACAGGAGGA AAACTTGATG GGCATTCCAA TCACCATTGT TTCTGGCTTT TTAGGCGCAG 60

GAAAAACAAC CTTAATTAAC CATGCGCTTG CGCAATCACC GTTTCCTAAA GAAGAAATTA 120

TTATTATTGA AAATGAATTT GGTCAAACAG GCGTTGACCA CGCATTACTT TTACAAACCA 180

CGGAACGAAT TATTCAATTA AATAACGGTT GCATGTGCTG CAGCTTACGA GGAGATCTTC 240

TTGCCTCTTT ATCAGCAATC TTAGAAGTGT ATCAAGAAGA ACAACAACCC ATTGCCCAAG 300

TCATTTTAGA AACAACGGGC ATTGCTGATC CGCAACCAAT CATTCAAACA ATTTTAACGA 360

CCCCGCACAT CAAAAATCAT TTCTATATAG ACAGTCTCTT AACGGTCGTT GATGGGCATC 420

ATTGGCAACA ACAGCTTCAA GAAGCCGAAG CCATCAAACA ACTGGCCTTA GCAGATCGCC 480

TCTTTTTTTC TGTCAAGGAA CCAGCAGACA GCACTCAGCT ACCAACGTTC CAAGAAACCC 540

TTCGTACCAT TAATCCTTTT GCTGATATTT TATCCTTTCA AGCCAACGAG CCTCTTTTGG 600

CTAGTGACTT TTTTCAATTA AATAAGTTTA CGGCAACACT TACAGAACAC GAAATGACTC 660

ATTCCCCGCA TGAGCACGCC CACGAGCATC ATCACCATAC CTTTCACTCC CTCACGCTAA 720

CGGCGTCCTC AGCCATCAAC GAACCTCTTT TCACTCGTTG GCTTGATTGG TTAATGTATA 780

CGCATCAGGA AAAGTTATAC CGTTTCAAAG GGATTCTCGC CTTACAGGAG CATGACTTGG 840

CCATTGCCAT GCAAGGCGTC AACCAGCAAG TGGCCTTTCA AATGACGAAT CAACCCCCTC 900

AAGAAACGAC AATTGTTCTG ATTGGCAAAG AATTAGAGAC AAAAAAAATT CAAGAAACCT 960

TTCAAACCCT CAACAAAATA GCTACTCCGT AAATTAATTT TTTACGAAAT GGAGGTCCTA 1020

TATATAATGG AAAAAACTGA CTTATCTAGC GCTTATCGTC GACTGAAAAG TCCCAACATC 1080

AAAACACGAA AACGCGCACT TAAAATCATT AAAGAACATA AGCGCAATAA ACAAAAGAAA 1140

ATTGCGTAAT TCACTAAAGG AGCACACCTA TGAAAAAATT TACTCTTCCC CTGTTAGCCG 1200

CCTTATCGCT AATCCTTTTC GGCGCTTGCG GCAAAACAAA CACCTCTGAT AAAACCGCTG 1260

ACGGTAAAGA AAAACTATCC ATTGTCACGA CTTTTTATCC TATGTACGAT TTCACTAAAA 1320

ATATTGTAGG CGATGAAGGA GACGTCAAAT TGTTAATCCC TGCTGGTTCT GAACCACACG 1380

ATTATGAACC ATCCGCCAAA GATATGGCTA CCATCCATGA TGCGGATGTT TTCGTTTACC 1440

ACAATGAAAA TATGGAATCT TGGGTACCAA AAGCTGCTAA AGGTTGGAAA AAAGGAGCCC 1500

CGAACGTCAT TAAAGGTACC GAAAACATGG TCTTACTTCC CGGCAGTGAC GAAGACGGAC 1560

ACGACCATGA CCACGAACAT GGCGAAGAAG GCCACCACCA TGAATTAGAC CCGCATACTT 1620

GGGTTTCGCC TCATCGTGCC ATCCAAGAAG TCACAAACAT CAAAGAACAA TTAGTCAAAC 1680

TTTACCCTAA AAAAGCCAAA ACATTTGAAA CAAACGCAGA AAAATACTTA ACAAAATTAA 1740

CAGCCTTAGA CAAAGAGTTC CAAACAGCTT TGAAAGACGC TAAGCAAAAA AGTTTTGTTA 1800

CCCAACATGC TGCATTTGGT TATCTTGCCT TAGATTACGG CTTAAAACAA GTGCCAATAG 1860

CTGGTTTAAC ACCTGAACAA GAGCCAACCG CAGGcGCTTG GCAGAGTTGA AAAAATATGT 1920

CACAGACAAC CAAATTCGCT ATATTTATTT TGAAAAAAAT GCCAACGATA AAATTGCTAA 1980

AACGTTAGCT GACGAAGCGA ATGTTCAATT GGAAGTCCTA AACCCGCTAG AAAGTTTGAC 2040

ACAAAAACAA ATGGACAATG GCGAAGATTA TCTTTCTGTA ATGAAAGAAA ACTTAACTGC 2100

TTTGAAAAAA ACAACAGATA CAGCCGGGAA AGAGGTTCAG CCAGAAACCT CTGAAAAAAC 2160

AGAAAAAACC GTGGCTAACG GATATTTCAA AGACAGTGAG GTGGCTGAGA GAACACTGAC 2220

AGATTACGCT GGAAATTGGC AATCCGTCTA TCCTTTATTA AAAGATGGCA CATTAGACCA 2280

AGTCTTCGAT TACAAAGCGA AACTGAAAAA AGATAAAACA CCAGCCGAAT ACAAAACCTA 2340

CTATGATGCC GGCTATCAAA CCGATGTCGA CCACATCAAC ATCACTGATT CCACCATTGA 2400

ATTTCTGGTC AATGGCAAAC CACAAAAATT CACCTATAAA GCAGCCGGTT ATAAAATTTT 2460

AAACTATGCA AAAGGCAACC GTGGCGTCCG TTTCCTTTTT GAAACAGACG ATGCCAATGC 2520

TGGGCGGTTT AAATACGTCC AATTTAGCGA CCACAACATC GCACCAACGA AAGCCGCTCA 2580

TTTCCACATC TTCTTCGGCG GCGATAGCCA AGAAAGTCTG TTCAATGAAA TGGACAACTG 2640

GCCAACGTAT TATCCAAGCG ACTTAAGCAA ACAAGAAATT GCCCAAGAAA TGATTGCGCA 2700

TTAAGCATTC AAAAGAGAGG AGGTCGGGAC AGAAGTGTTT AACTCCGAGA AATAAGAAGA 2760

AATTTCCGAA AATTGTTCTT TAATTTTTGG AGAATTTCGG CTTATTTCCG AAGGAGTTGC 2820

TTCTGTTCCC GCCGTTTATC AGTTTTTGAG CGTGGAGCAA AAATCCAAAG TGATTTTTGT 2880

CCCACGCTCC TCTCTTTTTT TATTTAGTCA AGCACACCTG kTCTGTTtCA AACGTCGCTA 2940

ACamTGGCGC AcTTCATCCG TTGATTTkGT ATTCATTAAT TGTGCTCTTA AATCACTGGC 3000

ACCTGGAAAG CCTTTGACAT AAATTTTAAA GAAGCGATGC AGCCCAACGA TGGAGCGAGG 3060

CACCAATTCC GCATATTTGT CTTGCAAATC CAATTGTAAA CGTAGCAAGC CCAGCAATTC 3120

TTGTGGCGTA TGTGTTTTGG GTTCTTTTTC AAAGGCATAA GGATTTTTAA AAATCCCCCG 3180

ACCAATCATG ATGCCATCTA CACCATATTG TTCTGCTAAT TCTAGGCCCT TTTGACGATC 3240

GGGAATATCC CCATTAATCG TAATTGTCGT TTGCGGTGCG ACACGGTCCC GAATAGCCAT 3300

AATTTGCGGA ATGACCTCCC AATGGGCATC CACTTTGCTC ATCTCTTTTC GTGTTCGCAA 3360

ATGAATGGAT AGATTCGCAA TGTCTTGCTC TAATAAATGC GTGATCCACG CCTCCATTTC 3420

CGCCATTTCA GTAAACCCAA TGCGTGTTTT GACACTAACA GGTAAGCCAC CTGCTTTGGC 3480

AGCGTCAATC AATTCAGCCG CGACTTCTGG GCGCAAAATT AGCCCGCTGC CTTTTCCACG 3540

CTCAGCCACG TTAGGCACAG GACAGCCCAT ATTTATGTCT AACCCTTGAA AGCCCATCTC 3600

CGcTACGCCG ATACTCATTT CACGGAAAAA TTCCGGTTTA TCCCCCCAAA TATGTGCCAC 3660

CATCGGCTGT TCATCTTCTG TAAAAACGAG GCGCCCGCGT ACACTATCTT TTCCTTCAGG 3720

ATGACAATAA CTATCCGAAT TAGTAAATTC TGTAAAAAAC ACATCTGGTG CCCCAGCTTC 3780

TTTCACGACA TGACGAAAGA CCACATCAGT GACATCTTCC ATCGGTGCTA AAACAAAAAA 3840

GGGCTTTGGC AATGTTGCCC AAAAATTCTT GCTCATCTTC ATATACTCCG TTTCTGATTT 3900

TTTCyTTAwT TTCAACGACA AwTGTCGACA TAATCTGGGG CGAAACAATG CTAATT 3956

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7053 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

AAnCAATTTT TGgATTAgCy GCTGGGTTTT CCAAnGnGTT TTTAgCkGCT AAAAGCGCTT 60

TTgGTAAcTC TGTAACGAGG AGGCAGTAAA nTGGTGCCCt TCACTTGGAG TGAACTTnyG 120

CGCTCGTtTG ACTACTTCTT GTAATTGGTC TTTATTTACA AGGGGTACTT CTTTTAATTG 180

AGCCATTGCA GTTTCCAATG CGTTTAGAGC CTCTGAGACT TCTTCTTTTG TCGCCGTTTC 240

TTTCGCTACG ATTGCTTCAG CCGCTTGAAT TGCTTCTTGC AATTGTTTTT CTGTTTCTTT 300

TGTAAATTGG TATCCTTCTT CCGGCTTTCT TCCTCGGGCT TGAGCGATTT GCTCTTTTAA 360

GGTTTCCTTA TCAACGGCGA CGGGTTGTTC TTTTAATTGT GCAAtTGcTT CACGCAATGT 420

TTTAGTCGCT TGGTCTATTT CTTCTTGAGT AACTTTTTCT TGAGCAAATA CTTTTTGAGC 480

TGCTGCTAAT GACTCTTCGA GAACTTTTTT AGTTTCTTCG GTGAATTCTT GCCCATCACT 540

TGGTTTCTTC GCTAGCCCCT CTTGAATTGC TTTTTCTAAA TCAGCTTTAT TCGCTACATC 600

AGGCATTTTT TCCAACTGAC TTGCAGCCTT TTCTAGGTTT TCCGCGGCAG CAGAAACGTC 660

TGTCTGTGTG GCTTGTTCAT TGGCTAGAAC AGCTACTGCC TCATCTAAGA TTGGTTTAAA 720

GATTGCCCAA GTTTCAGGCG TATATTCTTT TTCGACACGT GCTTGATTTA GCGTAATTAG 780

CGCTTGTAAT AACTCTTTAT CTACTAGCTC TGTGGGCTTC ACTGTTACAG AAAAATCTGA 840

AAGTGTGGCT TGTTGGAATC CTTGGCCTAC TAATAAGCCT TGACTCATAT CAAATTGTAA 900

TGGCTGCTCA ACAGTTGTCG CACGTTTTAC GCGAGTTGTT TCTTTTACTT TAACTTTAAA 960

GGTAACTAAG TCCATTGAAC CATTTAACAA TTGTTTATTT CCTTGATTAC TAAGAACTAA 1020

ATATAGATTC TCCACATCAT TTGAATGTTT CCGATATTTA GAATAATTTT CCATTTGTTG 1080

TGTATTATTC GTGGTTGCAG GCTGACCGAC AAGTTCAAAT TTGCTAGAAT CAAAGGACAT 1140

TTTTGTGCTT AACGCATTAA TGTTTTTTAA GTCTTGACCT TTAAGTATAA AGGTTAATGT 1200

ATCGCCAGGA AGATAAGTAT CCTTGCCATT TTCATTAACC ACAGCAAGGG ATAAACCGCC 1260

CGCAATTTCT TCGACATCAG GTATCTCAAT CCCACCATCT AATTGTCGTA AAACATAACT 1320

GATATCATAA GCATCAATTA CCCCATTTTT ATTTAAATCA CCTTTTTCAA CATAGCCATT 1380

AAAGTCACTG TCCACAGATT CTAAGCCTGT ATAGTTTCGA TAACTCATTG CATCATTTTC 1440

ATCAATCGTT CCATCGTTAG TAATGTCTCC ATGTAAGATT CCTTCTGTTC CTGGTTGTTT 1500

GTAAAAGAGC ATCTCACGAC CAGAGCCAAA ATTACCCACC GaTTTTAATA CCTTCATTTc 1560

AACAAAACGa TACGCTTGGt CCTTCGTTTc AAkGGTTTTT GtCAAAGCAT CTTGcTTCCA 1620

GTTAATAGGT TCTGAAAACT CTGTCCAATT AGCTCCATCT TTCGACGTCC GATACTGGAG 1680

CTGCAAAATA TTTCCGTTGC CGGcATTGTC TCTTGGCAAG TATTCAATTT TATCCATTTG 1740

ATACTCCGCA CCAAGATCAA ATTTTAAAGA CAAGAAATTT CCATCACTTG GATTAGCGAT 1800

GCCTGTTGAC CAGTTCGTAT GCCAACCTGT ACTCAAATCT TTATCTGTTA ATTTCTTCAA 1860

TTCTGCGCCT GGTTGTTCTG GTAAATTACT TGTTGCTTTG ACTTGATTAA TTGTTTCTTT 1920

ATAGGGATCA TCTTGTGTTT TTCCTTTAAT CGGTTCACTC CATTCAGAGA CCCCGTTTTT 1980

TCCAACAGCT CGAACTCGAA AAGTATGTTC GCTTAGGAAG GAAAATCCAT CAAATGTTGC 2040

TGTATTCGTT TGAATATTTC CGrAAACAGT CCCGTCACGC TCTACTTCAT AGCTTGTTGC 2100

CTCAGTTACT TGATCCCATT GAAGAGTCAA AGAAGATGGC GTTGTTTTTT CTTCATTGAT 2160

TGCGACATTT GCCGGAATAG TTAACTGGTC ATCGACCGTT GTATTTCCTC CGTCAACCGT 2220

TCCTTTATTA ATGAAACCTT TGACAGTTAT TTGAACATCT TTGGCGGTCA CAGATTGTTT 2280

CGCTAGTTTT ACTGATAACG CTGATTGGTT CAGTTTTTCT CCTGAAGCTT CAGAAAGATA 2340

CTGATTCACT TGGAATTCTT TATCAAAATA ATATAGATTT GTGCCAGCTA AAAATTCTTC 2400

TTTATTAGCT GCTtGTTTCA AGGTTACTTC TGTCCCCCCC ACCATTGCTG TTACGCTTTC 2460

GGsCGCTtCT GAAGCTAGCA AwTCAAGCGt GGTGCTtCGT tCATCAACAA AGnCTTtATA 2520

ACTTCCTTTt GTCGGTtCGA tAGTaACGGT CAGATCCCtT TTcGTTTGAT TTAGGGCCTT 2580

GACTATTTAT TTTAGTCGTA GCTGATTGCC CAGCTTCATA ACTAGTTGAA ATTCCGTCAT 2640

CTTCATACAT ATTAAAAGAA- GTGGTGCCAT TTGGATAAAT CAAGAAAGAT CGTTGATCTC 2700

GTTGGATTTC CTTTGGATTG TTGTTTGGAT TCGTCATTGG GATAATACTG CCATCTTTGA 2760

CGAAAAACGGG CACCTTCCAC AGCGGTGTTT TTACTCCATT TAAGACACGG CCACCTTGAT 2820

ATTTTTCGCC TGTAAATAAG TCAACCCAGA CTTGTTTTTC ATCTGGTAAG TAGATTCCAT 2880

CACGGATTGA ATTTCCTGCC TCATCTTGAT TGCCATTATA AATTGGGGCA ACGAGTAAAT 2940

TTGGTCCCCA CATATATTGG TATTGAGAAT CTTTTGTATA TGCTGTTCCT TCATTAGGAA 3000

ATTCTAAAGC CATGGCCCGT ACCATTGGTA GTCCATCAAC TGATTCCTTG GCAATACTGT 3060

AGTTATAAGG CATCATCATT GATTTTAATT TTAGATACGC TCGGTTCAAA TCTGTGGCTT 3120

CTTGATCAAA AGCAAATGGC GTTTTTGGAT TGGAGCCCCA GCCGTCCATA TTTAGCTGCA 3180

CTGGTGTAAA AGTTTTCCAT TGGAAATCAC GTATGTTGAT CTCTTTATTT TTTCCACCAA 3240

AAATTCCATC CATATCTGAA CCGACATTTG GTTGTCCAGA CAACGAAGTC CCGATATATG 3300

TTGGGATATG GAAGCGAATA TATTCCCATT GACCACCCGT TTGATCGCCT GTCCAAATAC 3360

CAGCGTGCCG TTGGGTTCCG GCCCAACCAT CTAATGAAAC AATCATCGGA CGAACTGCGC 3420

CGTCTGTTTC TTTTACAAAA ACATTCGCGG CATCTTCTAC ACCATTTAAG CCAAAGGAAT 3480

ACCCATACCC GACCCACGCC ACGTCAGTTT TGAGAGCTTT AACACCAGCG ACACTAACTT 3540

CTTTCGCAAT ATCCCGTTCT CCTTTTTTAG GATTTTTAGG ATCCGCTGGA TGGAGATTAG 3600

ATTGAGTCCA CAGCCCCACT TCAACACCAT TGGCTTGAGC ATACTCTGTA AATTCTTTTA 3660

AGTTTTGTAC ATCGCCGTCC AATGAATCTG TTTGGCCATA GCCAGCACCG TAGCCGTCAT 3720

TTGGCAAGAA CCACCCTAAC GGCATATCGT TTTTCTTATA TCGATCAATA ACTGCGCGTG 3780

CAGAAAATTG ATAATTCTCT TTTTCTCCAT TTAAAGATTC TAGCGTGCCA TTTAAGTTTC 3840

CTAGATCGCC TGGTTGGTAT TCTTTGTAAA AATTACCATC TTCAAACTTA ACTGCTCCCG 3900

CGGTTCCTTC TGCTACTTTC ACCCAATAGT CCCGATTATA CGCGTTTAAA TGCGCTTCAT 3960

AGAACCCGTA TTCTGGCATC AAAGCTGGTT TCCCAGTTAA TTCATAGTAA TCTTTTAAAA 4020

TACCTGCTGA GGAATCATTA AAGAAATAAA AAGCATCAAA ATCTGTTCCt TCGTGCGTTG 4080

TTGTGGTTTT TTGTGGATCA TGACTTCCAA AATCATAGTT GCCAGGTTTC CATGTATTTC 4140

GAACAACGCC ATACCCTGCA GTTGACCAAT AAAATGGGTT AGGCGATGCT ACTCCGCCAT 4200

CTACCCAATT GTTCGTATTG ACAATCTGGA TAGCTGTTCC TTTATGCGTA AATCGACCAT 4260

TTTGAGTTCC TCCGCCAAAA TAATTTTCTT GACTACTTTG TTTTAACGTT TGTGTAGCTT 4320

TATCATTTTT AAAGGATAGC GGAgCGGTTT CTTCTAAAAT AACTTGATTC TTTTTCTTGT 4380

CTAGAACCTT CATTAAAGCG CTTTCTTTTT CGAACATAAT TTTTAAACCa TTATTTtCTA 4440

GAATAAATTG GTTTCCTGAA TCTGTGACGT TCGTTTGCTC AAAGGCTTGT GTACCATAGT 4500

CGGCCATTGT TTTAGCCGTA ATTTTCGCAA CATGCTTTGG ATCATTTGGC GTTGGATATT 4560

CTTCAAATTT ACCAGTTGGA TCTAGATGAT AGCGAAAAAG ATGGTCGTTT AAAATACTAA 4620

TTTGTGCTTT TTTTCCAGAA GCATACGTAA TGACAAAGTT TTTCTCCGCT TTTtCAATCG 4680

ACACAATTGC TCCGACTGAC GTTGCTGAAT CAGCTACGTC TTCTTTTGCT GTTTGCTCTT 4740

GCGCATCAAC CACTTGAGAA TCCATCGTAA CAAACGAATG GAACGCCAAT ACAGACGAAC 4800

TCACTAATAA AAATGTCAAC CAGTGCTTTT TACCGTAAAT TTTTTTACCC TGATACTCTT 4860

TGTTCATTTT TTCCTCCTTA AAATTGCTCC CCATAGAATA GTTCCGACAA AATTACTCAC 4920

AAAATCCAAA ACCCAACAAA GTTTTATCCA GATGAAAAGT CATAACGATT TTCTTAGTTT 4980

ACTGAGACTG TTAGAAAGTC AGTTTCAGTA ACCCCTTCAT AAACACATCC tTTCAATAAA 5040

ATATTAAATG AAAACAATCT ATTTTCATCT TATATAAAAA TTTTAGTACA GAATCATAAT 5100

TTCCTAATGA ATGCGTTTTC ATTTGTTGAC TACCAATTTT TCACTTTTTT AATAACATAA 5160

TTATCGTATG ATGGTTTTTT TGTGTTTGTC AACGGTTAAT TATAGCAGCC ACCATTAACA 5220

AGTATTAACC CTTGATAAAC CTGTTAATTC ACTTATAATG ATGATTTTAT TGCGAAAATA 5280

TCTACATCAG AAAAAGAAAG ACTTTTTCAA AAATCTAATA CCTGTTGCTC AAATCTAGTC 5340

TATTTAAAAA TAATCAACAC CCTCCTGCTC AGCAAGAGGG TGTTAATTAC TCGTTTTTTA 5400

GTGATTGATT TGACTTGCTG GTGTGAAAAG CAAACTAATG ATGATGCCAA CTATCAGTAC 5460

CGCTAGCAAT ATAAAAATTA ACTTTCCAAA CCGGTGATAG GGATTGATCG AAAGCCCAAT 5520

CCCATAAACT TTCGGAACCC ACAAAGGAGG CGTTGGATTA TCTTTTTGAT CTTTATAGGC 5580

TATCTTTAAA ATCTGCCACG CAAAATACAA CAATGGTAAA ACCATAATTA CCCCAACTAT 5640

ACCTAGCCAT TTTGCTAATA AGTCACTAGT ATGCAAAAGA ATGATTGTCA GAATTGGTAC 5700

AAACAAAAGA ATAAAGCTAA TTGTGGCTCC CACTTGTGCT TGCTTAAGCA GTCGTCTCTC 5760

ACGGAGAGTA AGATTTGCAA ACATTCCATC CATCCTTTCA TGGTTTTCAG ATTAAATCAA 5820

CGAATCACTA AAAAACTAAA TTAACTGCGA AAATCAATGA CCATGCGCCC GCGAATTTTT 5880

CCTTGTTCCA TTTCTTCAAA AATAGCATTA ATTTCTTCTG GTTGACGTGT TTGACAAACA 5940

GGAACAACTT TTCCTTCTGC ACCGAATTGG AAAGCTTCTC TTAAATCTTG ACGGGTTCCA 6000

ACTAACGATC CGACTACGTG AATGCCATCA AGAACAAGAC GTGGAATAGA TAAATCCATT 6060

GTCTCAACAG GCAATCCAAC GGCGACAACC GTTCCTCCCG CACGGACTGC ATCGACTGCG 6120

CCATTAAATG CTGATTTAGC TACCGCAGTG ACTACTGCTG CATGGACACC GCCAATTTCT 6180

TTTTGAAGTT GTGCAGCAAA ATCAGGATCT TTAACTGGAT TAATTAAAAG ATCAGCACCC 6240

ATTTCTTTCG CAAAAGCTAA CTGATTATCA TTCACATCAA CAGCTACTAC TTTTGCATTA 6300

AAGACATGCT TGGCATATTG TAAAGCTAGA TTACCTAGCC CACCTAATCC TGAAATCATT 6360

AGCCATTGAC CTGGACGAAT CCCAGATTCT TTTATCGCTT TGTAGGTTGT TACCCCTGCA 6420

CAAGTAATAG AGGAAGCTGG CTCTGATGCT AAATTTTCTG GCACTTTCAC TGcATAACCG 6480

TCTGTGACGA TGCATTCTTC GGCCATGGCA CCATCAACTG TATATCCAGC ATTTTTTACT 6540

GAACGACAAA GCGTTTCTCG ACCAGAAGTA CAATATTCGC AATGCCCGCA GCCTTCATAA 6600

AACCATGCAA CAGATGCACG ATCACCAATT TTTAAACTAG TAACGTCTTT TCCAACCTCC 6660

GTCACAATCC CAATTCCTTC GTGACCTAAG GTACAACCGC TAACATCGCC AAAATCTGCA 6720

TTCTTAACGT GTAAATCCGT ATGACAGACA CCACAGACTT CCATTTTCAA CCGCGCTTCA 6780

TTATCTTTCA ATGGACGCGA TTCTTTTTCA ACAATTTTTG TTGTTTTTCC TTTTGCTGCG 6840

ACAATTGCTT TATACATGGT TCTTTCCTCC TTGTTTCTAT TGTGAATCAT TGAACAATAA 6900

TAGTATAATA CTACTTTCCA AAATAAGGrA CACCAATGAC TtGCGTAAAT TAACGCTAAT 6960

TTGaCAATTT ATCCTkGCTA GACAGCTAtc TTCCCTTATG TCATAGATAA TTCATACTGG 7020

CATGAnTTTC TCTGCATAAT TGnCAACTAT TTT 7053

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2213 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

ACTTATTTTA CTAACGTCGT GACGCTAAGT TTAGGTGGAC TTGTAGCGAT AGTGATGTAT 60

ATCGATCGAA TTTATACGCC GATAGCCATA TTTAATGTTT TATTTGTTCA ATATAATTTG 120

GnATAAAGTA GCCTATCAGC GCTTGGAAGA TTTTTATAAA AAAGAAGATG ATCCTGACTT 180

AGCGGTTTCC GGCAAAGCGT TACCTGAAAT TCAAACAATT TCATTAAAAG ATGTGTGTTT 240

TAGCGTAGAT AGTCAAACGA TTATGTCCCA ACAAAATCGA CAATTTTCAA TGAATAAAAC 300

CTATGGCTTG ATTGGAAAAA GTGGGACAGG TAAATCTACC CTGATTAAAT TGATTTTAGG 360

ATTATTGAAG CCGACAGAGG GGACAGTTTA TGTGAATCAA TTTCCTCTAA CACAATTTAA 420

TTTGGAAGAT TATTATGAAA AAGTTTTTTA TTTATCTCAA GATGTGCCAA TATTTCAAGG 480

AACGTTAAAA GAAAATATTG TATTTAATCA AGAAATTTCT GATGAACAAG TAATTGAAGC 540

TATGTATCGT TTTCAGTTGG GGGAACTTTA TGAACGGTTG CCGGAAGtTT AAACACTATt 600

GTGAGTGAAA AAGGAtGAAT TTTTCTGGTG GTGAAAAACA ACGGATTGCT TcACACGCTT 660

AGCATTTACA CAAGCAGAGA TCCTTATTTT AGATGAGGCG ACTTCTGCaT TGGACGAAAA 720

GnCCGAAGAA AAAGTGTTAC AAGAAGTACA AAAATTTACC CACAATAAGT TGACTATTTT 780

GGTTACTCAT CGACCTAAAA CCCTGCGATT TGTAGATGAA ATTATTGATT TAAATGAGTG 840

AACTTGTTAG CTCTTTGGCG CTTCTTTTGC TACACTTTTT ATAAGAATTG AGAGAAAGGG 900

GCAACTGTTG ATGAATAATA ATGAGCAAGA ATTAGCGGAA TTTTGGGATG AATTTGCAGA 960

AGAATATGAA GAAATTCAAC AAGAATCGCC ATTTCCAATT GCTCGAGAAT TACGTGATTT 1020

CTTGGTCCAA GAAGGCATTT TTCCGTGTCA GACATTTCTG GATATCGCAG GTGGGACGGG 1080

ACGATACTTA CCTTTTTTTC AAGAGCAAGT CACGGAGTAT ACGCTAGCCG ATATTTCTCA 1140

GAGAATGTTG GAGATTGCGG AAGCAAAAGC ACAAAGCAAC GTTGTTTTTC TTCACCAAAG 1200

CCAAGAAAGA CTGATTGAGA CTGGGAAAAA ATTTCAAGTC GTTTTTTCAG CGATGAACCC 1260

TGCTTTGGAT ACGCCGGAAA AAGTTAATGC ATTGTGCCAA TTAAGTGAAG AATGGTGTCT 1320

TATATTTCGG CTAGTGGAAG AGCAAGATTC GTTGTTTTCT CCTTTTGAGC AAGAGAGCAA 1380

CCCGCAGTTA AAATGGATGG CACAATACAA AGCATTTCTA AAAAAAGAAC AGCGACCTTT 1440

TTTCACAAAA AAATTTTTCT TTGAAGCATC AGAAGCTATT TCAAAAGACT TTTTCCGCAG 1500

CTATTTTGAA GAACAATGGT CAGTTCCAAT TTTAGAACAA CGCGTACAAG AAATTTTTGG 1560

TTCTCATGAA ATTAAGcAAA ATCAGCGTAC CATTATTTAT GAATTAATTG CGATTCCATG 1620

TAAAAAAACA ACCAGCGATG ACTAATTATC GCTGGTTGTT TTTCGGTTAT TTAATTGTGC 1680

GAAAGCTTTT TCTGGAATCT CTTTTTTATC AATTTTGACA GGAGATTGGA TGATTCTTTT 1740

TTTACTAATG GTTGCTTCTA TGTAAGCATT TGGTTCCAAA GGTTGAGGAT TTTTGTCTGT 1800

CCGTTCGTAG TCCATAACTT GCGTTTTTCC ATTACTTTTA ACAAAAGTTA GTCGATAGTG 1860

TAAAGAATAT TGGCCATCTA TGCTTTTTTG ATTGTCATCT ACTGTTTGTT TTCGTTCAGG 1920

AATTTCTAGA GGAACTTGTG CATAGGCTTT TTCACCCTTA TATGTATCTG CATAATAACG 1980

ATAGCCTGTA TAACCACCAA CAATGAATAA GCCAATAAAT AAAATAGCGC CGATTTTTTT 2040

CATTTTTTCT CCTCCTAACA ATTTACAAAA ATAGCATAGC AAAAGAAGAA GGAGATGGAA 2100

CATCGAAAAA ACGAACAGTT ATCTTACAAA ACTGTTCGTT TTTTTTGGTC TATTTAGCTT 2160

TCATCAGAGA GTAGAAAGGA ATCATTTTCT TACGATAACT AAAGAAAATA ACn 2213

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4808 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

ACGTTCAGGA ACTGATACCC TAGATTTGCG CAAGTTGCAT CGTCCGCAAa GATATCaTCA 60

ATCACCACCA ACGCCGTGGC ATTATTTTGA TGAAGAATGT CATCAAAAGC CATGTAATCT 120

TCAGGTTTTT CTAACGCAAC GCTTGGTAAA ACAGTTAATC CTTGCATCAT CATTGCTTTT 180

TGATAACCAA AGTAACGTTC AAAGTACAAA CTTTCATGCG TTGTATTGGT CACAAATAAA 240

ATATTTTTAT GGCCGTTTTC AATTAAAAAT TCCGTCGCTT GTTTCCCTAA TAATTGaTTA 300

TCATTATCCA CGTAAACTAT TTCCGTTTCA TGTTGGTAAG GTTGACCGAT TAAAGTAAAG 360

GGAATGTTGT TTTCGTCTAA GTAATCAATG ACGGGGTCCT TACTATCTGA ATACACAAGA 420

ATAAAACCGT CCACTTGTTT TTGAAGATGC ATTCTTTTAA CGTTTTCCAA TAGTACATCA 480

AAACTTTTGG CAGTTGCAAT AGCAGTAGTC ATATCGTAAA GACGGGCTTC CTCGTTGATG 540

GCTTCCATAA TTTCTAAATA AAATGGATTC CCTAACCGTT CTTTCGAATC TAAAGGAGGT 600

AAAATTACGC CAATGGCACT TGAAATTCGT TTACCCAAAt TCCGTGCCGT CATATTTGGC 660

ACATAGCCTA AGTCATCCAT GACTTTTCGA ACTTTCTTTT TTGTCGATTC AGAAATACTT 720

GGATGATCGT TAATGACACG CGAAACGGTT GAAGTTGCAA CCCCCGCTTT CTTAGCGACA 780

TCTTTTACTG TAATTGCCAT TTTTGCCACG CCTTTCTTAG TAATTATAGA AGGTTAAATT 840

GATAAACCGT TTGGTAGGTT TGCTTTTTAT TAGCTGGTAA TAAAATAGAG CCCCACTGAG 900

GATGATGAAC AATGTCTGGA TACTCTTGTG GTTCAATTGC TAAGCCATAG TTTGAATGCA 960

TCGCTTGTCC ATTAATAGAA AATGGAGCCT CAAAACCAGT CGTAGAAAAT AATACCATAC 1020

TTTGGTTCGT TGTTGCTATC GTCATTTTTC TTCCAGAAGT AGGCTCTTGT AAGCATAATT 1080

GTGGCTGTCT TAAGTCAGAG CCAACTAAGG AGAAAACATC ATCTAACCCA GTTGGATAGC 1140

GTTGGAGGAT TGTTTCGATA GGTGTCAGTT CATTAAAACG TATGGATAGT TCAGAAGCTG 1200

GTAAAATCCG TCCAGTAGGC AGCTTATTAT CGTCTAATTC TAACATACCT TGACAGTTCA 1260

AAGATAGCAA ATGTGACAGA ATGTCTAACT TTCCATCGCC AGATAAATTA AAGTAGGCGT 1320

GGTTCGTAGG ATTAAAAAGT GTTTCCTGTT TACTCCGTCC AGTACTGATC aTTCGTAAAT 1380

GATTATCCTC TAGCTGATAA GTGATGGTGG CCTCAATTGG TCCAGGATAG CCAGAAATGG 1440

TGTCGGTTAA ATGAAAAGCA ACGCCAATGA TATCCTTTTT TTCAAAGGGC TCGACTGTCC 1500

AATATTGAAA AGACCACCCC TTTGAGCCAC CATGGATATG ATGACCGTCG CTGTTTTGTT 1560

CTAATATTCC ATTTTGCCAA TGGCCATTAC GTATGCGTCC AGCAACTGGT CCAACAGTGG 1620

CACCAAAAAA GCTTTGATCC TTTAATACAT CTTGATACGA ATCATACGAT AAAAGGATAT 1680

TTTCCCAATG ATTATTTTTG TTTGGCATCA AAAGTTGATG GAGGCGCGCG CCATAATTGA 1740

GAAGAACGAC TTCTAAAAAG TCGTTCTTCA AAGTAATTTT TTTTAAGTTT GTGTTTGGAA 1800

TATCGGTTAC GGTGATACGA TTATCCTTCA TGATCAAGCC ACACTTTTTT TAAGTAATCA 1860

AACGTTAAAG CGCTCGTCTC TGGCACGATA GGCAACCCTT CGCCTAACTC TTCTTTGCGG 1920

CCAACGCCAA CAGGTTGCGC TCCACTAGCA AGAATCGCTT GAATCCCAGC TTTGGCATCT 1980

TCAATGCCAA GACATTCTGC TGGTGTCAAA TCAATTTCTT TCGCAGCTAA AATGAAAATA 2040

TCTGGAGCGG GTTTACCTGC CTGGACGTCT GCTGGATTAG CAATGGCATC AAAAAGGGGC 2100

GTCAATTGCA TTTTTTCTAA AAGAAAAGGC CCATTTTTAC TTGCAGAAGC CAAGGCGATT 2160

TTAATATTGG CTTCACGTAA TTCCGTTAAT AAAGAAAGAA TTCCTGGGTA GACATCTTCA 2220

GGCGTAATCG CTTGGATCAT TTCCAAGTAA TAATCGTTTT TCCGTTGCGC TAACTGGGCA 2280

AATTCTTCAG AAGAAAACGT GCCTTCTTTT TTTCCGTATT TTAACAGTAA CTGCAGAGAA 2340

TCTTCACGGC TTACACCTTT CAGTTGTTCG TTGAAAACAC GGTCAATGGA AATTCCAATT 2400

TCGTTCCCTA ATTTTTTCCA AGCGTGGTAA TGAAATTCTG CTGTATCAGT AATAACGCCA 2460

TCTAAATCAA ACAAAACGCC TTTAAACATG AACTTCTTCC TTTCTGATCG GTAAAACTAA 2520

TCGATCTTTT AACGTGTATT TTTGATCGTA AACAGTTAAT GGAATTGCAG GACCGTCGAG 2580

TAAAGTGAAG GCCACTTGGT CTGCTGCAAC TTCAACAAAT AGTAAACGAT TCCGATAATT 2640

AATGTGGAAT GAGTAGGCTG TCCATGTCGC TGGTAAAAAG GGTGCAAATT TTAGTTGTTG 2700

ATGGTCCGTT TTCATTTGGG CAAAACCTTG AACAATCGCT AGCCAACTAC CAGTCATTGA 2760

GGTAATGTGT AAGCCATCTT CTGTATCATT ATTGTAATTA TCTAAGTCAA GGCGCGCTGT 2820

TCGTTGATAC ATTTCCACTG CTTTTTCTTC CATGCCTAAT TCAGCAGCTA ACACAGCATG 2880

AATGCTTGGT GAAAGAGAAG ATTCATGGAC AGTCATTGGT TCGTAAAAGT TAAAGTTCCG 2940

GCGTTTTTCT TCTAAAGAAA AGGCATCATT GAAGAAATAA ATTCCTTGAA GAACATCGGC 3000

TTGTTTAATG AAACAAGACC GTAAAATTTT ATCCCATGAC CAATGTTGAT TTAAAGGCAG 3060

CTCACTAAGT GGTAAATCAG AAACAGGCAT TAAATCTTTA TCTAAGAAAG TGTCATGTTG 3120

TACAAAAATG CCTAATTCGT TATCCACTGG GAAGTACATA TTTTCTACAA TGTCTGCCCA 3180

TTTAGCTAAT TCGTCGTCTG CAATTTTGAG CGTTGTTTCT TCTTGGAATT TTAAATAAGA 3240

CTCACGGGTG TAACGTAGAA CCCACGCAGC GATGGTATTG GTATACCAGT TATTATTGAT 3300

ATTATTTTCA TATTCATTTG GCCCCGTTAC GCCATGAATC ATATATTGTT TGTGACGTTG 3360

GGAAAAGTGG ACCCGATCTG CCCAAAAACG AGCGATTGCT ACAAGTACTT CTAGACCAGC 3420

ATCTTTTAGA TAATCTTCGT CTCCTGTATA GTTGACATAA TTATAAATCG CATAAGCAAT 3480

CGCACCATTT CTGTGAATCT CTTCAAAAGT GATTTCCCAC TCATTATGAC ATTCGACACC 3540

AGTAAAAGTA ACCATTGGGT AAAGCGCACC CTTTAACCCT TGTTGTTGGG CATTATGAAT 3600

GGCTTGAGGC AATTGATTGT GGCGATATTT CAGTAGATTT TTAGTAACTT CGGGCTTAGC 3660

CAACGCCAAG TAAAGTGGGA CAGCATATGC TTCTGTATCC CAATATGTGG CACCTCCATA 3720

TTTTTCGCCC GTAAAGCCTT TTGGTCCGAT GTTTAAACGA TCGTCTTCTC CATAGTAGGT 3780

TGAGAAAAGT TGGAACAGAT TAAAACGAAT GCCTTGTTGG GCTTCGTCAT CTCCTTCAAT 3840

AACAACATCA GCTAATTGCC AACGTTTTGC CCAAGCGGCA GCTTGCCCAG CTTTTGCTTC 3900

AGGGTACAAG GTAGTCATTT CTGCAAATAA TTCGTTTACT CGTGTGATTT GTTGCGTTTC 3960

TGGCACATCA CGACTGGTAA CGACAAGCAC TTCTTTTGtG ATGGCGGTCG TTtCGCCtTC 4020

GTTTAAAGAC aCTGTCAATT CTGCGGTTAA AGCCAAGGGC TGTTGTGTAT ATGTAGGCAC 4080

AACAGAGGCT CCATCGATGA AATGGCGCAT GCCTGCAGTC ACTGTAAAAC GCTCTATTTC 4140

AAAGTTGTTA GGAATGGTCT TAGTCGTAAC AAAGCCAAGT GTTTCTTGTG TTTCAATAGC 4200

TCTTTCTTCC CAAAAATGTT CTTCATAATT GCTGTCTTCA TTTTGAACAT CGCCATCCAA 4260

TTTTGAGTGA AGAGTGATTG TGCCTGTACC TTCCAGCATT TCAATTGTTA AGTGGATGTA 4320

AGCTGCCTCT TTTTTTTCTA AGCTTAAAAA ACGTTCAAAA GAGAAGCGCA CTTTATTCTT 4380

CGGTGTTTGG ATGGTGAATT GCCGAGAAAG AACCCCGTTT TGCATGTCTA ATTCTAAAGA 4440

AAAATCTTCG TAAGGTGTAG TGGCTAAATC AATTGTTTGT CCATCAATTT GTAAGTCCAT 4500

TGCGATGAAA TTAATTGCGT TAATCACTTT ACCAAAATAT TCAGGATAAC CATTTTTCCA 4560

CCAACCAACA CGTGTTTTGT CTGGATACCA AACACCTGCT AAATAGGTTC CTTGATGGTG 4620

GTCGCCAGAA TAGTGTTCTT CAAAATTTCC CCGCATCCCC ATATAGCCGT TGCCAATGCT 4680

TGTTAATGAT TCTTGTAAAC GTAAATTTTC TTTATCTAAG TGTGTAGTAC GGATTTTCCA 4740

AGGATCAATT TGGAATAAAC GTTTGATTTG TTTCATGAGA ATCCCCCTCA GTGATTTTAT 4800

TGATAGGT 4808

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15747 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

GAGTAGGAAA CAGATATACT TAATATAGAA AATTCAGATG CACTTATTGG GAGGTTATTT 60

TATGCAACGA AAAGATGTTA CACCTAATTC AGATAAAAAA AGTTTATTGC AAAAATTTGG 120

GATTGGTTTA GCTGGCGGCT TACTTGGTGG CGCCCTTATT CTAGGAGGCG CTTATTCTGG 180

GATTATTCCC ACACCCAACG GCGGAAATAA TGCGGCGACG ACGACATCCA CTAATCACGG 240

CGACACAAAA GTCAGCAATG TAAGTTACAA TGTGTCTAGC GATGTCACAA AAGCCGTGAA 300

AAAAGTTCAA AATTCTGTGG TTTCTGTCAT TAATATGCAA AGTGCTAGCA ACAATTCTTC 360

GGCAGATGAT CCTTTTGGGG GATTGTTCGG TGGGAATGAA GGTACTCAAG ACTCTTCCGG 420

cAACAATGGT AACGATTTAG AAGCAGCCAG TGAAGGTAGT GGGGTTATTT ATAAGAAAGA 480

TGGTAAAACG GCTTATGTCG TGACAAATAA CCACGTCGTA GATAAAGCGC AAGGATTGGA 540

AGTTGTTTTA TCTGATGGTA CCAAAGTGAA AGGCGAATTA GTCGGAACCG ATGCTTACAC 600

GGATTTAGCC GTGATTAAAA TTTCTTCCGA CAAAGTTGAT CAAGTCGCTG AGTTTGGAAA 660

TTCTAGTAAA ATCACAGTCG GTGAGCCTGC TATTGCAATT GGTTCCCCTC TAGGCTCTGA 720

TTATGCTAAC TCTGTCACAC AAGGAATCAT CTCTTCTGTG AATAGAAATA TTACCAATAA 780

AAACGAGTCT GGTGAAcCAT CAATATTAAT GCTATTCAAA CCGATGCTGC CATCAATCCA 840

GGAAACTCTG GTGGTCCACT AATCAATATT GAAGGACAAG TCATTGGaAT TAACTCAGTA 900

AAAATTGTGC AATCAACTAG TCAAGTGAGC GTTGAAGGGA TGGGCTTTGC CATTCCAAGT 960

AATGACGTAG tCAACATTAT CAACCAATTA GAAAAAGATG GTAAAGTGAC GCGGCCAGCT 1020

CTCGGGATTA CGATGTCTGA TTTAACAGGT ATCTCTTCAC AACAACAAGA ACAAATTTTA 1080

AAAATTCCAA CTTCTGTAAA AACTGGCGTA GTGGTTCGTG GTGTTGAAGC AGCGACCCCT 1140

GCTGAAAAGG CTGGATTGGA AAAATACGAT GTTATCACGA AAGTTGACGG CCAAGACGTA 1200

AGCTCTACTA CAGATTTACA AAGCGCGCTT TACAAGAAAA AAGTTGGCGA CAAAATGGAA 1260

GTGACTTATT ATCGTGGTTC TAAAGAAATG AAAGCAACTA TCGATTTAAC CATTGATAAA 1320

TCAGCTTTAA CACAACAAAA TAATCGCAGC AATCAATAAA AGAAAGAaAC GACCTGTTGT 1380

TCAAACAACA GGTCGTTTCT TTCTTTTATT CTCCTACCGC AAATGCAGAA CCAACTTCTG 1440

CATCAAAAGG CAATTCTAAG ATTCCTTTTT TAGCTGGTGC ATTTGGTAAA CGTAATTCTT 1500

TTGCTGAACA AATCATTCCA AAACTTTCTT CACCACGTAA AACCCCTGGC CAAATCATTA 1560

AACCATCAGG CATCATTGCG CCTGGTTTGG CCACAACTAC TTTTTGTCCT GCTTTGATGT 1620

TCGGTGCGCC ACAAACAATT TGTAAGATTT CGCCGTTGTC TACTTCTGTT TGTGTGATAG 1680

ATAAATGATC TGAATCTGGG TGTTTTTTAC ATGATTTGAC AAAGCCAACC ACAATTTTAG 1740

GGGTGTCGTC TTTTTCAAGG GTTTCTGAAA ATCCCGCATC TTTAATCGCT GTATTGACTT 1800

TTTCGATTTG TTCGTCTGTT AAAGTCACTT GGCCGGTGCC ATTAATGTCA CCTAGAAGCG 1860

CTGAAGCATT GAAAATATTC CAAGCTACCA CTGTTTCTGG CGCCTCAACT AAACTTACAC 1920

GTGCCACATC GCCGACACGT TTTACTTGGT TTTCCGCCCC TTGATCGTCC GCAACGATGA 1980

CAAGTAAGAC ATCGCCAACG TGTTCTTTGT TATATGCAAA AATCATTTCT GTTCCTCCTA 2040

AATTTGAAAG CTATAGGCTG GaTAGCTAAA TTCTCTCCTG TCCTATAGTA ACAAAAAAAG 2100

TGGaAAGTTG CACTTACAAT TATTGAATTT TAGGCAATCC CTAATGCTGC CAAAAAAATA 2160

GATAATAGAA TAACCCCAAC TAAAAGTGTC GTTGTTTTCA CTTTTTTCCC TAATAACCAG 2220

TAAATAACTA AGAAGAATAA AGCCGGGAAA AAACCTaACA TGATTTGGTC GATATATTCT 2280

TGTAATGGCG TGGCGATTTT TCCAGAGCCA ACTTTGGCAG TTAACTCAAA CACGACATTA 2340

GAAGCCGTCA TCCCACCAAT AACCATTAAC CCTAAAATAG AGGCAGCTGT GGTAACTTTT 2400

TCGATTACGC CGCTATCTTC AACATCACTA AAAAATTTAG CACCTAATTT AAACCCTTTA 2460

TCCAAACAGA TATACCGTAA AGCAAAATGC GGAATATTGA TAATCAGTAA AAAGGCAATC 2520

GGACCAAAAA TATTTCCTTG TTTAGACAAT GAAATAGCTA CACCTGTTGC GATAACTTTT 2580

AAGGTCCCCC AAAAGAAAGA ATCCCCAATC CCAGACATTG GTCCCATTAA ACTAGTTTTC 2640

ACTGCATTAA TACTAGCTGA GTTAAATTCA TGATTTTTAG CATTTTCTTC TTCCATTGCT 2700

CCAGAAATTC CTAATAATAA GGTAGAGATA TGCGGGGTGA CAGACATAAA CTCTAAATGT 2760

CGTTCTAAGG CCGCTTTCTT CTCTTCATTA TTTTTATACA ACCTACGTAT AATTGGAGCC 2820

ATTGCATAAG CATAAGCTAA ATTTTGCATC CGCTCATAAT TCCATGATGC ATCTAACGTT 2880

AAAGAGCGAA AAAATACAGA CCAAAAATCT TTTTTAGTAA TGACAGATTT TGTGTTTTTA 2940

GAATTCATTG TCATCGACAA CAGCCTCCCT TGTTTGACCA TCACTCATAT TCATATATAG 3000

TAGGACAATC ACGATGCCAA ACAAAGCAAT CCCCACTACT GAAATTTTTA AATAAGCAGC 3060

GAATAAAAAG CCTAATAATA GAAATGGAAt CaTCTTTTTT GTTACCATCA TTCTTGCTAA 3120

CATCGCAAAA CCAATTGCTG GaATaATACC TGCCGCAATA TCTAATCCaT GCGTAATAAA 3180

GTCTGGGACA TAAGACAATA CCTTTTCAAT AAACGGCACG CCTAAGTAAA AGCCTAATGC 3240

CACAATTAGA GCTAACAAAA CTTTATTGCC AACTCCTGCA ATTAGTTGAA TTCGCTCTAC 3300

CCCTTGAATA TTGCCTCGTT TTGCTTCCTT ATCAGCTAAA TGGCTAGCGC CTGTAAGTAA 3360

AAATGTCATC TGAAGATTAT CGATTAAAAG TACTAATGTA GCGATTGGAA CAGCTAAAAC 3420

AACCGCACTA GAAATTGATT GCCCACTCAT AATTACATAA GCTGTTGCAA TAATCGTTCC 3480

TGAAACAAAA TCTGGTGGAT TACTTGCGCC AACAGGAAAT GCACCCATAA ATACCAACTC 3540

TAATGTTGCG CCAAGCTTAA TTCCCGTATC TAAGTCTCCT AAAGCTAACC CTACTAGCGG 3600

AGcGACAACA ATTGGACGGT TCCACATCGT TGAACCAAAC CAGCTATGCA TGTAGGCAAA 3660

AAAGGCAATT AAAAACATTA ATAGAACCTT GATAAATGTC ACTTCCATCT ATTTGACCCT 3720

CCTATTTATT TAATTTAATA ACTTGTTTAT AGCAACCTTT TGATCGTCAG GGACTAATTG 3780

AACGGTTAAT TCGATTCCTT CATTAGTTAG TTCTTTAATT TGTGCAATGT CAGTTTCTGA 3840

TAAATGAACA GCTTTTGAAA TTTGTTTGCG CGTTTCCCCA TTTTTCATTC CACCTAAATT 3900

AACTTCCTTA ATTGTTGGCA CTTCTTTGAC CAATTGATAC ATATCAGCAA CTGATTCACA 3960

TAAAATAAAT AGTTTATATT TATCTGTCAC CCCAGAATTT AACGCCTTAA tTGAATCATC 4020

AATCGTTTTC ATAACTAATT TAACACCTGA AGGTTTTGCC ATTTTCATAG CGCTCATTCT 4080

TAATGGATCC TCAACCAACG TATTACTAGC AATCAAAATA CAATCTGCTT CTAAAAATTT 4140

TGTCCAAGTA AACCCGACCT GACCATGAAT CAACCGATGG TCCACTCTTG TCATTTTAAT 4200

CATCCTTATT CACCTCTCAG TTCTTTCACA ATTTTCATAA ACGAATCAAC ATAAGGAGCA 4260

TAAATATCTC CCGTATCTTT TCTTGTATCT TTGAAGTTAC TCCCTACAAT TGCTGCATCG 4320

CAAATCGCCA ATTGTTCTTT CACACTCTTG TCATTTAACC CCGCTGCGAC AATTAATGGA 4380

AACTCAGGTA ATTGTTTTCT AAACAACTTA ATCTTTTCTA GTGATGTCTC TTCACCCGTT 4440

GCATTTyCTG TCACTGCAAT AGCATCACAA CGTTGCTGTG CAATTTTCAA ATCTTCTTCG 4500

ACTGATTTTT CAGACAACAT TGGTTGATAT TTGAAGCGCA CGCCACCAAT TAATTTTGCT 4560

GTTGTTTTTG CACGATATAA ATCAAAAAAA GCTTGTAAGG ATGCTTCATC CCGCGGTTTC 4620

ACATGACCAA CCACAGAATC AATCTGTAAA AATTGTAAAT GATATTTATT GGCTAAATGA 4680

AAACCTAATG GGTCAACATT TAACACATTT ACACCAATAG GAATCGGTAA ATCTAATGAA 4740

GTAACATATT GGaGCGCTTT TTCTAGCTGA ACATAATCCC CATAATAGTT CTCCATCAAA 4800

ATkGCATCTA CCCCATGTTC ACTGTAAATT tGAATCTCTT TTTTTGCTCG CTCTTGAATC 4860

TCTTGATCTG TTTTACCTTT TAAATGAATC ACCCCAATAA TTGGTTTTTC AACAGCAAAC 4920

AGCTCTAAAA AATCAGCTTT TTTCTTCATT AATACCCCTC CTAGAAATCA TCATTTTCTA 4980

TCTCTTCGAT ATGAACGCGT TGAATTCCTT GTCGTGCATT TCGAATAATT ACCTGAATCT 5040

TTTCATCTGT TAATTTTTCT GAACCTGCTA GCATAAGCTC TAGTGCCAAA TTCAAATTAA 5100

CGTCAGAGAA AACGATTACA TTTTCAAGTA CAGTTAGCCC CATCATAGCT GTAACCACAC 5160

TTGCACCAAA TAAATCACCC ATGAGGATAA ATTCATCTTC TGGCGCTTGT TCGGCTTTCT 5220

TTTTTAACTC TCCAAAAAAA TCAACTGCAC TTTCACCTGG ATACAGACTA TATGTATGAA 5280

CGGCAGCTGC AGAGTGTCCC GCAATCATTT CTAACGAGTG TTTAATCCCC TTAGAAAAAT 5340

CACCATGTGA AGCAATGTAA AACCTTCTCA TTTCATAAAC CTCCTTTATT TATTATATGA 5400

ACGAATGGCT ACTTTGTAGG ATTCGCTAAG GATATATTGC TCATTAAAAT ATAAAATTTG 5460

ATTATCATTA TCCACAACTG TTTGATTAAT TTTGACCAAC GCTTCTTCTT CACTAATCGC 5520

TAATGCTTCT GCCAACTTTT CATCGGCCAA GGCAATTTCA ATGATATTAT CAAAATAGCT 5580

AGCAATTACT CCTCTATATT GCCGTAAAAT ATCTAATAAT GAAGTATCTG TTAATTCTAA 5640

GTCCAACATA AACTGaTCTT CTTTTCTAAA ATAATCAATC TCGAAAATTG CAGGCATGTT 5700

ATCTAACGAA CGGAGGCGCT CAATAACAAT AACTTCTGTC GCTTTTTCAA CACCCAATGC 5760

TTCTGCAACA TGCCCTGTTG CTTTTTTCAA ACAAATTGAT ATGATCTTAG TTGTTGGTTT 5820

ACGATTCATT TTTAAACCTG ACtCAGTAAA ACTCCCACCA GCTGTCATGC TCTCAACATA 5880

GGATGGCATT GAAACAAATG TCCCTTTTCC ATGCCGTTTA ATTAAAATAT TTTCATCAAC 5940

TAATTGTTTA ATCGCATTAC GAACGGTTAC TCGGCTCACT CCGTAAATGG TCATCAATTG 6000

TTCTTCTGAA GGGATTTTTT CTCCTACTTG ATATTCACCA CTTCTAACTT TATTTAAAAT 6060

AGCATCGTAT AATTGTTGAT AAAgCGGTGC TTTTTTATTT GATTCTAATT TATCCATTAA 6120

TCATACACCT CATTTTAATA CGTCTTAATA CGTCTTAAAA ATAGCAAATA AAAAAGAGAA 6180

TGTCAAGAAA GAGTTTCCTT TCTTGACATT CTCTTTTTGC TTATTTTTCT TCAGAGTGAA 6240

CAATTAAGAC ATCGCAAGGG GCTTGACGAA TAATGTAGCT ACTGACGCTT CCCATGACCA 6300

CACGTTCGAn TGCGTTTAAT CCTGATTGTC CCACCATAAT TAAATCAACA TTGTATTTTT 6360

CAGGTAATTC ATGACTCATG ACTTCTTTGG GTGAACCAAA CATGGTAACT GTTTCAACAT 6420

TATGAAAATC AACGCTTTCA GCATACTTTT TACAATCGGC CATTAATTCT TCCGCAGCGG 6480

CTGTTTCTTG ATCAATTAAT TGATCATTTA ACGTCGAATA CCCCATCATT GTATACACTT 6540

GATTTTCTAA AATATTTGCA ACAATGACTC GGCTTCCGTT TCGACGAGCA ACCTCAATTG 6600

CACGTTCATA AGCTAAATTG GCTTGATCAC TACCGTCTAC GCCGACAAGA ATATTTTTAT 6660

ACATTTGTTC TTCCATTTTC GTCACCTTCC CGATGGAGGC TGATCTAAAA GGCTTCTTTG 6720

ACTAAGAAAC TAGAGTCAAA AATGAATTTA GGACCCTTGC GCAGAACGCC CTTTTGTCCA 6780

GCTGATTCTC TCTTAATTTA ATGGTTCCTC TGGTTCATGA AAATTAGCCT AAGCTAGTTA 6840

AAAATGCACT GATTTCTTCT TTTGTTTTTC GGTCTTTATT TACTAAGCGT CCTAACTCTT 6900

GACCGTTTTC AATTACGACA AAACTCGGAA TACCAAAAAT ATTCCACTCG GCAGCCACGT 6960

CAATAAACTG ATCACGGTCG ACTTCGATAA ATTGAAACGC TGGAAATGCT TCTTCAATTT 7020

CAGGCATGAC TGGTTTAATG AAACGACAGT CTCCGCACCA ATCAGCAGTA AAGAAGAAAA 7080

CATTTTTACC TTTTTCAACA TAAGTTGCTA GTTCTTCTAA TGAATTGGGa TAAATCATTT 7140

AAACATCCTC TACTTTCCGT AAACAATTGT ATTAACTGTT GATTTATCCA AGCCACGTAA 7200

TGCTTGTAAC AACATTTCCC GCGCTGCTTC ATAATCCGCA ATACTAAACA TTGTTTGATG 7260

CGTATGGATG TAACGACCAC AAACGCCGAT GACTGTACTT GGTACACCGT TATTAGTTGT 7320

GTGCGCTGCC CCAGCATCCG TTCCGCCTTT AGAAACAAAA TATTGATACG GAATATTGTG 7380

TGTCGCTGCT GTATCTAATA AATATTCACG TAAGCGAGGC AATGTGATCA AACCTGGGTC 7440

GAAAATCCGT AATAACGTTC CTTCACCTAA ATGACCATAC GTCCCTTTTT TCGTGTGAAT 7500

ATCATCTGCC GCTGAACAAT CCACCGCAAA GAAGATATCT GGATTAAATT TATGAACAGA 7560

TGGTTTAGAA CCACGTAAAC CAACTTCTTC TTGTACGTTT GCTCCAGCAA TTAATGTGTG 7620

GCCTAATTCT TCATTTTGTA ACGTTTCCAA TGCATCTAAT ACCAACGTAC AACCATAGCG 7680

GTTATCCCAT GATTTACTGA TAATATTTTT GCCATTGGCT GTTTTAATCG TTTCTGTTTG 7740

TGGAACAATG GAATCCCCTG GACGAACGCC AAAGCTTTCA GCTTCTTCTT TTGAAGAGAA 7800

TCCTGCATCA AATAACACAT CAGAGACTTC TAATTGTTTT TGACCACTCG TACCACGCAA 7860

CAAGTGTGGC GGAATAGATG AAGAAATACA TGGATAGTTC CCTTTACTTG TTTTTAACGT 7920

AAAACGTTGT GCTGAAACAA CATAGGGATT CCAGCCACCT AAAGGAACCA CTTGGAATAA 7980

ACCATTATCG TTAATTTGAG TTAACATGAA GCCAACTTCG TCCATGTGAG CTGCTACCAT 8040

TACGCGTGGC GCTGCTTCCA CTTTTGAACG TTTAATCCCA AAAATACCAC CTAAACCATC 8100

ATATTGGACA TCGTCCACTA ATGGGGTGAT ATGTTCTTTC ATGTACGCGC GAATGTCGTC 8160

TTCAAATCCG CTGGTTCCTT GTAATTCAGT TAATTCCTTG ATTCGTTGAA ATGTTTTTTC 8220

TTCCATAGAT GTATGTCCTA TTGCTTTTGA TTTGTTGGAA ATTACGCCAA CCAAGCGTCA 8280

CTTTTTCTCT AAAAAGTCAC TTTATCCTGT CGATAAAAAT AGGTACTCCT TTCATATGAT 8340

TTCACTCTTA TTATAACGCA AATACTAGAA AATGCACATC GCTTTCTCTT TTTAAACGTG 8400

ACAATGAGAA TGCCCTAAAA GAATGATTGA AAAACTTTTT TTATTTCTCA GCGGCCCCTT 8460

CTCTATGGTA AAATAAGAGT GATACTGACA AAAGTAATAG TAAGCAGAAA CCGACCAAGG 8520

GAGGTCATAC AAATGAACGA AGAAAATGAG TTGACATACT TTAAAGGCGG CTTAGCGCTA 8580

GGCGTTGGTC TTGGTTTAGT AGGCGGCATT GCCTCAACCT TGTTCTACCA TAAGAAAAAA 8640

ACAATCTCCG CTGATCTCGT CTTAGAAAAT GTAAAAGCTG CTTTTTTAAA AGAAGGACCG 8700

ATTGAAGGTT CGTGGATTGA GTTTGAGAAA AAGCCCTTGC GGAAATTCGC CATCCACTCC 8760

AAAACGTATA CTGGAGGAAT TTGCCGCATT GAAGACGACG GCATTGTTCA ATACGAATTC 8820

ACTGCAGATG CCTACACAGG AACTATCATT GACATTCAAC GATTAAAAGA TTAAAAAGAT 8880

TGTTACAAAA GCACTTATTT GCTTTTGTAA CAATCTTTTT TTCTTCTCAT TTGTTTTTTG 8940

TAAAAGGGAT GTTTATTTTA TGTCAAAAAA ATCAAATCAA TGATGAATAA ATACTCAAAA 9000

ATCGAATGCT TTCTGCATGG AAATCAGCTT AAATCTTTAC TTATCAACGT TTGTCCACTT 9060

TTTTCATAAA AAAGATTAGA CATTTTTTAA AAATTGATAA AAGAATAATC ATTTCTAACG 9120

ATTTTTTAAT TTCTCCTTTT CATTTTTACA TTATACTGAG AGTATTGCTA AAAATGAAAA 9180

GGAGCGTGTC AAATGAAGAA ATCTGTTTTA TTTACTTCAT TACTTGTATT ATCAAGCTTA 9240

GCTTTAGCnG CcTGCGGCGG TGGCAGTGAC GATAAAGGAG CTAGCAACGG CGGCAGCGAC 9300

AACCAAGTAT ACACAATGGT TGAATCCCAA GAAATGCCTA GTGCCGATCC GTCCCTTGCG 9360

ACAGATGAAG TGAGTTTTAC CACTTTAAAT AATGTCTACG AAGGAATCTA TCGTTTAGAT 9420

AAAGACAACA AACCCGCGCC TGCTGGTGCA GCCGAAAAAG CGACTGTTTC AGAAGACGGT 9480

TTAGTTTACA AAGTTAAATT ACGTGAAGAA TCAAAATGGT CTGATGGCAA ACCAGTTACT 9540

GCTGCAGATT ACGTTTACGG TTGGCAACGA ACAGTGGATC CTGCCACTGC TTCAGAATAT 9600

GCCTACATGT TTGAACCAGT AAAAAATGCT GAAAAAATTT CTAAAGGGGA ACTACCTAAA 9660

GAAGAATTGG GCATTAAAGC AaTCAATGAT CATGAATTAG AAATCACTCT AGAAACAGCA 9720

ACACCATATT TTGACGATTT ATTGGCTTTC CCTTCTTTCT TGCCGCAACG TCAAGATATC 9780

GTTGAACGTT TTGGTAAaGA TTATACaAAG AGTAGCGATA AAGCAGTCTA CAaTGGTCCC 9840

TTTACGCTAA CTGAGTTTGA TGGTCCCGGA ACAGATACTA AATGGTCTCT AACTAAAAAT 9900

GAAGAATATT GGGACAAAGA GACGGTCAAG TTAGATAAAG TCGCTATCAA CGTGGTGAAA 9960

GAAGCCCCAA CTGCCTTAAA TCTTTATGAA ACTGGTGAAG TGGATGATAC GTATTTATCT 10020

GGCGAACTGG CTCAACAAAT GCAAAACTCG CCTGACTTGG TCCAATTAAA AGCCGCTTCT 10080

TCTTTCTATT TAGAAATGAA TCAAGCAGAT GAAAAATCAC CATTGACTAA TGCAAACTTA 10140

CGTCGTGCTA TGTCTTATGC CATCGATCGC GATTCATTAG CTAAAAATAT TTTAGCTAAT 10200

GGCTCTCTTC CTTCACAAGG CTTCGTTCCT GTGGATGTCG CGAAATCACC AAAAACGGGT 10260

GAAGACTTTG TTAAAGAAGC CGGCAGCGAC AAATTAGTCA AATACGACAA GAAAAAAGCT 10320

GTGGAATACT GGAACAAAGC GAAACAAGAA CTTGGTGTTT CCAACTTAAC GGTTGATTTA 10380

ATGGTAGACG ACTCTGAAGG CGCTAAAAAA ATGGGCGAAT ATCTTCAAGG ATCACTATCT 10440

GATACTTTGG AAGGCTTAAA AGTAACTGTG ACGCCTGTCC CTATGGCTGT TCGCTTAGAT 10500

CGTACCTTAA AAGGGGATTT CCAAATCGCT GTTCGTGGTT GGAGTGCCGA CTATTCAGAC 10560

CCAATTAACT TCTTAGATTT ATTAGAAAGC TCAACTTCTA ATAACCGTGG ACGTTACAGC 10620

AATCCTGAAT ACGATAAATT CATTGCTGCG TCCAAAACCA CAGATGTTAA CGATCCTGAA 10680

AAACGCTGGG AAGATCTAAT CAACGCTGAA AAAACAGTAA TTGCTGATAT GGGTGTTGTG 10740

CCAATTTACC AAAAAGCAGA ATCACACTTA CGTGCACCAA ACGTGAAAGA AATTATTTAT 10800

CATCCAACAG GTGCTAAATA CGACTTCAAG TGGGCGTATA AAGAATAATC ACTCGATTAA 10860

ATTTAGAAAA AGATTTGGAA CAATAGCGAT TTTGGCTATT GTTCCAAaTC TTTTTTATAT 10920

AATCCCGATT AAAGAACCTA ACAACCCAAA AATCACAGTA AAAATAATTA AAGCAATCGG 10980

CGTGCGTCCT TTTTTCAACA TAATAAAACA AAACAGCGTA AACACTAAAG GTAATAAATT 11040

GGGCATGATT GTATCAAACA ATTGTGTTTG AAGGTCTAAC TTAGTGTCTT TCAAGCGAAA 11100

AACAAAAGTT GTCGTTAATT GTACACTCTT CGGAATCAAC GCGCCCACCA CGGAAATTCC 11160

TACAATTGTT GCTGAACGAG ATAATTTTTT AGTACTTTCG CTTAAAACTG AAATTGCTTT 11220

GACACCCATT TTATAACCTA GATTAATTAA CCCAATTTTG GCAGCCATAT ACGGAATATT 11280

AAATAACAAC AAGAAAACAA TTGGTCCTAA GATATTGCCT TCTAATGATA GTTGTGAACC 11340

AATCGCTGCT GCAATCGGCA AAACGGTTAA ATAAAATAGC GCATCCCCAA TTCCGCCTAA 11400

CGGGCCCATC AACGCTGTTT TAATCGCTTG AATGGATTTT GGTGATTCTT TTGACTCTTC 11460

CATTGCAATA ATCACACCAT GCATGAACGT CACTAATTGT GGATTCGTGT TCATAAATTG 11520

TAAATGATTT TCCAAAGAAA GAGCTAAGTC TTCTTTATCA CGATGAATCT TTTTCAATCC 11580

TGGAATTAAT GAGTAGGTCC AGCCGCCTGC TTGCATTCTT TCATAATTAA ATGAAGATTG 11640

TAGCAGATAC GAACGGAACA TTATTTTTTT TAACTCTTTT TTTGTTATTA CTTTTTCGGA 11700

TATCATATCA TTATATGTTT CTTCAGATGC CATCTTCTTC AAACCCTCCA TTATTATTTG 11760

ATGCAGTGAC TGTATTATTT CTTTGTGCGT AAAATTCATA TAAAGCAATT GCGACAGCTA 11820

AACCTGCAAC CGCTAATGTA TCTAACCCTA AATAGGCAGC CATAACAAAC CCAAAAATAA 11880

AGAACGGCAC GTAGGATGTT TCCCACATCG TTTTTAATAA CATGGCGAAA CCTACTGCTG 11940

GCATCATCCC ACCTGCAACG CTCAAGCCAT GAATACTCCA TTCGGGTAGA ATAGAAATGA 12000

AACTAGAGGC TTTATCAGCA CCAAAGAAAA TACATAAAAA TACTACAAAA AAATAACAAA 12060

TGAATAAAAC GATTGGTAAA AAACTAGACA TTCTTTCGAT GCCCCTTGTA TCTGCTTTTT 12120

TTGCATAAAC ATCAAATTTA TGCATAATTG GTGACATCGC TGTATATAAC AAGGTAATTA 12180

ACGCTTGCGC TGCAACTGCG AAAGGAACAG CCAAACcTAC GGTCACTTGT GGGTCAGATT 12240

TTGTGATAAT TGCTAACGCT GTTCCAATAA CTCCCCCAAT TACTACGTTA GGTGGTTGTG 12300

CACCGCCGAT TGCTACTGCA CCAATCCAAA CTAACTCTAG CGTTGCTCCA GTGATTAACC 12360

CTGTGTTGAC ATCTCCCAAT ATTAAACCAA CAATCAGCCC TGTAACAATC GGACGATGAA 12420

TATGTGTCTG TAAGACAAGC TTATCAACCC CTGCTATCCC AGCCCAAATT CCTACCAAAA 12480

TTGCCTGTAC TAACATTTTT TTCCTCCTaT TAaCTTCGTC TACTTTTTAT AAACTGTTTC 12540

ACCATTTACT AAGGTTTCTA GCACTTCATA TTTCTCATCA ATAATTGTAA AATTACCTAA 12600

AAGCCCTAGT TTGATTTCTC CCATAGATTC GTTCACATTC AACAAACGAG CTGGATTTTT 12660

CGTGACTGAA TTAATAGCAA TCACTTCTGG AAGTAATGCT TTTTCAACAA GGTTGCGAAC 12720

CATATTATTT AAATGGTTCG TACTTCCAGC TAACTTGCCA TTGGCTAAAC GGACCACATT 12780

TTGTTCATCA ATAACCATTT CGATCCCTTT TTCTGGTTTT GGATAAACGC CTGGTTGACA 12840

CCCTTTGGCC CAGATAGAAT CTGTCACAGC AATTAAATAG TCCTTGCCTT TTAGTTTGGC 12900

TAACACACGT ACTAATGAAT AATCAACATG TACCCCATCA GCAATAATTT CAGCAAACGT 12960

TTCTTCACTA TCCAAAGCAG CAACTACTGC GGTTGGTTTC CGGTGACTGA TATCTTCCAT 13020

TCCATTAAAC GTATGGGTAA AACTTTTTGC ACCTGCTTCG ACAGCTGCCA TTGCTTGTTC 13080

ATAGGTTGCA GCCGTATGGC CAATTGAAAT GACTACATCA TGTTCATGAC AATATGTAGT 13140

CAAGGCATTC TCTACATCAT TTTCTGGTGC CAACGTCATT AATTTTATGT GCCCTTTTGC 13200

TAACTCTTGC CATTTTTTAA ATGTCTCAAC GGATGGTTTT ACTAATAACT CTGGATTATG 13260

AGAGCCACGA AATTCTTCAG AAATCATTGG TCCTTCCGCA TGAATCCCTA AGATTTGTGC 13320

TCCGTTGGTT CCTTGGTCTT CCTCAATAAA TGAACCAATC ACTTCAAATG AATGTTCCAA 13380

GTCTTTTGGA AATGTTGTTG ACGTTGTTGG CAAAAAAGCT GTAATCCCTT CTTCTGGAAG 13440

GTAAGCTTGC CACTCTTTGA TAAATTCATG ATTGGCATGA TTGGCATCCC CACCATGCCA 13500

ACCATGATCG TGAATATCAA TAAATCCTGG AAGAATGCGC TTTTCTCCGT ATTCTTTGTC 13560

GACAGCTTTT TCATTGTACG GTAGAATTGC AACAATTCGT TGTTCTGACA CTTCAACTTG 13620

AGCTGGTTGA AATTGTTCGT TAATCCATAC GTTTTTACTT TGAATAATCA TTCTTCCCAC 13680

TTCTTCTCTT CAAAATTATT GTCCAACATT ACTTTTTTAT GAAAATAATG CATCTAAACT 13740

TTCTTCTGGT GTTGACGGTA CACGTTGAAT CGAAAGTTTA ACACCTAATT CTTGTAGCTT 13800

TTTAAAGGCG TCTACATCTG TTTTATTAAC GGCAACAGCG GTAGCTACTT GTtTTTGcCT 13860

TCTTCCATAT GCATATTACC AATGTTTAAT TTTTCAATTG GCACACCACC TTCTACCAAG 13920

CGCAATGCAT CAACAGGATT ATCAACTAAC AGAAAAATCT TTTGTCTTTC TGAGGCTTTA 13980

TGAATGACCT CAATCGTTTT CTTAATGGAA AAAAAACGGG TTTGTACTTC TGCAGGGGCC 14040

GCCATGTTTA ACAATTGTTG GCGCAGTGTA TCCTCAGAAA TTGCATCATT TGCGACTAAA 14100

ATTAAGTTGG CGCCTAAAGA ACCGTTCCAT TGTGTTGCGA CTTGACCATG AATCAAGCGA 14160

TTATCAATTC TTGTTAATAA AATATTCGGC ATAGTGCTTC CTCCTCTTTT CTAATTGGTT 14220

AGATTCCGTC TGCTTCTTCT TGAGTATTTG CTTGTGGTTG TTCTAGTTTT TTATGAACAA 14280

TCCCATCTCT GGCTGAATCT GTGACCATTG AAACAAAATC TTCTACGTTC GTGCTGATTG 14340

TGCGTAATGC AAGCCCTTCA AGAACCATCG GTAAATTTGT TCCTGCTAAC ACTTCGACAT 14400

TTGGATAATT ACTTGCGACA GTCATTGAAA TTCTAAACGG TGTGCCACCT AAAAGGTCAG 14460

TTAAAATGAC CACTGCTGGA CACGTCTCCA GCATAGAGTC AACGGCTTTA CGAATGTTTG 14520

ACTCTAATAC ATCAAGCCCT GTATCATTGA AAAAAGGGAC TGCCTGAACA TTTGCTTGCT 14580

CTCCTGCAAT CATACTAACT GATTGATGCA TTCCAGTTGA AAATACCCCA TGCCCCGTTA 14640

AAATAAGACC TATCATTTAT TCTTTCTCCC CTAACTATTT AsTrmGCTTA CTATATATTG 14700

GATAAGCAGA ACTTATTTCT CAAAGAGTGT AACAATTCTT GTGACAATTG AGTATGCTTT 14760

CCCATGATTT TGGACGATGa TTTCTATTTA AGAACATGTC CCTATACAGA TACAAGGCGT 14820

CCCAACTTAA AGATGCTTTT GAAAAAATGC TCGACTACAA CCGTGAGTCC AAACGAGCCA 14880

ACAAATGTAC CCATATACAA CACAAGAATA ACCAAGCTGA CTGGTTGATt GGTTAAAAAC 14940

GGGTGGACCA GTTTCAAGAC CACTTGAAAA ATAAAAACAT TCGCTAAATA ACTGCGATAC 15000

GCATAAgTGG CTATCCATTT AACCATTGGT AAAAGACGAA TCTTTTTGGC TATTCCTTGc 15060

GTTCCCAAAT AAAAAACCAA CCCGATAATC AATAAACTAT AGGCACTTTG AAGAGGATCC 15120

AAATAAGGTA GATCGTTCGG AGTCATCTGT GTTCCATCAA ATAAATTCCC TTGGACTGCC 15180

TGCCAAAAAA TAAAaAGaAA CAACAGAAAT AGTGCCACTC TCGTCTTTTT CAaGAAAGTa 15240

AAAATTTTCC CTTGATACTT CGCTGCGATG ACCCCGAAAA TACCATAAAT GAGATAACTA 15300

AGCCCTGAAC GATCTATTAA GTACCAAAGA GAAGTTTCTA CATATGGAAA AATCATTTTC 15360

TGGTACCAGA AGCACCAAAG GATATGCGCA CCGATTGTAA AAAAGACAAT TGGTAAAATC 15420

ACTTTTTTGT TTTTTGTAAT TCTGGTTGCC AACCAAATAA AATACGGCAT AAATAATTGG 15480

AATTGCAACA TCATTACGGC ATACCATAGA TGTGCCGCGC CATCACCACT GAGAACTTTT 15540

CCTAGAAATC CTAGGGACGT TTCATAGGGC ATTCGTTGTT GTAAACTAGG AAAAGCGTAT 15600

AGATAGAGAC TGGACCAGAC CACGTAAGGA AGCACCAATT CATAAAATTT TTCCTTTAAA 15660

AATTCAAGAT ACGATTCAGC TTGACTCTTT AACACCATGT TGTAAATAAT TGCAAAAATA 15720

AACATGGGTG CCGAATATTT AGACATT 15747

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1536 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

TGCGTGAAAA GTATAAAGCT TTTGAACTGC CAnATAAGGT AGGAACTTCC CAGCAAAATA 60

ACCTnGTAAT TGACCATAAG CAAACAATGG CGAAACATTT CTTGAAAAAA ATTGTAACTT 120

GTAAAATTCT TCATAAGGAT CCCCAAAATC CCATCGATTA AAATCAATCA CACCGACCTG 180

CCTAGTTGGC AAGTAAATAA AATTTCCCAC ATGGAAATCT CCATGCTGAT AAACCACAGG 240

CCGTTGTTGC AATAACGGTA ATTGTTTTTC TAAATATGAA ATTATTGGCT GCTGGTACGG 300

CATACAAAAT TGAGATTCCT TATACTTATT TAGTTGGCTC TGCTTTTTTT CGTATAAATT 360

TTGCGCAATC TCTCTTTGAC TAATCCTTTC AGGCAACAAA AGCTTATGGA TTTTGTTTAA 420

ATAGCGGCCT GCCTCTACGC CTAAGTTCAG TTGTCTTTTG GGAGAAAGAG CGGGTAAAAC 480

ATCGCTTAAG TCTTCACCTT CTAGATACGT TAATAACATG TATCCTTGTT CACTTTCAGG 540

AATAGGTGCA CATAAAAAAG GTTTAGAACT GGGAAATCCT AACGCCGCTA CTTTTTTGAT 600

AAATTGATAT TCTTGTTGTT TGGCTAATAA CTCTTCTTGT TGAAAAATTC GCAATAAAAA 660

GGTTTCTTGA TTTTTGGTAA CTTTAAACTT TAAATCAGTT GACCAACCAC GGTGAATAGC 720

TTCAACTTTT TGCCATTGCT GAAACGCTGG AATCTTCTCA AACGCCCTCA TTACTTTCGC 780

CTCCATTTCC CTATTGGTTT TAGTATACCA AAGAAAAAAG GCTGGAGACA ATTGGGGATA 840

AATTGTCTCT AGCCTTCACT GGTTATCATT TTCTAGTTTT CCTTATATAA AtCGTTCCAC 900

TCGTTACTAA GAAAGCTCCT CCTAAAACAA CCCAGAAAGT ACTAGGTACT TCTGCACCTG 960

CTTTTGGTAA TACTTTTTTG GGAAGCTGTT CTTTTTTCCC AATCATATTT TTATCAGAAA 1020

AAGCATGAGT ACTTTCTGTT GATTTGGTTC CCTGATTTGC TGATGAATTC AGCGAATTAT 1080

CTAAGGATGA CTCACTAGTA GTTTTTCCAG ATTTGTCTGG GTTTTGGTTT TGGTTTTGGT 1140

TTTGATCTTT TCCTTCGATT GGTCCTTCAC TTGTTTTTGT TTCTTTGTCT TTATCTGGTT 1200

CGCCACTTGT ATCTAGTTCT TTATCTTTAT CTGGTTCGCC ACTTGCATCT GGGTCTTTAT 1260

CTTTGTCTGG TTCGCCACTT GCATCTGGGT CTTTATCTTT ATCCAGTTCG CCACTTGCAT 1320

TTGGTTCTTT GTCTTTGTCT GGTTCGCCAC TTGTATCTGG TTCTTTGTCT TtGTCTGGkT 1380

CTTCTACATG TGGTTGATTT ATCTTGTGGT TATTTTCAAA TAAAGCACCT AATTGTAGCA 1440

ATGTTCGGTC TTCATTTAAT GCACTATTAA ATTGAATCCC CAAAGGCAGT CCTTCTTTTG 1500

AAACGTAAGT TGGTACACTT AATGCTGGAT GTCCAA 1536

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17250 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

GCCAAATTTG GTGGATCGTG TCCGCTTGGG AnnGCCTTAT GGGTGGACCC TACGTGCCAT 60

TTGGGTGAAA ACTAGGCCAT ATTTTAAAnG GAATATCCGT GCTTTCCTTG GCCACCGTTC 120

ATAATCCATT TCCACTGAAC CGCAAAAAGA ATCGTATCCA GAGAGATGGC AAGGCCAnAC 180

AGAGAAaTTT TyCTTkGGaA CAACTTCAAG AAACAATTCT TGAAAAAGCC AATCAACAAG 240

TAGGCAGTCG CCGGaCGACT AAAGAAGTCA TTAAATTTAC TACTAAAAAT AAAATTGGAG 300

ATAAAATGAC TCTTGGTACG AAGAGCTTGT TAATTTTGCT ACTTTGTATG ATTTTTTATG 360

TAAGTCTTAC CTTTGTAAAG GGTATTTCAT TAACAACAGC TTTCATTCAT CCAGCGGAAA 420

CAATTCAACG TATATTTGTT CCTGAGAAAG TAGCAGGGAA GGATACAAAT GCTGTGAAAa 480

CGACTAAAAC TAAACCAGCT TCAACAAAAG AATCTAGGCA GCAGGAGGAA GTTCAGCCTA 540

GCCCAACGAA CGTACCAGAA AATAATAATA GCGAACAAGC TGTGTCAACG TATACTGTGG 600

AAGTTGGAGA TTCAGTCAGT TTGATTGCTG AAAATCACGG ACTCACTATA GAACAACTTC 660

AAACTTTGAA TCCAGAAATC ATTGAGGTAC CTATTTATCC TGGACAAGTG TTGAAATTGA 720

AAGAGGTGAC TGAATAATGC GCGAGGTAAA ATTAATTTTA GCTGaACAGT TTTCaTACTT 780

AGGCsTTGCa CAACGAATTG CTAAATACAG TAATAAAGCA TTATATCAAa GCCACATCTT 840

aGgAAATTTA TGGCAAATTT TAAGTCCTTT AATTCAGTTA GGTGTCTATT ATTTTGCTTT 900

TGGGATTGCC CTTGGTGGTG CTCGAACCGT AAAAGGAGGA GTATCCTATA TTGCTTGGTT 960

GATGGTAGGA TTAAGTACAT GGATTTTTTT AAGTACAGTA ACAAAACAAG CTTCTAGTTC 1020

TGTTTATATG CAAGTTGGCA TGGTCTCTAG AATGAAGTTT CCAATAAGTA TTTTGCCAAT 1080

GGTTAAAATA TTTTCTGAGT TACCAAGTTA TTTTGCATTC ACTACGTTAG CAGCGATTGT 1140

GTCAATCTCG ACAGGTGAAA AGATTTCTAT TTATTGGATT CAGCTACCTT ACTATATTGT 1200

AGCGATGATT ATCTTTTTAT ATAGTTTTAG TCTAATTAAT TCAACTATTG CTGCATTAAT 1260

TAGGGATTAT CAAATTTTCT TGAATTCTAT TATCCAAGTA TTAATGTATA TGTCAGGGGT 1320

CTTTTGGGAT TTGAGTACAA AGAATTTACC GTCTTGGTTA TCAAAAATTC TTATGTTGAA 1380

TCCTTACGCA TATATTATTA ATGGATTTCG AGATACATTT TTCTATAAaa GATGGTTTTT 1440

TGAAGATAAG CCCCAGCTTG TTGTTTTTTG GCTAACAACA TTAGTAATAT TTATCGTAGG 1500

AGCTCATTTA CATGTTAAGT TCAGAAGTAA ATTTGTTGAT TATATGTAAC ATGTGAAGAA 1560

GTAAAACCAC CGTATACTGA CAAGAATACG GTGGTTTTAC TCTTTAATTG ACAATAGCCT 1620

TCTCTTTAAA GAGAAGGCTA TCTAACGAAT CAATATTATT ATTTTAAAGT TGGATCAGAC 1680

TATTTCTAGT ATTCAAGTAT TTTTCATTTA AGCGATGTAT TCCCTACCTC TAATACATCC 1740

TTGTATCTCT CTAGACACGC CAgTTTTGAA TAAAATCATC AAAACAAyTG AATGGTTCGT 1800

TTAAAAATAT TTTTCGTGTT CTAATTATTA TAACATAAAA CTCGTCAACT GACCACTAGC 1860

TGgACGAGTt CATAATTTkT CCAGATAAAT AGACAGCCGT TTTCCGCCGA CTGTCTAAGA 1920

TTAAGATAGA TTTGTAAAGA AGTGgAATAA ATGgCTTCCT ACTATCTCtT TATAACAkGA 1980

GGrAAAgAAA ACkGCaAAaG ATAGCTGTGT TTTTTGAAAA AAAGATCAGC TGGTTTCCGC 2040

CAGCTGATCA TCCGTTGAAA AGAGATATAA AGAAGTGAAT AAAGCTTGCT TTACCTCTAT 2100

TTTATACCAT GTTTTGATTC AAATAACTAT TCAGTATTAG TTATACTTAA ATAATAAATA 2160

TGTATACGGA TGTATTGTTT TGAGGTATAG TCATAGAGAA AAGAGGGAAC GTTTATGAAA 2220

AAATTTTGTT TAAGTTATAG CTCTGGTAAA GATTGTTTAT TGGCGATGGA TCGATTGGTA 2280

CAAGCAGGAA ATCAGCCAGT TGCGTTAGTG ACAACATTAA GTGATGAAAT TAATCGTTCT 2340

TGGTTTCACG GCATACCAAT CTCCGTCTTA GAAGCAGCAG CCGAAGCGTT AGATTTACCA 2400

CTTGTTATAT CACATAACAA TGAAACAAAT TACACAGAAA AAGTGGTTGA GGCACTGCAG 2460

GAAACCAAAA AATTAGGTGC AGAAACGGTT TGTTTTGGCG ATATAGATAT TGAACAAAAT 2520

GGTGCTTGGG ATCGCCAAGT TGCGTTAAGT GCTGGACTTG AACCACAATT ACCTCTTTGG 2580

CAGGAAAATC GCGAacACTT GTGAAAGAAT TCTTGGCAAA AGGGTACACA GCAATTATCA 2640

AAACGGTCAG CAAAGAGGCA GGCATTCCAA TAAAATTTTT AGGAGAACCG TTAAATGAAA 2700

CATTTATCAC GTACCTTAAA GAGCATCAAT TAGATATTTG CGGAGAAAaC GGGGAGTACC 2760

ATACGTTGGT AATTGACGGC CcGCTATTTA AAAAGAGGCy CAtTATTATm CTcAGGAATT 2820

aTGAAgcCTT ATGCTaTyCT TgATtATtGA TGCTTAAaTG AAaGTAAaaG GACACAAGCT 2880

GCcACTTTTt CTAGGGATAT gCTATAaTTG mCTtAaGAaA AGTAGAGAAA GTGAGTGGTT 2940

AATTATGGAA GAAACTGTCT TTTTTAATTT AGGAAATGCC CTTGCTTCAA AAAGAGATCA 3000

AAAAGAATTA ATTAAAGAAG CCCAAATTGC GAAAGATACT CGAAAGATTC CAGGAAAACT 3060

GATTATCGTT GAAGATGAAG AAAATGGTAC CCATATTTTA TTTGAATTAG CTGATCAAAC 3120

AGAACCGTCT GCTGAAACAA AAGAATTCAA AGTGAAAAAA GTGATTGATT AATAAAAAAG 3180

CAACGAATTT TTCTAAATTC GTTGCTTTTT TATATTTTTT TACACAAATG ATACAAAAAG 3240

AATATCAAAA TCAAATAAAT GATTAGTATA CTTGAAGaAA CAAATAGGAA GCAAAGCAAA 3300

CGTTTTGAGG TGTTCTTTCA TGGAAAATCG AGGAATTTTA TTTGATAAAG ATGGTACGTG 3360

TATTCGTTTT GACACTCTTT GGCAAGCAGG TTTGCAAGCT TGTTTTGAAA CACTAAGTAT 3420

GTTaGCCCCT CATCATTCAG CAGAAATAAA AAAgATATTA GCTATTCAGG AGCAACGTTT 3480

TTTGCAAAAA CATTTACTTG ATGAAGTCCT TTATCAGGAA CTTTATCAGG AATTGGCGCA 3540

ATTTGAGGAA TTAGTCGAAC AGGGAATCAG CAGTCGATGG CTGGAGCAAT TTTTTTATGA 3600

TTATTTACGA AAAAATCTGA AAAAGATCGA ACCGATTGGT GATTTAAAAC AGTTATTTCT 3660

TGAGCTGAAA CGGAAGAACT ATAAAATTGG ATTAGCAACT TCAGATACTT TGCCAGCGAC 3720

TATGTTGATT ATGGAATATC TTGGTTTAAC AGAAATGTTT GATTTTATTG CGACAGGAGA 3780

TCGTTACTTA CCGAAACCAG ATGCGGACAT GCTCCAaGCC TTTTGTCAGT CATGTCAATT 3840

GAAGGCGACA GAAGTAATTA TGGTAGGTGA TTCGCTCGTG GATGTTTTTA TGGGAACGTG 3900

TCATGGCAAA GCAGGAATTG GTGTTTTAAC CGGCAACTGC CAGTCAACTG ATTTTGAAAA 3960

GTTTGAAGTA GCCTATTTCC GTGATATCCA TGAAATACCT TATCAAGAAT TATGGGAAAA 4020

TACCAAGAAG AAAAAAATCT ATTATTTTTG ATATATGTAT TGTGAAAAAC GTTTCTTTTA 4080

TTCGAGTATC TAAGCAAGCA TACTAGTGTA TTTCAATAAA ATACCATGAT GTTCATAGGT 4140

TTTATAATTT CAGCTCACAG ATTTTTGTTC AATAAGAAAA AAATTACAAC TAAAATAGAA 4200

AACGGGCGTG GGACAAAAAT CACTTTGGAT TTTTGTCCCA CGCTCAAAAA CTGATAAACG 4260

GCGGGAACAG AAGCAACACT TTCGGAAATA AGCCGAAAGT GACTTTCTCA CACAAGTGTT 4320

CGGGATATCT TAATCCTTAC CTACGTGGTG CTCGGAGTTA AACACTTCTG TCCTGACTTC 4380

GTTGTTAATA TTTTACAATG ACAAGTTTTT CTTATTTTAA TCCTGAACTT GATTTAGAAA 4440

TTGCACTTCC ATTAGAAAAA GAAACAGTAA CAGTAGCACC TAAATCACCA TCTAAGTTTT 4500

TCATCCAGAT TAAAGTATCA TTTTTTTCAC CATTGATATT TGTAGAAGTA ATCCCTGTTG 4560

GGTCTCCTAA ATCTTTTCTA GCTTGTTCTT CAGAATATGA ACCATCAGTA GCAATGTTAT 4620

TTACTTGGTC TGCAGTAACT TTATCATGTT TTGCAACCTT CAAGCCACTT ACAGATTTGC 4680

TCGCTGCTTT GCCATCAGAG AATGAAACAG TAATAGATGC aAGAAGATCA CCACCCTTGA 4740

CATTTGTCCA AGATAGAGTA GTGGTAGAGA TGCCTTGGAT ATCTGTTGTA GAAGAAGAGG 4800

CTGGTTCACC TAAAATTGCT TTTACTTCAT CTTCTGTAGA ACCGCCTTCG CTGCTGTTCA 4860

TAATATCACC GACATTAATT TTATCGTAAG TkGCTTTTAA ATCAGAGTTA TCTTTCGATT 4920

TATCTTCTGr TTTCTTTTCA GAGCTTtCTT TTTTTGTATC ATTTTTTGAA CTATTCGTAG 4980

CTTTTGAAGA AGAGGCTTCT GTTTTCACTT CTTTGGTATC ATTGCTTGCT TTTTTATCCT 5040

CTTTGCCAAG ACCACTACCA GCAATCCCCA GTACAATAAC TACTACAACC CAAAACCAAA 5100

TTTTCTTATA AAATGGTTTT TTTACTTTTC CTTCATTTGA ATCATTTTTT TTCATTTTCA 5160

CTTTCCCTCC CATAGTTTAG TTAATGATTT GTAGCGCTTA CTAAATAGTT GAAAACAATC 5220

CCCAATTAAA TAACTTTGGA CTACAAATCA CTTAGTTTAA ATGAAAATAA GTATACCGCC 5280

ATGTGTAAAA ATAACACACC ACAGTATACT TATTTTATCA TTATTTCATA AAAATAACAC 5340

CATTATTTTT ATGAAATAAT TAAAAAAGAT AAAAAATCTA AAAACTTCGA CCGCCGCCAC 5400

CGAATGTGCC GCCGCCAGTT GAGTGGGTGG TGCTACCACC ACCGCCCATT CCGCCACTGC 5460

CGCCATTGTT TTTAGGAATA CGACGCGTAG TGATGAAAGA GTTGGTTAAC TGATCTGTGC 5520

GGGAGGTTAA GTTTAAAGTT GTTTTTTCCC TAAAGGGATA TTGATAACTT GAAAATTTTA 5580

ATTGATATTT AGAAATATTA ATGCCTAAGA AGACCAAACT GAGTATCAGC GCAgCAGCAA 5640

AAGCAATTAC CATTTCCAGC GGGGTAATGA CTTTATAACG AGTGATTTTA CCTGTTTCGC 5700

TGTCCACACG ATAGTGCCCC CCAGGAACCC CTTTATTAAC AAATGCTTGA GTTTCCTGAA 5760

CAAAGGTTTG AGCAGCCGCG AAATAATTTC CTTGACTCAT ATTATCCCAT ATTTTATCTA 5820

AGGTATCATC AATTCGTGCA TCTGTCATAT AATCAATCAT GTTTCCAGAA GTAGAGATGT 5880

AGATTTTCCG TAAGTCCATA TCAATGAGAA AAAGAATCGC ATTTTGGTCC TTGCCAACTT 5940

TATTTAAAAG ATAATGATCT GCATATTCTT GTTCATCGCC ATAGGTATTA TTATTTGTGG 6000

TTACAATAAA GACAGAGGCT TTTGTTTTTT CTTCTAAAGG TTGTATCTCT TGCTTTAGTT 6060

GGTTGATTTG ATCAGGTGTC AGTAATTGTG CGCCATCATC AATTGAATTT TCAGCCGCCG 6120

AAACGGGTAG GGCAAGGCCA AAGGTAAGAA GTATTAGGAA AAAAATAGGT AAAAGTCGTT 6180

TTTTCATAGG AAATAACCTC CTAATAAAAA GAGTGCTAAG AGTCCAGCGA AAATGCCTAA 6240

CGTTGCTAGC CCTAAACGAG TGTAACTAAT TGGAAGTACA CCGCTTACTT TACCTGTTTG 6300

TCCATTCATG GCATAATAAT AAACTTTTTT TGATTGTTCC TGACTGCGAT AGGTGACTAA 6360

CCAAATTGGC AAAAGCATAT AGTGATTATT TTCTTGGTTT AGAGAAATAT CCGTGCGAAC 6420

ATTTGTCAAC GTAGTATAAC CAGAAGCAGA GTCCCGTAAC AATGATTCCG AATAATCTTG 6480

TAATTCTTGT TGGACGTGTT TTTTGATTGC TTCATATTCA ATATCGCGTT TTTCTGCTTG 6540

GAAACCTGCT AGGTATTGGC TTTTGAAAGG CACCGCTTGA TCAATTAAAA ATGGTTGAAC 6600

GCCTTCCACC ATTTTTTGTT GGACGTTTTT TGACAAGGCA TTTTTAACTA ATTCTTTAAA 6660

CGATAATTTT CCTGAACGTC CCACATTGAA ACGTTTTGTT TCGGTGTATT CAATATCACC 6720

TACGCGCCAA ATGCGAATAC TTGTGCCAGT AGCTTGCAGT TGTCCGTCGA CTTCTGCATC 6780

AACGACCCAA TAGGGGAAGT AAACCCCAGT TAATTTATCG ATTTGATTTT TGCTGAAAAA 6840

ATCTCGGGGA ATGAACCATT TTTTCTTGGT CCAAGCGAGA AATTTTTCAA TCGCTTCGTC 6900

TTTTTCAATT GTAAAGGGTA AGACATTTTC TGGTAAAAAT TCGCCACTTA ATCGACCAGC 6960

CAAGACAACT GGATTATGGC AATAATAACA GTATGTGGCA GCCGTTGTTG CATCGGTGAC 7020

AATTTGAGCG CCACAGCTTG GACAATTAAA AATTTCCATT GTGCCTACAT CTGCAGTTGT 7080

TTCATTCTCA GTGGTCTCTG TTTGTTGCGT ACGGAgGATT CTTCTTCCGC TGTGAAAGTC 7140

AAATTTTCCT CGGCAGTCGT TGTTGAGGGA TTCTCTAAAT GCGCCTCATG TTGCTGTTGC 7200

TCATAACGAG TCACTTCTTC TTCCGTGTAA ATATTTAAAC AATATTCACA ATGAAATTTT 7260

TGATCTTTGG GATTAAATAA GAGCGGGCCG CCACAGTTGG GGCATTTATG GGTAAATGTT 7320

GTTTCCATTG CAACCGTCCT TTCTAATAGA GTGAGTAAAT AACAGGAATC TGATTCAAAG 7380

AAAAAGGGTA GACTTGTGCC TGACCCTTTT CTTTAATTGT TCACTACTTG CTTATAAAGG 7440

AACCCCTTGG AATGGTTTAC CACAATGTGG ACAGAACTTA GGAATACCAT TACTTAAGTC 7500

AACAATTTCG CTACATTCAC TACATTTCAT TTCTAATTTT GGTTTCTCGT TTGTTGGTTT 7560

TGGTGTACCA CAGTTTGAAC AAAACTTCCC AGTGTTTTCA GTCCCACATT GTGGGcAAGT 7620

CCAAGTGTCT GTAGCCCCTT GTGCGTTTTG CTGGTTTTGT TTTTCTTGCA TTTGTTGTTG 7680

GTTTGTTTGA GACGctTGCG CAAAGTAGTT GCCATTCGCA TTCATGCCCA TACCAACGCC 7740

CATAAAGCCG CTCATAGCGC CACCTTCATT TTTACCAGCA GCCTCCATAC CACGAGCCAA 7800

GGAACCTTGA ACATAGCCTT CGCGAACACC TGGATCGCTG AGCATAGCAC CTTCATTTCG 7860

CATGTTGATT AATTTTGTTG AATCATCTGT ATAAGAAATG CTCGCAACTG CTACAGAAAC 7920

AATTTCCATG CCGCGATTGG TGCGCCAATC CTCATCAAGC GCAGTTTGCA TATATTTGCT 7980

TAATTCAACA CTTTTTGAGG GAACATAAGA AATACGTTGC CCATCAGCAG ACATTTGATT 8040

GATTGCTGCT TGCAATGCTG TAATGAACTC TGCCAAATAT TGTTCATTAA TATCATTAAT 8100

ATCTACTTGC GTTTGGTTTT TAGGAATCGC ATTAGTATAA AATAGCAAAG GATCTGTAAT 8160

TTTAATTGAG TAATTTCCGT GTGCACGTAA GAAAAGTTCG GCATTATAAA AATTATCAAA 8220

ATAGTTTAAT GGCGCAGGTG TTCCAAATTT AATGCCTTTA ATTTCTTGCA AATTAATATA 8280

AAAAACTTGT TGTTTTTGTG GAGTTACGCC ACCAAATTTG AAACGACTGA AGGTCTCTGC 8340

AATCGCATCT TTTAAAGAAC CGTTAAACAT TGAAGGCGCA GAATCATTTT TGACTGTATA 8400

ATAGCCTTCT TCAGCAGTGT AATCGATGAT TTTGCCACCA TCGACAAGAA GCATCATCAT 8460

GTTGGGATAG ACATGTACAA CAGAACCATC TGTTAAAACA TCGTCCGTTC CTTTACGATT 8520

```
AGAACCACGT TTATCATCTT TACGTACAAA AACACCTTTT GTCATGACGG TTGTGTCCCC    8580
CATATTATCA GGTTCAATAA CCTCTAGCCA TTGATCGGCC AAGCCACCGC CAACTGAACT    8640
TGTTGCTGCT TTAATAAGTC CCATGTTCAT TCCTCCTAAT AAAAAACTAA TTGTTAAAGC    8700
TAGTATACCA CACAATGAAT GCAGAGAAAT CCAACGAAAG CCTGATGAAA TGGAAAAAAT    8760
TGATCAAACT TTATTAAGAA ATTAAAAAAA GTATAGGTGA AACACAAGCA TTTTAAAAAA    8820
AGTTACGTAT AATAGAAGAA AAGACATATT ATATAGTTAG GGGATTATTG AATGGAACGA    8880
AGCAATCGTA ATAAAAAATC TTCAAAAAAA CCACTTATTC TTGGTGTTTC TGCCTTGGTT    8940
CTAATCGCTG CTGCCGGTGG CGGGTATTAT GCTTATAGTC AATGGCAAGC CAAACAAGAA    9000
TTAGCCGAAG CGAAGAAAAC AGCTACTACA TTTTTAAACG TATTGTCAAA ACAGGAATTT    9060
GATAAGTTAC CGTCCGTTGT TCAAGAAGCT AGCTTAAAGA AAAATGGCTA TGATACTAAA    9120
TCTGTTGTTG AAAAATACCA AGCAATTTAT TCAGGGATTC AAGCAGAAGG AGTCAAAGCT    9180
AGTGATGTTC AAGTCAAAAA GGCGAAAGAC AATCAATACA CATTTACCTA TAAATTATCG    9240
ATGAGCACGC CTTTAGGCGA AATGAAAGAT TTGTCTTATC AATCAAGTAT CGCCAAAAAA    9300
GGCGATACCT ACCAAATCGC TTGGAAGCCA TCTTTAATTT TTCCAGATAT GTCAGGAAAT    9360
GATAAAATTT CGATTCAAGT AGATAATGCC AAACGTGGAG AAATTGTCGA TCGTAATGGT    9420
AGTGGGCTAG CAATTAACAA AGTGTTTGAC GAAGTGGGCG TAGTGCCTGG CAAACTCGGT    9480
TCTGGCGCAG AAAAAACAGC CAATATCAAA GCTTTTAGTG ATAAATTCGG CGTTTCTGTT    9540
GATGAAATCA ATCAAAAGTT AAGCCAAGGA TGGGTCCAAG CAGACTCCTT TGTACCAATC    9600
ACAGTCGCTT CTGAACCAGT GACAGAATTA CCAACAGGGG CTGCGACAAA AGATACAGAG    9660
TCACGTTATT ATCCGCTGGG GGAAGCAsTG CGCAATTAAT CGGGTATACG GGCACCATTA    9720
CTGCAGAAGA TATTGAAAAA AATCCAGAGT TAAGTAGTAC GGGCGTAATT GGTAAAACTG    9780
GCTTAGAACG AGCGTTTGAT AAAGAATTAA GAGGGCAAGA TGGTGGCTCA CTAGTTATTT    9840
TAGATGACAA AGAGAaTGTC AAAAAAGCTT TACAAACCAA AGAGAAAAAA GATGGTCAAA    9900
CGATTAAATT AACGATTGAC AGTGGCGTAC AACAACAAGC TTTTGCTATA TTTGACAAAC    9960
GCCCtGGTAG TGCGGTCATT ACCGATCCTC AAAAAGGCGA TTTATTAGCA ACGGTTAGTT   10020
CACCTTCGTA TGATCCTAAT AAAATGGCCA ATGGCATTTC TCAAAAAGAA TATGATGCTT   10080
ACAATAACAA CAAAGATTTA CCATTCACAG CACGTTTTGC GACAGGCTAT GCACCAGGCT   10140
CTACCTTTAA AACAATTACC GGTGCCATTG GTTTAGATGC AGGGACATTG AAGCCAGATG   10200
AAGAGCTAGA AATTAATGGT TTGAAATGGC AAAAAGATAA ATCTTGGGGC GGCTATTTTG   10260
CGACACGTGT AAAAGAAGCA AGTCCAGTCA ATCTGCGTAC CGCTTTAGTT AATTCAGATA   10320
ATATTTATTT TGCCCAACAA ACATTACGGA TGGGAGAAGA CAAATTTAGA GCGGGTCTTA   10380
ATAAATTCAT TTTCGGTGAG GAACTAGATT TACCAATTGC CATGACGCCT GCTCAAATTT   10440
CAAATGAAGA TAAATTTAAT TCTGAAATTC TTTTAGCAGA TACGGGCTAC GGTCAAGGTC   10500
```

AATTGTTAAT TTCGCCAATC CAGCAAGCTA CCATGTATAG TGTTTTCCAA AATAATGGAA 10560

CATTAGTCTA TCCAAAACTA GTTTTAGACA AAGAAACGAA GAAAAAAGAC AATGTAATCA 10620

GCGCCAACGC AGCGAATACA ATTGCCACAG ATCTTTTAGG TAGTGTGGAA GATCCTTCAG 10680

GTTATGTTTA TAATATGTAC AATCCGAACT TCTCATTGGC TGCTAAAACA GGAACAGCTG 10740

AAATAAAAGA TAAACAAGAT ACAGATGGCA AAGAAAACAG TTTCCTTCTA ACGTTAGATC 10800

GTAGTAACAA TAAATTCTTA ACAATGATTA TGGTTGAAAA TTCAGGAGAA AATGGTTCTG 10860

CAACAGATAT CAGTAAACCA TTAATTGATT ACTTAGAAGC AaCCATTAaA TAAaAAGAGa 10920

AAATGAaCGA AcGAGGTTtC TGTAAAAAGA AcCTCGTTTT CCCTTTTTAG AAAAAGGAAA 10980

GTGTGCTACA ATGAAGCGAC TAAATTGTAG AAGGAAAGGG ATTCGTGTGC GAATTTTAGC 11040

AATAGATACA TCTAACCAAA CATTGAGTAT TGCTGTGTGT GAAAATCAAA AAATTCTGGG 11100

AAGTTATACA GCAACGGTTA AAAGAAATCA TAGCTTAACA TTAATGCCGG CAATTGACTA 11160

TTTGATGAGC CAATTGAACT TGGCGCCAAC GGCAATTGAT CGTTTTGTCG TAGCAGAAGG 11220

ACCTGGATCA TACACTGGCT TACGTTTAGG CGTGACGACT GCCAAAACAT TGGCCTATAC 11280

TTTAAAGAAA GAGCTAGTAG GCATTTCCAG TCTACAAACG TTAGCTGCCA ATTGTGTTGG 11340

ACAAACTGGT TTAATTGTGC CTCTTTTTGA TGCGCGTCGT AAAAACGTCT ATGCTGGTGC 11400

TTATCGGTTT GTCAATGGCG TGTGGCAAAA TGAGTTACCG GACCAACATA TTTcCCTGCG 11460

GGAGTTATTA GAACAACTAA AAAATGAACC AAATCTTTTT TTTGTTGGTG AAGATGTTGA 11520

AAAATTCACT GAAGAAATAG CTCAAATCAT TCCTCATGGA GAAATTTGTG ATGTTCCACA 11580

ATGGCAAATT CCTAATGCCG CAGTTTTGGC TGCGTTAGGT AGTGTGGCTG AACCTGTTGA 11640

GAATGTTCAT GGGTTCTTAC CTGCGTATTT GAAAAAAGTA GAAGCGGAAG AAAATTGGTT 11700

GAAGACACAC ACGCCTAATG GAGAAAGCTA TGTGGAAAAA CTTTAATTTT GTGACAAAAT 11760

ACTTTACAAA AAGACGGCAA CGGTATATAA CTAAAACCAT TGCAATGAAC CAGCAAGAAT 11820

TTCAATTGCG TGAAATCACC TACCAAGATA TTAAGGATTT ACTAGCAATT GAACGTGAAG 11880

TTTATGCTGG GGAACTTCCT TGGACGATGT CTGCTTTCAT GGCAGAACTT GGCTCAAAAG 11940

CGCCGCATCT CTATTTGTTA GCGGCAATTA ACGGAAAAAC AGTTGGCTTT ATTGGTTGTC 12000

GGATTCAAGG AACGGATGCA CACATTACGA ATGTCGCTGT TCACACAGCA TACCAAGGCT 12060

TAGGTTTAGG TCGTTCTTTA ATAGAAGAAA CGCGAATTTT TGCCAAGAAA AATCGGTGTG 12120

AAACCTTATC GCTAGAAGTA CGAATGAGTA ATGTTAAAGC CCAACGTTTA TATCGAAAAA 12180

TTGGCTTTGT CTCACAGACG GTGAAAAAAG GCTATTATGA CGAAAATAAT GAAGACGCAC 12240

TAGAAATGAT AATTGATTTA AAAGAAGAGT GAAGAAAAAA TGTTAGTAAC AAAAAAAGAA 12300

ATATCTGTAG CAAATATTGG TGAAGAATTA TGGCGTTTTA GTAATTCTTC TTATACAACT 12360

GGCTCCCCTT GGaCAGAAAC ACAATTCGCG GAAGATTTTG CTCAGGaAAA TAGCGAATAT 12420

CTTTTTTTAG TTGAAAATGG CCAGTGGTTA GGCTATATTG CCTATCATTT TATTTTAGAT 12480

GAAGCAGAAA TTAGTCACGT TGTTGTGAAT GGTCAGAAAC AACACCAAGG AATTGGTTGT 12540

ccAGTTAATG aAAGCTTTTA AAGAATATGT TAAAAGTCgT GACATTACAC AAATTTTcTT 12600

AGAGGTCAGA GAAyCGAATA TACTAGCyCm AAAACtTTAk GrGAAAaCAG GTTTtCGGAA 12660

AGTAGCTGTT CGAnAAAATT ATTATAAAGA ACCGCAAGAA AATGCCTTTG TAATGTGTGC 12720

GAAATTAAGG AAGGAAGCCC AATGACAATT TTTACAAAAG AACGTAAATT ATTATTAGCT 12780

GTGGAAAGTA GCTGTGATGA AACAAGTGTG GCTGTCATTG AAGATGGCGA CAAAATTCTT 12840

TCCAATATCG TAGCTTCGcA AATTAAGAGC CATCAACGTT TTGGCGGAGT CGTTCCGGAA 12900

GTTGCGAGTC GTCATCATGT TGAACAAGTG ACGATCTGTA TTGAAGAAGC ATTAACAGAA 12960

GCAAAAGTCA CACCAGAAGA ACTCAGTGGC GTTGCTGTGA CCTATGGGCC TGGGTTAGTC 13020

GGTGCTTTAT TAATAGGACT TTCTGCAGCA AAAGCATTTG CTTGGGCGCA CCAATTACCA 13080

CTTATTCCAG TCAATCATAT GGCTGGGCAC ATTTATGCTG CCCGTTTTGT GGCACCGCTC 13140

GAATTTCCAT TGATGGCCTT ACTCGTTAGT GGAGGACATA CGGAATTGGT CTACATGAAA 13200

GAAGATGGCT CTTTCGAAAT TGTTGGTGAA ACACGTGATG ATGCAGCTGG TGAAGCGTAT 13260

GATAAAGTCG GACGTGTGTT AGGGTTGCCT TATCCTAGTG GCAAAGAGAT TGATGCTTTG 13320

GCCCATGAAG GCACAGATAC CTATCAATTT CCACGGGCGA TGCTAAAAGA GGATAACTAT 13380

GACTTTAGCT TTAGTGGTTT GAAAAGTGCG TTTATTAACA CGGTTCATAA TGCAGAGCAA 13440

CGTGGTGAAG CACTGTCTAC AAAGGATTTA GCCGCaAGTT TyCAGGCGAG TGTGGTAGAA 13500

GTATTAGTGA CTAAAACCAT TCGTGCTTGT CAAGAATATC CTGTTAAGCA ATTGCTAATC 13560

GCTGGGGGCG TTGCTGCGAA TCAAGGATTA CGAGAAGCAA TGAGACATGC AATCAGTGaA 13620

CAATTACCAG CAGTGACATT ACTTATTCCG CCATTAAAAC TTTGTGGGGA TAACGCGGCG 13680

ATGATTGGCG CTGCCGCCTT TATCGAAGCT GcAAAAGAAT CACTTTGCTT CGTATAACTT 13740

AAATGCAGAA CCTGGCGTTA GTTTTATGAC AATTAGTGAA GAAGGCTAGC CAGCTAGACA 13800

AAAATTTGCT ATAATGAAGC CAGTTTAACA ATGACATTCG ATGATAAAGA GAGTACTTAT 13860

TTCAACGTCA GCTTAGCGAT CTTGGGAAGG TGGGAGCCAA GAGTGATCGG AATGAGGAAG 13920

CGCACTTTTA GAATTCTGCT GAAAGCAGAC GGCAATCACC GTTATCGATG AAGCAGGTCT 13980

ATCCAAGTAG TATAGACAAT TAGAGTGGTA CCGCGGGTAA ATTCkGGGnT CTGACAAAAC 14040

AAATTATTTT GTTTTGTCAG AGACGTTTTt GTTTTTTTAA TACTTACAAA TTAnTCAGGA 14100

GGGTTCTTAT GAaCAACAAA GAAATCGTAG CAAAAGCGTT ACATGATGtT TTAAATGAAG 14160

AATTAACGAT GGATCAAATC GAACAATTAT TGGAAAaTCC TAAATCTGTT GATCATGGAG 14220

ATGTTGCTTT TCCAGCCTTT TCATTAGCTA AAATTTATCG TAAAGCACCA CAACAAATTG 14280

CAGCAGAATT AGCAGAAAAA ATTGATGGCA CAAATTTTGA AAAAATTGAA GTGGTGGGAC 14340

CTTATTTAAA CTTCTTCATG AATAAAGAAG CAGTCAGTCA AGCTGTCATT GGCGAAGTAG 14400

TCAAAGAAAA AAATAACTAC GGAAACAGTA CCTTTGGCAA TAATGGCAAT GTTCCTATTG 14460

ATATGTCTTC ACCAAACATT GCTAAACCAA TTTCTATGGG GCACTTACGT TCTACGGTAA 14520

TCGGGAATTC AATTGCCTTT ATTTTAGAAA AAATTGGGTA TCAGCCAATT CGGATTAACC 14580

ACTTAGGTGA CTGGGGTACA CAGTTTGGTA AGTTAATTGT AGCTTATAAA AAATGGGGTT 14640

CTGAAGAAGC AGTCCGCCAA CAACCAATTA ATGAATTATT GCGCTTATAT GTGCAATTTC 14700

ATGAAGAAGC AGAAGAAAAA CCAGAATTAG AAGACGAAGC ACGTGCATGG TTTAAAAAAT 14760

TAGAAGAAGG AGACCAAGAA GCTAATGAAT TATGGAAATG GTTCCGTAGC GAGTCTCTAA 14820

AAGAATTTGA TAAAATCTAT TCCATGTTAG AAGTTGAATT TGATTCTTAT AATGGCGAAg 14880

cyTTTTATAA TGATAAAATG GATGAAATTG TTACTTTATT AGAAGAAAAA CATTTATTAA 14940

CAGAAAATCA AGGTGCAGAA ATTGTTGATT TGACAGAATA TAATTTAAAT CCAGCACTTA 15000

TTCGTAAATC AGATGGTGCG ACGTTATATA TCACACGGGA TTTAgCTgcT GCGCTTTACC 15060

GTAAACGGAC GTATGATTTT GCCAAATCGC TTTATGTAGT CGGCAACGAA CAAAGCAATC 15120

ACTTCAAACA ATTAAAAGCA GTCTTGAAAG AACTAGGGTT TGACTGGTCA GATGATATGG 15180

AACACATTCC GTTTGGTTTA ATTACTCAAG GCGGTAAAAA ATTATCCACA CGTAAAGGGA 15240

AAATCGTTTT ATTAGAAGAA GTATTGAATG AAGCTGTTAC TTTAGCAGGA AATCAAATTA 15300

ACGAGAAAAA CCCAGACTTA GCAAATCGTG AAGAAGTGGC GAAACAAGTA GGTGTAGGTG 15360

CAGTTATTTT CCATGACCTT AAAAATGATC GTCTAAATAA TTTCGATTTT GTATTAGATG 15420

AAGTTGTTCG TTTTGAAGGA GAAACTGGGC CATATGTTCA ATATACACAT GCTCGTGCGA 15480

TGAGTATTTT ACGTAAAGCA AACTTTACAC CAGACGCTAC ACAACGTTAT GCACTAAATG 15540

ACAAGGATAG CTGGGAAGTC GTGAAGTTGT TACAAAAATT CCCTGAAACT GTGATGCAAG 15600

CTGCTGAAAA ATATGAACCA TCTGTTATTG CGAAACATAG CATTCATTTA GCACAAGCGT 15660

TCAATAAATA TTATGCCCAT GTTCGGATTT TGGATGAAGA TGCTCAAAAA GAGGCACGCT 15720

TAGCGTTAGT TTATGCAGTA GCTACTGTTC TAAAAGAAGA TTTACGTTTA CTAGGGTTAC 15780

ATGCACCAGA AGAAATGTAG AAGTTTTTGT GAAAATTTTA CTGCGTTGAA CGCTAAACTT 15840

ATTAAGCAAA TATTAAAGTT AAGTTACAGA TTGAGAAATT CATAGTAATT AAATAGCTGG 15900

GGTGCTATAC TCGAGTCAAG TTATAAAAGG ATTAGAAAAG GTGGCACCCA ATTATATGAC 15960

AGAAAAATGG CGTGAAGACG aAGAATATCT TTCTTATGTA GAAGATTTAT TAGcAACCGA 16020

AGAAGTCCAA AAATTAGCAG AATACACACA ACATATGCAT TCAACTCGTT TAGAACATTC 16080

AATTAGTGTT TCCTATCATA GCTATAAATT AGCAAAGCGT TGGAACGGCA ATGCTCGTGC 16140

AACAGCTCGT GCTGGTTTAT TGCATGATTT GTTTTACTAT GATTGGCGCA CAACAAAGTT 16200

TGACGAAGGC TCTCATGCCT ATGTTCaTCC aCGAATTGCA GCCAAAAATG CTGAAAAAAT 16260

TACAGAGTTG TCTGATTTAG AACGAGACAT TATTATTAAA CATATGTGGG GCGCAACTAT 16320

TGCACCACCA AAaTACAAAG AAAGCTACAT TGTAACCTTT GTTGACAAAT ATTGTGCTGT 16380

CAAAGAAGCG gCGTTGCCGA TGACGACAGC CATGAAAACC AAATGGCGTC AGTATTTTGG 16440

TCAACAGTCA ATCTAAAAAA AGCTTATCCG TAAGGATAAG CTTTTTTTGT TTGACGTTGA 16500

CTAAAAAGTC TAGCATTCAT CTATTTAAAA TAAATATTAA AAGTAATTTT GTTCTGTTAA 16560

TTTCTTTTGA GCGTATCGAA AGACTGCTTC TTGATTTTCA GTTGTTAACT GTTTGTAAAT 16620

GTTCATAATA TTGGTTGGGA CAACAAATGG CGTTTCTTCC GTTTCAGGAA TACCTAAGAT 16680

TTCAAGAGGT GATACATCTA AAATTTTAGA TAAGGCGACA ATTTTATCCC TCCCCATATT 16740

TTCAATCATT CCAGTTTCCC ATTTCCTGAC GGTACTTTTT CCAACACCAA CAAAATTTCC 16800

AACTTGTTCT AATGTAAGAT TTTTCTCCAT CCTTTTTTGC TTCAAAAAAT TTTCTTTCAA 16860

AAAAACCACC TCCGTAAAAA CATTCTAGCA CAAAAAGTGT CGCAAAAGAA ACATGTTTTT 16920

tTCAAAATGA AAGTAACGAA TATAATAGnA AAATCAGCAT TTTTTAATGA TATTGACGTT 16980

TTTTTCACAA AATGATATAA TTTTAAGTGT CTTATTAGAC ACTGAAAmGG GGGAGAGTAA 17040

AATGGAATCA AATTTAATTC ACTCAAAATT GaAAATGAAA GGAAAAAATG TTGaGTGGTT 17100

GGCAAATGAA ATGAcTGCaT ACGGGCAACC AATATCCGCT TCAACGATTT ACAAAAAATT 17160

AAAAGGAGAA GTTGTTTTCA AAGCATCGGA TATTAAATTG ATTTCACAAT TGTTGGATTT 17220

GAATAATGAA GAAATTATGG ACTATnTTTT 17250

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7947 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

AGATTAnCGG AAnCTACCTA TCTGAACGTA CAGATAATAA TATTACmATG CAAGCAATTC 60

GTGATCATTA TGAAGTATCC AATCGATTTG GCGCAAAAGA AGTCTTGGCT TTTTCTTCTG 120

AACGAAAATG GGGAGCTATT GAATTTCCAG AAATTGGGAC AGTTTATTTA GGGGCTCCTG 180

AGCGTTTGGT AGACGATAGT CGGTTGCCAG AAGCTGTTTT TACAGCACAA GAGAATGGGT 240

ATCGCGTCTT GATGTTAGCT ATTGCGGAGC AACAACCGTT AAACGAAACC AAAATGCCTT 300

ATTTAGAGCC ATTAGCAATT CTAGAAATTG ATGATCCAAT TCGCCAAAAT GCCAAAGAAA 360

CACTGGCTTA TCTAAAAGAA GAAGGAATCG ACTTAAAAGT AATTTCTGGT GATAATCCGG 420

TGACTGTTTC AAATATTGCT CGTCGTGCTG GGCTACCCGG TTATGAGTCT TATATTGATT 480

TGTCGACCAA AACGACAGAA GCAGAAGTGC GTGAAGCAGT TCAGCAATAC ACAGTGTTCG 540

GACGTGTATC GCCTCAGCAA AAACGAACCA TTGTGCGTGA GTTGAAAGAC ACAGAGCATG 600

TGGTGGCGAT GACCGGCGAT GGCGTTAACG ATGTTTTAGC GTTACGAGAA GCGGATTGTA 660

GTATCGCCAT GGCAGAGGGA GATGGCGCTA CTCGGCAAAT TTCTAATCTA GTTTTGCTAG 720

ATTCTGATTT TACAACGTTA CCAGATGTCT TGTTTGAAGG ACGACGTGTG GTAAATAATG 780

TAACCCGAGT TTCCAGCGTT TTCTTCATTA AAACGATTTA TTCGTTTATT CTCTCAATCA 840

TTTGTGCCTT AACGGCCATT GCGTTCCCGT TCATTCCGAT TCAAGTGACC TTGATTGATT 900

TAGCAATTGA AGGATATCCA GCCTTCTTCT TATCTTTCGA AGGGGACAAA CGAAAAGTTG 960

TCGGGAAGTT TTTACCAACA GCTTTGAAAA ATGCGTCTGT TAATGCGCTT TTAGTTGTAG 1020

CGAATATCAT TGCTGTTTAT CTGATTGGAC AAAATCAAGG ATTTTCTTCT TTAGACACCA 1080

CGACGTTAAT GTATTATTTA TTAGTTGGAA TTAGCTGTAT GGCTGTTGTT CGAGCATGTC 1140

TGCCTTTAAA TCCACTACGC ATTTTCTTGG TTTTCAGTAC CATTATTGGT ATTTATGTAG 1200

CAGCTATGTT ATTCCACAAC ATTTTAGAAA TTGGTTTCTT AACTTCGCAA ACAATGGGTC 1260

TCTTCTTTAT TATGATGGCC ATTAATATTG TGGTGCGTGT AACAATTGGT TTCGTACAAA 1320

TGAAACGTGC TGGAAAAACA ATTAAAGATT TATAAAAAGC AGAGGTTGAG CTAAAAGCAT 1380

TCGTCTCCAG AAATGCAGTT CTTCTTTAAA TCTTGAAGGA AGTCGGAGCA AGCTTGATGT 1440

TTAACTGATT ATTTTAAGAG TTAGTGTCTG GGGCAGGAGT CAGGAAGACT TTTGTCCTAG 1500

ATACTTTCTT TTTTTGTTGA GTTGAAACAA GATACTTGAA AAACAGCGAT TATGTGATAC 1560

AATAAATCAT GTGTAAATGT GTCCATTGGA ATTCGCACTG AAATGACAGT GCGTTTTTTA 1620

AACGGTCAAT CACATGAGTT AGCACAATGC TATTCTTAGG AGGTGAGGAC AATGGCACAA 1680

CGCATTCCTC AGGAAGTCAT TGAGGAAGTT CGTCACCGAA CGAATATAGT AGATATTATT 1740

GGGCAGTATG TCCAGTTAAA AAAATCCGGT AAAAATTATA TGGGTTTATG TCCGTTTCAC 1800

GAAGAACGTT CCCCTTCATT TTCAGTAGCA GAAGACAAAC AAATTTTTCA CTGTTTTGGC 1860

TGCGGGAAAG GTGGAACTGT TTTTAATTTC TTGCAAGAAA TTGAAGGCAT TAGTTTTCCA 1920

GAATCCGTGA AACGTGTTGC AGATTTGGAA CATTTATCTG TGGACTTTGA TTGGTCAGAG 1980

CCGCGTGAAG TAGCGGATAC CCCAGAAAAC CAACAAAGAC GGAGTTTGTT ACAACTGCAT 2040

AGTAAAGCCG CAGAACTTTA CCATCATATT TTAGTGAATA CTAAAATTGG CGAACCCGCA 2100

TTAAACTATT TATTAGAGCG GGGGTTAACA CAAGAACTAA TTGAGACCTT TCAAATTGGA 2160

TTTGcGCCGC AAAAAcGTGa TTTCTTAAGT CAAGTTTTTA AAAACGAACA ACTGGATGAA 2220

ACGCTTTTTG AGCCTTCTGG GCTGTTTGTC CAAAGAGATA ACGGTACgTT TTTGGaTCGT 2280

TTCTATCAAC GAATCATGTT TCCAATTAAT GATCCCCAAG GCAATGTCAT CGCCTTTTCT 2340

GGTCGCTTGT TaAAAACAGC TGATTTTCCa GGGGATGAGA tGCCTAAATA CTTGAATAGT 2400

CCTGAAACAa CACTTTTTAA TAAGCGAGAA ACGCTGTTTA ACTTTGATAG AGCCAGAAAA 2460

GAAATTCGTA AAGAAAATAC GGTCTTGCTT TTTGAAGGAT TCATGGATGT TATCGCTGCT 2520

TGGCAATCAG GCGTAAAAAG TGGGGTGGCT TCAATGGGGA CTAGCTTGAC CAATGAACAA 2580

ATTCGGCGTT TGGAACGGGT CGCTAAGGAA GTAGTTATTT GTTACGATGG CGATAATGCC 2640

GGTGTTCAAG CCACAAACCG TGCGATTCAA TTATTGCAAG AAAACAGTCA TTTTGACTTG 2700

AGTATTGTCA GCATCCCTGA AAAGCTGGAT CCTGATGAAT ATGTGCGGAA ATATGGCGCA 2760

GAAGCTTTTC AAAATTTAGC CAATCATGGC CGAGAAACAG TTTTTAGTTT TAAAATGAAT 2820

TATCATCGGT TAACTAGAAA CATGAATAAC GAAAAAGAAC AATTGGATTA TGTGAACGAA 2880

TTGTTACGGG AGTTAACCAA TGTTCAATCT CCCTTGGAAC GTGATCGTTA TTTAAACCAA 2940

ATTGCTCAGG AATTTCAACT TTCTGTTCAT AGTTTGGAAG AGCAATTCAA TCAATTGAAG 3000

CAAGAGCAAC GTTCAGTCCA ACGGCAAGAA AGGCAACAAT TTTATCAAGA TGAAATGATG 3060

CCACCACCAA TGGAAGAACC TGTTTTTGAA GAGAATCACG TCCAGAACAA GTTACCGTTA 3120

ACCCAAGTTC AAAAAGCAGA ACGTTCTTTA TTATTTCGTT TAATGAATGA ACAAGGAGTA 3180

CGGCAAACCA TTCAACAATT GCCAGATTTT TCATTCGCTC ATGATGAATA CCAAGAATTA 3240

TATTTCTTGC TAGAAAGTTA TGCGACACTG CATCAAAGTT TCGACATTGC CGATTTTATT 3300

AATTTCTTGC AAGATAATCA AACCaAACAA TTAGCAATTG AAATTGCGTA TCAGAATTTA 3360

TCGGAAGAAA GCTCTGAGCG AGAAGTTGCC GATTTATTAC ATGTAATTGC CTTATcAAGT 3420

ATTGCAGAGG CAATTGAGCA GAAAAAGATT CAACAGCAAG AAGCAAAACG AGTGGGCAAC 3480

CAACAGCTTG AAGCCGAATT AACAATGGAG ATTATCCAAT TGGCTCGACA GCTTAAAGCT 3540

CAACGAACAT TTACTTAAGC AATATAATGA ATAGAACTAA GCGTTTTGTT CATTTTATTT 3600

TGATAAGTCC CAGTGATTTA TCAGTTATTT CAAGCGTGAA CCGAAAAATG AAGGGGGCCT 3660

TCTTTCATGG AAAAAGAAAC AAGTAAAAAA TATGAAGCTG CAGTGGCAGC ATTTATCAAA 3720

GAAAACAAGC CTAAAGGAAC AGTGGTTTAT GATGATTTAG CGAATCAATT AGCCACACCG 3780

TATACTTTGG ATGCAGAAGC TATGGAAAAA TTGATCCaAA AAGTTGAAGA TGCTGGTATC 3840

AGCGTGGTCG ATGAAAATGG AGACCCAAGT GAACATAGCT TGAAAAAAGA TGAAAAAGAA 3900

GCGGAAAAGG CCCAAGCAGA AGACTTATCT GCGCCAACAG GTGTTAAAAT TAATGATCCC 3960

GTGCGTATGT ACCTAAAAGA AATTGGGCGC GTTCAATTAT TAACTGCCGC AGAAGAAGTT 4020

GAATTGGCAC TTAAAATTGA AGAAGGCGAT CAAGAAGCAA AACAACGTTT AGCAGAAGCT 4080

AACTTACGTT TGGTTGTTTC AATTGCCAAA CGATATGTAG GTCGTGGTAT GCAaTTCTTG 4140

GATTTAATTC AAGAAGGTAA TATGGGCTTA ATGAAAGCTG TTGAAAAATT TGACTACCGT 4200

AAAGGATTCA AATTCTCTAC GTATGCTACT TGGTGGATTC GTCAAGCGAT TACGCGGGCA 4260

ATTGCGGACC AAGCTAGAAC GATTCGAATT CCTGTTCACA TGGTTGAAAC AATCAATAAA 4320

TTGATTCGGA TTCAACGCCA ACTATTGCAA GATTTAGGTA GAGAACCAAC GCCAGAAGAA 4380

ATTGGTGCGG AAATGGATTT ACCAACAGAA AAAGTTCGTG AAATCCTAAA AATCGCACAA 4440

GAGCCAGTCT CTTTAGAAAC ACCAATTGGT GAAGAAGATG ATTCACATTT AGGTGATTTC 4500

ATTGAAGACC AAGATGCTAC CAGTCCTGCT GAACATGCAG CTTACGAATT GTTAAAAGAA 4560

CAACTAGAAG ATGTTCTGGA TACTTTAACA GACCGTGAAG AAAATGTTTT ACGTTTACGT 4620

TTTGGTTTAG ATGATGGTCG GACACGTACG CTAGAAGAAG TCGGAAAAGT TTTCGGTGTC 4680

ACTCGCGAAC GTATTCGTCA AATTGAAGCA AAAGCGTTAA GAAAACTAAG ACATCCTTCT 4740

CGTTCAAAAC AATTAAAAGA CTTTTTAGAA TAAACATACC GTTTAGgGGC TCTTTTCGTG 4800

CATGCGAGAA GAGCCGCTTT TGATTTTTAA AAAGTCTCTA AATGTCATTG CTAGTTATTG 4860

GAAGAAGGCG CAAGTTGTGG TAAATTAGAA GTAATGGAAT ATTAGCAGAA TGTGAGGCGG 4920

AAGAATGACA AAAAAATGTC CGAAGTGCGG GAACGAATTC GATGCAGAAT TAACGACCTG 4980

CCCAACATGT GGCTATTCAC TAACTGACAC AACAGTTGAT AAAGAAGAAG CGGAAACGAC 5040

TTCAACAAAT ATAGATTTTG AAACACAGGA AAATGAAGAA CACGAAGATC AATTAAATGA 5100

GAATATTGAG TGGTCTGAGT TAAAAGATAT GAGCTTAGGC CATGTGATGG AACTGTTTGG 5160

CGAATCACCA GAAGAAGAAA GTAATGATGA CAAAAAAGAA GAGTCGACAG AAGATAACTT 5220

AATAGTCTCT GATTCAGAGG ATGTTTCTGG TCTAGAAGCA TCGCTTCAAG AAGGGGCTAG 5280

TGAGGAGACT CATGATTCGG TGGAGGAGTC GATACCAGCG ACAGAAGAGA CGCCTACTCA 5340

TTCAACAGAA GAAAAACTAG CAACGGACGA AGCAGTTAAT TTAGAGGAAA CAACAGAAGA 5400

AACAACGGAA GAAACAACTA CTGTTGAAGC AGAAACAGCT GAGGTATCTG AGACGGTGAA 5460

GTCAGAAGAG GAAGCTTTAA CAGAGATTTC AGGAACTGAA GTTATCTCTA CCACTTCTGA 5520

AGAAGAAATT TTTTCACAAC CACCGATAGA AGATCAAGAT GTAACACCAA ACGAAACTTT 5580

GCAAGCTTAT ATTCAAGCGC ATCGAGCAGA TACAGAAATG TCTGAAAATC CATCAGAAGA 5640

AACAGCAGAG ACCCAAGAGT TAGAAAACAG CGGTGAGGCA GTATTAACCC AAGCCGAAAC 5700

ACCAACTGAA TCCATTTCAG ATTCTGAAGA AGGGTTAACA AGTGCAGCGT CGCTAGAGGC 5760

ACCGAGCGAC GCTGATGCAA CCGAGGGAAA AGTTTCAGAT TCAGGTGTGA TACCACCGAT 5820

GAATGAAACT GGGAATGCTC AACCAGCGCC TGCTCCGAAG AAACCATCAA AAAAAGTAGT 5880

TTTCGTTGCG CTTGCGGTTG TTTTATTAGC TGGAGGTAGT GCTTGGGCGT ATCATGATCA 5940

AACGCAAAAA GCTGCCGCGC AAGAGGCGGC GGCACTAACC AAGAAAACAG ATACGCTGAA 6000

AGATGAACTA GCAGCCTTTT ATACCACGAA GGAGCAAGTT TTTATTAAAC CCGATATGGT 6060

GACTGTTAGT CCAGAAAAAC TCTCAAAACA AGTCGCGGAA ATTAAAGATT CAGAAGAGTA 6120

TTCTCAATTA AATAAACAAA TTCAAACATT GAAAGAGAAG CAACAAACCA TTCAACAAAT 6180

CAACCAATTA TTTGAGGCGC CGATTGTCAA TGGCAATGAG TTGAAACCGG CGATTCTAGC 6240

TGCTGATCAA CCAATTTCCG TAAAGAAATT AACAGGAAAT GATCCATTTG ACCAATTGAT 6300

GAACCAAGCA ATCGATCAaG CGAACCaACA ATACAACCAA TTACAAAAAG CTAAAAAAGC 6360

GGTTGAGGTC ATTTATAAAG ATGGAAAAAC TACCAATCAG CTGAATCGAG ACACGTATCA 6420

AGCTGCTAAG GCAGAAGTAG ATAAAGTAAC AAGTGATAAG CTGAAAAAAG AACTGGTCAA 6480

ACAAGTCACG ACAGCTGACC AAGCGTTAAC CAAGGTTGAA GAAGAACAGA AAAGAATTGC 6540

AGAAGAACAA GCTGCCGCAG AGCAAGCAAA ACAAGCAGAA GAACAAGCGA AACAAGCTGc 6600

TGCCGCAAAA AAAGAGGACG CTAAGAAGGA AGAAACAGCA AAAACAGAAG CGAATGGCTA 6660

CACGGCGCCG AATAGTGACG GGGTTTACAC TAGTCCGCTT TATGCGCCAG ATGCTGCCGA 6720

TATCGCTGAT AGTAGCAATC CAGCGTGGAC ATGGGCACCA GGTGTGAAAG AAAAAGTCTT 6780

AGACACAGTG ATTGCTCGTG GCTATGTGGT TCCTGGCGGA TATTCTTTAG AACCTGCCAA 6840

AATTGTGAAT GGCGAAGGTT ACTACAATCT TTATGCAACA AACAATCAAT CAAAATTATT 6900

AGAAGGCACT ACAGAGAAAA ATGTTCACAT GTATTTAGTT ACAATCAATG CTAAAACAGG 6960

TTGGTTCAAA GGGAACGCTT CTCGTAATGC AGGGCAATAA AAGAAAATAG AAACAATCGA 7020

ACAAAAGCGG AAATCTTTTG TTCGATTGAT TTTCCTTGAG GAGTTAATGA TAAATGAGAG 7080

AATTACTAGC AACAACATTT ACCGGCATTG TGGTCGATGA AAATGAACAA TTTTATTTTG 7140

TTCAAAAAAA TGGCATCACA TTCCGTTTAA AAAAAGAAGA AGGTACGCAT GCTATAGGGG 7200

AGGCTGTAGA AGGGTTTGGT TATTTAAATC AAAAACAAGA GCCAGCACTG ACAACTACAA 7260

TTCCTACTGT GCGCATTGGT AGCTATGGCT TTGGTACTGT CACAGGAACG AGACGTGATT 7320

TAGGTGTCTT TGTAGACGTT GGCTTACCAG ATAAAGATGT CGTTATTTCG TTGGATGAAT 7380

TACCAACTAT GCGGGAACTT TGGCCTAAAA AAGAAGACCG AGTAATGGTT ACTTTAAAAG 7440

TGGACAGTAA AGACCGCATT TGGGGCGAaT TAGCAGnTGA nAAAGTCTTT AAAGCAATGG 7500

CAAAACGTGG GAATGCTGAA ATGCAAAACC AAAACTTAAC CGGGCATGTG TATCGTCTGA 7560

AACTTGCAGG AACGTATATT TTAACTGATG ATTTATACAT TGGATTTATC CATCCTTCTG 7620

AACGATTCCA GGAACCACGT TTAGGTGAAA AAGTAGTGGG GCGTGTGATT GGCGTGCGGC 7680

CGGACGGCGT ATTGAATCTT TCATTAAAAC CAAGAAGTCA TGAAGTAATC AGCGATGATG 7740

CGTTAATGAT TCTAACTTTC TTAGAAAGAG CAAACGATCA CCAAATTCCT TTCACCGATA 7800

AATCAAGCCC TGATGAAATT AAACAAACCT TTGGCATTAG TAAAGCTCAA TTTAAACGCG 7860

CCTTAGGCCA TTTAATGaAA CAAAAATTAA TTAAACAAGA AgATGGTAAA aCAATTCTAA 7920

TTGGcTCGTT GGrACAATCm AAAGAAA 7947

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7865 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

GCACCAGCGT GGTTGGATCA TTCTGATCAC CCGTTTCATT GGCCCGATAC TGGTGTTTAT 60

TTATATGAAC CTGGCCAAGA AGTTGTTTTT CAATTCGATC AATTGGTAAC AGATATTCGT 120

GAATGACCCA GAAACCAGTT AAAAAGCAAT TGCGTTTAGG GTTAACTACT TTATTTGCTG 180

TACAATTTtAT GmAAGAAaTT TCACGATTTT TAaCAACCCA TCCACATGTG AaTTTAATAT 240

TACAGCArGA CGGCTCGCCA AAaTTACAAa CGATGTTAGC AAATAAAGAA ATTGATATGG 300

GACTAATTTC TTTCCCAAAT ACCTTACCTG AAATTATTCA TATTGAACCT TTAGAAACGA 360

CTACCAAAGG CTACCATGTT TATGTCGTAG TTCCAGAATC AAATCCTCTC TCCCAATATG 420

AAAAATTAAC TTTTAAAGAT TTGAAAGATC AACgATTTTC ATCGTTAAGT GATAATTTTA 480

TGATTGGTCG CCTACTACTA GATCGGACTC GGAGCTTCGG TTATGAGCCG AATATCATTT 540

TGCATAACGA TGATTTACAA GTACTTCTTT ATAGTTTGCA AAAAAATAAT TCGATTTGTT 600

TGCTGCCGAT TGAATATTAT GAAGTGGGAA AAAGTCAGGG ACTGAAATGG ATTCCtTTAA 660

AAGATAAGTT TGACTATTTC CCCaTCGGCA TTGCTTTGCG CCGCGATTTT AGTCTGACAG 720

AAGATGTTAG AGATTTTATC CaAATCATTA AAGAAAATTA ATGGATAATC TAGCCTTGAT 780

AAGCAAAGAC GATTTTACTA TCATAAATTT TCAACACTTT TTCAAATAAG AATTTTGAGA 840

ATAAACCCGC GACAATTGGA ACAACAAACC AAGAAAAGAG TAAGGGTAAG ATCCCTAAGC 900

CCGCATCCAA TGAAGCAAGT GGTCCCACGA GTCCAACCAA GCCAAATCCA GCCGATTGTG 960

GGGTTCCTGA AATGTTGAAT AGAACCACAG GAATTGCTGA AATAGTTGCT GTAAATAAAC 1020

AAGGAATTAA AATAATTGGG TATTTAAATA AGTTAGGCAT CATCATTTTC ATGCCGCCCA 1080

AAGAGACAGC CAGTGTCACT CCCGATTGGT TGACATTCCA AGAATTAATC ACAAGAACAA 1140

CAGTTGTGGC TGCAATTCCC ATTGCAGCAG CACCTGCAGA CAAGCCATTC AACTGAATCG 1200

CCAAACCAAT TGCCACAGTA GTGATTGGCG AAATAATCAG TGCAGCAAAT GAACAAGCAA 1260

TTAAAATACT CATGATAATT GGCTGAAAAT CAGTAAAATT GTTAATCACT TTCCCGATGG 1320

CAACCGTAAT TTGTGTCACA TATGGATAAA TCAACATGCC AAACAAACCT GCACCTACTC 1380

CAACAACGAT CGGTAAAGCA ATGATTTCCA CTGAACCGAA TTTTTATCG ATGACTAATA 1440

AAAGTAAGAC TGCAACAGAA GCGGTAATCA TAATATTAAT GATGTCACCA GTGCCTGCAC 1500

CAATAAAGCC TTTCACTTCT GaATTAAATT TAATCACGCC AGAGCCAGCA AATGCTGCAC 1560

CGCCAGCAAT CATCATTTTC TGTGGCGTTA AGCCAAATTG AAAAGCmATC AGTCCACCAA 1620

TAATTAAAGG GGTGGCTAGT TGGAAAATTT GCGCCGCATG AATAATCATT TCGATAATCT 1680

TGTATTCTGC AAAATATTTT AAAATGGCTC CTAATACAGC ATTGGGAATT AAGGCGATAA 1740

TGGTGCCGGA AgcTGTTCCA GCTAACACTT TGTTAAAAAA TATTCTTGGG GTTAATTGAG 1800

CATCCAGTTT ATCAGTGGGC ATTTTCAACA CTCCTATTGA TTATTTTTGT TTTTTTAGAA 1860

CAAGTTTGAG AATGATTGAA CTTTCTATTA GATAAAGAGC GGCCTCAAAG AAGAGCGCTT 1920

ATTTCTTTGA GGCGCTTTTT TTTTAACTAA AGCAAGGAAC GATTTTTAAT ATATATCCGT 1980

ATCCGCATTA TATTTCGCTA AATTATCTTT AACATCTCGc AAGAACTTAC CTGcTTGTTG 2040

TCCATCTAAA ATTCGATGAT CAATAGAAAG ACATAAATTA ACCATATCTG CCACTTTAAA 2100

GCCTCCATCT GCGGTTGGCA CAAGGCGTTT ATTAATTGAC TCTACTTGTA ATATCGCAGC 2160

TTGTGGATGG TTAATAATTC CCATCGATTG GACGGAACCT AGTGTTCCAG TATTATTAAG 2220

AGTAAAGGTC CCTCCTTGCA TTTCTTTGCT TGCTAGCGTT CCTTGTCGGA CTTCTTGTGC 2280

CAAGCGATTG ATTTCTTTGG CTAGCCCTGC AATCGAATAG TTATCCGCTT GTTGAATCAC 2340

TGGCACATAA AGATGCTCGT CTGTCGTTAC AGCTATGGAT AAATTAACAT CTTTATGATA 2400

AATAATTGAG CCATCATCCC AAGATGTATT GATTTTTGGA TTTTTCTTTA ACGCTTGAAT 2460

GACTGCTTTG GCAAAAAATG GGAAGAAGCT TAGAGATAGG CCCTCTTGTT GTTTAAATTC 2520

ATCTTTTAAT GAATTTCTAA GTTGGACTAA GTTGGTCACA TCCGCTTCCA CCATTAACCA 2580

AGCGTGAGGG ATTTCGTTGA CGCTTTGGAC CATTTTTTTA GCAATTGCTT TACGTACAGG 2640

ATCTGCAGAG ACAATTTTAT CTGTGCTTGT TTCAGTTGGA CTAGCTGTTG CGGCGCTTTC 2700

ATTTTGCGAA gCAACTGGCT CCTCTGAAAT TGAAGGGCTA GTTCCTGGAC TCACTGTTTT 2760

CTCAGGTGTT CTTGCCTGTG TTGGTGTAAA GTTTGTAACA TCTTTTCTAG TAATTCGACC 2820

ATCACGACCT GTCCCTGTAA CTTGTGTTAA ATCAATCTTT TTTTCTTGGG CAATTTTAAG 2880

GACTGCTGGG GAATAACGAC CATTATTTTT TTGATGGGAC GTCGCAGTGC TTGTTGTTTC 2940

AGCCGTTTCA TGTTCTTGCG CTTGTTCCGC ACTTGCTTCT TTCACTGGTG CTAAAGTAGC 3000

TACTTCTGTT TCTTCTGTTG TTTCTTCTGT CTCTAACGTC ATCACCGCTG TTCCGATGGG 3060

CACATCTGTA TCTAGAGAAA TCAGAAATTC TTTTACTACA CCATCAAAAT CTGATGGGAC 3120

TTCCGTGGTG ACTTTATCAG AAACGACTTC CATTAAGGGA TCATAgCGTT TCACTGAATC 3180

TCCTGGTTTA ACTAACCATT GGACAATGGC TGCCTCTGTG ACACTTTCGC CTAGATGAGG 3240

CATTTTGATT TCTTTTGTGG CCATTTATTT TTTCCTCCTT GCTTCAATTA AAATTCTGCT 3300

AATTCTTTCA TTGCTGCTAA TACTTGCTCT TCATTGATTA AAAATTCTCT TTCTAAAGGT 3360

AATGCATAAG GCATGCTTGG ACAATCAGGT CCTGCAAGTC GTTGAATCGG TGCATCTAAA 3420

TCAAAGAGAG CATCTTCTGA AATCATTGCC GCAATTTCAC TCATCACGCT GCCTTCTTTA 3480

TTGTCTTCTG TTACGAGTAA GACCTTCCCT GTTTTCTTAG CTGCAGCCAC TAATGTTTCG 3540

CGATCTAAAG GATATAATGA GCGAACATCG ACGATTTCGG CATCAATTCC TTCAGCGGCT 3600

AGTTTTTCCG CCGCAGCTAA AGCTAATTGC AAGGTCATGC CGTAACTAAT CACGGTTAAA 3660

TCACTACCCG TTCTAACGAC ATTGGCTTTG TCAATTGGTA CGATATAATC ATCTGCAGGC 3720

ACTTCATCTT TTAGTAAACG ATATAAACGC TTATGTTCgT AAAAAATGAC TGGATCATCG 3780

GAACGAATCG CTGCTTTAAT CATCCCTTTT GCATCATAAG GATTAGAAGG GGTCACCACT 3840

CTCAATCCTG GTTGTCCACA AAAGACTTTT TCTGTAGACT GAGAATGATA CAGCCCACCA 3900

CGAACACCGC CGCCATAAGG GGTCCGATAA ACGATTGGCG CAGTCCAATC GCCTTTTGTT 3960

CTATAGCGCA TTGTCCGAGC CTCTGATAAC AATTGATTGG TTGCAGGCAA AATGTAATCC 4020

GCAAATTGAA ATTCACCAAT TGCGCGATAG CCCATTAAAC CTAAGCCGAC AGCTAAACCA 4080

CCAATTAAGC CTTCTGTTAA CGGTGTATTA AAACAACGCT CGTCACCGTA TTTAGCAGCC 4140

AAGCCCTTTG TTACACCGAA CACGCCGCCT TTGTCGCCGC CGACATCTTC TCCAAAAATA 4200

ACTACTTTTT CATCACGAGC CATTTCTTCA GAAATTCCTA AGTTAATTGC TTCTAAATAA 4260

GTCATCTCAG CCATTATATT TCTTCTCCTC TTCTCTTATT TTGCATACAC TTCTTCTAAA 4320

ATTGATGTCG GTACAGGATC TGGCATTGCT TCTGCTTCAT CCGTTGCTTG ATTGATTTCT 4380

GCACGAATTT CTTCATCAAT TTTGGCAATG TCTTCGTCTG TTAAATAGCC CTCTTCTAAT 4440

AATTGTTTTT CAAAAAGCTT CACTGCATCG TTCTTTTTCA TTTCTTCAAT TTCTTCTTTT 4500

GAACGATAAA CAGATTGATC ATCGTCAGCG GAATGAGACG TCAAGCGCGA AACCATTAAT 4560

TCAATCAATT TTGGTCCTTT TTTCCCGCGA GCCGCTTTTA CTGCTTCTTT AAATGCTAGA 4620

TAGACTTCAG TAAAATCACT ACCATCAACG GTTACACCTT CAAAGCCATA AGCTTTCGCG 4680

CGATCGGCCA TTCGTTTATT GGCATACTGT TCTTCAATTG GGACAGAAAT CGCATATTCA 4740

TTATTTTCAA CAACAAAAAT AACTGGTAAT TTTTTTACGC CTGCAAAGTT CATAGCTTCT 4800

TGGACTTCTC CTTGATTGGC AGAGCCTTCC CCAGTGGTGG TTAATGCAAC AAAATCAGCT 4860

TTTTGAAGTT GCGCTGCATA ACCAACACCT GTCGCTAATG GCATTTGTGT ACTTACTGTT 4920

GAAGAGAAGG AAACAATATT ATGCTCTTTT GAACCATAAT GATTCGGCAT TTGACGACCA 4980

TGGGAAGAAG GATCCGCTTC TTTTCCAAAA GAACCCATTA AAATATCTTT GGAGGTCATG 5040

CCCCAAACCA AGCACGCGGT CATATCACGA TAATACGGTA AAAAATAATC TTTTTGAGGA 5100

TCAAAAGCCA TCGCCATCGC TACTTGTGCA ACTTCTGCCC CTTGACCAGA AATATTGAAA 5160

GAGGTCTTAC CAATCCTTGT TAATTGCCAC AACCGTTCGT CTAAACGaCG tCCTCTTAGT 5220

ACTTcAcGAt ATGCCTGaAT CAaTtCTTCT TTTGATAAAC CTGATTTTTT AAGCGCTTTC 5280

ATTTTCATCA ACCTTTCTCT ATTTGTGAAT CGCTAAGCCA TAAGTATCTA AGGCTGCTTC 5340

TTGCAATACT TCTGTCATTG TTGGATGCGC ATGGATTGCT TCACCAATTT CAATGGGCGC 5400

GGCATCTaAA TACATGGCAG TACTTGCTTC GGCAATTAAA TCCGTGACAT GTGGGCCAAT 5460

CATAGAAACC CCTAATAAAT CATCGGTCTT CTTGTCACGA ATTACTTCTA TAAAGCCATC 5520

TGTTTCTCCA TAAACAAGTG ATTTACCATT ACCATTAAAA TTAAAGGTGC CAATTACTAC 5580

TTCTTTTTCT GCTGGTAAAG TTTCTCTAGT ATAGCCGACA CTTGCTATTT CAGGATTTGT 5640

ATAAACACCT CGAGGTACGT TTGTATAATT TAAAGGTTCA ACTGTCTCGC CTAAAAGATG 5700

CTGAACAGCT AACTCTCCTT CTTTCATAGC CACATGTGCT AGTTGGAGTG TATCGATACA 5760

ATCACCAATC GCATAAATAT GTCCTtCTGT GGTTTGATAA AATTCATTTA CTTCAATGCC 5820

TTTGTCAGTA TATTTCACTG AGGTATTCTG TAACCCTAAC TTATTGATAT TGGGCTGACG 5880

TCCAATTGCT ACCATGACTT TGTCAACGGT GAGGGTTTCT TGTCCTGCAA CTTCTACTTG 5940

AACCTTTTGT CCAGTAACTT TCGCTTCTTG GACTTTGCTT CCTAATAAAA TGTTAATTCC 6000

TCGTTGCTCT AAACGTTTCT TTAATTCTTT AGAAATTGTT GCACTCTCAT TTATAAGTAA 6060

ACGATCCAAA AATTCAATGA TAGTTACGTT GACACCTAAA CTATTTAATA AAGAAGCCCA 6120

CTCGACACCA ATAACGCCAC CACCAATAAT TGCAATTGAT TCTGGTAGCT CTTCTAATTC 6180

AAGCATCCCA TCAGAAGATA AGATAAATTC TTCATCCAAT GGCAAATTAG GTAACGTTTT 6240

AGAACTAGAG CCTGTCGCAA TAATGACATT TTTAGGAcGA TAATTTCTtC TTCTCGTGTT 6300

GGAnCATTAA ACGTAACTGC CACTGCGCCA GAGACTGGTG AAAAAATGGA GGGACCTAAA 6360

ATCGCACCTT CGCCTGCTAA AATTgATTyT ATTTTTCTTG CATAAGCCTT CAACACCTTT 6420

GTGTAACTGC TCAATAATTC CTTCTTTCCG TTGTTGTATT TTAGAAAAAT CAATGGACGC 6480

TGCCTCTGTT TCAATCCCAA AAGAAGCCGC TTGTTTTAAG GTATCAAAAA CCTCTGCGCT 6540

TCTTAATAAA GCTTTTGTAG GGATACAGCC TTTGTGCAAA CAAGTGCCAC CTAATTTGTA 6600

TTTTTCAACA ATTGTGACGT TTAGCCCTTT TTGTGCAGCT CGAATTGCTG CTACATAGCC 6660

TCCTGTTCCG CCACCTAAAA TCAGTAAATC TGTTTGTTCC GCCATTTTGA TCACTCCTGC 6720

TCAATTTTTG CTTCACTATA ATTTAAAGCT TGTTCTTCCC CTGTTAATAC TCGATTGACG 6780

CCTTCATATA AAGCAGCCAT CTCCATTTCA CCTGGGTAAA CCTTAATTGG TGCAATCCAA 6840

GTGACTTTTT GACTAATTTC TTGAACAACG GTTTGCGAAT AgcTGCGCCC CCTGTTAAAA 6900

TAATTGCATC AATTGTGCCT TCAAGAACCA CGGCCATCTC GCCAATACTT TyyGCGATTT 6960

GGTAACACAT TCCTTTTAAA TAGTAATTTG CTGTTTGATC ACCTGCAGCT ATTTGTGCTT 7020

GGATATGCCG TAAATCTGTT TCACCTAGGT ATGATTTAAG ACCGCTATTC CCCGCAATTA 7080

GCTTTTTCAC TTGGCTAATC GTTAACTCCT GTTCAAGTAT CCATTGCGCA AATTCAACGA 7140

GCGGTAAGGC ACCACTGCGT TCTGGCGTAT AAGGaCtTCA CCATCAgCCC ATTTACCACA 7200

TCCACawGCG nCcTTTTgAT GCGCCCCCnA GCTAATGCCC CCACCTAAAT GAACAACGAT 7260

AAAATTGCTT TGCTCATATG TTTTCCCtAA ATCCTCAGCG ATTTTTCGAG CGACCGCTTT 7320

TTGATTTAAG GCGTGACCcA AaCkaCGCCG CTGAATGCCT TTTAATCCAG AAATCCTGGC 7380

TAATGGCTGT AGCTCATCTA CAACAACAGG ATCGACAATA AAGGCTGGCA CGTGATATTT 7440

TTCTGCAAAC TCATTAGCTA AAATCGCCCC TAAATTTGAA GCATGGGTAT TAAAGCGTTC 7500

TGTCCGCAAG TCTTCTAGCA TTTGTTGGTC GACAAGATAC GTACCGCCAG GAATAGGCTT 7560

AAGCAAACCA CCACGCCCTA CAACAGCGGC TAATTGTATG ATGTTATGTG TTTCTAAAAA 7620

TTCCGCGATC ATTTGTTTTC GAAATGATGT TTGACTGACA ACATTTTCAA AAGGTGCCAA 7680

TTCTTGGACG CTATGTCTAA GCGTTTCTTC TGCCAAGCAA TCATGATTAG CAAAAAGCGC 7740

TAACTTCGTG GAGGTCGAGC CAGGATTAAT AACCAATACT GTTTCCATTG TtTTCTGCTC 7800

CTTkTTTtAA ACTTGTCTCA TGGCAAATCT TAATGAATGG AATTTGCTTT CTGTCGAATC 7860

GCTTC 7865

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4023 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

TGCTCCnGnT GTGCTCGATG TATTACAAGA AAGATTCCAA ATTCTTCGCA ATATTTATTG 60

GATGCAACCT ATTGGACGTA GAAGTTTATC AGAAACGATG GGAATTACAG AGCGAGTCTT 120

GCGTACAGAA ACAGATGTTT TAAAGCAACT CAACCTGATT GAACCTTCAA AAAGCGGGAT 180

GACATTAACA GAACGCGGTT TGGAAGTCTA CCAAGGCTTA GAATTAGTCA TGAATCAATT 240

GCTTGGCATG CACCAAATCG AAAAAGAAAT GACACAGTAC TTTGGCATTC AACGTTGTAT 300

CGTTGTTGCA GGAGATAGTG ATATCCAAAA GAAAGTTCTT TCGGACTTTG GAGATGTCTT 360

AACCAATACG CTGAATTTAC TCTTACCAAA TGGTGAGAAT ACAATTGCGG TTATGGGTGG 420

GACAACAATG GCCATGGTGG CCGAAAATAT GGGATCACTA GAAACTGAAA AACGTCATAA 480

TTTGTTTGTT CCAGCCAGAG GAGGGATTGG TGAAGCAGTT TCTGTCCAAG CCAATTCAAT 540

TAGTGCTGTA ATGGCCAACA AAACGGGTGG TAACTACCGG GCGTTATATG TCCCAGAACA 600

ATTGAGTCGT GAAACGTATA ATAGTTTATT ACAAGAACCC TCGATTCAAG AAGTTCTTAC 660

GTTAATTAGT CATGCCAACT GTGTGGTGCA CAGTATTGGT CGAGCGTTGC ACATGGCGGC 720

ACGCCGaAAA ATGTCTGATG ACGAAATGGT TATGCTCAAA CAAAAAAATG CCGTGGCAGA 780

ATCGTTTGGT TATTTCTTTG ATGAAGAAGG AAAAGTGGTC TATAAAATTC CACGAATTGG 840

CCTTCAATTA AAAAACTTAC AAGAAATTCC CTATGTCGTC GCAATTGCGG GCGGCAAAAC 900

GAAAGCCAAA GCCATTCGCG CATATATGAA AAATGCACCA AAACAAACGT GGcTCATCAC 960

AGATGAAGCT GCCGCAAACG AGATTTTAAA AGGGGTAACC CTTTAAAATA AAAAATTTTT 1020

GATTTTCATA AGGAGGAAAT CTTTAATGAC AGTTAAAGTA GGTATTAATG GATTTGGACG 1080

TATCGGACGC TTAGCTTTCC GTCGTATCCA AGATGTAGAA GGAATCGAAG TAGTTGCAAT 1140

CAACGACTTA ACAGATGCTA AAATGTTAGC ACACTTGTTA AAATATGACA CAACTCAAGG 1200

CCGTTTCAAC GGTACAGTTG AAGTTCACGA AGGTTCATTC AACGTTAACG GCAAAGAAAT 1260

CAAAGTTTTA GCTAACCGTA ACCCTGAAGA ATTACCATGG GGCGAATTAG GCGTAGACAT 1320

CGTTTTAGAA TGTACTGGTT TCTTTACTTC TAAAGAAGCT GCTGAAAAAC ATTTAACTGC 1380

TGGTGCAAAA CGTGTAGTTA TTTCAGCTCC TGGTGGTAAC GATGTACCAA CAATCGTTTA 1440

CAACACAAAC CATGAAACAT TAACTGGAGA AGAAACTGTA ATTTCAGGCG CTTCTTGTAC 1500

TACAAACTGC TTAGCTCCAA TGGCTAAAGC TTTACATGAC AACTTTGGTG TTGTTGAAGG 1560

TTTAATGACA ACTATCCACG CTTACACAGG TGACCAAATG ACATTAGACG GACCACATCC 1620

TAAAGGCGAC TTCCGTCGTG CGCGCGCTGC TGCTGCAAAC ATCGTACCTA ACTCAACTGG 1680

TGCTGCTAAA GCAATCGGCT TAGTAATCCC AGAATTGAAC GGTAAATTAG ATGGCGCTGC 1740

TCAACGTGTT CCTGTAGCAA CTGGTTCATT AACTGAATTA GTTACTGTAT TAGACAAAGA 1800

AGTTACTGTT GATGAAGTAA ATGCAGTAAT GGAAAAAGCT GCTAACGAAT CTTATGGTTA 1860

TAACACAGAC GAAATCGTTT CTTCTGATAT CGTAGGTATG ACTTACGGTT CATTATTCGA 1920

TGCAACTCAA ACAAAAGTGA TGACAGTTGG CGACAAACAA TTAGTTAAAA CTGTTGCTTG 1980

GTATGACAAC GAAATGTCAT ACACTGCTCA ATTAGTTCGT ACTTTAGAAT ACTTCGCTAA 2040

```
CTTATAAGAT TCATTCTTAT CATTAGTGAA ATTTGAACAA TTGTGAAATA ATAAGCGGGG    2100
AAGCAACGCG CTTcTCCGCT TTCTTTTTTT AAAAAAGAAA AATCCTAAAA TTGACAATAC    2160
AGGAGGTCTA CTCATGGCTA AAAAGACAAT TAAAGATGTA GATTTAAAAG ACAAAAAAGT    2220
CCTTGTCCGT GTTGACTTTA ACGTTCCGTT AAAAGATGGC GTGATCACAA ACGATAATCG    2280
TATCGTAGCG GCTTTACCAA CAATCAAATA CGTGATTGAA AATGGCGGAA AAGCAATTCT    2340
TTTCTCTCAT TTAGGTCGTG TAAAAACAGA AGAAGACAAA GCTGGCAAAT CATTAAAACC    2400
AGTGGCTGAA CGTTTAAGCG AATTACTAGG TCAACCAGTT ACTTTTGTTC CAGAAACACG    2460
TGGTAAAGAA TTAGAAGACG CTGtTAACAA TATGAAAGAT GGCGACGTTT TAGTATTTGA    2520
AAACACTCGT TTTGAAGATG TTGACGGCAA AAAAGAAAGC GGCAATGACG CTGAATTGGG    2580
TAAATACTGG GCTTCTTTAG GCGATGTATT TGTCAATGAT GCTTTTGGTA CTGCACACCG    2640
TGCACATGCC TCAAACGTTG GAATTGCTTC AACTGGTATC CCAACAGTTG CTGGTTTCTT    2700
AATGGAAAAA GAAATTAAAT TCATCGGCGA AGCTGTTGAA AATCCAAAAC GTCCATTCGT    2760
AGCAATCTTG GGTGGCGCAA AAGTTTCTGA CAAAATCGCT GTCATTGAAA ACCTAATTGA    2820
AAAAGCCGAT AAAATTTTAA TCGGTGGCGG TATGGCTTAC ACATTCATGA AAGCACAAGG    2880
CTATAGTGTT GGTTTATCTT TACTTGAAGA AGATAAAGTG GACTTAGCGA AAAGCTTAAT    2940
GGAAAAAGCT GGCGACAAAT TAGTTTTACC GGTTGATACA GTTGTCTCTA AAGAGTTCAG    3000
CAATGACGCT CCTTTCCACA CGGTTCCTTC AACTGAAATC CCAGATGATG AAGAAGGCTT    3060
GGATATTGGT GAAAAAACAA TTGAATTATT CGCTAACGAA TTACAAGGTG CGAAAACAGT    3120
TGTTTGGAAT GGACCAATGG GCGTATTTGA AATGAGTAAC TTTGCCAAAG GAACAATTGG    3180
TGTTTGTGAA GCGATTGCTA ATTTAGAAGA CGCAACAACA ATCATCGGTG GCGGGGACTC    3240
TGCTGCTGCA GCAATTCAAT TAGGCTACGA AAACAAATTC TCTCACATCT CAACAGGTGG    3300
CGGAGCAAGT TTAGAATTAT TAGAAGGCAA AACTTTACCA GGTTTAGCTT CTATTAACGA    3360
TAAATAATCA GTTTAAAAGA CAATATTTTT AAAAAAAGAA AAGGATGTGC TTTTCATGCG    3420
CAAACCAATT ATCGCTGGTA ACTGGAAAAT GAACAAAACT TTAAGCGAAG CACAAAGCTT    3480
TGCAGAAGCA GTAAAAAATG CTGTTCCATC AAACGACGTT GTTGACGCTG TGATTGGATC    3540
TCCAGCATTA TTCTTAGCTC CTTTAGCTTG GAACCTAAAA GATTCTGAAG TGAAATTAGC    3600
TGCACAAAAC TGCTACTGGG AAAACGCTGG TGCCTTCACT GGTGAAAACT CACCAGCTGC    3660
AATTGCAGAT TTAGGTGTAG ACTATGTGAT TATCGGACAC TCTGAACGTC GCGAATATTT    3720
CCACGAAACA GATGAAGATA TCAATAAAAA AGCCAAAGCA ATCTTTGCAA ATGGTATGAC    3780
ACCAATCTTC TGTTGTGGTG AAACTTTAGA AACGTATnAA GCTGGTAAAA CAGCGGAATG    3840
GATCGAAGGT CAAATCACTA AAGGCTTAGT TGGTTTATCA AACGAACAAG TTGCTTCAAT    3900
GGTTATCGCT TACGAACCAA TTTGGGCAAT CGGAACTGGT AAATCTGCAG ATGCTAATAT    3960
TGCAGATGAA ATCTGTGGTG TTGTTCGTTC AACAGTTGAA AAATTATACG GCAAAGAAGT    4020
```

TTC 4023

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2288 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

GGGCGGTGTC ACTGCGTCTA TTTTCGGTTA TCGGCCAACC TTTTTCATTA CTGGTACGAT 60

TTTATwACTC GTTTTTGTTT TGAGTCTCGT CTTTGTCCAT GAAGAGTTTG TGCCAATTGA 120

AAAAAATCAA GCCGCATCTG GGAAACAAAT TTTAAAAAAA CTAGAACATC CCCACGTGAT 180

TCTTGGAATG TTTATTACGA CATTAATCAT TCAAGCTTCC AATAATTCAA TTAGTCCCAT 240

CATCAGTTTA TATATTCAAC AATTGTTGGG TGGTCACGGA AATGTCACCT TAATTAGCGG 300

AGTCATTGCT TCTATTCCAG GGATTGCAAC GTTAATAGCC GCCCCTCGTT TCGGCCGGTT 360

AGGCGATCGT ATTGGCAGTG AACGCATCTT GACAATTGGC TTAATTTTAG CCATTTTTGT 420

TTACCTACCA ATGGCCTTTG TTCAAAATGT CTGGCAGCTT GCCATGCTTC GTTTCTTAGT 480

TGGGATTTCC GATGCCTGTT TACTACCAGC TGTTCAAACG TTGATCACTC GTTATTCCCC 540

AAGTGACGCT GCAGGTCGTA TTTTCAGTTA CAATCAATCT TTCCAAGCGA CAGGAAATGT 600

GATTGGTCCT ATGATTGGTT CAAGCGTATC TGCCGCTTTC GGTTATCGTG GGGTCTTTAT 660

TTCCACTTCT TGTCTGGTTC TCCTTAACCT TCTATGGGTT CGTCGAAGCA CCGCCGAATT 720

AAAAAAAGAG AAAAATGATG ACTAAGTTAT ACACAGAAAA GTTATCCACA ATTTGAATTT 780

TTAATTTTTC TCTTTTTCTT ATTTTAAAAA AAATAGGCAT ATAAAACAAA AAAGAAAAGC 840

GGATATTTAC TGGTTTTATG ACCAGCATAT CCGCTTTTCT TTTGGtTAAC TATACGCTTT 900

TTGTTTCACA ACTCGTGAAA CTTTTTAGAC GTATTGTAGT TCTTTTTCAT CTTGGTAGGC 960

GCGATCAATT AAGCCGCCCC CTAAGCATTC CATCCCATCA TAAAAAACAA CAGCTTGACC 1020

TGGAGTGACT GCACGCGCTG GTTCCTTAAA CGTAACGGTT GCTTTTGTGC CATCTTCAGA 1080

TAAGGAAACA TGCACTGGAA TATCTGCTTG ACGATAACGG AATTTAGCCG TACAGTCAAA 1140

CTCTTTTGGc ATTGGTGTAT CAACAGTAAA ATGAACCTCA CTAGCTTCTA AATGTGtTGc 1200

ATATAGTTTT TCATGGTGAA ATCCTTGtCC CACATATAAT GTATTGGTAG TTAAATCTTT 1260

CCCAATCACA AACCATGGTT CTTGTGTTTT ACCGCCGCCA CCAATTCCTA AGCCTTGGCG 1320

TTGACCGATA GTATAATACA TCAGACCATC ATGTTGTCCT TTGATTTCGC CATCTTCTGT 1380

CACCATGTTG CCTTTTTTAG CTGGTAAATA ATTACTCAAA AATTCTTTGA AATTCTTTTC 1440

ACCAATAAAA CAAACGCCCG TTGAGTCTTT TTTCTTGGCT GTGGCTAATC CAGCACGTTC 1500

AGCAATTGCA CGAACTTCTG ATTTTTCCAT GCCACCCAAT GGAAACATTG TTTTTGCCAA 1560

TTGGGCTTGT GATAATTGGC TTAAAAAATA GGTTTGGTCT TTATTATTGT CAATTCCACG 1620

TAACATATGC ACTACGCCAT TTTCATCTGT TTCTACTTGC GCATAGTGGC CAGTkGCAAC 1680

ATAGTCAGCG CCTAGTTGCA TAGCATAATC TAGGAATGCT TTAAATTTAA TTTCTTTATT 1740

ACACATGACA TCTGGATTTG GTGTCCGTCC TAACCGATAT TCTGCTAAAA AATATTCAAA 1800

CACACGGTCC CAATATTCTT TTTCAAAATT GACCGAATAA TACGGAATCC CAATTTGATC 1860

CGCTACTTTT GCGACATCTT TATAATCCTC AGTTGCGGTA CAGACACCAT TTTCGTCTGT 1920

ATCGTCCCAA TTTTTCATAA AGATCCCAAT TACATCATAG CCTTGCTCTT TTAAAAGCAA 1980

CGCTGTGACA GATGAATCGA CACCGCCACT CATGCCTACC ACGACACGCG TTTTGCTGTT 2040

ATCTGCCATT TTATCACCAT CATTTCAATT AGATTCTGAA ATCATCCACG ATTTGAAGGT 2100

CGTAACTTTT CAGTGCTTTT CATTATACTA TTTTTTTGAA ATAAGTAAAG CAAGCCGGGC 2160

GTATAATGAG AAAACACTTG GAACAAAGAC ACAGCCAACA AGCTATTGCT TTCAAATAAA 2220

CAGAAAACcT GmAGGAACAA GCTACTTTCG CTTGTkTCTT CAGGTTTCTC TAAATTATTC 2280

TACTTTCT 2288

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6691 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

TTTAATTTTT GGGAAaAAAT TwCCaAAGGC TAAGGaTaGk TAAAtTGgGG ACCwTATTTg 60

GAATaTTGaT TAAATTCTGA CCACTTAATT GtCTTTkGAA AAggTTATGC TAAAGTTACT 120

TTGTAATcGA ATGaGGAGGG CTATCACATG GaAGAAAAGA ATTGGCCTGA GTTATTTGCT 180

CGTTATCTTA TTGTCGAAAG AGGCTATTCA GAAAAAACAA AAAAAGCATA TCAAGAAGAT 240

ATACAACATT TTTTTTCGTT TTTAAAGACT TCAGGTAATG ACAATTATTT GACTGTGGAA 300

CATTTGGATG TTCGGGCTTA TTTGAGCGAA TTATATGATC AGGAATATAG CCGCAATTCG 360

ATTAGTCGAA AAATTGCCAG TTTGCGTTCT TTTTATCAAT TCTTATTAAA AAATGAAGCA 420

ATCCAAGAGA ATCCTTTTTC CTATGTTCAT ATGAAAAAGA AACAGTTACG GTTGCCCCGT 480

TTTTTTTATG AAAAAGAAAT GGATGCATTA TTTGAGAGCG CACAAGGGGA GCAACCGCTT 540

GATTTGAGAA ATCAAGCATT ACTAGAAGTT CTGTATGGGA CTGGTATTCG GGTCAGTGAA 600

TGTGCGAACT TAACGGTGGA CGCAGTTGAC TTTCAAGCTT CTGTTTTATT GGTTCACGGA 660

AAAGGAAATA AAGACCGTTA TGTGCCTTTT GGTTCATTTG CACAAGACGC CTTAAAGGAT 720

TACTTAGAGA ATGGCCGCGC GCTATTAATG ACAAAATATC AAAAAAAACA TCCGTATGTC 780

TTTGTTAATC ATCATGGTGA ACAAATCACG CCAACAGGGA TTGAATATGT CTTAAACCAA 840

CTAATTAAAA AAAGTTCATT AAATGCGGAA ATTCATCCCC ACATGTTACG CCATACATTT 900

```
GCGACACATT TATTAAATAA TGGCGCCGAT ATGCGGACAG TTCAAGAATT ATTAGGACAT     960
GCTAATTTAT CAACAACTCA GATATATGCG CATGTGACAA AAGAAAGTTT ACAAAAAAAT    1020
TATCGAACGT TTCATCCTCG TGCATAATGG AGGTAAACGT CTAGGAGGAA CAATAATGGT    1080
TGAATCACAA TTTCACTCTA CAACAATTTG TGCAGTTGAA AAAGATGGAA AATTTGCAAT    1140
GGCTGGTGAC GGCCAAGTAA CAATGGGCGA ATCTGTCGTT ATGAAAGGTA CGGCCAAAAA    1200
AGTGCGCCGT ATTTACAATG ATGAAGTCGT TGTCGGTTTT GCAGGAAGTG TGGCAGATGC    1260
CTTTACATTA GAAGAAAAGT TTGAAGGAAA ATTAAATGAA TACAACGGAA ACTTAACGAG    1320
AGCTGCGGTT GAGTTArCCC AAGAGTGGCG TACACAACAG TCTATGCAAA AATTAGAAGC    1380
TATGCTGATT GTGATGAATA AGGAAGAAAT GCTCTTAGTT TCAGGAACTG GGGAAGTCAT    1440
CACACCAGAC GATGGTATTT TAGCAATTGG tTCAGGCGGT AATTTTGCTT TGTCCGCTGC    1500
AAGAgCCATG AAAAACTTTG GCGATAAAGA AATGCCmGCG AAAGAGATTG CTAAAAATGC    1560
CTTAAATATT GCTGCAGATA TCTGTGTATT TACGAATCAT AATATCATTG TAGAAGAATT    1620
ATAAGCGAGG GAAAATCCAT GAATGAATTA AACAAAACAC CAAAAGAAAT AGTCAAAGAA    1680
CTAGATCAAT ATATTGTTGG ACAGCAAGCT GCCAAAAAAT CAGTGGCGGT AGCTTTACGT    1740
AACCGCTATC GTCGCTTGCA ATTAGAAGAA AATATGCAAC AAGATATTAC GCCTAAAAAC    1800
TTACTAATGA TTGGACCAAC AGGTGTTGGT AAAACTGAGA TTGCTCGTCG TCTCGCAAAA    1860
ATTGTCAATG CGCCTTTTGT AAAAGTTGAA GCAACCAAAT TTACAGAAGT AGGCTATGTT    1920
GGTCGGGATG TCGAATCAAT GGTGCGGGAT TTAGTAGAAA ATGCGATCCA AaTTGTTGAA    1980
AAACAACAAT ACAGTCGCGT GTATGCGCAA cATTAAAAAA GGCCAATCAA CGTTTAGTCa    2040
AAGTATTAGT GCCTGGAATT AAAAAAGAAC AAAAACAGGC TGGCGGTAaT CAATTTGAaC    2100
AAaTGATGAA CATGTTTAAC ATGGCTCAGC AACAGCAAGA AGCACAAGAA GAAGTAACGG    2160
AAGATATTCG AACGAATCGC CGAACAATTT TGGAACAGCT GGAAAAAGGT CTATTAGATA    2220
ATCGTGAAGT AACAATTGAA ATTGAAGAAC CGAAGAAAAC GATGCCAGCT ATGAACAATG    2280
GCCTAGAACA AATGGGTATC GACTTAAATG AAACGTTAGG CGCTCTGTCA CCAAAGAAAA    2340
AAATCGAACG TACTGTAACG GTGAAAGAAG CACAAGAATT ATTAGTGAAA GAAGAATCAG    2400
CAAAAATTGT CAATGACGCT GATATTCATA GTGAAGCTAT TCGTTTAGCT GAATCAAGCG    2460
GAATTATTTT TATCGATGAG ATTGATAAAA TCACCTCTAA AAGTCAACAA AATTCGGGCG    2520
AAGTCTCTCG TGAAGGAGTA CAAAGAGATA TTTTGCCGAT TGTTGAAGGC TCCCAAGTTA    2580
ACACGAAGTA TGGTCCTTTA CAAACGGATC ACATTTTATT TATCGCTTCA GGTGCTTTCC    2640
ACTTGTCAAA ACCAAGTGAC TTGATTCCAG AATTACAAGG CCGCTTCCCA ATTCGAGTTG    2700
AATTAGATGA TTTAACGGCG GATGAtTCGT AAGTATCTTA ACTGAGCCAA ACAATGCTTT    2760
AATTAAACAA TATGTAGCAT TAATTGGCAC AGAAAATGTT TCAGTCATCT TTACAAAAGA    2820
AGCAATTGAA CGGTTAGCAC ACATCGCTTA TGATGTAAAC CGTGATACAG ATAATATTGG    2880
```

```
GGCGCGTCGT TTACACACAA TTTTAGAGCG TTTATTAGAA GATTTATTAT ATGAAGCACC   2940
AGATATGCaA ATGGGTGAAA TTACGATTAC CGAAGCATAT GTCAATGAAA AATTGAATGA   3000
CATTGTTCAA AATGAAGATT TAAGTCGTTA CATCTTATAA AAAAGGCTGG AGGAGAAAAG   3060
ATGGCTACAT TATTAGAAAA AACACGTCAA GTGAACGAGC TTTTGCAAAA AAACAATCTG   3120
TTTGACGTAC AAGCAGAACT TCCGTACAAC AAAATGGCAA TGATTTTAGG GGATATTTTA   3180
GAAAGTAATG CGTATATTAT TAGCAGTTCA GGGGATTTGT TGGGGTATAC TGAAAAATTA   3240
GATGTTAATA ATGCCCGTAT CAAAAATATG TTTAAAGAGA AAAAATTTCC TCAAGGATAC   3300
ACCGAGGCCG TGGATATGCT CAAAGTAACC GAAGCCAATA TTCCAATAGA TAGTGATTTA   3360
ACTGCATTTC CTTTTGAATC ACGAGAGCTG TATCCGTTTG GTTTAACCAC TATTGTGCCT   3420
TTGTACGGTG CTGGCAAACG CTTGGGTACG ATTATTTTAG CGCGTGTTGA AAAATCCTTT   3480
AACGAGGATG ATTTAGTGTT AGCTGAATAT AGTGCGACCG TTGTCGGCAT GCAAATTTTG   3540
TATCATCAAT CCAGAACAAT TGAAGCAGAA GTTCGCAGTG CGACTGCCGT GCAAATGGCC   3600
ATCAACACTC TTTCCTATAG TGAATTAAAA GCCGTCCATG CCATTTTTGA AGCCTTAGAT   3660
GGCGAAGAAG GTCGTTTAAC TGCTTCAAGT ATTGCAGATG AAATAGGCAT TACTCGTTCC   3720
GTAATTGTGA ATGCCTTAAG AAAATTAGAA TCAGCTGGAA TTATTGAATC ACGTTCTCTT   3780
GGTATGAAAG GGACTTATTT AAAAGTATTA AATCAACAAT TTATTAAAGA GTTGGAAAAA   3840
TAAGCAATTG CTCGTATCAA CTAGGAGGAC AACGCAATGA CTGTACAAAT TGAAAATGAA   3900
TTTCTAATCG CAACATTTGC TGAAGAAGGA GCAGAATTAG TAAGTTTACA ATCAAAAGAA   3960
ACCGGCATTG AATATATTTG GCAAGGAAAT CCTGAATTTT GGGCCCGCCA TGCGCCCGTA   4020
CTTTTTCCGA TTGTCGGTCG TCTAAAAGAA GATACGTATA TGTACCAAAA TCAAGCGTAT   4080
CATTTAACTC AACATGGTTT TGCTCGTGAT CAAGTATTTG ATGTGATTGA AAAAGGTGGA   4140
GAAGAAGTTT CTTTTTCACT AAAATCAACA AAAGAAACCA AGAAGAAGTA TCCTTTTGAT   4200
TTTGAACTAG TGATTACTTA TACATTAGAA CATCAAGAAT TGACTGTAAA CTACCAAGTA   4260
GAATACACAG GGAAAGAAGA GATGTATTTT GGGATTGGTG GTCATCCAGC CTTTAATGTA   4320
CCTTTAGAAT CTTCACTAAC GTTTGAAGAT TATTATCTAA GTTTTTCACC TAAAAAATCA   4380
CGGACACAAA TTCCCTTAGC AGGCCCGTTT ATTGATTTAG CGAATAAAAC GCTAGCCCAA   4440
ACGAATACCA GCTTTGACTT AACGCATCAA TTATTTGAAA ATGATGCAAT GATTTTTGAA   4500
ACAAAAGGTC AAACAGCAAT TACGATTGCC ACAGACGAAA GTGACCATCA AGTCACTTTG   4560
AGTTATCCAG AAATGCCTTT TGTGGGGATT TGGTCACCGA CACCGAAAGA AGCGCCTTTT   4620
GTCTGTATCG AACCTTGGTG CGGGATTGCT GATGCTGTCG ACGCAACTGG CCAACTTGCT   4680
GaAAAAtTTG GCATCaATAA GTTmCCAGCT AACGAACTAT TTAAAmCACA GTATATGATT   4740
AGTGTAAAAT AAAGAAAAAG GCTGTGAACA TTATTTGTTC ACAGCCTTTT TCTTTATTTT   4800
TTTTGCACGT AAGCCAAAAG AAAGACGACT TTCTGTTCCA TTTTTAATTC GCTGAATGTT   4860
```

TTCACGATGA CGGACAAAGA TAAAAGTAGT CAAAGCAATC GCAATCACGG TTAATAACCA 4920

GTTAAATGTT GGTAAAAATAG CTGGGACAGT GAAAGGCAAA ATAATAGTAG AAAGTGTAAT 4980

TAGTACAGCG CTAATCATAC TTGTTAAACT CACCATACTG GTTAGATACA GACAAATCAC 5040

AAAAATAAGA GCAGAATAAA TAAAAAATGT GGGGCTGTAT GCTAATAACA TGCCAGCGCT 5100

AGTGGCTACG GCTTTACCAC CTTTGAAATT GGCAAAAATA GGGAAGGTAT GCCCTAAAAC 5160

AGCTGCTACC CCAAAGAAGA GCGGATTCAC GCCTTGTAAA CCAAACAAAT AGGGTAATGA 5220

AGTGGCTAAC GTTCCTTTCA AGATATCCAT TAATAATACC GTAATTCCGG CAGGTTTCCC 5280

TAAGACACGA AATGTATTGG TTGTTCCTGT ATTCCCACTC CCAAATTGGC GTATATCTTT 5340

TTTAAAGAAA AGTTTACCAA TCCAAACACC TGAGGGAATC GAACCTAATA AATAGGCAAC 5400

AAGTAACAAA ATGACGATTT TCATCAAGTT AACCTCTTTC TTATTCAAAC TAGGATTTAT 5460

TTTATCATGA ATAGAAAAAC GGAACAACGT CTAATAACTA AAACAAGAAA AGTTTTTCTT 5520

TTGCTTGAGG AAGCGTTGAC AAATGGTTGA TAACCGTTAT AATAAATGTG TGTCATTAGA 5580

CAAAGACGTC TAATCACTGG TTTGAaAATT AGTTATCAAC GAAGATACAA TATATATGTA 5640

CTTATGAAGT GTCAGGGAAA GAGGTGCGAA TCCTCTACAG ACCTACCTAC TGTATGGTGG 5700

ATGAAACCAA TAAGACCACA GATTATTCTG GAAGGATTGG GAGTAAGAAG AAGCTAAGTC 5760

AGGATACCGG CTTGATAAGT CTAATCATTC TTTTGTAGGA AAAAGAAGGG GCTTTGAAGA 5820

TAACTGGCTG AGTATATCAT TAGCATAGTT AAGTGACTGC TTTTTTTCTA TTGGAGAAAA 5880

AAGTGGTTTT TTTGTATTGT TTTGACGTTG AGACAAAGGA GGTTCATTTC AGAAAATTTT 5940

CCCCAAAATA AAATAGACGA ATGCGAGGAT GAAAAAATGA AAAAATTTAC TTTAACAATG 6000

ATGACTTTAG GTTTAGTAGC AACACTTGGC TTAGCAGGAT GTGGTAAACA GGAAAAGAAA 6060

GCAACTACCT CTTCTGAAAA AACAGAAGTA ACGTTACCAA CCAAAGACCG TAGCGGCAAA 6120

GAAATTACTT TACCCAAAGA AGCAACCAAA ATTATTTCCC TAGTGCCATC AACAACAGAA 6180

GTGATTGAAG ACTTAGGTAA AACCGACCAA TTAATCGCAG TTGATACTCA AAGTAGTACA 6240

ATGATGACTG ATTTAAAAAA ATTACCACAA ATGGATATGA TGGCTGTCGA TGCCGAAAAA 6300

TTGATTGCCT TGAAACCACA AATTGTTTAT GTGAATGACA TCAATTTAGC TAGCTCAGAA 6360

AGTGTTTGGA AGCAAGTGGA AGATGCTGGA ATTACAGTCG TTAATATCCC CACTAGTACA 6420

AGCATCAAAG CAATCAAAGA AGACGTCCAA TTCATCGCTG ATAGCTTATC TGAACATGAA 6480

AAAGGACAAA AGTTAATCAA AACAATGGAT CAAGAAATCG ACGAgTAGCG AAAATTGGTA 6540

AAACTATTAA AArACCAAAA ACTGTTTTGT TTGAAGTGGC TGCCTTACCA GACATTTATA 6600

GTTTTGGTAA TGGGACATTT TTAAATGAAT GGATTGAACC AATTGGGGCA AAAAATGTTT 6660

TGGCCAATGA AAAGGCTGGT TnCCAGTGAC A 6691

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20072 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

TTAnGGTAnC CAATCGTTCC CCGGTAAATT TGGACAGTTT TAATCTGTTC CACAAACGAT 60

nAACGGCTCC GTTTGTTAAT AACTGCCGTT CTAACATCTT TTCTAAATTG TGAAACAATT 120

TCTTGTTTGG CATAGGGCGG ATCCAATAAG ACCAAATCAA ATTGTAACTT CTCTTCATAA 180

AACTGTTCTA AGGCACGATT AGCATCCATT TTTCTAACTT CAAATTTTTC TGGTTCTTTT 240

GTAATCGCAA TATTTTCTTT AATTACTTTT AGTGCAGCAA AATTCTTTTC AATACAAATA 300

CTTTTGTCCA TTCCACGAGA AACGGCTTCA ATCGCTAACC CGCCACTTCC TGAATATAAA 360

TCCAGCGCCA TTCCGCCATC GAAATATGGA CCAATCATAT TAAAAATAGA TTCTTTTACT 420

TTATCTGTTG TGGGCCTAGT ATTGTCGCCA TCTAACGCTT TCAAGCGCCG GCCACCATAT 480

TCTCCAGAAA TAACACGCAT TGTTTCGACT CCTTTATCAA TCATTCTTTT ATAATTCATT 540

ATAACAAAAA GAGTGAAAAA TGTTTGAAAT TTCTCATAAC ATTTCTCACT CTTTTGCATT 600

AGACATTCTC TGCTAAGCGA ATCTCTGTAG ATAATTCCTC TACTACATAG CCGTCGTCTT 660

CTTGTTCTTG ATAAGCCGCT TTGGTACCAA TTTTTTCAGC AAAATTCATT TCGACATCTG 720

GTCGATACGA TTTATCGACA CTACGAACAA AATGTAACTG ATTAATTTTT TCCATGTTAC 780

TTTCAATATT TTCTTCATTC ATGTAAATCA CCACATATTT CATTTTTCGT GAAACATAGT 840

GAATCAGTCC ATATCGGCGT AACTGTTTTA ACGGTTTTAA TGTGTATACC CAAACAATTA 900

GGCAACGACG CTCTTGAAGC GTAAATTCTT TTTCTTCTTT AGCTTGCATG GCAGCTTCCT 960

CCACAACCTG AATGTTTCTT TCCTCGTTCA AAGAAAGGAT TTCCAGCATC TACTTTAATA 1020

TCTTCAGAAA TTAGTTGCGC AATCTTTAAA CCAATTGTAT CTAGCAACGT TTGAACGTTT 1080

GTCTCAGCAA AACGATAATT TGCAACGATT TCATTTAAAT CAAGCGCCCG CTTCGCTTTA 1140

CGTAGGGCTC GTTGTTTCTC CCGAAAATCG GGTGCATGAA TTCCATAAGC CTCTACTCGT 1200

TCAAAGGCTG ACTTAGCTTC TAAAAATGCT TTTTGTAACT GAGCAACTTC CGCCGAAGCA 1260

TACATAGCTT TTTTTGTTTC TTTATAGTTT TGAACAGCAT CACTTTCTAA AAGTGCTGAA 1320

ATTAAACATT CTGTTTGATC TTCAATTGCA AATAAATGTT CTGTTACAAT CATAACTTTT 1380

CACTACTCTC CGTGATTTGC CTATTTTCTG GTCGTTTCTC TAATAAGACT AACACACCTT 1440

CATTGGAAAA GGCAACGCCC AATTCTGTCG TTTCTGACTC AGCAAAATAG CTATCATATT 1500

GCGGGTCCGA AAACCATGaC AAAATAGTAA AGGTACTATC ATAAACAGGT TCTTGGAAAA 1560

TACCtGcTGG TTCTCTCAAC GAAAGTTkGC TCGTTTtAAC TAATAAAGrA AATTCGkCCT 1620

TAGTCmAATG AAaGGGATGA ATGGCTGTAT GTTCCTCTCT CGTTTGTAAC AATTCTGCTT 1680

GTCGATTTGC CGACAAAATT GTTGGTGTAT GGTTAGCCAT CGTTTGATAC AATGTTTGGG 1740 cATCTTTTTG GCTTTGTGTA GAAACTGAAA AAGCAACTTG ACTTTCTTGc TGTTTAACCA 1800

GATTCATGAC ATAAATCGCT TGGTTACTTT GAACAGGATC GAAGGTCATT GACTGCGTTA 1860

TATCAACTGG CAAAGGAGAA CCTTCATTTA ATTGATACGG CATTTCCGTG AGCAACGTTT 1920

CTTTATTGAC ATAATTAACT GCCATTAAAC CGCCTGTCCG CTGATTAAAA AATAACACCG 1980

CAAAACTGCC ATTATCAAAA GCAACCAGTG GTCGATAGTT CATATCTTCT TCCATCAGCT 2040

CGACATCGTA AGCTTCATGA TGAAAATTAA ATGCAAAATT CGAATAAATC GTCATTAAAT 2100

CAGACAAGTC AGCAAGCTTC ATATCTATTT TAAACGGTTC TATCATTTTG TCATTGTTAA 2160

AGGCCTTAAT TGCGATCACT TTATTTTCTC GAACGTTGAC TTCTAAGTAA TCATTCATCG 2220

TCAAACCGAA CGTCCATAAT TCATACGTAC GACCTGTTTT AATTTTTTCT TTAGGATTGC 2280

CAAACTGGCG GATAAAGATT TCCACATCTT TCCCTATGTA TGCTGCGTAC CAGAAGCTGG 2340

CAATTCTTCA TGAGGAACAG CGGTGTGAGC GCTGGATAAG CGCCCACTTT GAGCTGTTTC 2400

GGTTGGTTTT TCAGCAGGGA AAAAAACTGG CTCTAAATAT CCAATGAGTA GAACTGTTAA 2460

AAAAACAGCA GTAAAGGCCC AAAATCGCTT CATTTTTTCC CCTCTCATCT TATTTTTCTT 2520

TCACTTCTAA TAACAAATCA CCTGAGCTGA TTGcTTCGCC TTCTTCAACA TAAATGTGAT 2580

CTACCGTGCC CGCAAAACGT GCTTCGATAG TCGTTTCCAT TTTCATGGCT TCTGTAATCA 2640

GCAACGGCTG ACCTTTTTCT ACTTTATCGC CACGTTTGAC CAATACTTGT AGAACAGAAC 2700

CAGACATCGT TGCGCCGATT TGTTCTTTAT TAGTTGGTTC CGCCTTTTGT TTTACTTGAA 2760

CAGCAGACTT AATAGAGGCA TCTTTAACTA AAACCTCACG ACGTTGCCCA TTCAAATTGA 2820

AGAACAAGAC ACGGTTTCCG TCAATATCAG GTTCGCCAAT TTCATCTAAA CGAATAATTA 2880

ACGTTTTTCC TCGTTCAATT TGAACTTCCA ATGTCTCTCC TTGACGAATA CCATTAAAGA 2940

ATGTTGGCGT ATCTAATAAG GTTATATCTC CAAAGGTTTC GTATTTTTGA CGATACTCTA 3000

AAAATACTTG TGGATACATT AAATAACTTa AAACTTCTTC CAATTTAGGT TGaTACCCAA 3060

TTTTTTCGGC TAATTCTTCT TGTACTTTCG CAAAATCAAC AGGTGCTGcT AAGTCACCTG 3120

GcCGCTCTGT GAAGGCTGGT CGACCTTTTA GAATAATTCG TTGCAGCTCT TTCGGGAACC 3180

CACCAACTGG TTGACCTAAA TCACCTTGGA AAAAAGTCAC AACTGATTCA GGAAAACTTA 3240

GTTCTTCACC ACGCGCATAA ACATCTTGTT CTGTCAGATT ATTTTGAACC ATGAAGAGTG 3300

CCATGTCTCC CACGACTTTT GAAGATGGCG TCACTTTTAC AATATCGCCA AACATCAAAT 3360

TCACTGTGTG ATATATTTTT TTGATTTCAT CCCAACGGTG CCCTAAGCCT ACCGCTTTTG 3420

CTTGCTGCTG TAGATTAGAG TATTGCCCAC CAGGCATTTC ATGCATATAG ACTTCTGTTT 3480

GCGGGGCATT TAACCCATTT TCAAATGGTT GATAATACAT GCGCACATCT TCCCAATAAT 3540

GATTGATTTT CTGTGCATTA TCAATGTTAA TAGTTGGCGT CCGTTCACCA TTGACTAAAG 3600

CATAATATAG ACTGTTCATA CTTGGCTGGC TAGTTGCGCC ACTCATCGCA CTCATTGCGA 3660

CGTCAACAAT ATCAACGCCC GCTTTAGTAG CTGCTGAATA TGTGATGATC CCATTGCCAC 3720

TAGTGTCGTG AGTGTGGAGA TGAATTGGTA AATCCGTGGC TGCCTTTAAT TCACTAATTA 3780

AACGATAAGC AGCTTGTGGT TTCAATAAGC CAGCCATATC TTTAATCGCA ATGATTTGTG 3840

CACCTAAATT TTCCAATTCT TTAGCCATAT CAAGGTAATA TTGAACATTA TATTTTGCTC 3900

GGGCTGGATC ATTGATGTCC CCAGTATAAC AAATTGCTGC TTCCGCAATT TTTCCGGTAT 3960

CCCGAACGAC TTGAATACTT TTTTCCATTT GAGGAATCCA GTTTAAGCTA TCAAAAATTC 4020

GGAAAACATC GACTCCTTGG CGAGCGGATT CTTTAATAAA TTCTTCAATA ACATTATCAG 4080

GATAATTTTG ATAGCCTACC GCATTAGATC CTCTGAACAG CATTTGTAAA AGTGTGTTTG 4140

GCATCAACTG ACGAATTTTT CTTAAACGTT GCCATGGGTC TTCGGTTAAG AAGCGATAGG 4200

CAACATCGAA CGTAGCGCCA CCCCACATTT CACTAGAAAA CAGCTCAGGC AGGGCTGCAT 4260

CAGTTAGACC AGCAATTGCT TTAAAATCTT GTGTTCTCAC ACGAGTAGCT AATAAACTTT 4320

GGTGCGCATC ACGGAACGTG GTATCTGTCA TTAACACACT TTCTTGATTC TTAACCCAAT 4380

CGATGACGGC GGTTGCTCCT TGAGCGTCTA GAATATTTTT AGCTGTAATC ACTTTTTCTG 4440

GAACTTCAAT ATCTGTCGGC ACGCGCGGGG CTTCAAAATA TTTTTTCTCC GTTCTTTCAA 4500

TGCCAGGGAA CCCATTGACC GTTACTTCTC CAATGTATTT CATCGTTTTA TTGCCACGAT 4560

CGCGCATACG AGGAAATTCA AATAATTCAG GTGTATTATC AATAAAGGTT GTTTTGGctT 4620

CGCCAGATTG AAACGCTGGA TAGCTCACAA CATTTTGTAA AAACGGAATA TTTGTTTTCA 4680

CGCCACGAAT CCGAAATTCT TTTAAGCAGC GTTGCATTTT ACTGATTGCC TGTTCAAAAG 4740

AGAAACCATG CGTACAGACT TTAACCAATA AAGAATCAAA GTAGGGAGTC ACAGCGTAgC 4800

CAGAATAAGC ATTCCCTACA TCTAAACGCA CACCAAAACC ACCTGGTGAA CGATACGTAT 4860

CGATTTTCCC TGTGTCTGGC ATAAACTGGT TCAGGGGATC TTCAGTCGTA ATCCGACATT 4920

GAATAGCCGC GCCTTTTAAT GTCAATTCGT TTTGTTTTGG CAAATGCATA TCTTTATGCA 4980

AATCAAGCCC TTGCGCAATT TGTAATTGAG AGATCACAAT ATCAATATCT GTAATCATTT 5040

CTGTGATAGT ATGTTCTACC TGAACACGAG GATTCACTTC AATAAAATAA AACTGATCCC 5100

CTTCTACTAA AAACTCAACC GTTCCCGCAT TCACGTAGCC GACATGTGCC ATTAACTGCA 5160

CAGCAGCCGA ACAAATGGCT GCTCGTTGTT CTTCATTCAT TGATACACAT GGTGCGACTT 5220

CTACCACTTT TTGGTGGCGC CGTTGAACGG AACAATCACG TTCAAATAAA TGCAAGACGT 5280

TCCCATGATG ATCGCCTAAA ATTTGTACTT CGATATGTTT AGGATTAGAA ATATACTTTT 5340

CAACATAAAC CTCGTCAGAA CCAAAGGCTG CTTTCGCTTC ACTTTTTGCT CTTTCGTAAC 5400

CTTCTCGTGC TTCTTTGGCA TCGTGGGCAA CGCGCATCCC GCGACCGCCG CCACCTAAAG 5460

CAGCTTTAAT CATGATAGGA AAGCCATGTG TTTCACCAAA AGCTACAACC TCTTCCACCG 5520

TTGCTACTGG CCCGTCTGAG CCTGGAATCG AAGCAATCCC CGCAGCTACA GCTGCTTCTT 5580

TCGCTTTAAT TTTATCGCCA AAAATATCTA AATGATGTGT TTTAGGTCCG ACAAAAATGA 5640

TTCCTTCTTC TTCACAACGT TCGGCGAAcG TAAGTTCTCT GATAAAAAGC CATAACCAGG 5700

ATGAATGGCA TCTGCTCCCG ATTTTTTGGC AATTTGGATA ATATTTTCAA TGTCTAAATA 5760

GGCTTCAATT GGTTTTTTCC CCTTACCAAC TAAATACGCT TCATCTGCTT TGAAACGATG 5820

AACAGAATAC TCATCTTCTG CAGCATAAAT CGCTACTGTA CGGATATCTA ATTCTGTACA 5880

AGCTCTAAAA ATTCGAATCG CAATTTCGCC ACGATTGGCA ACTAATACCT TTTTCATAAC 5940

GTCACTCCTA CCATCAATTT CCTTCAAATT TATTGGTTCT GTTCTAAAGA TAAATAGTAC 6000

TCATTTTCTA ATTTCTGACG TGTTTCATCT GCACTGATAT TCAGGACAAA GGCCACAGCA 6060

ATCGAAATAA TCAGCAAACT ATTCCCACCT TGACTCAAGA ACGGGAAAGT AATTCCTGTT 6120

AATGGAATAA TCCCAGTAAT CCCACCAACA TTAATAAAGA CTTGAATTAA TAACATTGTA 6180

CCAATTCCAA TACACATTAG TGAATTAAAA GGTTTCTTCG AGCGTACACC GACTAAAATA 6240

ATTCGCGCAA TCATAAACAT CAACAATCCT AAAATCGCTA GACCGCCGAT AATGCCTAAC 6300

TCTTCTAAAG TAATCGCAAA GATAAAGTCC GTATGTGCTT CTGGTAAAAA GCCTTTTTTC 6360

TGAACACTAT TCCCTAGACC TTTGCCAAAC CAGCCGCCAT TACTGATTGC ATAATAAGAA 6420

TTAGCTAACT GATGTCCTAA GTTCCGTTCA TCTAAAAATG GATTTTTAAA AACAGCAAAA 6480

CGATTATAAA TATATTGATA ACGCGCTGGG AAAATTTTTC CTTTAGACAT AATTAATAAC 6540

TGAATAGCCG TGATACTACC TAAAATGCCC AAGCCACCGA CCAAGTAGGT ATACATATAA 6600

TTAATCCCAC TTGCTAAAAC CATCACGATG GTAATTAATG TTAGAATGGC CGCATTCCCG 6660

AAGTCAGGCT GGATTGCAAC TAGAGCAATT AATACAAAAA CGAGCATCAA CGGGCGCCCG 6720

GCAGCTTGTT TAAATTGATC CATGCCTCCA TTAATTGTCT TTTGTCGCCG TGCTAAAATA 6780

TAAGACAGAT ACCAGACAAC CATAATTTTT AAATACTCAG CTGGTTGCAT CGAAAATCCA 6840

CCAATCTCAA TCCAACCACG TGCGCCGTTA ATTTCTTTCC CGATACCAGG AATCCGAACG 6900

GCTAAAACCA TAACCGTAAT TACCGCAATG GCAAACATGA TAAAACTTCG ATTCTGAAAA 6960

ACGGATGTTT TCATCTTATA AATAAAAAAC ATTGCAACTA ACCCTACTAC CCAAAAGGCA 7020

ACCTGGTTAA TAACGAATGA GGTAGGTGAA AAACCTTTCA TTACTTGTAA TGCGGATGTG 7080

GAACTATATA CCATAATCAA TCCCACGATA CTTAAAATTA AATATGGAAT AAAAATACTA 7140

TAGTCTAGCA AATGCCGTTT CTTTACTTTG TTTGGCAAAG CTTAACCCTC CTTGCCAATC 7200

TCGTTTTTTT CTTGTGTTTC TTCATAAACA TCTGTATATA ATTGATTCAA TTCGCGTTCT 7260

AAATCACTTA GTAAGCGATG TCCTTCATCA TTAGTTAAAA TCCCCAAACG GGTCGCATAG 7320

GTCACTTGTC TTGAAAAACC ATACATTTGT GTATCCACCA CTTCTTCAAA TGCTTTACAT 7380

TGAGAAATGC ACAAACTGTT TTTTTGATTA CGAATAAGCA TTTTAATGCG TTCTGCATCG 7440

TCTTCTAATA ATTGAATGGC AAATTTCTTT GAAATTCCTT CGTCCATTTG TACAGAACCT 7500

CCCTATCTTA ATTAAATAAT TATAACATAA TTCTCAGTTA GATACATGAT GAkTTCAGAA 7560

AATTTCCTTT AAGAAATTTT CTGAAACCAA GCGATATATT AAGTAAGAAA TCTCTTTATA 7620

TAACAAGCTT ATTTTATCAC TACAACAACA AATACATTAA AAGACTTTTT TCTTATTTTC 7680

TGCTTTTAAT TTTAAATAAA ATAACAAACT ATATCATTTT TTTGTCTCTT GATATTCTTT 7740

AATTACTTCT AAAAAGGCGG TGCCATATyT ATCTAATTTA TTTTGGCCGA CCCCTTTTAT 7800

TkGCAATAAT TGAATCGTTG TTTGCGGTAA TTTTTCACAC ATTTCTTTAA GTGTACTATC 7860

GGAAAAGACT ACATATGGCG GCACGCCTGC CTCTTGAGCC ATGTCCATAC GTAATTCGCG 7920

TAATCGTTCA AACAAGGCAT CATCTACTGC TACTTTTCGA ACTTTTTGGT CTTCTTTTCG 7980

GAAGACTTTC CGTTCCCCTA AAAGCACCTG AACACCTTCT GTCGAGACTG ATAACAACGG 8040

AAATTGCCCA TCAGAAGGAA TTAAATAACG TTCTGCTGTC AAATAATCAA TCAGTTGTGT 8100

GACTTCTTTT TGCGTGCGAC CTTTCATTAG TCCGTATGTT GGCAAACGAT CAAAATGCCA 8160

TTGATCAATT TTTTGATCTT TTGAGCCAGT CAAGACTTTA CCAACTAAGC CCTTACCGAA 8220

ACGTTCTCCC ATACGTTTTA CACAAGACAA AACTTTTTGG GCATCGACCG TAATATCGAC 8280

TAATTCTCGA TTATCCAGAC AATTCGAGCA GCGACCGCAG TCCGTCCCTT CTTCACCAAA 8340

ATAGCGTAAA ATATATTTTT GTAAACACAT TTGTGCATTC GCATATTGTG ACATTTCCCG 8400

TAATTTCAAG TACTCTTTCT GCTTATAATC AATGGTCATT TCAGATTGTT CAATAAAGTA 8460

TTGTTGAATT TGCAAGTCTT GTGGAGCAAA CATTAACACA GCATCACTAG GCAGACCGTC 8520

ACGACCAGCT CGGCCGGCTT CTTGGTAATA AGATTCAATA TTCCCTGGTA CTTGTGCGTG 8580

AATGACAAAA CGAACATTAC TCTTATTAAT CCCCATACCA AAGGCATTTG TTGCTACCAT 8640

CACTTGCACT TGATCATACA AGAATGCTTC TTGATTTTCA CTACGTAACT GTTCACTCAT 8700

GCCCCCATGA TACATCCCAG CTGCAATTTT TTTACTTTCT AATAAATGAT AAATTCGTTC 8760

AACTTCCTTC CGCGTGCTTG CATAAATAAT ACCTGATTGT CCAGTATTCA TCTTTAAATA 8820

TTCCAAAAGA AAAACGTCTC TATTTTGATC TTTCACGACT TGAAAAGATA AATTTTCTCG 8880

AGCGAATCCT GTTTTGATTT CATTTTCTGA TGGAATTCTT AATTGCTTGA CAATATCCTC 8940

CGCTACTTGT GGCGTTGCTG TTGCTGTTAG TGCAATCACA GTTGGTTGCT GTTGAAATTG 9000

ATCGATAATC TCTGCTAAAC GCAAATAGCT TGGACGAAAA TCATGCCCCC ATTGTGAAAT 9060

ACAGTGTGCC TCATCTACTG CCAATAAATC AATCGGTACA TGTGTCAGCA TCTGTTGAAA 9120

ATCATAAGAT TCTAATCGCT CTGGCGCAAC ATACAATAGC TTGACTTCTT TGTTAACCGC 9180

TAATTGAATA CGATGATTCA TCTCTTGATA GGAAATGGTA CTATTGATAT ATGTAGCTGG 9240

AATCCCCATT AAATTCAGGG CGTCTACTTG ATCTTTCATT AAAGAAATTA ATGGTGAAAT 9300

CACCAATGTT AAATTATCCA GAAGTAAGGC TGGCAACTGA TAACAAATCG ATTTGCCTCC 9360

ACCTGTTGGC ATAATGCCTA AGACATTTTC TTGTCGCAAT ACATGACGAA TAATCGTTTC 9420

TTGTCCTGGC CGAAAATCAT CATAGCCAAA TGTATCTTTT AATAATTCTT GTAATGCACT 9480

CATTTATTGC GCTCCAATCT TTGTTCTACA AGTCATTTTA ACACGCTAAT CTAACAGGAA 9540

CAAGGTGTGT TAAAACAAGA CACAATCGTG AAACACCAGT TAACGTGAAA CATCAAAAAG 9600

AATGAAAAAT TGTGAAAGCT TGCTATCAAA AGCTTTCAAA AATAGCGGCA TTTGCTTGTT 9660

GTGGATAACG AGTAGTTATC CACACATAAA AAAATAACCT TATCTTAAAA TACGCATTAA 9720

TTCGTTTTTG CCATCAAAAA AACCAACTAT TTTCGTAGTA AAACGAGCAT AGTTGGTTTA 9780

AATGTTTATT TGTCGAAAAA TTATTCAGCT TTTTTCTTTT TCTTGCTTGC TTTTTCGCGT 9840

TCGTTTTTGT TAAGAATTTG TTTTCTTAAA CGGATGCTTT CTGGTGTTAC TTCACAGTAT 9900

TCATCTTCAT TCAAGAATTC TAATGATTCT TCTAACGTTA ATTTTTTCGG TTTTTTAATG 9960

ACTGATGTTT GGTCTTTCGT TGCCGAACGA ACGTTAGTCA TTTGTTTAGC TTTTGTGATG 10020

TTCACAGTTA AATCGTTGTC GCGGTTGTTT TCCCCAACAA TCATTCCTTC GTAAACTTCT 10080

GTCGTTGGTT CAACGAAAAC TGTACCACGT TCTTCAATAC TCATGATTGA GTAAGTCGTT 10140

GCTTTACCAG TATCGATTGA AACTAGTGCA CCTTGATGAC GGCCACCAAT CGTTCCTTGA 10200

ATCATTGGTA AGTATTGATC GAAGGTGTGG TTCATAATAC CATAACCACG AGTCATTGAT 10260

AAGAATTCTG TTGAATAACC AATTAAGCCA CGAGCTGGTG CTAAGAAGAT TAAACGCATT 10320

TGTCCGTTAC CGGCATTAAT CATATCTTGC ATTTCGCCTT TACGTAAACT TAGTGATTCA 10380

ATCACGCTAC CCATGTATTC TTCAGGTGTG TCAATTTGAA CACGTTCAAA TGGTTCACAT 10440

TTAACTCCAT CAATTTCACG TTCAATAACT TCTGGACGAG AAACTTGTAA TTCATAGCCT 10500

TCACGACGCA TGTTTTCAAT TAAAATTGAT AAATGCAATT CGCCACGACC TGATACAGTC 10560

CAAGAATCTG GGCCAATTGG ATCAACACGT AAAGATACGT CTGTTTGTAA CTCAGCCATT 10620

AAACGTTCTT CGATTTTACG AGCGGTGATG TATTTTCCTT CACGTCCCGC AAATGGAGAA 10680

TTGTTAACTA AGAAAGTCAT TTGTAAGGTT GGCTCATCAA TGTGTAGAAT TGGTAATGCT 10740

TCTTGATTGT GAACATCTAC AACTGTTTCC CCAACGAAAA TGTCTTCCAT TCCAGAAACG 10800

GCAATTAAAT CGCCCGCTTT TGCTTCATCA ATTTCCACAC GTTGTAAGCC AAAGAAACCT 10860

AAAATTTTCG TTACACGGAA ATTTTTCACG CTGCCATCTA ATTTCATCAA CGCAACTTGG 10920

TCGCCGACTT TCATTGTGCC ACGGAACACA CGGCCAATCC CAATACGTCC AACGTAATCG 10980

TTGTAGTCTA GTAATGAGAC TTGGAATTGA AGTGGTTCGT CTGAATTGTC AACTGGAGCT 11040

GGCACATGTT CAATAATTTT ATCAAAAATT GGGGCCATTG TTGGCTCTTG ATCTGCTGGA 11100

TCATCTGATT CACTTGAAGT TCCGTTTAAA GCAGAAGCAT AAACAACTGG GAAATCTAAT 11160

TGATCGTCGT CTGCACCTAA TTCGATGAAT AACTCTAAAA CTTCATCTAC TACGTGTTCA 11220

GGACGAGCAG AAGGTTTGTC AATTTTGTTA ACAACCACGA TTGGTGTTAC TTTTTGTTCT 11280

AATGCTTTTT TCAATACGAA ACGTGTTTGA GGCATTGTTC CTTCATACGC ATCGACAACT 11340

AAAACAACAC CGTCTACCAT TTTCATGATA CGTTCTACTT CACCACCGAA GTCCGCGTGT 11400

CCTGGTGTAT CTAGAATGTT GATACGTGTA CCGTTATAAT CTACGGCTGT ATTTTTTGCT 11460

AAGATAGTAA TTCCACGTTC ACTTTCAAGT GCATTGGAAT CCATTGCACG TTCTTGTAAT 11520

TGTGTGTGTC CATCTAAAGT GTCAGATTGT TTTAAAAGTT CATCTACTAA GGTTGTTTTA 11580

CCATGGTCAA CGTGGGCGAT AATTGCCACG TTACGAATAT CATCTCTGTA TTTCAATTTA 11640

ATTGCTCCTT ATCTATATAC GCTTATGTTA CTTAAGTCCA TAAAmGCCGA CTGCTTATTA 11700
TmGCAAACTT TTTTCATTAA AAAAAGcTTT CGCCTGCAAA AATAATGAAA GTTTTAGAAA 11760
TTATTCGGCT TTTTCTGAAA ACAAGGACAT TTTTTtACAA ATTTtyCTCT TTTAAAAGAG 11820
GAAGAGCCTT ACCTTTCAGC AACGCTCTTC TTTGATTATT CCGATTCATT TAAATATCGG 11880
GTAAATACTT CATCATAGGT TTCAGGAGTT GCTGCAATAA AGTATTCACG ACCCGCAAAA 11940
GTTAATGGTT TTCCTGTAAT ACCAGAGTAT TTAAACCCAA ATTCTTCCAA AAGTACTAAG 12000
CCTGCTGCAT AATCCCAAGG ACTTAGATTA GATAAATAGC CATGATGGTT GCCTTTTAAC 12060
ATAGCAATAA TTTCCAAACC TGCACAACCG CTAATACGCA CGCCCATGCT TGCTTGTCCG 12120
ATTTCCGTTG CGTGATGAAT ATTTTTTCCA TGCATGTAAG CATTCATGCC CCATAATCCA 12180
TCTGCCAAGG CTTTCATAGT AGGCGCTTTT AATAATTGAT TATTCCGATA AACGCCAAGC 12240
CCTTTGCCGC CCCAATAAAG TTCTTCTCGC ATCACATCAT AGATAAAACC TAATTTACCA 12300
ATGCCGTCTT CATAAACAGC TAACATAATA CAAAAATTTT CTCTTTCCAT CACGAAATTC 12360
ATTGTGCCAT CAATCGGATC GATGATCCAC ACGCGACCAG CAAAACTTTT TAAAGTATTG 12420
TAGCCTTTTT CTTCCCCTAA AATTTGATCT TCGGGAAAAT TTGTTTGAAT TTTATTCATT 12480
AAAAATTCTT GGGTTTGTTC ATCCATATTT GTCACTAAAT CGGTGCGGCC ATTTTTCTGT 12540
TGAACGGTTA ACTCTGATTC CAAGTTCATT CGAATCACAT CGGCTGCTTC AAAAAGCCAA 12600
CTTTGAATTG TTTGAATAAA TTTTTGCGTT TCAGTCATCT AATCATCCCC CAGCTTTAAT 12660
TTACCGCTTC CCTGTTTTTT AGCTTGTTGA AGCGTACGAT ATAACGAATA TCCTGAAGCA 12720
TTTTCAAATT CACGACCTAA GCGTTTTTCT TCGCCGATAC TTTTAACGAC CGTTTTAAAT 12780
TGTTGATAAG TTGTTAAAAA TTTGTCAACA GGCAAGCCTG TTTCGTTGGC TTGCTCAACT 12840
GCTGTCCACA TATTAGkGAC AATCACCATT TCCtCTGTCG TCCAATCTAA ATCTAATGGA 12900
TATTGATAGT CTTTCATGTC TTCACCTCAT CTTCTCTAGT TTAGCATATT TTCATAAAAn 12960
AACATCTTAT TTACGTCGAT TTTTTAATGC ATTAAAGAGA GGATATTCTG CTGTTCCTTT 13020
GCGATACAAA CtTTGTTGTG AAGCATAAAT ACCGCTATTT CCTGAACACA TAAAACTGAT 13080
TAAGCAAACC ATGAAAAAAT AGACCGCCGC TTCACTACCA AATAGTTCAA TTCCCATAAT 13140
GAAACAAGCA ATCGGCGTAT TCGTGGCGCC AGAAAAAACA CCAATAAATC CTAAACCTGC 13200
TAAAAAAGGA ATTGATAGAT GTAACAAAGG GGCCAAAGAA CTCCCTAACG TAGCCCCAAT 13260
TTCAAACAAT GGCGTCACTT CTCCTCCTTG GAATCCGGCG CCCAAGGATA ACACCGTAAA 13320
GAATAATTTC CCAACAAAAT CAAAAGGCTG CGCATTACCT GCAAAAGCAT CTTCTAGTAA 13380
AGGTAAACTC AAGCCTAAAT AACGTTGGTT GTTTAAGACG AGCACAAACA GAACGACAAT 13440
GGCAGCGCCA AGAAACGCAC GAAGGTAGAC ATTAGCAAAC CAATTTGCAT AGGTCTTTTT 13500
TAAAAAGACA ATGGAACGAC TAAAAACCCA ACCTGCCAAA CCAAAACAAA TCGATGCCAA 13560
AAATAATTTT ACAAATAGTT CAACCGACCA AGTCGGAATT TTGCCCATCG GATAATGCGT 13620

ATGTGTTACA CCAAAACTTT CAGTTACAAA ATTAGCAAAT AAGCCTGCAA AAAAACTTGG 13680

GAAAAGTGCC TCTGCACGAA CTTTTCCAAC CGCCAAAACT TCTAATCCGA AAACCGTTCC 13740

CGCTAATGGT GTGCCAAAAA CAGAACTAAA ACCTGCACTA ATTCCGCTAA TAATAATTAC 13800

TTGTCGTTCC AAAGCGGACA GTTTAAAGAC TTTTCCTACA GCATTTGCGA TTGCGCCCCC 13860

CATTTGCACC GCCGTCCCTT CACGACCAAC AGAACCACCA AATAAGTGCG TAGTAATCGT 13920

GCCAAACAAT GTTAAAGGAA TTAAGCGCAA GGGAATATCT TCTTCCCCGC CATTGCCTTG 13980

TTCAATCACT AAGTTATTTC CTCGGCTAGC ATTTTTCCCA AAACGCGTAT AGAAATACGT 14040

AAAAACGACA CCGCTAATGG GGAGGATAAA GAGTAACCAA GGATTTGCTA AACGAACATC 14100

TGTTACATAG GTGAGACTCT TTAAGAAAAA GGCGGATAAC GCCCCCATAA ATAAACCAAT 14160

AATTGCGGTA ATGAGCAACC ATTTTGCCAT ATAACCGCCG ATTGATAGTA GTTGCTTTGT 14220

TTCTTTTTTC ACTTTACATG CTCCTTTACA ATTTTCTTAT CGCCAATCCT AGCGTACAGA 14280

AAAAGCCTAC CAAAGGGTAC ACTTCTCCCG TTTGkaGGCG TCATTAATTT ATTCGTCCmA 14340

TAAATAGTTC GTGGTGAACA CCATCACCAT GTTGCTATTA AACTAATwAa ATwATACCaC 14400

GCAaCTCTGG AAAGTGCAAT AAAAAaTCaA TTGCGTTCCG TTTTtATCCA CGTGCTATTA 14460

CCTTTGATAA ACCGATAATA AGCCATCACC ACACTAACGA ATAAAATTAA GTACATGTAG 14520

GAAATAGTAA AAGGCAAAAT AATCAGCCCC AGTTTTGCAG GTACTTTTAG TTGAAAATTT 14580

TTCAACTCGC GATAATAGTT GATTGTAAAA AAAATGCCGA AAACTAACGA AATACTGACG 14640

GCCAAAACAA AACTAATGAC AAGCGATAAC GAAAAGCCTA ACCTATACAA GTGTTGCAGT 14700

TGAAGCCCAG TTAAGACAAA AACGAGCATC CCCGCTACGA GATTAATAAA GGGTTGCGCT 14760

AAAAAATAAT AGATTCCACA CTTCTGCCTT AGCGTTAATG ACTTTGATTT AATCACCCTG 14820

GGTAAATACG CTAGACAATC TAAATTTCCT TGAACCCAAC GACTTCGTTG ACGAATAAAA 14880

CGTTTCACAT CTCTCAGTGC TTCTTGAGCC ATAAAGGCCT CGTCAATATA AGCAATCGAT 14940

AACTCTTTCA GCATTAATTT AATGGTTAAC TCATAGTCAT CCAACAAAGC ATTTCCCCAA 15000

GGCTGATAGC CTAAGTGTTC GGTAATCATT GATAACCGAA AAAACTGACC ATTGCCACCA 15060

AGTCCGATGG TTTTTGTTGC TCGACGCGCT TTTTGAATAA AATTATTAAT CGTAAAAAAC 15120

TCAACATCTT GTGCAATAAA CATAAAATTT GTGACATCTT TCATTTTGAT CCTTGTTTGC 15180

GCAscACAGA CTGTTGTATC TGTGAAAACA GCAGTTAGCT CATTGACAAT ATTTTCACTA 15240

AAAAAACCAT CTGCATCTAC CACACCCACT AAAACTTGAT CCCAAGAATA GTGTCCTTTT 15300

TCTGCATCTG TTAATACAAT TTGGAAAGCG TGGTTCAAAG CGGCCCCTTT GCCTTGTTGC 15360

GCATTGGGAA ATTGACGAGA CAGAACTTGG ATTGGACCAG CAATACTGTC CATCATTGCT 15420

TTGGTTTGGT CCGTTGATCC ATCATCAATC GCAATTATTT TTTTCTTAAT TGGTAAAGCA 15480

ACCAGTTTTT TTAATGTTTC TCCGATGACT TTCTCCTCGT TTACACACGG TACAAGAAAG 15540

TAGCAGAAAA ATTCTTCTGC TACTTCGACT TCTCTTCTTG TTTTTTTCTT GTTCGAAAAC 15600

TGAGAAATAA ACTTGTAAAT AAATAGGTTA AGAAGATGAA CATAATTACT TTGTATAACT 15660

GATACAAATA AATCATGTAA TTTGACCTCT TTTAAAAAAA TAACGGTATA GAATGTAGGC 15720

AGTGACAATC GAAACTAGAA TAACGCCAAT TAAATCAACC ACAAAAAAGG CGCCTGTTGA 15780

AATTACCCCA AGTTTCGGTA AAGCTGATtG GCTCATtAAA TACCCCTTTC TCTCTATTTG 15840

TTTTTACCAA CACACCACCA CGATGACCAT CACACACCCC CTATAATAAT ACTTTGGaGG 15900

ATTATAGTAC ACAGGTGGAT TATAGTACGT AAmTGGcGGA ATATAyCCTG GcCAATATGT 15960

TGGTGGTGGC GGATTATAAA CCGGTAwTTT TCGGTTTTCT CGAACAATTG cAATTGTTTG 16020

ACCCGCATGA TTTTGCGTCA GACTGACGCG TGTTGGCGTA GTATTTCGTG AATAACCAGT 16080

CGGCGCTTGA ACTTCTCTAA TTTCATATAC CCGGTTGACA GATAGCCCAC TAATAGTGAC 16140

TGTACCATCC CAACCTGTTG TTACTCTTTG TGTATAGCCA TTAGTTGTAT CTTTCCATTC 16200

AAAAACAGCG CCCGCTAAAC GTGCTTTTGT TTCAACATCT TGTTTAATAA TTCTGATAGC 16260

ACCATAGTTT GCTTTATTGT TTAACACTAA CGTATACGTT TTTGTGCCTT GACCTCGTAA 16320

ATCAATCCGA TGGACCCTAC TATCCAACGC ATACCCAGTT GGCGCTTGTA TCTCGGTGAT 16380

TTCATAAATA CGATCCGACG ATAATCCTGG AATCAGTACT TCACCATTTC CATTTGCTGT 16440

AAATGTTGCC GTATAGTTAT TGGTAGTATC GCGAAATCTG AATTTCGCTC CTGCTAAACG 16500

TTTTCTTGTA TCCGTATCTT GCTTAATAAT TTTGATGGCG CCTTTTTGTG CTTGATTATT 16560

TAAGACTAAT GTATAGGTTT TACCACCTTG TCCACGTAAA TCTACTTTAT GAACACGTCT 16620

ATCTATGATA TAACCATTCG GCGCTTTGAT TTCTGTGATC CCATAGATAC GATCACCGGC 16680

TAATCCAGTA ATCAGTACTT CACCGTTCCT ATTCGCTGTA AACGTTGCAG TATAATTATT 16740

GGTTGTATCT CGGAATTTAA ATTGTGCGCC TGCTAAACGT TGTTTCGTAT ATCCATCTTG 16800

CTTAACAATT TTGATGGCAC CTTTTTGTGC TTGATTATTT AAGACTAATG TATAGGTTTT 16860

ACCACCTTGT CCACGTAAAT CTACTTTATG AACACGTCTA TCTATGATAT AGCCATTCGG 16920

TGCTTTTATT TCTGTGACCT CATAGATACG ATCACCGGCT AATCCAGTAA TCAGTACTTC 16980

ACCGTTCCTA TTCGCTGTAA ACGTTGCAGT ATAATTATTG GTTGTAtCTC GGAATTTAAA 17040

TTGTGCGCCT GCTAAACGTT GTTTCGTATA TCCATCTTGC TTAACAATTT TGATGGCGCC 17100

TTTTTGAGCT TGATTATTTT TGGTGAACGT GTAGGTTTTG CCAGCATAAT TCGCAGGTAA 17160

AAGAATTTTA TGCACCGTTC TGTCAAGTAT ATAACCATTT GGTGCTTTTA TTTCTGTAAT 17220

TTCATAAATC CGGTTGACGG AAAAATTATA CAAAGTTGCT TCACCATTAG CATTAGCAGT 17280

TACTACGCCT GTATTCCCAT TCACTGTGTC ACGCCATCTA AATTGTGCGC CTGCTAAACG 17340

TTGTTTTGaG TAACCATCTT GTTTAATTAA TTTTAAATTG CCACGTAGCG CAACATTTTC 17400

AATGGTATAA TAACCAATCT TGTCAGCAAA ATTAGTCGTT AGGGTAACTC GGTAAGTCTT 17460

GTTATTTAAG ATATAGCCGT TTGGTGCTTT AATTTCTGTA ATTTCGTATA CACGATTGAC 17520

TGGCAAATTC GTTATCAGGG CTTGTCCGTT TTTGTCAGCT ACAACTGTTC CTTCTTTTCC 17580

TGTCACCGTA TCACGCCATT TGAACTCTGC ACCTGCTAGA CGTTTTCTTG AATCTTTGTC 17640

TACTTTGATT ATTGCTAGAC TACCTTTTTG CGCTGTATTT TCAATGGTAT AATAACCAAT 17700

TTTATCwGCA AAATTAGTCG TTAAGATAAC TCGGTAAGTC TTGTTATTTA AGATATAGCC 17760

rTTTGrTGCT TTAATTTCyG TAATTTCGTA TACACGATTG ACTGGCAAAT ctGTTATCAG 17820

kGCTTGwCCr TTTTTGTCAG CTACGACTGT TCCTTCTTTT CCTGTCACCG TATCACGCCA 17880

TTTAAACTCC GCACCTGCTA GACGTTTTCT TGAATCTTTG TCTACTTTGA TTATTGCTAG 17940

ACTACCTTTT TGCGCTGTAT TTTCAATGAT TACTGTCACT ACTTTATTAG CTTGTGCAGT 18000

TGTTAGAGTG ACTTTATGCA CAGTTTTATC TAATACATAC CCCACTGGTG CTTTCGTTTC 18060

TGTGATTTCA TACGTACGAT TGACTGATAA GTTGGGAATT GTGATGGTTC CATCGGTTCC 18120

AACTGTCACT rTTCCTGTTT TTCCTGTCAC TGTATCyTTC CACTGAAATT CTGCGCCTGT 18180

TAAACGTTTT TTrCTkTCTT TGTCTTGTTT AACAATTTTG ATGGAGCCTT TTTGCGCTGT 18240

ATTTTCAATG ATTACTGTCA CTACTTTATT TGCCTGTGCA GTTGTTAAAG TGACTTTATG 18300

CACGGTTTTG TCTAGTACAT AACCTGTGGG TGCTTTTGTT TCTGTAAGTT CATACGTACG 18360

ATTAACTGCT AAATTGGGAA TTGTGACCGT TCCGTCGGTT CCAACTGTCA CTGTTCCTGT 18420

TTTTCCTGTC ACTGTATCCT TCCACTGAAA TTCTGCGCCT GTTAAACGTT TTTTGCTTTC 18480

TTTGTCTTGT TTAACAATTT TGATGGAGCC AATTTGTACT TGCTGCTTAT TCTTCACCCG 18540

ATATGTTAAA GTGTAGTCGG CTTTTCCAGC AAATGAAACA GGATGTTTAG TTTCGTCTAA 18600

TTCATAGCCA TTAGGCGCCT TCGTTTCAAC TAAATAATAT TCTCCATCAG GTGATAAATT 18660

TGAAAATAAC AGCTCGCCAT TTTCATTCGT GACACCTTTT TCTATTAGAT TATTTTTTkG 18720

GTCATACAAC GAAAATTCAG CACCTGcTAA ACGTTTCACC TCATTATTTT CATCTACTTT 18780

TTCyACTTTT AACTGGkTTn CTAAACGCGt ATTTcTTTTt tCAAGCGAAT CGTCTGkGTT 18840

TGTTCTAcTT AATTnCAAAT GGAaCTGGTG TTGCATCTAA CTTGTAACCA GTTGGTGCTT 18900

TTGTTTCAAT GAATGCATAC TGACCAGCTT CCAAACCAGC TAAACGAATT TGACCTGCTT 18960

GATTGGTGGT TAATGTTTCT TGAACAACTT GTTGTGTTTT TGTGTTCACG ACTTTGAAAA 19020

CAGCACCTTC TAAAACTTTA GCAGTTACTG CAcAATTTTT TCyAAAaCGA CACCCCyTCG 19080

ACCTGTTCCA TCACTACTCT CGTCTGCTGT GATGTAGGCA ATATCATCAT CTTCCCAaGC 19140

ATCGAATCCC TCACTAGAAA TACGCATATA CGTGTAATTT GATAGCACGG CATCACTTGG 19200

AATATTTTCA TTTGGAATCG TTTCTAAATA AATATAATAG GCATTTTTAT TGGAAAGATA 19260

ACTGCCCGTG GCTGTTGGTA TTCCTTCAAA TTTCTCTGCT TCATAATTAA ATCCAGATTC 19320

AGAGGCATTT TTTTGGAAAA ATTGACTGGC ATCATAGCTA TATTCGTCAT GTTCCAAATT 19380

CCGAATCATT TCACCATTTG TTTCTAAAAT AGGCTGCTTG GTGAGATATG CAGCAATACT 19440

GTCAATAGCT AATGTTTGGT TTGGACCTAC ATTATCTAAA AATTTAAAGG TTGTTTtTTC 19500

TGAACCTGAA AGACCTGCTT GATTATAGTT AAAATAAATG ACCCAAGCAA TTCTTCCCTC 19560

TGAGGTAAAT CCTAACGGGA CTTTTCGAAC CAGTTCAGTT GGCGGAACTT TCTCCACGTA 19620

TTGCCCGATA TTAATTTTGA CAATTTTTTG TGAATCCCCA AAAGTAAAAA CGACTTGATC 19680

TTCAAAATCT TCTGATAAAT TTTGATTATA TCGTAGCTTC AGTTGCAAAT CTGCTGAGTA 19740

ATCTGTAAGC GTTTGTAAGG CTTTAATCGT CCAAGTACCA TCATCGTTTG GAATCACTTG 19800

AAGTGCGTCT CCAATTTCAA AAGGTGTCAT TGAATAGGAA TATTTTTCGC TAGCCATCCT 19860

TAATGTCATC GTCTCTCCGC TTGCTAATTC AGAGATTCCT AATGACAAAG AAACTGTCGC 19920

CCAATCTTTT GTAGCAACAA TTTCCGTTTT TTCCCCCGGn yATTCAACAG CTATAGACCG 19980

GATTTGTGTG CCACTTTCAT TATCCATTGC ATAAGAAATA ATACTTGTAT TAGATCCTAT 20040

AATAGTAAAG ATAATTAAAC AAATCGAAAA TA 20072

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 688 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

TTGAnACAGA TGGAGTnnGC ACTGGCTATA CAGTAAGATA TGATCTTGGT GGAGAAAAAG 60

AnAAAGACAT AACAGTCAAT TTACCCGGAG CAGaATTACC ACTTAGTTTA AAAGTAAAAG 120

CACATTACAC AAATGGATTA CCACTTGAAG CGGGAAAAGA AGTACGCTTA AGAAATTTAA 180

CAGATGGATC AACAGAAGTA mTTAAGAAAC ArGTAGACGC TAACGGAGAA GTTGTATTTA 240

CTGAGCAAGA TGGTATAAAA AAGGAAGTAA ATTATGGTAT TGAAACAGAT GGAGTAAGAA 300

ATGGCTATAC AGTAAGATAT GATCTTGGTG GAGAAAAAGA AAAAGACATA ACAGTCAATT 360

TACCCGGAGC AGAATTACCA CTTAGTTTAA AAGTAAAAGC ACATTACACA AATGGATTAC 420

CACTTGAAGC GGGAAAAGAA GTACGCTTAA GAAATTTAAC AGATGGATCA ACAGAAGTAC 480

TTAAGAAACA AGTAGACGCT AACGGAGAAG TTGTATTTAC TGAGCAAGAT GGTATAAAAA 540

AGGAAGTAAA TTATGGTATT GAAACAGATG GAGTAAGAAA TGGCTATACA GTAAGATATG 600

ATCTTGGTGG AGAAAAAGAA AAAGACATAA CAGTCAATTT ACCCGGAGCA GAATTACCAC 660

TTAGTTTAAA AGTAAAAGCA CATTACAC 688

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10194 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

nGATAAGGTC CTATAAATAG nCCGGATATn TTTTTATTAT TAAGATCTAG TACTTTCCTC 60

TTACTATGAG TGTACTAGAT TCTTTTTTTG CCAACTTTAA AGAACTAGAC CAATTTTCGG 120

TTTTTAGATT TTTGTTGACA AGTACCTACT TTTAATTGTA TGATTGTTTA TGTTTCTACG 180

GAATATATTT ACAAAGGAGG TATATAGACC AATATCGGAT AGCGCCAGAC CTGAACGTTC 240

AGGTGACGAG GAGAGAGCTT ATCGAAGATT CGGCGGGTGG CTCTAGGGAC TGCACTCTAC 300

AGATAACAAA GAAAAACTAA TTGTGAAGTT AGAACAAAGC GGTTATCACG CAGGTAGAAA 360

CATATAATGG GATTTGCTAG TTAAACAGGA CCTAGTAAAT ACAGAATTTC AATTGGAAAG 420

TAGGAATTAG CATTATGTGT GGAATTGTCG GAATGATTGG AAAAGAAAAT GTCACGGAAA 480

GTTTGATTCA AGGATTAACA AAATTGGAAT ATAGAGGTTA TGATTCTGCT GGGATTTGTG 540

TGAATGrTCm AAGCCAACAG CCCCATGTAG TTAArGCAGT TtGGGGgCAT TkCCsATTTa 600

GCAGCAAAAT TAACGCCAGA AATCGATgGA ACAGTTGGTA TTGGTnCATA CTCGTTGGGC 660

CACTCATGGC GAACCAACTG TCGCCAATGC ACATCCCCAT GTTTCAAGCG ATCAACGCTT 720

TGCATTAGTG CATAATGGTG TCATTGAAAA CTTTGAAGAA TTAAAAGAAC AATTTTTAAG 780

TGGTGCTCAC TTAATTGGTG ATACAGaTAC TGAAATCGTC GTTCAATTAA TTGcTCATTT 840

TGCTGAAAAT GGTTTTTCAA CAAAAGATGn CTTTTAAAGC GGCTTTAGCm GAwATTaAAG 900

GTTCATACGC TTTTGCATTA ATGGATAAAA TGGaCCCCAA TATTATCTAT GTAGCGAAAA 960

ATAAAAGCCC GTTACTTATC GGTCTAgGTG mAGGCtTaAT GkTTATtTGT AGTGaTGCGA 1020

TGGcGATGAT TAAAGAAcCA ATCAATTtGT TGAGATtGTT GATGGTGAGA TTGTAACTGT 1080

TACAGCGGAC AAAGTAGTAA TCGAAACGCC AGAAGGACAA GTAATCAACC GAGAACCTTT 1140

TGAAGCGCAA TTAGATTTAA ATGACATTGA AAAAGGAACA TATCCGTTTT ATATGCTCAA 1200

AGAAATTGAT GAACAACCTG CAGTGATGCG CAAAATTGTT CAAGCGTATA CTCAAGGTGA 1260

AAAAGTGACT TTAGATGAGG CGTTACTGGA AAAAATTAAT CAATGTAACC GAATTTACAT 1320

TATTGCTTGT GGGACAAGTT ATAATGCAGG TTGGGTCGGA AAATCTTTAA TTGAAACGAT 1380

TGCCAGTATT CCTGTGGAAG TTCATTTATC AAGCGAATTT GGCTACAACA TGCCTATTTT 1440

ATCTGAAAAA CCGTTCTTCA TTTTCTTAAG TCAAAGTGGT GAAACGGCAG ATAGTCGCCA 1500

AGTACTGGTA CAGGTTAATG AATTAGGGCA TCCATCACTT ACATTAACAA ATGTTGCTGG 1560

CTCAACATTA TCACGTGAAG CGGACGATAC CTTGTTATTA CACGCTGGAC CAGAAATTGC 1620

TGTCGCTTCC ACCAAAGCTT ATACTGCTCA AATCGCTGTC TTAGCTATTT TAGCAAAAGC 1680

AATTGGCGAA GAACGCCAAA CATTAGCAGC CGTCTCATTT GATGTGGCTC ACGAATTGAG 1740

CGTGGTAGCA GCTGTCATGG AAACCTTGAT TGATGAAAAA CAACAAATGA AAGCCTTGGT 1800

GGAAGATTAT CTGACGTATA CTCGAAATGC CTTCTATATC GGGCGAGGAG TCGACTATTA 1860

TGTGTCGATG GAAGCGGCCT TGAAGTTGAA AGAAATTTCC TACATTCAAG CAGAAGGCTT 1920

TGCGGCAGGT GAATTGAAGC ATGGAACGAT TGCCTTAATT GAAGAAGGAA CACCTGTGAT 1980

TGGGATTGTG ACAGATGCCA AAGTCGCTGC CCATACAAGA GGCAATTTAA AGGAAGTCGA 2040

AAGTCGTAGA GCAAGAAATA TTGTGATTGC TTTGGAATCC TTGGCAAAAC CACAGGACCA 2100

ACTGATTATT CCAGATGTTC ATCCGTTACT TTCTGCCTTA GTTAGCATTG TTCCAGGTCA 2160

ATTATTAGCG TACTATGCAA CCTTACAAAG AGGCTACGAT GTTGATAAAC CAAGGAACTT 2220

AGCCAAAAGT GTAcTGTTGA ATAAATCAAG AACCTGAGAG AAAGATGTTC AACGGGTTTT 2280

CATGAGTTTT AATATCTGAG ACAAAAATCA CTTTAGATTT TTGTCTCAGA TATTTTTTAT 2340

AAAAGAAAGA TATTCAGTCA AAAATGAAAG ATTCCAAAGA AATTCTTTTT TTCCGTTGAA 2400

AACTGTTTGC ATTTTAGCAA AATTAGATAT AATTAATAAG AACCGAaaAA TAAGAATTAG 2460

CCAGTGAGGT AACTTTGTTT GAAAGAAAGC AAAAAGTAGG TTAAGCTGAA AAGTGCTAAT 2520

TTATTTAACA AAAAAAGATT GTGAGTTGTT TGATATGGAT CAGTTTAGTA CATTAACTGA 2580

ACAAGAATTT ACTACCTTTG CAATGACACA TCCTGCAGGT AACTTTTTAG AAACACCCGA 2640

AATGAAACAT TTATTGGAAC GACGTGGCTG GCATTGTGAA TACGTTGGCG TCAAACGAGA 2700

AGGCCAACTA ATTGCTGCGT GTATCCTAAG CAAGAAAAAA GTCAAAATTG GTTATGCTTT 2760

CGACATTGAC GGCGGAATTT TGATGGATTA TACAGATAAA AAATGTGTAG AAGCCTTTTT 2820

CACAGGCTTA AAAAAATATG TGAAAAAAAA TGATGGCCTT TATTTAACCT TTACTCCAAA 2880

TAAACAAATT TGTTTACGTG ATTTTAATGG CGGTGAAGTT GAAAAGGTCA ATCAAGAAAC 2940

ATTTGACTAT TTCACATCAA TTGGTTTTGA ACACCAAGGG TTTGATGTGC ACAATTTTGA 3000

TGGTGCGCCA CGTTGGCTGT TTGTCAAAGA TATGGCTGGA TTAACAGAAG AAGATCTTTG 3060

GAAATCTTAT GGAAAAGATG CTAAATATGA CATCAAGAAA ACATGGGAAT ATGGAGTAAC 3120

AATCCGAGAA CTTCgTTACG AAGAGCTACC TCTCTTTAAG AAATTAACGG AAGAAACAAG 3180

TGCACGGCGT AATTTtGAaG ATAaGGATTt AGCTTATTAC CAaGCTGTTT ATGAaGAGTT 3240

TGgCGAaCGA gCGAAtTTAT GGTGcGGACT AAcTTTGCTA CTTATTTaGA AAATCTTCAT 3300

GAGAAATTAC GCAAATTACA AGAGACGCTC AACGAAgTCA ATGAAGCGTT AATTGCCAAT 3360

CCTAAAAGCC GTAAAAAAAA TAACCAGAAA CGTGAATTTG AAGATGAAGT TCGTACCGTG 3420

CGCAAACGGA TTGACGAGGC TAAAGAAATG AAAACATCAG ATGAACCTGA AATTTTAGCA 3480

GGTGCCTTAT TCATTGTTCA TCCACAAGAA GTTGTTTACT TGTTTAGTGG TACGTATGAA 3540

AAATATAAAC AATACTATGC ACCGTATTTA ATTCAACATA AAATGTTAAC TTACACAGTA 3600

GAAAATAACA TTCCCAAATA TAATTTTTAT GGCGTTGACG GTATCTTTGA TGGAAGTGAT 3660

GGGGTTCTTA AATTCAAACA ATCCTTTGGC GGTCATGTCG AAGAATTAAT GGGCAATTTC 3720

CAATGGAAAG CCAAACCAAT GAAATATGCC TTATATCACG CATTAAAAAC AATTAAAGAA 3780

AAAGTCTAAA ATTAAAACCC AATCCTTAAA AAGGAATGGC AAAGTGAGGC GTTTAAATGG 3840

AAAATCAAGT CGAGGTTATG ACGTATGCAC ATTTAAAAGA AATTATGCAG GCGTTAGAAG 3900

CAAACGAAGC GATTACAGAA GATACAAAAG TCTTTATTGA TACAGGATGG GATAGCGTGC 3960

AAGAAGTAGC CCCAGATGCG GTCAGTATTG AAAAAGTTGC TAAATTTACA GTAGCAGATG 4020

TTTTAACAAA CGAAAgTTTT GCTGGGTATA GTTTAGAAGA AAAGGCGGAG AAGATGAACG 4080

CTGAGGGCGA CCTAGAAACG GCAATTATTA TTCGTAATCT TTACTAAAAA AGTCCGTTTG 4140

AATAGTTAAA ATCGAGGTCG GGACAGAACC GTTTAGCTCC GAGCAACGCG TAsTGTTGGT 4200

TCTGTTCCCG CCGGTTATCA GGTTTTGAGC CTGGAGCAAA AATCCAAAGT GATTTTTGTC 4260

CCAGGCTCAT CTGCTTTAAA TAAGAAATAA AGGAGGTTAG ATATGAGTAC CATTCmTGrA 4320

TTyCCAAAAA ACTATGAACG TTTTATTGCA CAAGGAGAAG AAGCACTTGT AGAACATAAT 4380

CAGATCGCCG CTTTaGAAAA TTTTCAACAA GCCTATCAGC wACCAmCAAc GCCGCCAGTA 4440

AACCAAAAAA TTGTCCAATT ATTGTTGGAA ATGGGtGAaG CTGATGAAGC ACTAGCATTG 4500

GCAGAAGCGT TCCAAGAGTC TTATTTTGAG AACCTTGAAA CAGCGGCAAT TTATATGCAA 4560

ATTTATAGTC AAAGTCGTCG CTTTATTGAA GGGTATATTT TATTGAAGCA ATTACTTCAA 4620

ACTAAAAAAA TAACGCTAGC ACAACAAAAA ACGTTAGAAC AGCAATTAAT GCAAGTGGAA 4680

GAGGCGTATC AACAATTAGA AACGCAGCAA ATTCAAGCAA TTAAAAGGrA CTTGTTAGTA 4740

AGCGATCAAT TGCCCGTTTA TCAACAATTA GCGAACATTA AAACCAGTTT ATATTTGCCT 4800

AAACCAGTAT TTGTGGAAGT CGCAAAAGAT TTAGTTATGA ACCAAGCACT TAGTTATTTT 4860

GCTCGGGAAT GGTTCATTGA AGAATTAGCA CTTTTGCAAT TTTCAGAACC ACTGACTTTT 4920

CTTTGGTATG ACAATCAGCC TCAGACAGTC TTACTAGAAG GAAAAACTGG ACCGTTAAAT 4980

ACGCCGATCT ATTCCGAAAT TTGTACTGAA CTTAGGAATC GTTTAGAAAA TGATGATCCA 5040

ATTATGCTGC AGCATTTGGA AGAAGAAATT CGATTGCATC TTGCCTATTT GTATCCATTA 5100

GCAGAGACGG TTATTTCAGA TCCAACTATC TGGGTCTTGG GTTATTTGGC AACCTATTAT 5160

CCTGAATATA TTGAAAAAGA ATTGACTGAA GCAAAAGGCC ATCAAATCGA TGCTGTTCAA 5220

AAAGTTCAAC AGGCGATACG AACGGCATTT ACACAAATAA TGCTTTAAAA AATGGCACGT 5280

CCTAAAAGCG AGGACGTGCC ATTTTTTCGG TTTCAGAAAT TTGCTGAATC TGCTCTAATT 5340

GCTTTTCTTG ACTGATTATT GCGGATAATA GAAAAAAGCA AGACAGAAAT AAGTAAAAAG 5400

AATTCCTTGA ATCAAGCAAA TCACAACGAA CAATGGCGTT GCTCCTGGAA AAGAATCAGA 5460

AATCCGTGTA ATTAAAAATT CAACTAACCA TAAAATCATT GAAAATAAAA ATAATTTTTT 5520

CTGACGAGTT CTCATAGTCC GCAACAACCT TTCTGGTGAT TTTCACTTCT CTATCATAAC 5580

ATAGAAAACG AGCTGTTTTT TCTCAAAAAT TTTGTTACTT CATAAGAAAA ATGGGAAATA 5640

AGCAAGTTTC AGGAAGTACT AGCTTTACCG AAAAGAATAC TTAATCCATC TTAATCAAAC 5700

CATACAGTCA CGTTTATTGC AAAAGAAGTG TTTTTGGATG GAACAAATTA TTTCATAAAC 5760

ACAAATATTT CTTTGTTGTT TGTATAAATG TCATAAAAGT GATTACTTAC TACTTTTTGA 5820

TTTTTAGACA TATCCGTATC GGTCAGTGGT GCCATCTCTA ATCCTAAATT CGATGAAGTA 5880

TTGAACCAAA CGTAATTAGG GAAGTATAGA CCGTCATTTG GTGGCACAAT TTTTGAAAGA 5940

ATTGGGTAGT TTCGACTTGT ATTAGCATAA ACAGTTGAAT TTTTGAAAAA CGTATTTAAA 6000

TAAACTTTTT TTCGTTCGTT TGTTACAACA TTTTTTAAAT CATCTGTTAA AACAGCTGCT 6060

TGGCGATTGA ATGCTTCTTT TTGATAGTTT AATGTGGAAG AATACACCAA AACAAACGAT 6120

AAAATATAAT AAGATAACAA GCTGGCCACA AGTAACGAAG TTAGATAAAT TGCTTTCACT 6180

TGTCGATTGT TAAATAAACT TAACATGCTG ATTGTAACGA AAACAGCAAA TCCGTAAATA 6240

TAGCGAGGCG CATCACCTGC GATATTTCTT GAATAGACAA TAAAAATGCC AAAACTTAAG 6300

AGAGAAGCAA GTCCTAAATA CAAGATACAA TAAAGGAAAC TTAAGCCTAA GTTGATTTTG 6360

CTATTTGATA GATGACTAAT AAAAAATAAA ACAAGCAGTA GGAGCAGTAA TAAAAGCCAT 6420

AAACGATTAC TCTGGTTCAA TAGCTCTGAA AAATAATGGT AACTATTTTG ATAAGCCACA 6480

CTGGGAATAT CGTGCAAACT TGGTAAGGCC TGACTGCCCC CAATGTTGTT TGTTGCGGAA 6540

GGTACAAATT TTAACTCGAT AAGGTAGCTG ACTATAGAAA GTAGATAAGC GATAGCAGCA 6600

AGTGCGACTT TTTTGCAAAG AGCAATGAAA TTTTCTCCAG CAAGTAACTG CTTGAAAGTA 6660

AGGAATAAGA CGATAACTAT ATAAATTCCT GACGAGGCTT GGTAAGTATT AAACATAACA 6720

AACAGTGAAA AAACAGAAAC CAAGAAAAAC GTAAAAGAAT TACGCTGCCA CCAATAAAAA 6780

GGAAGAAACG AACAAAAAAT ACTCAATGCC ATATAGGGAC TATCAAACCT GAAACTAAGA 6840

CATTGGAGGA ACCAAGGATT GAGGCCAATT AAAGTAGCTA CTAAAGAGGA ACCCCAGCTA 6900

ATTTTTTGGT CATTGATGAT ATAAATTGCC aGTATGCTTG TGAGAGATAA GATAAGACCT 6960

GTTAGAATCG GTGTGGTCAA GCCCAGATCT GTCAAATGAC GACTTCCTTG AACAAACCAT 7020

GAGAGAAATT CACTGCCCCA ACGTGCGTAA GTCGTACCAA AATCTGTCAT TCCAGCGACT 7080

TGGCGAGCAT TATCATCTAA GTAAGGGAAA TTGATAATGC CGATGGATAA AACTGCCACT 7140

TGATAAATTA AAATGATCAA TAGAAGATAG GCTTTATTTT CAATACACCA ATTCTTTAAA 7200

TCAAATTTTT TTGATTCkGT AGGTAATTGT TTCATCATTG TACTCCACTC CAATAATTAT 7260

TTAmCaATTA AAATATmCCA TAACGAGTGG AATGAAGAAA TTGCTTTTGT TATTTTCTCA 7320

TCTTGACAGA ACACTAAGGG TAGGCAAAAC AAAAAAATCC CAACTCTTTA TTAAAGAGTT 7380

GGGATTTTAC CAAAAAGCCA CCTGCCAGCT TCATAGATAA TTATTCAAGA ATATTGCTGT 7440

TTAATTAAAA ACAATTCTAT ATGTTTTAAT ATATACATCA AGATAACAAT TTAGATATAC 7500

TTTGGGTAAT TTTTGGGTAT TTACAAAATA TCTTTGTATT TGTTTTCATA TTTTTCTTTG 7560

TyCTTTTTTC TATTTGCTTC AGTAGCTGAT GCGTATTCAT ATGACTGTAT TATCTCTTTA 7620

TGTCCTAGCT GTGTTTTGaT ATATATTGGA TCAGCTTCAG CTAGGaTTAA TAAGTCAGAA 7680

CGGGcATGTC TAAAACTATG AGATGAAATT CtTTCAATAT TTGcTAGTTC ACATGTTTCT 7740

TTTATATATT TATTCACAGA TGAATTTGAG ATGGGCCATG AATTTTTTTT ACTAACAAAA 7800

GCAAAATTTT GGGGATTACG TAAATCTCCT CTCAATATTT TTTTCTTTTG TTCCATTTCA 7860

TATTTTTTTA AATAATAGAA AAGAGAAGAA GGGACATCAA TATTTCGAAC ACCTGCTTCT 7920

GATTTTGTAG GGCCAATTCT ATACTGGCCA GCATCATACT TAtAGGCATT TTTTATCTTA 7980

ATAACTTCGT TACTTTCATC AATGCTATCC CAAGTCAACG CAGCAACCTC TTGATAACGA 8040

GCTCCAGTAT ATAGTTGTAC TAAAATGAAG TATTTACTTG TGGTGTTTTC ACGAGATAAT 8100

AAAGCTTCTT TAAATTTAGT TAAATCATCA ATAGATATGG TTTTATTTGA TTTCTTTTCT 8160

GGTACTTGGC CAGTTAACTT TATATTTCTT GTTGGATTTG AATGAATATA TTGCATGTCC 8220

AACGCTTCTA AAAACATAGA TTTTACAATG TTATGTCTGT TGCTCACCGT CTCAACTGCT 8280

AGTCCTTTTT CATTGATATT GCCAGGTTTT GCAAGCCAGT TAATCCATTG CTGATATTTG 8340

ACTTTGGTGA CGCCTTTTAG CAATTGTTTG TCCCCAAAGA AATCATGTAA ATTCTTTCTT 8400

GTTAAAGCTA TTTTTTGCAT ACGGCCAGCA GACACTTTAT TCTTTTTAAA TGTATCAACC 8460

CAAATATCAT AAAAATCTAA AAGTTTAATT GCTTCAGTGT TTTCACCACG CATAATTTTT 8520

GATTTAAACT CTGTTTCAGC ATCTCTACAA GCCTTCTTAG TAGATCTAGT AATCATTTTA 8580

TTTTTCCATT CATTTGTAAT AGAGTCTTTA TATTTAAATC TTAACTGCCA TTTTCCATTA 8640

GCTAATTTCT TTGGTTCTGC CATAAATACC AACCTCTCGG ATTTTTTTAA TTTAAATACA 8700

TTTAAGAACG ATTTTTTATT GAATATGTTA TAAGTAAGTA TCAAGTTATC GTATATCATA 8760

AATACGAATA CGAATATATG TtCGTTTTTT GTGaTAAAAA GAAAAACCCG AAGGTTTCTC 8820

TCTTTGTTAA TAAAATAATA AGAAGATATT AATAATCAAT AATTTCTACA AGTCCGCTAT 8880

CTGCACTTTG TGGAGTTATG TGTATTTCCA CATTAGCTTC TCTTTTTGCT CCATATGCAT 8940

TTACTATAGT AGCACTTACT TTTTGATACC ATTTACCATC TTTTTGCGTC CATGGTTGGg 9000

ATGACGCCGA TAATTGAATG CATTTTGCTT CCTTTATAAG GATACATATC ATTAAATTGA 9060

TGCTTAGCTA AAGTATCTAA GACGGTTTGT TGTTCTAAGG TTGGCTCATC TGCTAGCTTA 9120

GTAATGTCAT TAGTATTTTT ATCTTCGCTC GGTAAAGTAT TAGTAGCTAC ATCAGACTTC 9180

GCTTCAGTAC TTACTTCTTT TTTTTCAGAT ACATCAGATT GTTTATTATC TTGGCTACTA 9240

CTTTTAGAAA CATCAGAGTT TACTGTTTTk TGTTTGCTAG TTTCAGTACT kGATACTGGa 9300

TTTTTCaTAt ATTCTTGAAA ATCTTTTTCA TTTAAAAATC TAACAACAAA ATCTTTGTCG 9360

GCAGTTTCGA ATGTGAGAGT ATCTCCCTTT TTAAGAGATA CTTTTTTCTT GTTTCCCACT 9420

CCGTATAAGT CATGGTTACT ATTTTCTCCG CCTGATTTGA CGGTAATAAG AACTATGCTC 9480

GATTTATCAT CTAGTTGAGT TAATACAGCA TAGTAATCTC CAGGATCAAT ATCTTTTCCA 9540

aCTGAATATT CCCCTGACGC AACAAAGtCT TTTGAATTCG GTTGTTTAAC TGTGCTGCTA 9600

GATTGACTGG TTGACTTTTC AGTATCTTTA TTATCATTAT TTCCCGrACA AGCAGTAAGT 9660

AGTGAAATAC CACAAACACA AACCAATCCT AAACTAATTA ATTTTTTCAT CATAATTCCT 9720

CCTAAATAAT TTATTCAAAC CATTCATATA CCCCAATTGT ACCATCACAC CTAAAAATTA 9780

TTTGATAATT TTTATAGATC ATTCCATTTT CATATAATTT TGAGTAACTT TTGATTGCGT 9840

CTTGAAGATA TTCAGTAGTT ACTCCTAGGA ACTCTGCACA TTGCCAAACA TATATTAATC 9900

GCTCATGATA ACAATCTATA AAATCTTGTG GAGTTATTAA CAATTTACAT CCACAATCTC 9960

TAGCTCGTTG TTCTTGTTGT CTTTTTTCAT TACTATTTTG CTCTATAATA TCCCCAACAC 10020

TTGTTAAATG ATGACCAATT TCCTCTGCGA GAGTACCCAC CATTTCTTCG TTTGATTGAT 10080

TAGGGkTTAA ATATACTACA TTATTAATAT AGAGACCTTT CTGATTTcTG GGCATCATAG 10140

GTCTATTCTG TAGTTTAGTT CAGGATAATT AGCCATCAAG CTTTCCAATT TATn 10194

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2053 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

nATTGGTTGG TACAGGTATT nGAATTGGTT nGCTCCCTTG GCATCAAAAG GTTTGCmAAT 60

tCGtTAAwTT mTyCAGATTG GTAATATCTG aTAACTCGkT GtAAGAAcAA TCaATAGAAA 120

aTAGATTTGT ATTTTgGCTG CTAACAtCTA GAACAGTTAA TTGATTGCTC aAACAAACTA 180

AAATGGCtAG ATCAACATTA TTAGTAACAT CTAAAACGGT TAATTGATTG TAAGAACAGT 240

CTAAAAGATG TAACCTAGTA TTTTGGCTGA CATCTAAAGT AGTTAATTGA TTGCTAGAAC 300

AGTCTAAATT CCTTAAATAG GTGTTCTGAG TAATATCCAA GTCTGTTAGA CGATTTCCAC 360

TACAAGATAG AGTCTCAAGC ATTCTATTCT GACTAAAATC CAGATTGGTC AAATTATTAA 420

ATTGACAGCT TAAATCAGTT AATTGTTCTG CATACTTGAT ACCTTTCAAG CTATCTAAAT 480

TTTTATTAGT GACAGTAATT TTTGTTAAAC TTGACATATC TGCAATAGTC AATTGATCCT 540

TATCTTTCCC AAAGCCTGAC CAGTTATTTC CTTGGCTTTG TACAGTAGAA TCAGTCCTTA 600

AAAGTGTTGC GAACCTAGCA TCGATATCTG TTCCTTCAAT TAAATTGGTT GTTTTAGGAT 660

CTAATTCTTT ATTTTttTTG TCCAGTTTTG CCGATTCGTT TTCTTGAGTA GGAGTAGTCT 720

TTGTAGATAA AGTTGAATTT AATGTGCTCG TTGTTGTTAA TTCCATTGAG TTTTTTGTTC 780

CTTGTGATTG ATCCGTATTT GCATCTATAG AGGAATCTGT GCCCGAAAGT AAACTTGAAG 840

ATGATGTTGA TTCAACTTTT TTGGATGAAG TAGTTTTTGC CGATACCATC ACACCCATTA 900

TTCCTTGTGA AAGGACTAAA CCTAAACACA AAAACAATCT AAAGGATTTA CTCATAGTTA 960

TTCTCCTTTT CTACATAATT GTCTAAATAG AAACAGCCAA AAACTACTTA TTAAAAACAA 1020

TATGCCAAAC CAGCTTAGCC AAGTTATTCG ATTGTCACCA GTTTTAGGTA ATCTTTTTAT 1080

TTGCGTTCTG CTACGAACGA TTGATAGCGA TTGAAtGGTT GTTTGCGGCA ATACATtCAC 1140

AACTGGATTT TTTTTTGATG GTATGTCTGA AGTTTTTGTA AACCCGATTC CGACTTCTGT 1200

GGTCTCCCTA TTTTCTGCAT AAGTTGTTTG AGGAATATAA ATTAATAGAA CCAaCAaTAA 1260

TAAAAaCkGA AGATTTCyTC yTCTCATTAT GTCACTTCCC CTTCGTTTTT CCACTTCTCT 1320

CCACGACGAA CTAAGTAAGC AATCAATCCA GCAAGTGCTA CTGTCAAAGA TAATGCTACC 1380

CATGGTACCC ATTGTGGGAT CACTACTTGA TACGCTGCCG ATTGATTCAC TTTTTTCGCT 1440

TGCGCTTCTC CTACTTTAAA TTCTTTTTCC CATTTCCAAG TTCGACCATC ACCTGTTGCT 1500

TTTCCAGTAA AGTGATAAGT TCCTGGTTTA AACGTATTCA ATCCAATAGG AATGTTAAAA 1560

TCGAAATTCG AATTTGGTGC TACAGAGAAT TTCTCCATCT TATGTTTATA TAAAATTTTC 1620

TTTTCCCCTT TTAGACGTAT CTCCCCTTGA ATCGTTAGTT CTTGAAGAAC CTTTGGTTGA 1680

TCGTTTTGAA TACGTGCTGC AATCACTTTT TGACCATTCG CTAATTGTAA GCTGGCTTTC 1740

TTTAAATGAA GCATCGCTCC CTGATTAAAG GGTTCTTCAT CCTCCGTAAG CATAATCGCA 1800

ATCACGCGCG CATATTGAGG AGTTAAGCCA GATTTCTTTT GTTTCACTTC CGTACTATTC 1860

TTTCGAACAA AGCGCAGAGA GCCCAGCTTT ACTCCGGAGA AAGCCGATTC AGGTGGCGTG 1920

GATCTCAAAT GTnATCGGTT TCTCTTCTTT CCCCTTGGAG GGGGAATTTC CTTAACGGCC 1980

ACCTTTnATT TTTTACTAAC TCACGGAATG GGGGCTTTCA AACGGGCATC GAACTTAnGG 2040

GGGAACATTG GnA 2053

(2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4178 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

GCAGTCGAGC CTGTTGGTCC ATAAAAATCG GTGTATGAAA GCAATTCTTT TTAGATATTT 60

TTGACTTCCC AATTTGATAT ACTGCGGAAA GTATCTGAGA ATTTATTTAA CACATTTTTT 120

ATTTACTAnA GAAAGTAAAG ATCaGAGCTG TGCaAATAGA AATGATTAAA AACATCCCaA 180

TATCATCATC TAATGTTTCA AAAACTTCAA CAGGTTTCAT AAAATGTTTA ATTATTTTTT 240

TCATTGtATT CaCCTCAAAA AGAGAATAGA ATGAAATATT CTACTTATCC ATTGATATAT 300

CTAACGATTG ACTTACAAAA TTGAAATAAA AGTCTAAAAG TCAGTTGGTG TGATTATTAT 360

TATAATACAA AAAAGACCTC CATTCTTTTA AGAATAAGGT CTTTTTTTTC GTAGTCGCCT 420

TCTAGGCAAA CTAATTCAGT TAGCCGTATT CTAAACTCAA TAAAATGTCA AATACTCATA 480

GGTGTATTTT TATCACAAAA TTTACCACAG ATTTGCGGTA TTATTTACAT ATCTTAGATT 540

ATTAAAAATG ACAAAACAGC TATTTAATGC GTATTGTATG ATTTTTTTGT TCCTTAAATT 600

GTTATGGAAT TTCAGATGGG GAGTGGCAAA TATACGAAGA AAATTAACTA TTTATAGCGG 660

TTTTATTTTA TTAGAGAGTC TCGTTAGTTT TAGTATTGTA TGTGTGATAG CAGGGATATT 720

CTCCCTGACT ATTACACAAT TAATTCAGCA AAATTATCAA CGGGAGCAAG AATTAACGCG 780

CACTCGTCTA GGATACGAAG CAATTTTATT TTTAGAGCAG ACGGGTGATT TACGTTTCAA 840

AGAGCGGATA TATCAAGGAG AAACTTACCG ATTTTCTTTG ATGGAGAACG ATGGAAGAAG 900

GATACTAAAA GTCACAGATT CAAGAGGAGC GATTTTGATT GGTCAATAAA AAATATGCGG 960

```
GTTTTACGAT GTTAGAGTGT TTAGTAGCAC TAGTCGTTTT AAGTTGTATG TGTCAATTGT      1020
TTCAGTTGAT GATTCAACAA AGTTTTATCG GAAATCAATA TTTGAAAAAC aATGATTCGA      1080
AATCCTGGCa TATTTTTTTA ATTCAATTAG AAAAAGAATG TCAAAAGCTC GTTTTCCAAA      1140
CAGGCTCTGC TCcAAGAGAT TTCTTTCCTC GACTCAGAAA CAAACAAGAC TATTTCAATT      1200
CAAATCaAAG AAGATAAAAT TATTAAGCGA GTCAATGGTA AAGGCTATCA ACCGCTGTTA      1260
ATTGGTATCm AAAATGGACA GTTTAAAAAT GAAGGACAAT CGTTTACTTT GGAAGTTACT      1320
TTTACTTCTG GAAAAACATT TGATAGTTTT TTTCCGATAA AGGGAGAATT TTAATGAGGC      1380
AAAAATATTC AGGAAACTTA TTGTTCACGG CCATGGCCAT TGTTTATTTG ATGAGTTTTC      1440
TCGCCCTTCA GTTACTAGAA GAACGTCAGT TAACACAAAA ATTTACGCAA GCTACCCAGG      1500
AATACTATGC AGGGAAAAGT ATCTTTCATT TATTTCTTGC AGATGTTAAA CAAAATAGAC      1560
GAAAGTTAAA AACAGAAGAA AGGCTCGTAT ACGCGCAAGT GACCCTCGAT TATACATACA      1620
AAAATGAACA ATTAAGAATA ACTGTTTTAT TAAACAAATC TGGTCGAAAA TACCAATATC      1680
AAGAGAGAGT TTCTCATCAA AAAAAAGCGG AAACAATACT GGAATAGCGT TGTTTTTCTA      1740
CAAATTATTA GATTTTAGAA GAAAGACAGT AAAGGCATTG ATTATTCTTT CTTTTTTcGT      1800
TAAAATAAAA CAGGTATGGA AGTTGTTACA TTTTGAATAG AATAGAGGTG CTAATTTGTT      1860
CCCAGAAAAA aTTGAAGAGG CTTTCGAGCC AATGGAACAA GCAATCCAGT TATTGCAACA      1920
GTCATTGGAT ACTTCTTTTT TAGATGCGTA TATTGAAAAT GGTGAAAATA TTTTAGATGA      1980
TTTTCAAGTA CGAGTACTTG ACGGAGTACC TAATCCAGAG ACAGTGAAAC AATTAGAAAC      2040
GCTGTATCAT ACGATTAAAA AAATTGATTT AGCTCCTGAA GATGTGCGAC GCTTGTCCCA      2100
ATTACTTTTG CTAAAAGGAA CGAGAAAAGA GCAGCTTCAA GCAAATCATC AATTGACGCC      2160
AGACGGTATT GGCTTTTTGT TTGTTTATTT AGTAGAACAA TTAACAAATA AATCAGAGCC      2220
ATTAAAAATT TTAGATCCTG CTTCAGGAAT GGGAAATTTA CTGTTAACGG TTCTGCTGAA      2280
CCTTGAAACA GCTGGTTACA AAGTTTCTGG TTATGGTGTG GATATCGATG AAACACTTTT      2340
GGCAGTATCT TCTGTGAACA ATGCTTGGTC ACAAGCCAAT ATTCAGTTGT TCCATCAAGA      2400
CGGATTGCAA GATTTGTTAT TGGATCCAGT TGATTTAGCA TTAAGTGACT TGCCGATTGG      2460
CTATTATCCA AATGATGAGC GAGCAAAGGG ATTTGCTGCT GCTGCGGAGG AAGGCCATAG      2520
TTATGCGCAT CATTTATTAA TGGAACAAGC CATGAAATAC GTGAAGCCAG CAGGGTTTGG      2580
CTTATTCTTA ATCCCGACCA ATATTTTAGA GACTGAGCAA AGTGAATTCT TTAAAAATTG      2640
ATTGACAAAA AATGTCTATT TACAAGGAAT GATACAGTTA CCTGATGAAT TGTTTAAGTC      2700
AGAACAATCA CGTAAAAGTA TCTTGCTTGT TCAAAACAAG GGTGCAGATG CTGAGCAAGT      2760
GAAGGAAGTT TTGTTAGCAA AACTTGCCTC TTTGAAAGAT ATCAATAAAG TAACGGAGTT      2820
CTTTAAACAA TTTGAAGCTT GGAAAGCTTC AAATTTAAAA TAAAATAGTA ACGAAGAGGA      2880
GAAAATTATG TCTAAAACAA TTGCAATTAA TGCAGGTAGT TCAAGTTTAA AATGGCAATT      2940
```

ATACCAAATG CCAAATGAAG AAGTGATCGC TAAAGGAATC GTTGAACGTA TTGGCTTAAA 3000

AGATTCTATC TTCACAATTA AATATGGTGA AGGACAAAAA TATGAAGTAA TTGTTGATAT 3060

TGATAACCAT GAAGTGGCAG TAAAAATGTT ACTAGACCAA TTAATCGACT TAAACATTTT 3120

AGGTTCTTAT GATGAAATTA CTGGTGTTGG TCACCGTGTA GTTGCTGGTG GAGAAGAATT 3180

TAAAGATTCT GTTGTTATTA CAGACGAAGT ATTAGAAAAA ATTGAAGCTT TATCTGAATT 3240

AGCACCTTTA CACAACCCAG CGAATGCGAT GGGGATTAAA GCATTTAAAC ATATCTTACC 3300

AGAAATCATT AGCGTGGCAG TCTTTGATAC GTCGTTCCAC ACAACAATGC CAGAACACAA 3360

CTATTTATAT AGTGTACCGC GTGAATACTA TGAAAAATTT GCTGCACGTA AATATGGGGC 3420

ACATGGAACA AGTCATCGTT ATGTCTCTCA ACGTGCTGCA GAAATGTTAG GTCGTCCAAT 3480

TGAAGACTTG AAAATTATCA CTTGTCACTT AGGAAATGGT GCTTCAATCA CTGCAGTTGA 3540

TGGTGGTAAA TCAGTCGATA CATCAATGGG CTTCACTCCT CTTGCTGGCG TTACAATGGG 3600

GACTCGTTCA GGGGATGTTG ATGCGTCACT ATTACCTTAT TTAATGACAA AACTTGGCTT 3660

GACTGATGTG CAAGACATGG TTGATATTTT AAACAAAAAA TCTGGTTTAT TAGGTTTAAC 3720

AAATGGTTTA TCAAGCGACA TGCGCGATAT TCAATCAAAT CTAGATAAAC CTGAAGTTCA 3780

AACAGCATAC AACATTTTTA TCGATCGTAT TCGTAAATAC ATTGGTAGCT ATGTCACTGT 3840

TATGAATGGC GTTGATGCGA TTGTCTTTAC TGCTGGAATC GGTGAAAATG CAGTTGGCGT 3900

TCGTAAAGAT ATTATTGATG GCATGACTTG GTTTGGTTGT GAAATCGATG ATGACAAAAA 3960

TAATGTACAC GGTGAAGAAG CAGTTATCTC AACTGATGAT TCAAAAATCA AGCTTCTATT 4020

AGTACCAACT GATGAAGAGT TAATGATTGC CCGTGACGTA GAACGTCTAA AATAATTAAA 4080

CGAACACACC TTGTGAGATG CTCTCACAAG GTnTCTTAAT AGGTAAATAA GAAGAAACTG 4140

AGCCAACTAT CAAAAGTGAT AGTTGGCTCA nTTnCCTT 4178

(2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21252 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

CAAAATCAGC CACCTATGGC CAGTCTATCC TGCCAAAGCG TTTTTTATTC TTCATCGGTT 60

GCTTTTTCTw GTTCACGATA kTTtcTCaTT TyTTTGCTGA ATAATTCACC GTTACTAGGG 120

GGCGTATATG TAATAGCATT CGCTCCGGCG ATAATGGTTT CATGAATACT CGATTCAGTT 180

GGTCCTCCTG TTGCAATAAT CGGTAATTCT GGATACTTGC CACGGAAATA GCGTACGGTG 240

GCTGCAGTGT CTTTTCCTGC ACTGATATTA ATCATATCCA CGCCAGCCGC AAGTTTTTCa 300

TCAATTGGTG TATGaATAGA AGTTAcTGTG CTGATAATAG GAATATCGAT TGCTTCATCC 360

ACCATTTTAA TTGTTTCAAT TGGTGTAGGA CCATTCAAAA CGACACCCAT TGAGCCTTGT 420

GCTTCTGCAA ATAAACTCAT ATAAGCAGAA CGCATGCCTT GCGTTAAACC GCCGCCGACT 480

CCTGAAAAGA CAGGAATATC CGCTGCTTCA ATAATGCCTT TTGTAATCGC TGGATGCGGT 540

GTAAAAGGAT ATACAGCAAT CACTGCCTGT GCATCCGTAT TGCGGATAAT CGCAATATCT 600

GTCGTAAAAA TAATTGATTT TATTTTACGG CCAAAAATTT TAATGCCACT GGCTTGCCGA 660

ATAATCTCTG GCATTTTTAC AATATCTTGT TTCAGCGGCG TCATAATTTC TGGAATCCAC 720

TTTTCGCTCA TTGTCATCAT TCCTCCATGC TTAACGGTCG TATTTATTTA AATTGTATTC 780

TTCAATTACA TGAATGATCT TATCAGCGTA TGTAGGATCC GTTGCGTATC CAGCATCTTG 840

AAGCGCTTGT GCAGCCTCTT TATAATTTTT GGCTAACAAG ACTTTTTCAT ATAAATTTGG 900

ATCCCAATCG ACCCCATCAA CAAATAACTT GGTATGATCT CGCATGGATT CTTCCCAATT 960

GTCATAAACT CGAAAATCAC CCTGAATCGT AATCCAGACC TCATTGACAT ACTCTTTTGT 1020

TTCTAAATTG AsTTTTGGTT GATTCCCATA TGCTTTAATC CCAAAAAGAT TGTTATATTT 1080

CGATGCTAAT TGACTkTGTC CAAAATTTGA TTCTAAAACT GCTTGCCCCA TAATAATACT 1140

TGGTAATACC CCGTAAGATT CTTGTAATTC ACGTGCATaC GGTGTGaTTT CAGCAAcAAA 1200

TTCTTCTTTT GAGATGTTTT CATTTTGGCG ATTGACTGAC TCGCTTTGTC CCTTATCTGA 1260

ATCaGAAAGG ACAGATAACG AAAAGACAAA GGCCAAACCG ACGACAACTA CGCCTAAGAA 1320

TAACTTTGGC AAATTCAGCT GCTTTTTTCT TTTTTTCTGA TAaGTTGgCT CcACTTCATT 1380

AATTCCCCaC TTCaTCTTTT aTTTTCCgAt AAAATGTTTc aCTAtCTTTc ACTTTcgtGA 1440

tTGatTTCaT TTGtTaCTAt aACCGAATAA TTGTATCATT TTCACAGAAA ATTATCAACT 1500

TTCATCACTG CTTACGCTCG TGATTTCAGT ACTTGATTCA GTGGTAGCAG TTGAAGAAAC 1560

TGTACTACTT TCAGAAATTT CAGTGGATGA TTTCACTGTT GTTTCAGTTG ATGTGTGAAC 1620

AGATTCTTCA ACTGGCTTTT TGCCAAAAAA GACAAAAATT ACCGCACCAA ACATAATAAT 1680

CACAGATAAA ATCCCAGCAA TTGTGTATTT CAAAAGACCA TCCTTCCTAT TCGTTATTCA 1740

CCTACGTAAG TATCCATCAG ATTTTCATTG ACCCAATCTA ATAATTCAAC TTTTGAACCT 1800

TCCAATTGTC GTTGAATCGT AGCTGGTAAC CGGAATGTAA TAATATCATC AAAACCCCAT 1860

TCACCATAGG CCAGCGTCAC TTTCAAATCA ATATAACCAG CAGCAATTTC TTGTTCTTGT 1920

TTATTGTTAG GAATTAATTC TACAACGCCC AAATCAGATT GTAGCCATTT TTTGGCAATT 1980

GTTGTTTTTA ACCGTTTATA TTCCGAAACA GCTAGGTTAC GTTTTTCCGG AAATAAGATG 2040

GTTTGGCGGA GTATTTTTTG AGTCAGCATC CACTCATAAT CCGTAAACAA ATTCTTAATT 2100

TTTTGCATGA CTTCATTGTC CAAACCAAAT TGGCCTTCTT TCCAAGCTTG CCAATCGGCT 2160

GCTTGAATTT GATAAGCGGA GTAAAAATTT TCTTCTGTTT GATACTGTTC AATAATTGCG 2220

TCAACAATAA TATCTACATA AACTTGaTGC ATAATCATTC GCCTCCATTT TCTATTTTAC 2280

TGAAATTCTC TAAAAAAAGC ATCTTTTTAT CTCTGGaTTT TATCATATTT ACAAAATTGT 2340

CCTTGTTTTC TAATTTCTCT TGTTTACAGT ATAATAAATA ATGAAGAAGA TTTGAAGAAA 2400

TTGCATAGAT TAGGAGGACT TCTTTTGACA AAAAAATTAT ATCAATGGGG CATCGTAGGA 2460

TTAGGTGGAA TTGCCcACGA ATTTGCTTCA ACGTTTAAAC AAGAAACTAs CCAaCTTGCA 2520

GCTGTTGCTT CACGGACATT GTCTAAAGCA GAACAATTTG CCGCTGATTA TTCAATCCCT 2580

AAAGCCTATG GTTCTTATGA AgAAtTGCTA gCaGATGaAA CAATCGATGT TGTCTATATA 2640

GCAGTGCCAA ATAAACAACA TGCTGAACAC ATTTTAAAAG CGCTCCAGGC CGGAAAACAT 2700

GTGTTGTGTG AAAAAGCTAT TACGATGAAC GCGGCAGAAT TAAAACAAGC CTTGGCAGTT 2760

GCAGAAAAAA ATGACCTTAT TTTAGCAGAA GCCATGACTA TTTTTAACAT GCCACTATAC 2820

GACGAATTAC GTCGCTTGAT GGATGCTGGC AAATTTGGTC AGTTAAAAAT GATTCAAGCA 2880

CCATTTGGAA GTTACAAAGA ACCCGATCCT AAAAATCGCT TTTTTAATCC TGACTTAGCA 2940

GGCGGCGCAT TGTTAGATAT TGGTACTTAT GCGGTTTCGT TTGCTCGTTG GTTTCTCAGC 3000

TCACAGCCAG ACGTTTTAGC TTCTAGCATG GTTCCCTTCT CTACTGGCGT TGATGAACAA 3060

TCGGTGACAT TGCTTCGGAA TAAAGAAAAT GAGTTGGCTA CTATTTCATT AACATTCCAA 3120

GCAAAAATGC CAAAGCAAGG AATTGTTGCT TTTGAAGATG CCTATATTAC CATTACAGAT 3180

TATCCACGCG CAGATGAAGC GATTGTTCAC TACAACGACG GTACGGTTGA GACCATCGTT 3240

TCTGGGTACT CAGCTGAAGC CATGAATTAC GAAATCTACA ATATGGTTGA AGCAATCAAA 3300

GGTACACAAC CCAATCGCTC CTTATTTTTC ACGACAGAAG TCATTGATAT TTTGGACCAG 3360

ATGCAAAAAT TATGGCAGAA TTAAAAAAAA GAACTTCGGC TTTTGCTGAA GTTCTTTTTT 3420

TTATTTGTTC TCTATGGATA AATACGGTTT AACATACGTG GGAATGGAAT CGCTTCACGA 3480

ATGTGTTCAT TACCAGTAAT GAACGTAACA GCACGTTCTA AACCTAATCC GAAGCCAGAA 3540

TGTGGCACAG AACCGTATTT ACGTAAATCT AAATACCAGC TGTACTCTTC TTCTGATAAG 3600

CCAAACTCAG CTACTTTTTC TTTCAAATAA TCATAATCAG TTGCACGCTC TGAACCACCG 3660

ATAATTTCAC CATAACCTTC TGGTGCAATT AAATCGGCAC AAATTACAAC ATCTTCACGT 3720

GTCGGATGTG GTTTCATATA GAAGGCTTTA ATTGCTTTTG GATAATTTAA AATGAATGTT 3780

GGAACGCCAT AGTAATTTGA AATAAAGGTT TCATGTGGGC TACCAAAGTC TTCGCCCCAT 3840

TCAATGTGAT CATAATCATT GTTTGCTTCT TCTTTTTGTA ATAACTCAAT CGCATCATCG 3900

TAAGTAATAC GTTTGAATGG CTCCGAAACA TATTTTTCCA GTAACGCTGT ATCACGTTCT 3960

AAAACATCCA ATGGATATTG TTGATTTTTC AATACCGATT TAATTAAATG TTTTACGTAG 4020

GCTTCTTGAA TATCTAGTGA TTCTTCATGA GTTGTGTATG CCATTTCTGG TTCAATCATC 4080

CAGAATTCCG TTAAGTGACG ACGAGTTTTG GATTTTTCCG CACGGAAAGT TGGTCCAAAA 4140

GTAAAGACTT TTCCAAATGC CATCGCACCA GCTTCTGCAT ATAATTGACC TGTTTGTGAT 4200

AAGAATGCTG GTTGTCCGAA GTAATCTGTT TCAAATAATT CTGTTGTGCC TTCTGGTGCT 4260

GAGCCTGTTA AAATTGGGCT ATCAATTTTA ATGAATCCTT GTTCGTTAAA GAACTCGTAA 4320

GTAGCACGAA TAATTTCGTT ACGAACTAAC ATGATCGCAT GTTGTTTGCT TGAACGTAAC 4380

CATAAATGAC GATGATCCAT TAAGAAATCT GTGCCGTGTT CTTTCGGCGT AATTGGGTAA 4440

TCTTCTGAAG CGCCAACGAC TTCTAAATCA GTGATATCCA TTTCATAGCC GAATTTAGAA 4500

CGGCTATCTT CTTTGATAAC ACCAGTAACC ATGACTGCTG TTTCTTGkGA TAAATGTTTA 4560

ATTTCTTGGA ATTTTTCTGT ACCAGCTTCT TCACCAAAGG CTTCAATGAA ATTTGGTTTG 4620

AACACGACAC CTTGGAAAAA GGCTGTCCCG TCACGTAATT GTAAAAAGGC AATTTTTCCG 4680

CTTGAGCGTT TGTTAGCAAT CCAAGCGCCA ATTTTTACTG TTTCTCCAAC ATGTTTGTTT 4740

GAATCAATAA TTTGAATTTG TTCCACGTTT GTCTCTCCGT TCATTTcTTA GTTTATTCTT 4800

TTAGGCGTTG TGCTGAATTC GTTTCTTTTC TATAAAGTCA TTCATTCGCT CAATCGCTTT 4860

TTCTAACGTC ATTAAATCTG TCGCATAGCT CATCCGCACA TGTTCTGGTG CACCAAAGCC 4920

CTCTCCCGTT ACCAGAGCAA CATGCGCTTC TTGTAATAAA TCTTCCACCC AGTTAGTCAC 4980

GTTATCATAC CCACAAATTT CTAACGTTTT TTTGACATTC GGGAATAGAT AAAAGGCGCC 5040

TTCAGGTTTT ATTAGCGAAA CGCCTGGCAA AGCTGCTACT TTTGGATAGA TATGATTTAG 5100

GCGTTTTTCA AAGGCCTGaC GCATGTCTTC AACTGTCGAC TGTTCACCTG TTAAAGCCTC 5160

AATTGCCGCA TATTGGCTAA CTGCGACTGG ATTACTGGTG GATTGTGAAG CGAGTTGTGT 5220

CATTGCTTGA ATAATTTTTT CATCCCCGAC TGCAAAACCA ATGCGCCAGC CTGTCATTGC 5280

ATATGtTTTG GAAACACCAT TAATAATGAT TGTCTGTTGA CGAATtGCTT CTGAGATGGT 5340

TGCAATTGGT GTGAACCGAT GTCCATTGTA CACCAAACGA CCATAAATAT CATCTGCAAT 5400

AATGAGGATA TTATGAGCAA CTGCCCATTC ACCAATTTGG CGTAATTCTT CTTCCGTATA 5460

AATAACACCA GTTGGGTTAG ATGGTGAATT TAAAATGATT GCTTTTGTTT TGTCCGTTCG 5520

TACAGCTTCT AATTGTGCGA CTGACACTTT AAATGATTGC TCTTGCGTGC TTGATACAAA 5580

TACTGGCTTG CCTTCTGCTA ATTTGACTTG TTCACCATAA CTTACCCAAT AAGGGACAGG 5640

GATAATCAcT TCGTCCCCAA CGTTCAAAAT GGCTTGAAAT AATAGATACA GCGCATATTT 5700

GGCACCATCT GTCACAATAA CTTGCTTGGG TTGATAACGA AgTTGATATT CTCGTTCCAC 5760

ATAAGAGACA ATTGCTTGCT TTAATTCAGG AATGCCGCCA GAAGGTGTGT AATAACTTGC 5820

TCTTCCATCC TCAATTGAAG CAATTGCTGC TTTTTGAATA TTTTTCGGGG TTACAAAATC 5880

TGGTTCCCCA ACCGTTAAGC TTAAAACATC TACACCTTTT GCTTTTAATG CTTTCGCCTT 5940

GGCGGcAGCT GcTAATGTAA CAGAaGGAGT CAATTGCTGT GCTCGATTTG ATACGTTCAT 6000

ACACCTTTCA CTCCCACTTT CATCAAAACC TATCTAAAtT AAATATCTTT GACAACGTTG 6060

ATCTCTTCGC CTTTATCAAA GCTCATCAGA TAATACGTCA GTACTCCTTG ATCATTTTCG 6120

GTGACCACTT CCCATGCAGG GCGCTTATCA TAAAGCCCTA ACGTAGCTTT CATCACCGAT 6180

TCATCGGGAT GGGCTTTAGC AACAATGTCT GTTATTTGTT GTTCAGTTTG ACCATCTTTT 6240

TGATTTAAGA CCGTGACTTT TTCTCCCGAT TTAGGGACAA TCACGACAAT CTCTGTGCCT 6300

TTATCATTTT TGCCGGTTAA AGTAAAATAC GTATTTTTGC GTGTAAACCA GTAAAATTTA 6360

TCAACTGTTT CTAAATTAGC AGAAGTTTTG GCAATTGCAG TTGCTTCTTT TTTGGCTTGG 6420

GcCATTGGTT GATTGGtTCG TACaAAAAAG ATAATTATTG TTACAAGAAT TGCTAAAAGG 6480

ATTGCCAATA GGGTTAAAAG GACAGTAATC CGTTTCTTCT CTTTCTTTTC TTCTAATTCC 6540

TCCAAAAATC CTTCCATCCT CTCTAAACTT ACTGGGTCTT ATTATAGCAG AAAAATAGCG 6600

GCCTTTCACA AATATGACCT AAAAATAATC GATTATTTGT TTAAAAATTC CTTTAATTCT 6660

TCTAAAATTT CAGGCATGGG CGCTTCTTTT ACTGAGAGTC CCTTGGGAAA CGCTTTGATT 6720

AATCGTTTGC CATATTTTGC AGTGACTAAA CGACGATCCA GAATCAGCAA TGCCCCTCTG 6780

TCTTCTTCAG AACGGATTAA ACGGCCCAGT GCTTGTCGCA AACGTAAGGC TGCTTTCGGT 6840

ACCGCTTCTT GAGCAAATGG TTGTATCCCT TCTGCTTCTA AATACGCATA TTTTGCTTGC 6900

ACAAGTGGTC GTTTCGGGTT TTCAAAAGGC AAACGCGTGA CTATCAATAA CTGTAAAGCT 6960

GTTCCTGGTA AGTCCACACC TTCCCAAAAG CTATCGGCTC CTAGCAAAAC ATTTGGTTTT 7020

CCATTAGAGA ACCGTTTAAG AATTTTTTCA CGACTGCCGC TAATCCCTTG GGCGAGGACC 7080

TCTTGTCCTT TTtCCAATAA TCTTGGCTGG ACTTGATGAT ACACGGTGGA CAATAATTCA 7140

TGCGAAGTAA AAAGAACCAA CATCGAACGT TCCTCTTGCG CCATCAATTT TTCAATGGTA 7200

TCTGAAATGT AcTGACCCAT tTCATAGGmA GAGGCATTTT GAATGGCTAC TGCTTCTTCC 7260

GGTACATAGA CTCGCGCATT TTCTGCATAG TTATAAGGCG TTGGCAGGGT TTTGAACGTC 7320

ACCTCTGTTA AACCTAACCG TTGCGGTAAA TACTTTTTAT TGTTGCCAAA CTTCAATGTG 7380

CCGCTTGTAT AAATAATTCG TTCATACCGC GGATACCACT TTGTTTCAGG AAGCAAACTA 7440

GCTTCTAAGT CATGAATTTC AATCAAGCCA TGGTCTTCAT TTTGTGGTAA ATATTCTTTT 7500

ACCCAACGTG GACGCCAATC ATCAAGATAA ATGGTAAAAC AGTGATGAAT ACGTTCAATT 7560

TCATCAAAAA ATTGAAACAA GCGAACAAAA ACAATGCGAT CAGCCAAACT AAAGGTTTCC 7620

AACTGACTCA TAATTGTTTG CTGTAACTTA CTTTGAATAG TTAACAGCTC CTGCAGTAAA 7680

AGCTGGAGCG TTTGACTATG TTTTTCTCCC GCAGCAGATA AATAATCAAT ACGTTCTTTG 7740

GTTAAAAGCA ACTCCTGATG GTCTATCCGC TTCTGTGTTT TACTAAATAA ACGATGCAAT 7800

TCAAAAAGAA AATCTTGCCA TTCATCCACA AAATCCAGTA AACCTTGCTG ATAAATTTTT 7860

ACAAACCGCT GCTCTTCAGA AAACTGTTGA AAAATTTGCT GTAAAGCAAC AAAAAGTTGT 7920

TCTTTTTCAA TAAAATGCGT GACTTGTTTT TTTAGTATCA CACTGTTCAC TCGTTCGCTC 7980

GCCATCCGAC TAGCGATGTC TGGTAAATGA TGTGCTTCAT CAATAATTAG CCACGGACTT 8040

TCAGGTAATA ACGGTTGTTT TCTTAGCGAC TCTTGCGCCA AAAATGCATG ATTAACAATC 8100

AACACGTTAC TTTGAGCGAC TTTTTTATAC AGATGTCTTA AAAAGTCAAC TTCATAAAGT 8160

TCATCTTTTT TTGAAAGAAA ATCCAATCCG CGATGTGTAA CTTCTTTCCA GAATAAATGA 8220

TTTAAATTCG TCAATTGTAG TTCATCTAAA TCGCCCGTTT CTGTTTCTAG CAACCAGACG 8280

AGAACGCCCA TTTGATAAAG AGCGTATTGT TTTTGCGGAG TTGGCTGATT CAACGTTGCC 8340

TTAAAGCATT GTAAATCAAG ATAGTGTCGA TGGCTTTTAA TAACAATGCC TCGTAATTTC 8400

CCTTGGCAGA TTTGATTGGC TAATGGTAAA TCTTTTTCCA CTAATTGATT TTGTAAAACA 8460

ATCGAAACTG TTGAAATGAT GACTGGTTTT tCTGGCGTTG CTAAATAGCT CATCGGCAAC 8520

AGATAGCCTA ATGTTTTTCC CGTTCCAGTC GCCGCTTCAA TAAATAAATC TTTTGTCGTT 8580

CCTTCTGTAA AATGATCATA GACTAAATTC ATCATCCGAC TCTGTTCCGC ACGATAAGCG 8640

AAACGTTCGC CAAACAGTTT TTCTTTGGCT TTTTTAGTTT TGGGATACGT GCTAGTCTGA 8700

TAGAACGTTT CTTCAAAAAG AGGAACTTCT TTTTTGCGTA AAGCAATGCC ACTGACCACT 8760

TGATAAGCTG GATTTAATGG TGTGACTTGT TTTTTCATTT GTTCATACGT TTGTTGAATA 8820

AAGGCACTGG TTTCCCGCGC TGTCTGTTGA CTTAATTCCG CAATTTTTTC CATAGTGACT 8880

AATGGTAAAG ATTTCATTTT TTCTTGAATT AATAACAAAA GTTCgGCTGT AACCTGCGCA 8940

TCACTATCTG CTTGATGGGG ATTCTCATGA GACAAACCTA AGCTTTCCGA CAAATCACTT 9000

AAACGGAAAC TTTTTTCTGT CGGTAAAAAA ATCTGTGCCA ATTCCACTGT ATCAATGGCT 9060

GGAATAGTAA GTGGCGGCGT GCCACACCGT ACCAGTTCAC GAGCTAAAAA ATTATAATCA 9120

AAGTGGACGT TATGCGCCAC AAAAATAGTA TCTTCCAGAT AATGATAAAT TGTATGTGCT 9180

ACATCTTCAA AATAAGGCGC TTTTTGAACT TGCGTATTAC TAATCCCCGT TAAACTTTGA 9240

ATTTGTTTTG GTACTACTTG ATTGGGATTC ACATCTGTCG CAAAGTTCGC AATAATCTTT 9300

CCATCTTGAA CGAGAACGCA TCCAAACTGA ATGATCCGAT CACTTGTAGG ATCTGTTCCA 9360

GTCGTTTCAA GGTCAACTAC TGCATACACT TGTTTGTCTT TCAATGACTT CACGCCTTTC 9420

ATTCATCTTT GAAAATTATA CTACAATTTA TAGGATTTTC ACTAGTATTT CAAAAATTCT 9480

CTTGAAATCT TTTCCAGAGA TAATAAGAAA GrGCCTGAAA CAAAAATCCA AAGTGATTTT 9540

TGTTTCAGGC yTAAAAGCTG ATAAATTGAA CGGTTTTGTC TTACCTTCTC CTAAATAAAC 9600

GATTAATAAC AGTCAAGGAT TTTCTGCATA TACTGGTTTA ATTGCCCATA GCCTTCCAAT 9660

TCTCTGATTA CTGGGAGCGG ACAACGAACC TTTTTATTAA ATAGTGTCCG CTTTTCAACA 9720

TATTCTGTAA tTCGTCTAAA AATTTATCTA ATCCTTTTTC TAATTTCATA CAACTTAACT 9780

GATTGCAAAT TCGCACGTCT CCTGCCTTTT TCTTTTGAAC TTGTTTAAAT GATTGTGAGA 9840

ATGCTTTTGC TTCAAACATT TGTGCTTCTA ACTGTCTTTT TTCAATAGTT ATTTTCGGTA 9900

CTGCTTGATT TCCATCAGTA TCCTCTAAAT GGCGACTCAT AAAGCGACGA TAAATTTCCT 9960

CTTCTAATTC GTCCAATTGT CGTTGAAAGT CCGCAGGAAT TGTCCGACTT GATGAATTTT 10020

CTTTCTCACT TTCCATAGAG ACAGCTTCTT TTTCTGGCTT CTGCTCCTCT TCTTTTTCGG 10080

CAGACTTATT CACTGAATTT AGCGAGCTTT TTTCTGTTTC TTGTGGAAAA GTAAATAAaG 10140

GTTTTGCCGC GGTACTCTGC TGGTTTTCTT GTTTTTTTTC TGGCAGCGCT TCTTTAATTG 10200

GCGCTGGTTC ACTTGTTGCT ACAGTTGTTT CACCTTCTTG ATTTTCCCGA TTATCAGGCG 10260

AGAGATTGAC GTCTGGTTGT TCCAGATTAG TATCCTCTGA TACTTTCAGC GATGCCTGTT 10320

CTGCGGCTAT CACAGGAGAA GCTCTTTTCA AATGGATCGT TTTTATTTGT CTAGAAATGA 10380

CGGGATTTCT CGTTTGATTA ACGGCTTCAT TTAAGGAGGC TAAATTTTCT TTTAAAAGTG 10440

TCAACATTTC ACGTAAGACA GCAATTTCAC CAGCTAATTC CGTTAATTGA TTTTCTTGCG 10500

GCACTACTTG CACATTGGTT TCATTCATAC GGACTGTTTT TTTATTTTTA GTTGTCGGTA 10560

TTTCATACAT CCATTGTTTC AGTGCGGATC GTAAGTCACG AGTTGTCGGC TCAAAcTTCC 10620

AAAATAAACA GTGTAGTACA TTAcTCCGTT TCAACATAGA CATAAATCTT TTTAACAATT 10680

AACTCTTCAA ACTGTTCTGA TTCATCCCAA CTTTCGCATA AAGCCTCAAA AAATTGTTCC 10740

CCTTCTGCTA AATTTTGATA CTCTTTCGCT AAATAAAGCT TGCGrATTAA GTAAAAaTTA 10800

TACGGTGTTT CAACATTATC TGCATAAATT TCTCGTCCTA AAAAATCTGG ATTCGTTTCT 10860

TTACAAAGAA AGACCGTTTT CTCTTCTGGT TGTGTTTCAA GTGTTTTAGT CATGAGATTC 10920

ACTTCCTTTT TCTCCGTTAA TAACCATTTT TTAATGATAG ATACCTTTAG TATACAGGAG 10980

AAATTCTTCT GATTTTCTCA TATTTTTAAG AGACAATAGA AATTTTAGAG AAAAATCGAC 11040

AGAAAAAAAA CGTGAGATAA ACAGTCCTAT TGACCATTTA TCTCACGTAA AATACTTAAA 11100

CTGACTATTT TATCCTTGGA CATGTAACCA ATTATTCTGC TGAATCAATT GGCAAAATCG 11160

TGTCTTTTTT CTCGCGACGT GCAATTGCTG GTAAAATCAT TCCTAATAGA ATTAAGACGA 11220

TTGGTGTCAC AATGTTAGAA GCTAATTGGA ACAACCAGCC TTGAGGGTCT GAGGCATACT 11280

CAAATTTAGG CACCATTCCT AGGATACAAG CAAAGGCTGT AAATAGGAAA CACCAAAGAC 11340

CAAACACAAA ACCAATTTTA GGATTTTTAA CAAATTTATA TTCTGAATTA AATTGTTTGT 11400

ATGCTTTGTT TAGAAGCATA TAGGCAAAGA AGACCCACAA GTAACGCATT GGCATAACTA 11460

CAGAATTTAA GTTAGTTAGC CATTTAACCA CTTCGTTCAT GTCCTTCAAC CCAAACATTG 11520

GTAAAGCGAT AATAATACTT ACTAAAATAC CTGTTAATAA GTAACCATTA ATTAACGTAC 11580

CTTTTTTCGT ACGTTTACGT AACCAACTTG GAACAAATTC TGGATCTGCT TCAGCTAAAA 11640

GAATTTTCAG TGGAGCATCA ATTGAAAAGG CTAGCGCAgc CACTTGACCC ACTGTATTCG 11700

TCAAAGCATA AACAATCATT AAGAGATTAC CAACACCATA ATATGCGCCT AATTTTTGGA 11760

AGGCACTGTA GGCACCATTT GCCATTAAAT CTTTAGGAAT ATTGTCACTT GAGAATAACA 11820

TGCCCATTGC AAATGAACCT AAAACAGCCG ATAAACCAAC CATTGCCGCT AGGAAAATCA 11880

TTCCTTTAGG GAACTCTTTG GCTGGATTTT TAGTTGCATT TACGTAAGGA GAAATTTTTT 11940

CAGCGCCACC TACCGCAAAA ACTAACATCG ATACGGTTGT AAAATAACTA AAGTCAAATT 12000

TAGGAATATA GGTGCTGACT TGTGACATGT TTGGCGTCGC AAATTCAACC GTTGAATTAA 12060

TAAATGGTGC GCCTACCGCT AATAAAATGA ATAAAATAGA CATCACAAAC ATGGCGGTTC 12120

CAGCTAAACC ACCAATGACT TTTAAGGTGG TTAACCCTTT TGTTGATAAT AACAAGAAAA 12180

CCAAGAAAAT TATTAATGAA ATAAATGATA CCGTTACGAT TGGCATCGTA TTGACTAGGT 12240

TTCCATTCCC TTGAACTGCC CAACCTAAGG CAATTAAAAT TGCTTGTGGT TTTTGTGCCA 12300

AGTAAGGAAT GTGCACTACC CAATAAGTCC ATGCTGCATA ATAGGCTAGA CGTTTAGTAC 12360

TGGTATTTTG CACCCACGAA CTAACTCCAC CATTACTGTC TTTAAATGTT GAACCTAACT 12420

GTCCTACGAT TAACGCATAG GGAATAAAAT ATAAAACTAA AATCAAAAGC CATGAAGTGA 12480

CCACTGAAAT CCCTTGTTGT GCATAATTAT TCACAACATT TCCAAGACCC CAGACCATGT 12540

TAAAGGCAAT TAACCCGACG GTGAACCATC GTAATTGCTT TTTGTCCATA TAACCAATTC 12600

CTTTCTTATT AAAAAAAATA AAGCTCTTCA TTTATCATAA GCATTTTACT TTTTATGCAC 12660

AAGCTCAACA TAATTAAATA AAGATGGCTT TTTAATAAGT AATTAAAACG CTTCTAGCAA 12720

TGATTTTCTA ATTTATTTGA TGACTAAAAA AGGCCGGTGA ATAAGTTGCT TCACCAGCCT 12780

TTTTCTAAGA GATTTTTCTT TTATAACATA TCAAATAATG ATAATTGGTT TTCATCTGGC 12840

AAGTCTTTTA AGACACCATT TTCATTCATG TATTCAATTA ACGTTTTAGA AACTTTGCCT 12900

CGTGTTGCCA AATCTTCTTT CGATAAGAAC GGTTGCTCTT CTCGCGCAGc AACAATGGCT 12960

TTCGCTACGT TGAGGCCTAA ACTTGGCACT GCTCGGAAAG GTGCAATCAA GGTATCCCCT 13020

TCAATCACAA AATTTTCAGC ATCTGATTTG TACAAATCAA TCATGCCAAA TTTATAGCCC 13080

CGTTCCAACA TTTCATTACA TAATTCTAAA ACCGTTAATA AGTTTTTCTC TTTAGTGGAG 13140

GCGTCCATCC CTTTGtCCGT AATTTCTTTC ATTCGTGCTT TCACAGCTTC TTTTCCTTGT 13200

GACATTGCCA CTAAGTCGAA GTCATCCGCC CGAACGGAGA AAAACGCACA GTAATATAAG 13260

ATTGGGAAAT AAACTTTAAA GTAAGCGACC CGCAACGCCA TTAATACGTA TGCTGCCGCA 13320

TGGGCTTTCG GGAACATATA TTTAATTTTA GAACAAGAAT CAATATACCA GTCAGGGACG 13380

TTGTTTTCTT TCATCGCATT TAAATACGTA TCCCGTAGTT CATCAGGAAT CTTATTCCAT 13440

TGTCCTTTAC GTACCGTTTC CATAATCTTA AAGGCCATCC CACTGTCCAA ACCAGCATGA 13500

ATTAAATAAA CCATGATATC ATCCCGACAA CCAATTACTT CAGCAAGATT GgctTCGCCA 13560

CGACGAATCA ATTCTTCTGC ATTCCCTAGC CACACATCGG TTCCGTGAGA CAGTCCAGAA 13620

ATTTGAAGTA ATTCTGCAAA AGTTGATGGA TGGGTCTGTT CTAACATACC TCGAACGAAA 13680

CGTGTCCCAA ATTCAGGAAT TCCTAAAGTT CCTGTTTTAG AATAAATTTG CTCTGCATCC 13740

ACACCTAGAA CATCAGGTCC TTCAAAAATA CGCATCACTT CTGGATCGTC TGTTGGGATG 13800

GTTTTTGGCT CAATCCCAGA TAAATCTTGT AACATCCGAA TCACGGTCGG ATCATCATGT 13860

CCTAGAATAT CAAGTTTTAA AATATTATCG TGAATGGAAT GGAAATCAAA GTGCGTCGTT 13920

TTCCATTCTG CCTCTTGATC ATCAGCAGGG TATTGAATTG GCGTAAAGTC ATAAACATCC 13980

ATGTAATCAG GAATTACAAT AATACCGCCG GGATGCTGCC CAGTCGTCCG TTTAACCCCT 14040

GTTGAACCTT TGGCTAAGCG ATCAATTTCA GCACCACGCA AATGTAAATT ATGATCTCTT 14100

TCATATCCTT TAACAAAGCC ATACGCTGTT TTGTCGGCTA CTGTACCAAT GGTTCCTGCC 14160

CGGTAGACAT ATTCTTCACC AAAAAGAACT TTCGTGTAGT TATGGGCTTC GGCTTGATAA 14220

TCGCCAGAAA AGTTCAAATC GATATCGGGA ACTTTGTCGC CATGGAACCC TAAGAACGTT 14280

TCAAAAGGAA TATCGTGACC ATCTTTAAAC AATCGAGCGC CACATTTTGG ACAGGCTTTT 14340

TCAGGCATAT CAAACCCTGA ACCATAGGAA CCATCTTCAT AGAACTCAGA ATATTGACAC 14400

TCAGGACAAT AATAATGTGG CGCTAAAGGA TTTACTTCAG TAATCCCTGT CATTGTCGCA 14460

ACAAAACTCG AACCGACAGA GCCCCGTGAA CCAACTAAAT AGCCATCTTC ATTACTTTTA 14520

TGCACCAGTT TTTGAGAAAT TAAATAAATG ACCGAGAAAC CATTCCCGTT AATCGAATTC 14580

AGTTCCTTTT CTAACCGTTT TTCGACAATT TCTGGTAAAG GATCGCCATA CAATTCGCGC 14640

GCACGATTGT AGCTTAAATC AGTGATCTCT TGTTCAGACC CAGGAATTTT CGGCGTATAC 14700

AAATCGTCTT TTACTGGAAT AACTTCTTCA CATAAATCAG CGACTTTGTT GGGATTTTCT 14760

ACTACAATTC GCTTAGCGAT GTCTTGTCCT AAAAATTGAA ATTCCGTTAA CATTTCATCG 14820

GTGGTTCTGA AATGAACCTT TGGTAAACTA TGGCGATTTA ACGGATTAGC GCCGCCCATA 14880

GACCCGACTA AAATTTTCCG ATAGATGGCA TCTTCTTCAT TTAAATAATG AACATTTCCA 14940

GTGGCGACAA CCAGCTTACC AAGTTCGTCA CCGATTTTCA CCAGATTACT AATAATTTCT 15000

TCCAAATCTG CTTCATTTTT CACCAATTCT TGTTCAATTA ATGGCGCGTA AACTGGTTTG 15060

GGCATAACTT CAATATAATC GTAGAATTTG GCACGATTTT TTGCTTCTTC GACACCTTTT 15120

TGCATCATTG CTTCAAAAAT TTCGCCGTTT GAACATGCGG AACCGATTAA AAGTCCTTCT 15180

CTTAACTTCG ATAATTGCGA ACGAGGAATC CGTGGTACTC GGAAAAAATA ATCTACATTG 15240

GACATGGAAA TCAATTTAAA AAGATTTTTT AATCCTGCTT GAGTTGTTGC TAAAATCGTT 15300

GCATGGAAAG GTCGAGCTCG TTTGTATGAG TCTCCTTCGC CAATATGACG ATTCAAATCA 15360

TCATGGAAAT GCATATCATG ATTCTCTTTC GCTTCTTTTA AGAAAATCCA GCACAAATGA 15420

CCTGTTGATT CCGAGTCATA GATGGCGCGG TGATGTTGTT CTAAGTTGAC GCCAAATTTT 15480

TTCGATAAGG TATTTAGACG ATGGCTTTTG AAATGTGGAT ACAAGAATCT GGATAATTCT 15540

AATGTATCGA TGACGGGATT AGCCGCTTCA GGTATTCCAT GTTTGCCATA ACTTGTATTT 15600

AAGAACCCCA TATCAAAAGA AGCGTTGTGG GCCACTAAAA TAGTTCCTTC TGAAAACTCT 15660

TTAAACATGC GTAAAACTTC TTCTTCGGAT TTTGAACCGC GCACCATTTC ATCTGTAATC 15720

CCAGTTAAAT TGATCGTTGT TTGAGATAAT GGATGGCCAG GATCAATAAA TTGCTCAAAA 15780

GTCTCAATGA CATTTCCTTT GTGcATTTTG ACAGCGGCCA ATTCAATAAT TGTATCATAA 15840

ACAGCGGAAA GCCCCGTTGT TTCCACGTCA AAGACCACAT AGGTAGCATC TGTTAATTCA 15900

ATATGTTCTT CGTTGTAAGC AATGGGTACA CCATCATCAA CAATATTGGC TTCCACGCCG 15960

TATAAAATTT TAACGCCTGC TTTTTTACCT GCAGCATGGG CATCAGGAAA CGCTTGGGCG 16020

CCGCCATGAT CCGTAATTGC GATGGCCTTA TGACCCCATT TCCCCGcTTG CGCCACCAAG 16080

TCACCGACTT TATTGGTCGC ATCCATTGTA CTCATGTTAC TATGCATGTG TAATTCTACA 16140

CGTTTTTCAC CTTCTGGCGC GTAGTCTTTT CTTGGCGCAT GGGCGACTTC CATCAAATCT 16200

TGAGCATTCA TGACTAAATC ACGCATGAAG GTATCTTCTT GTACACTACC ACGAACACGA 16260

ATCCAACTTT GCGGTTGAAT CGCATCAAAA ACTTGTTCAT CTTTTTCCCC ATTCGAAAAT 16320

TTCTTCACCA CGAAAGAGGA AGTATAATCA GTAATTTTCA AAATTAAAAT TTTTCGTTTA 16380

GAACGCAATT CACGCACTTC TTTATCAAAA ATAAAGCCTT CAATCGTTAT CCGACGTTCT 16440

TCTTCCAGTA TGTTTCCCAT TGGCATAATG GGTTCATCAT TGGGAATATT CCGACCTAAA 16500

CGAATTGGAC CATCAAACGC AGGCGCTTGT TGTTGTTTTT CTTTTTTCAT TTGTTCATGT 16560

TTAACAAGCG ATTCAGCAGC TTGTTGTTGA AAGGCTGCGG CTTGTTCTAA TTTTTGCTCT 16620

TCAAATTTTT TCAGCACTTC TGCAGCTTGC TGTTGATCCA TTTTTGGTTC AATATGAAAT 16680

TTAGGAAACC CGTAAGAGAA ATAAAGTTCC TCAATAATTG GTAAATATTG TTGCTTCAAA 16740

TAAGGAATAA CTGCTTCATT ATCAACAGGT AAGATAATTT TCCGATCTTC AAATTGTGGT 16800

GTCTGTGTTT TTAAGACTTT CTGAACTAAC GGCGTATTGC ATTGACTATT TAACAAGGCC 16860

AATTGCCAAT AATCCGTTAA TTGTTCCTCT GTAAATGTTG TCTGATTTGT GGTTATTTGA 16920

ACGGAAACCT GCGCAATTTG TTGAAAAGCT AGTTCCAATT GTTGCATAAA TGTTTGATAT 16980

AGCATCACTG GCATCAATTC ATCAAAACTT AAATGAAATT CCCATAAGCG ACTTTGTTGG 17040

TGAACAATGA CCTTGTCAAT CCGGCCATTT TGAATTAAGG GATGTTCTTT CGCTGTGTCA 17100

TCTAACTGAA TTTGCGTCAT CAGTTTATCA AATAAATCGC GTGCTTTTTC TGTCACAAAA 17160

TGAACCTCCT TCTATCATTG AAAAAAGGCC CTCAGACAAA TCTGTCAAGT GCCTTTTTTC 17220

GGTTTCTCTG CATTCGCTTC TCTATTAAGT ATACTCGTTT TTTGTAAAAA AACAATCTTG 17280

TTTCTTTTCT AAAAAGAGCT AAGAAGAGCT TGAAGCAAAA ATCACTTTAG ATTTTTGCTT 17340

CAAGCyTAAA CGCTGATAAG TGGCGGAAAC AGTTTATCCT AATCATTCTG TCCGCCCTTT 17400

TCATTATTAT TCCACTTCTG ATGTTGTGTT CATCArAATA GAAAGTGTAC TTTCTAAttC 17460

TTCTTTGCGT ACTTCCaACA TTTCACCAGT ACGTTTGATT TkGACTTCTA CCACACCATC 17520

GACTGCTTTT TTCCCGACTG TGATACGGAT TGGACAACCA ATTAAATCAG CATCAGCAAA 17580

TTTTACGCCT GCGCGCTCGT TGCGGTCATC AACTAATACT TCGTAGCCCG CTTCGGTCAT 17640

CATTGCTTCA ACTTCTTGTG ATAGTTTTGT TTGGTACTCA TCTTTCACAT TCATTTGGAC 17700

CACGTGTAAA TCAAATGGCG CAATTCCTGT TGGCCAATTA ATACCACGCT CGTCTGCATT 17760

TTGTTCAACG ATTGCTGAAA GTAAACGACT AACGCCGATG CCATAGCAAC CCATGATCAC 17820

TGATTTTTCA CGACCGTTTT CATCTAAAAC TGTTGCGCCC ATAGCATCAC TATAACGCGT 17880

TCCTAATTTA AAGATATGAC CAATTTCGAT TCCTTTAGTA AAGGCTAAGA CACCATTGCC 17940

ATCAGGAGAA GGATCGCCTT CTTGAACAAA GCGTAAATCT TCGTAACTAA TTGGTTGGAA 18000

ATCACGATCT GGATTTACAT TTGTCAAATG ATAGCCATCT TCGTTAGCTC CAACAATGGC 18060

GTTGGCTAAA TCTTGAACAG CTAAATCTGC ATAAATTTTA ACATCTTCGG AAACATTGAC 18120

TGGACCAATT GAACCAAAGC CCGCTCCTAA GACACGGCGA GCATCTTCTT CTGTTGCTTC 18180

ATCTAAGAAA TCAGCGCCTA AGAAGTTTTT CAATTTGACA TCATTGACAT CATGATCGCC 18240

ACGAACCAAG ACCATTACTG GCTCTTCATC GGCAATAAAT AGTACTGATT TAATAATTCG 18300

TTGTGGCTCA ACTTCAAAGA AGTTTGCCAC TTCGGCAATT GTACCCACTT CTGGTGTTGC 18360

AATTTTTTCT AAATCTAATT GTGTTTCATG TGATTTTTTC GGCGTATAAA GACTTGTTGC 18420

CATTTCTAAA TTGGCAGCAT AATCACTTTC TGTTGAATAA CAAATAGTAT CTTCCCCAAT 18480

TTCAGAAATC GCCATAAATT CTTTAGAATC TTTACCACCC ATGGCACCAC CGTCCCCAAT 18540

AATGGCACGG AATTCTAGTC CACAGCGTTC AAAAATGCGT GAGTAAGCTT TTTCGTAATC 18600

ACGGTATGAT TGATCTAAAC TAGCTTCATC GGCATGGAAA GAATAGCCAT CTTTCATGAT 18660

AAATTCACGG CCTCTTAATA AGCCAGAACG TGAGCGTTTT TCATCACGAT ATTTTGTTTG 18720

AATTTGGTAT AAATTCAAAG GTAATCGTTT GTAAGAATTA ATTTCATCAC GAATTAATTC 18780

TGTGAAGGTT TCTTCATGCG TTGGGCCTAA GATATAGTCA CGGTCATTGC GATCTTTTAA 18840

GCGATACAAG TTAGGTCCAT AAGTTTCATA ACGGCCTGAT TCTTTCCAAA GTTCTGCGGG 18900

TAATAACGCA GGCATTAACA TTTCAACTGC ATCAATTtTC TCAAATTCTT CTCGCATAAT 18960

CGTTTTTAAT TTTTCTAAAA CACGATTGGC TAAAGGTAAA TAAGAATAGA TCCCTGCAGC 19020

GACTTGACGA ATATAGCCTG CGCGCAACAG AATTTGATGA CTTAAGACTT CTGCATCATT 19080

TGGTACTTCT CTTAACGTTG GAATTAACAT TTTTGACTGT TTCATTAAAA CTCTCCTCTA 19140

TTTTAAAATC AACTATTTTT GATTTCTTTA CACACTTTTT TACTTAAACT TTCTCCTTAA 19200

AAGAAAAAGC GTTGAATATC GTTCCAAGTA ACTaACACCA TTaACACCAT GACAAACCCA 19260

AAGCCAATTA ACGTAATGAT GCCTTCTTTT TCAGGACTAA TTGGTTTTCC ACGTACACCT 19320

TCAATAATGT TTAAGACAAT TTTCCCGCCA TCTAAAGCTG GGATCGGCAA CAAATTAATA 19380

ATCCCTAAGT TCATTGACAA CATGGCCATT AAGAATACAA CTGTACTTAC TCCAGCATTG 19440

GATGCTTCTT CCGATAATTT AAACATCATG ACTGGCCCAC CTAGTTTGTT TAAACTAAAG 19500

CcTGTGAATA GTGAGCCGAG TGCTTTAAAA ATCTGTGTCG TACTATTCAA AGTATCCTGA 19560

ATACCGCCCA TCAATTTTGA CGGTAAATCG GTTTTCATAT AAGGATAAAC GCCGACTTTA 19620

CCAATCGTTT GTTTTTCCAC TTTTTGTTTT TCTGGTGTCA CTGTTAGTTG CTCTTCTTTG 19680

CCGTTACGCT CAACTACGAA CGTTAACGGC TTTTCGGGGT TCTTCTGCAC AATGGTTGTA 19740

AAATCTTCGT ATTTTTTGAT TTTTTGATTA TTAATCGATA AGACTTTATC GTTTTCTTTC 19800

AACCCAGCTT CTGCGGCTGG GCCATTAGGA ATCACTTGTC CAATTTGATT CGTGTTTAAA 19860

TCAGTGACGC CTCCTTGTAG AAAGACCGCT AACGTAAACA GAATAAACCC TAAGATAAAG 19920

TTATTCATGG GTCCCGCAAA GTTCGTCAAA ATGCGTTGCG ATAATTTCGC TGATTGAAAT 19980

TGAACGTCAA GTGGCGCAAT CCGCACCTCG GTTCCATCAC TTTCAATAAT CGTTGCATCA 20040

TGGTCAACTT TATAAACGGT TTCTTCTTCC TCGTTTCCAT TGACATAGCC CTTGATGAAT 20100

AATTCTTTTT CAAGATCAAA ATCAATGACT TCCATCGGAA TACTATGAGG TAATTGTACT 20160

TTTTTACTTG TATTAATTTT AACCACATTA CCCACGGCAT TTAACTCAAC AGATAGAGGC 20220

ATACCTGGTG TGATTTCTGT CATGTCTTCG CCCATCCCAG CCATTCGCAC ATAGCCACCA 20280

ATTGGCAATA AGCGAATCGT ATAAGTGGTG CCATCTTTTC CACGATGCGC AAAAATCTTT 20340

GGTCCCATTC CGATTGCAAA CTCACGAACT AAAATACCCG CTCGTTTTGC AAAATAAAAG 20400

TGGCCAAATT CATGTACGAG GACAAGAATA CCGAAGACAA TAATGAATGT GATAATTGTT 20460

TTCATAGATG TCCTTCTTCC TTCATTCACT GACGTTCAGT GCTTTTTATT ATTACATGGG 20520

AAAGTCTCGT AAAACTTGCT CTGTTCCCTT CCCGATAAAC CAGTGCAACT GTTTACTGAG 20580

AACAGTCGCC ACTGTTAGTG AACAAATCTT TCCTTATGTA AGACAAGTTT TATTTTAGCA 20640

TATTCACTCA TAAATAACCA TTGGACTTGA CCACTTCGCC TAGATATTTT GTGTTTTTTT 20700

AGAAAACTGT ATAAATGAGA GGmaTTGATT ACTACGACCA GTATTTATTT TAAAGACCTT 20760

CAACAATAGT AGCTAACTTT AGACCATAAA CTTTTCCGAC AGTTTGTCGT ATTTTATCAA 20820

CACGCTGATT GACATGTGCT ACCATGAAAG TATTAATAGT AATTTTAGGG GGAATGAATC 20880

AATGTATGGA GAATTACATG GTAAAGTAGC CGTTGTCACT GGTGCAGCTA CGGGTTTAGG 20940

ATTATTTATT AcTCTGCGTT ATATCTTAGA GGGCATGAAT GTTGTTGCTG ACTATGTCGG 21000

TGAGCTGCCT AAAGAATTTG AAGATGTCCA AGCCAAGCAC GCAGATCGCG TAAAATTCGT 21060

GAAGGCTGAT GTCTCTAATG AAGAAGATAT TAAAGCGTTA gCTGAAACCG CTTTAAAAGA 21120

GTTTGGTCAT GTTGACATTT GGGTAAATAA CGCCGGTGTG GAGGCTTCTT TTCCGACAAT 21180

TGATATGCCA TTAAAAGAAT GGCAACGTGT CATTGATGTC AATTTAAACG GTGTCTTTTT 21240

GGGTTCTACG AC 21252

(2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8726 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

GCATATCAAC AAGTCAAAGA ACGATATGAA ATTTTACAAA AAGAAGAAAA AGCCTTAAAG 60

CAACAACTGG CTTCTTATCA TGATttCTAT GAAACGTTTT GCTTGGCTGA AGAAGTCTTA 120

TTGGCGGTTA TCCAAAGAGA ACAAGCACGT CAACATTTGT TAAATGCTTC ATCATTTTTC 180

ATCTTGCAAA CGTGGATTCC AGTCGAAGAA AAAGCCGAGA TTTTGACAGC GATTGAAGAA 240

AAAGTTCCTA AGGACGAAAT AGCTCTTACT TTTGAAAATC CTACGAAGGC GGAAATTGAG 300

ACGGATATAC CAGTTAAACT AGCGAATAAT AAATTGGTGC AGCCGTTTGA AATGTTAACA 360

GAGATGTATA GCTTACCAAA ATATGAAGAA GTTGATCCGA CACCTGCGAT GATGCCGTTT 420

TATCTAGTCT TTTTTGGTAT GATGGTAGCG GATATTGGCT ATGGCTTACT GATGCTACTC 480

TTGTCTATTA TTGCTTTGAA AGCATTTGTT TTACCGAGAG GCATGAAACG TTTTGCCGAC 540

TTTTTCTTGA TTTTATCATT TCCAACAATT ATTTGGGGCT TCATTTACGG TTCGTTTTTC 600

GGTGCAGCAT TACCGCCAAT AATGTTTGGA ATCAAATCAC CATTTCCGAT TTTATCAACG 660

ACCGAAGATG TAAATACGAT TCTTATTTTA TCGGTTATTT TTGGGTTTAT CCAATTGGTT 720

GTCGGTTTAA TGATTAATGG GATTCAGTTA TCAAAACAAA AGCGCTATTT AGATAGTATT 780

AACGAAAGTT ACGCTTGGTT AGGGATTTTG TTTGGTTTAG CATTGTTAGT CGTTGGTAAA 840

TTAGTGGTTA AAAATGAAGG CTTATTTACA GCAGGTGCGA TTTTAGCTAG CCTTTCTGCC 900

ATTGCAATTA TTGTGATTCC AATGATTCAA TCGAAGGCCA AATTAAAAGG CTTAGCCAAA 960

GGACTGTATG GACTATACGG TGTGACAGGT TATGTCGGTG ATTTAGTTAG TTATACACGT 1020

TTGATGGCGT TAGGAATTGC TGGGGGAAGT ATTGCTTCAG CGTTTAATAT GCTCGTAGAG 1080

TTCATGCCGC CAGTTGCCCG TTTTAGCGTG GGCATCTTGC TGTTAATTGT TTTACATGCG 1140

TTAAACATAT TCTTATCGTT ATTAGGTGCT TATGTTCATG GCGCACGTTT GCAATATGTT 1200

GAGTTCTTTG GAAAATTTTA CACAGGCGGT GGCCGGGCGT TTAATCCGCT AAAAACAAAA 1260

GAAAAGTATG TCAATGTTGA GAAAAAATAG AAAATTATTT ATTGGAGGAA AATAGAGATG 1320

ATGGATTACT TAATTAATAA TAATGGTGGA ATTGTTTTTG CAGTATTAGG AATGGCAATG 1380

GCGACAATTT TTGCTGGGAT TGGATCAGCC AAAGGGGTTG GCTTTACTGG GGAAGCAGCA 1440

GCAGCTTTAA CGACGGAGCA GCCAGAAAAA TTTGGTCAAG CGTTAATTTT ACAGTTATTA 1500

CCAGGTACAC AAGGACTATA TGGTTTTGTT ATTGCTTTCT TAATTTATAT TAACTTAGGA 1560

AACGATATGA GTATGGTTCA AGGATTGAAC TATTTTGTAG CAGCTTTACC AATTGCTTTT 1620

GCAGGATTGT TCTCTGGGAT TGCCCAAGGT CGTGTAGCCG CTGCGGGGAT CCAAATTTTA 1680

GCGAAGAAAC CAGAACATGC AACAAAAGGA ATTATCTATG CTGCCATGGT TGAAACTTAT 1740

GCCATCTTAG GGTTCGTAAT TTCCTTCTTA CTTGTATTGA ACGTCAAATA ATTGAGGAGG 1800

CAGATGGATG GATGCCATTG AAAAAATCAT CAGTGAAATT AAGCAGCAAG GAAAGCAAGA 1860

AGTCGAAGCC TATGTCACCA GTGAACAAAC GCGCATCGAT CAAGAATTCC AAGCAGCACA 1920

ACAAGAAATT TTGCTGAAGC AAGAACATGA AATAGAAAAA CGACAACAAC AATTATTAAA 1980

AGAATTTAAA CAACGCCAAC AACGGCAAAC ACTAGAAATT CGTCAGGATG CATTAAATAA 2040

AAAACAaGCA TATCTTAACC AaTTATTTGA CGAaGTGGtC TTAAAGATGA GTGAATGGTC 2100

AGCGGAAGAa TTTCAACAaT TCATGAAAGC ACAATTAGGT TCACTTGAAT TAACAGGAAA 2160

AGCAACCATT CTTTTAGGCG AGTATTCTCA AACGAAAGTG ACGCAAGAAT GGCTGACCGC 2220

CCTTTCTGAT GCTACTGTTC AATGGGAGCT TTCAGAGGAA GTTGTTCCAA AAGAAAGTGG 2280

CTTTATCGTT GCAAAAGACG GTTTGGATTA TAATTTTCTC TTTTCAGCAT TAGTTGAAGA 2340

GATTCAAAAA ACAGAAGGCT TTAAAGTAGC AGAAAAACTT TTCAGTTAAG AAGGGAGGAC 2400

GTAGCATGAG AAAACCAACG TATCATCAAA TTAATCCGCT GATTCGTTTA AAAGAAACGG 2460

AGCTGCTTTC TGAGCAACAG TTTCAACAAC TTTTAGAAGC GGAAACAGTT GAAGACGTTA 2520

GAAATATGTT AAAAAGTACG GTTTATCAAC CTTATTTAAC CGAAACTTTT GAAGAAAAAT 2580

TTGACTACCA TTCTTCTGAA GCGCTGGGAA GTCTTTATCG CTGGTTATAT GAAATGGCTC 2640

CAGAACCAGC TGTAGTAACG TTGTATACGA TGCGTTTTAC CTTCCATAAC TTGAAAATTT 2700

TGACGAAAGC TGAGCGAACA GGTAAAGATT TTGATTATCT TTATTTAGAA GATGGTCGTT 2760

ATTCTCTAAA CACAGTCAAA AGTGCCATTC ATACAAAAAA CTCTTCCGAA CTTGAGCCAG 2820

CATTATTAGA AGTGATTTGT GATGTTTTTG CCTATCTGGA AGAGGGCGGT TTACCACAAG 2880

CAATTGATAT CCTTTATGAT CGAGCATTTT TACAGCAACA ACGAAAATTG GCGGATGAAT 2940

TGGGCTATGA AGAGTTAACG AAGGAAGTCA TCGCTTTTAT TGATTTAACC AATCTATCAA 3000

CGATGGCTAG AGGGATTGTT CAACATCAAA ACAGCAACTT TTTATCAACC GTACTTTCTA 3060

GCGCTGGCAG TATCACAAAG AAAGAACTGT TGACCTATGC GGAAAAGTCA CTTGTAGAAT 3120

TTACAGCATT TGTCAGAACA ACGAATTACG GACAACTATT AGAAAAGATT ATCAATAAAG 3180

AAACCAtGAa TTAAATTTAT TGGCTTTtGA AAAACTCAAA GATGATTACC TGACCTCAAt 3240

GTATGAaAAC GGTCGAACGG TCGCTTTtGG ACCGCTCCCA TTACTAGCCT TTTTGAATGC 3300

GAAAGAGGTT GAATGGAAAA ATCTACGTTT AATTTTGGTC GGAAAACATA GTGGGTTCTC 3360

AAATGAAAAA ATTCGAGAAA GGATGCGAAA AGTCAATGGG GTATAAAATT GGTGTCATTG 3420

GAGACAAAAA TTCAATCTTA CCATTTAAAC TGTTTGGCTT TGAGGTCCAT CATGCCATTT 3480

CAGAAATGCA AGTTAGAGAA GCCATTGAAA CTATGGCGAA AAATAAGTTT GGGGTCATTT 3540

TCATAACGGA AGAAGCCTCT ACTTTAGCAG AAGAAACGAT TGAACGATAC AAAGAACAAG 3600

TGACACCAGC GATTATTTTA ATCCCAAGTC ATAATGGTAC AATCGGAATC GGTTTAAGTG 3660

AGATTGAAAA AAATGTTGAA AAAGCAATTG GACAAAATAT TTTATAATGC AAAGGAGAAT 3720

GAGCGTGCAA ATTGGAAAAA TTGTCAAAGT TTCAGGTCCT TTGATTTTAG CTGAAAACAT 3780

GTCAGATGCT AGTATCCAAG ACATTTGTCA TGTAGGAGAT TTAGGCGTTA TCGGAGAGAT 3840

TATTGAAATG CGAGGCGACG TCGCTTCGAT TCAAGTATAT GAAGAAACAA CAGGCATTGG 3900

ACCAGGAGAA CCAGTTATTT CAACAGGAGA ACCATTATCT GTTGAATTAG CCCCAGGTTT 3960

AATTGCCGAA ATGTTTGATG GTATTCAACG ACCATTGGAT ACATTTCAAG AAGTAACCCA 4020

CAGTAACTTT TTAGGCCGTG GCGTTAAAAT TGATGCGTTA GATCGTGAGA AAAAATGGAC 4080

GTTTGAACCA ACTGTGGCAG TTGGTGAAGA AGTGTCGGCA GGTGACATCG TCGGTGTGGT 4140

TCAAGAAACA CCGATTATTC AACATAAAAT TATGGTGCCT TTCGGCGTTT CAGGAACGAT 4200

TGCCGAAATT AAAGCAGGTG ACTTTGCCAT TGATGAAACA GTTTACTCAG TGGAAACGGC 4260

TAAAGGAACG GAAAGTTTTA GCATGATGCA AAAATGGCCC GTTCGGCGGG GACGTCCCAT 4320

TTTAGAAAAA CTAAGTCCCA AAGTACCGAT GGTGACCGGA CAACGCGTAA TTGATACCTT 4380

TTTCCCAATT ACGAAAGGCG GAgcGGCAGC AGTTcCAGGA CCATTTGGCG CTGGAAAAAC 4440

AGTCGTTCAG CACCAAATTG CTAAGTGGGC CGATGTCGAC TTAGTCGTTT ACGTTGGTTG 4500

TGGGGAACGC GGGAATGAAA TGACAGATGT TTTAAATGAA TTTCCAGAAT TAATTGACCC 4560

AACAACTGGT GAGTCTTTGA TGAATCGGAC GATTTTAATT GCGAATACGT CAAATATGCC 4620

GGTAGCGGCA CGGGAAGCCT CGATTTATAC AGGGATTACC ATTGCAGAAT ATTTCCGTGA 4680

TATGGGTTAC TCAGTCGCAA TTATGGCGGA TTCTACTTCT CGTTGGGCAG AAGCGTTACG 4740

AGAAATGAGT GGTCGGTTAG AAGAAATGCC TGGTGATGAA GGCTATCCAG CCTATTTAGG 4800

TAGTCGCTTA GCTGAATATT ATGAACGAGC AGGACAAGTC ATCGCGTTAG GAAAAGATCA 4860

TCGTGAAGGA AGCATTACAG CGATTAGTGC GGTTTCGCCA TCTGGTGGGG ACATATCAGA 4920

ACCTGTCACT CAAAATACGT TACGCGTTGT TAAAGTATTC TGGGGCTTAG ATTCCCAATT 4980

AGCACAAAAA CGTCATTTTC CTTCTATTAA CTGGTTGCAA AGTTATTCTC TTTATTCCAC 5040

AGAAGTAGGG CAATATTTAG ACTTGGAATT GCAAGGAAAC TGGGCCgCTA TGGTAGCTGA 5100

AGGGATGCGG ATTTTACAAG AGGAATCTCA ATTGGAAGAA ATTGTTCGCT TGGTTGGGAT 5160

TGATTCCTTG TCGGATAAAG ACCGTTTAAC gTTGGAAACA GCCAAATCAT TACGGGAAGA 5220

CTATTTGCAA CAAAATGCTT TTGATGACGT GGATACGTTC ACTTCTCGAA CAAAACAAGC 5280

GAAAATGTTG CAATTGATTC TAACTTTTGG TGAAGAAGGT CAAAAAGCCT TAAGTTTAGG 5340

CACTTATTTC TCTGAGTTAA TGGCGGGAAC AGTTGAAATC CGCGATCGCA TTGCTCGTAG 5400

CAAGTATTTA CCAGAAGAAG AATTAGAAAA ATTGGATCGT TTACAAGCAG AAATTAAAAC 5460

AACGATAAAA GAAATCATTG CTGAAGGAGG AATGACGAAT GATTAAAGAA TATCGTACAA 5520

TCAATGAAGT CGTTGGTCCT CTGATGATTG TTGAAAAAGT GGCAGGCGTA AAGTACGAAG 5580

AATTAATTGA AGTACGCATG CAAAATGGCG AAATTCGCCA AGGGCAAGTT TTAGAAATCA 5640

ATGGAGATAA AGCGATGGTC CAAATTTTTG AAGGAACGAG TAACATCAAT ATTCGTGATT 5700

CAAAAGTTCG CTTTCTGGGA CATCCTTTAG AATTAGGGGT TTCGCCAGAT ATGATGGGGC 5760

GCGTTTTTGA CGGCTTAGGT CGCTTAAAAG ATAATGGACC AGAATTATTA CCTGAGAAAA 5820

AATTAGATAT TAACGGCGAA GTTATCAATC CAGTTGCTCG TGATTATCCC GATGAGTTCA 5880

TCCAAACAGG GATTTCAGCG ATTGACCATT TAAATACCTT AGTTCGTGGT CAAAAATTAC 5940

CTGTTTTTTC AGCATCTGGC TTACCTCATA AAGAATTAGC GGCACAAATT GCCAGACAAG 6000

CCAACGTGTT AAATAGTGaA GAAGAATTTG CCGTAGTTTT TGCGGCCATT GGGATTACCT 6060

TTGAAGAAGC GGAATATTTT ATGGAAGATT TCCGTCAAAC AGGCGCAATT GATCGTTCAG 6120

TCTTGTTTAT GAACTTAGCG AATGATCCaG CCATTGAACG GATTGCAACA CCTAGAATGG 6180

CTTTGACAGC CGCTGAATAT TTAGCTTATG AAAAGGGCAT GCATGTCTTA GTTATCATGA 6240

CGGATATGAC AAATTATTGC GAAgCGTTGC GAGAAATTTC AGCAGCACGC CGTGAAGTTC 6300

CAGGACGTCG TGGTTACCCA GGTTATCTTT ACACGAACTT AGCAACGTTG TATGAACGGG 6360

CAGGCCGAAT TCGTGGCTCA AAAGGTTCCG TAACACAGAT TCCTATTTTA ACAATGCCAG 6420

AGGAGGATAA AACACATCCA ATTCCCGATT TAACTGGCTA TATTACAGAA GGGCAAATTA 6480

TCTTGTCCCG GGAACTATAT AAGAGCGGTA TCCAACCACC AATTGATGTA TTGCCATCAC 6540

TTTCCCGTCT TAAAGACAAA GGAACTGGCG AAGGCAAAAC GCGCGGGGAT CATGCaGCGA 6600

CGATGAATCA ATTGTTCTCA GCCTATGCAC AAGGGAAACA AGCCAAAGAA TTAGCTGTCA 6660

TTTTAGGAGA ATCAGCTCTT TCCGATGTCG ATAAAATTTA CGCAGCTTTT GCCCAACGTT 6720

TTGAAGAAGA ATATGTCAAT CAAGGATTCG ATACAAACCG TTCGATTGAA GAAACGTTAG 6780

ATCTTGGGTG GGAACTATTA AGTATGCTAC CGAGAACAGA ACTAAAACGA ATTAAAGAGG 6840

ATATGCTCGA TCAATATTTA ACTGAAGGGA AGTAGAAAAA TGGCTCGATT AAATGTTAAT 6900

CCAACCCGTA TGGAACTTTC CCGATTGAAA AAACAGTTAA CTACGGCCAC GAGAGGTCAT 6960

AAGTTGCTGA AAGACAAACA AGATGAATTG ATGCGACGGT TTATTGCGCT AGTCAAAGAA 7020

AACAATGAGT TACGTATTCA GGTTGAACAA GAAGTGACGG ATGCGTTATC AAATTTTGTA 7080

TTGGCAAACG CAACGTTGAA CGAAGCCTTT ATTGAAGAGT TAGTGGCAAT TCCAGCTGAA 7140

AAAGTTGAAT TAGAAATTAT TGAACAGAAC ATCTTAAGTG TGCCAGTTCC TAAAATGATT 7200

TTTGATTACG ATGAATCGGT TCAAGAAGCA CCCTTAGATT ATGGCTACGT GAATTCAAAC 7260

AGCGAGCTTG ATCAAGCTTT TGCTAAAATT TCAAGCATTT TACCGAAATT ATTAGCACTT 7320

GCAAATGTTG AAAAGACGTG CCAATTATTA TCAAAAGAAA TTGAAAAAAC GCGTCGTCGT 7380

GTCAACGCCT TAGAGTATAT GACGATTCCA CAATTrGArG AAaCAaTTTA CTmCaTTCmA 7440

ATGAAATTwG rAGrAAATGA ACGTGGcGAA ATTACACGTC TAATTAAAAT TAAAAGCATG 7500

AACAAgAAAA TTAGTGGTTA GGAAGAACGA GTGCGAAcCa GTTGGTAGCT TTCGTGCTCG 7560

TTTTTAATAA AAAAGGGAGT GGAAAAAATG AAAAATGAAG AACGCCGCAA AGCGATTGCA 7620

CTGAATTGTC AAAAATACGA AAGTGACTAT GCTCGCTTGG TGGAACCAAT CAATGaGCTA 7680

CTGTTAAACT TGGGAGCGGC TATTTCAGAA GAAGCCgCGA AcAAATCATT CTGAATGTAA 7740

AACGATATCA TCACGGCGTG AAATATCTTC CTGAGTGTCA TTTAGATGAA AGTAACCAAT 7800

TTATCGAAGA TGGGCTTGAA GCTCTTAAAA AGGGAGATTT AGGCAACGGT GCGTTACAAC 7860

TATTTGGGGC CGGCTTAAAC TTTGCTAGTT TTGCAGCAAA AGCATAAGGA ACCAAAAAAA 7920

TTGATGCCCA CCAAATGTTG GCAGAACGTT TtACGAAGTT GTTGTCTGTA CaAACGGATA 7980

ATGATAaCAA ACAATAAGrA AAATACCCTk GGTTTATTAA GTCGACTTAG TAAACCAAGG 8040

syATTTTTTT AACTGTTATT GATTAGGGTA TTCAGACACC TGTAATTGCC ATAACTTATC 8100

TGTCACTGAA GCAATAAACG CAGCACTCTT AGCATTATCA GGTCCTTTAG TAAAAATTGC 8160

TAAAGCAAAG GGGTGAGGTG TTTCTAAAAT CCCCATATCA TGAATAAATT CATCATAGGA 8220

ACCAATTTTA TGAGCAACTT TTCCTTGAGT TGTGGGAGTT TCCATTCTTT CATGGAAAAC 8280

ACTTTGTTTC ATAGAATCAT AAATCGCTTG GTAATCTGGT TTTGTTGCTT TTTCAGTATA 8340

TAATTTTTGT AGAATAACAA GCGCATCTTC AGAAGAAAAT TGTGGCTCTT CaATCGAAGG 8400

TGTCTTGTGC AAATAACGCT GATACATTTC TCGTTTTGCT TTTGCATCAC CACCCAATGT 8460

GTCATAGAGC ATATTTTTAG CGATATTATC AGAATAGGTG ATGGCATATT CTTGTAACGT 8520

TTTTAATGGA TATTCTGGTT GGATATTATA GGCAATGATC CCcAGTTCCT TCTTCATAAT 8580

CTTCTTCGGC GTTATAGGGA ATCAGGTCTG TCCATTTTTT TTGACCAGAA GCAACCGTGT 8640

CGGCAACTAG CATCGTCAGG GGCACTTTAA TAGTACTAGC TGTATAAAAT TCATGTTGGC 8700

CATTCACGGA ACGCTGTTCG CCAGTT 8726

(2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2531 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

GTAAAGACTT TTCACAAGTA TCAAAGAAAT CCGTCGTTAA TCTTTACTAT TCGACAAACG 60

AACACCCTCT TGCTCAGCAA GAGGGTGTTT TTTTGTGTAA TTAAGTGGAA AATCAACATT 120

GACTTTTTTT GAAGTCACAT AAGTGTGAAA ATTAGGTTTT TGGACAATGA TTCAGAAATC 180

TTTTGAAAAT TGAAGGTGTG AGGAAATTTA GAAAGATAAA AAGAGGGAGT ACTGAAAACC 240

GAAATTTCAG TTAAGTGCTT GGATCAAGAA ACAAAAACTA CGTTAkTTTG TCAATATCAG 300

TTGAAATTGG AGAAGAAGCA GAATGGGGTG ATAGTATTGG CAAATTAATT ATTTAAGGTC 360

ATCGTTACAG GATTTTTGTT TTTTATATyC aATCmAAAAG AAATAAATAC AAAGaACACA 420

GATAAGTATT TTAGCAAGTA ACCAATAAAT AGCAGGACTG TTGGTGCTAT AAGAATACAA 480

GAAAATAGTA GTAAACTTAA TTGAGAATCG TTTGCTATTT AAATAGTGTT CTTTTCCTTT 540

TAATATATAT ACGATAAAAT AAGTATATTA ATTGTAAaAG GACATGATAA TTATGGAAGA 600

ATTAGAATGG AATwwTTTTT aTGTTTGCTA TAGTTTTGAG GGTAGTGATT CTTTTAATAA 660

CTGTCAAGAT AGTAGATTAC TTTGnGAATA CTCCTGAAGT ACAAATACAG AGAAAACTTC 720

CTGAGTTTGA AAAAGTGCCA CCTTATGATC CAAGAGATAA TTCATTTCTG ATACAAATTA 780

TTCTTTAAGT AATACAACGA CTAAGTTTAT TTACCAAACC ATTACAATAA GAAAAGGTGC 840

AAGTATACCT AATTACTATT GGTATGATGA TGGTCAATTT AGAGGAAATC TAAAACTTCA 900

ACGTTGGCAA GACTACGGAG AAGATCCAAA CATAAGAGGA TATATCTGTA CTTATGGAGG 960

CTACGTTACC AATGGTTCAA GCCCGGCGCA AAGAATTGTC GATATTAATG AACAATAGCT 1020

ATTCTAACTT TATGTAATAG GAGGGAGTAA TCATGAAAAA AATCGGGTAT TTTAGTTGTA 1080

TTATTTTTTT CATGTTTTTG GTAGGTTGTA GTAATAACAA AAAAGAAAAC GGCAATCTTT 1140

TGAATGCCAG TTCGTTTCCT TTAATACTCA CCACGATTAT TGAAAAAGAA GAAGACCTAA 1200

CGAAAGGTTC AATTTTTTTC AACAAGGATA AAACCATGAC GCTTGAAAAA GAATATTTAG 1260

TTAATCCCAA TAATGAAGAC ACAAAAAAAA CAAGTAGAAC AGAAAAAAAG GTATATAAAA 1320

ATATTAAAAT ACAAGAAAAT AAAGAGAGCT ATGAAATTAT AGGTCAATTG GACAAAAAAA 1380

CGAAAAAAAT AGAGTTTAAA AAAGTTGATG AAGGTAAACG TATATCTGAT GCAGAAGGTA 1440

ATGTGTATGG TGATTTTGGT GGTAAATAGG ATAGTTAAGT CATAAAAAAG AAGGCTAAAA 1500

ATTTTTATTT TTTTCGTCTC TATACTTTCT GATGTTAGTT TTTTTACCAA CATCAGAAAG 1560

TATAGAGACG TTTTTGTGTA ATTAAACAAA AAATCAACAT TGACTAAGAC CAGAAATACT 1620

TCCTTTAGTG CTCCGAGTTG CTCTAATATA AGACACCAAA TTGATAACCA TTTGCTTGAT 1680

AAAACTCTAT TAATTGTGGC AATGCTTGTT TTGTTAATTC CTTATCTACT GAATCATGCA 1740

TTAAGACTAC TCTAATATTA TAATCCGGAT AAACCTCAAG TGAATGTTGA TGGAAAGCGA 1800

GCATTTCTGC GACTGTAGTA GGCTGTCTGT CTAACGGCTC CGCATCTCCA ACCATCGCAT 1860

TCCAGTCAAT CCAATGGATG CCAAGTTGTT TGAATAATTC ATCAGATTTT TCTGTACCTC 1920

CCCATGACAT ATGACCTCCT GGGTATCGCC AGAGAGTTGT TTGAAATGTT TCGCCCAATA 1980

TTTTTTTAAG AGAGTTCTCC ATGGTTGCTA TTTCCGCTTG TATTTCTTTA GTGTTGACAA 2040

TCCCGTTTGG ATATAACGTT GCATAATCAT GAGTTGAACT ATGAAAGCCT ATTGAATGTC 2100

CTTCTGCAAC TTCGCGTTTC ACTATCTTTT GATTTTCGGA AGTTAGGGTA TTTCCTACAA 2160

GAAAAAAAGT TGCGTGGACG TTATACTTTT TTAATGTATC TAGGACTTGT GGGGTtATTT 2220

GATTATTAAC GCCATCATCA AAGGTTAAGA ATACTAATTT TTCACCTTTG TAAGGTTCTT 2280

CTCCAGTTAT ATATTTTCTG ATTTTGTCTG CTGAGTAAGC ATATTTTTTT GATGAAATAG 2340

TCATGTGTGT TGATTCCTTT TCTAACTTGG TCGGGGAATC TTTTTTTGCA GTTGAATTTT 2400

TAGTCGATGT CTGCGTAGTA TCATCATTTT TACTTGTTAT CTTGGGTTGA GTTAATAGGA 2460

ACGTTTGAAT GAAAACAAGA ACAATGAGGA TTAACGCACC TAAATnATAC ATACCCGTnT 2520

ATTTTGAATT T 2531

(2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2717 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

ACCGCCGCcG CCAArGAtGa AcGAAAATAT CTcGAATCCA CCAAAACCAC TACTGCTAAA 60

TCCGCCGCCA CCGAAACCGC CAAAGCCaCC aGCGCCGcCA CCGwAATTAG GATCAGTTCC 120

TGCATGGCCA TATTGGTCAT AGGCCGCTTT TTTCTGTGGA TCACTTAATA CTTCATACGC 180

TTCAGAAACT TCTTTAAATT TTTCTTCCGC ATCGGCTTCT TTGTTAATAT CTGGATGGTA 240

CTTTTTGGAA AGTTTGCGGT ATGCTTTTTT AATTTCATCA TCTGAAGCAC CTTTTGCTAA 300

TCCTAGCACT TCATAATAAT CTCGTTTCGT CGCCATTGAC TTCCCTCCAA CTTTCTTCAA 360

ACTTTACAAT GTGTTTAGAA CCAGCCAAAT TCATCGATTG CTTTCGGTCT CTCTGGTTGG 420

ATTATCTGTC ACACCTTGAA TCAGTGTTAC TAGCTGTGAA TAGGTATTTT CAGCTTGTTT 480

AACATCGTTG TAATCATACC ACAGTTCTTT CAAAACCTAA ATAGTTTCTT GCTAACTTGA 540

TTGTGGTTGA TTTTCTCACA AAAGACAAGG GACTGGGACA TTAGTCATTG TACTAATGAT 600

CCAGTCCCTA AAATCaATTG TTATTGATTA TTTGTCATCA CCATTTACTT CTTCAAAATC 660

AGCATCAACA ACATCATCAG CGCCACCTtG gAGCAGCTTC AGGgATTTTC TTGTGCTTGT 720

TGTTGCGCat TGgTTCGTAA AGTTTTACTG TTAAGTTTTG GACGATTTCA TTTAATGAAT 780

CACGTTTTGC TTTCATTTGT TCGATATCGT TTGCTTCGAT GGCTGCTTTT AATTCATCAC 840

GCGCATCTTC TGCTTTTTTC ACTTCTTCAG CATCTACTTT GCCTTCTAAT TCTTTCAATG 900

TTTTATCAAC AGTGAATAAT AAAGCATCTG CATCGTTACG TAAGTCAACT TCTTCTTTAC 960

GTTGTTTATC TGCTTCAGCG TTTGCTTCAG CATCTTTCAC CATACGTTCG ATTTCATCAT 1020

CAGATAAACC TGAAGAAGAT TTAATCGTGA TTGTTTGTTC TTTTTGAGTG CCTAAGTCTT 1080

TCGCACGAAC ATTTACAATC CCGTTTTTAT CAATGTCAAA GCTTACTTCA ATTTGTGGCA 1140

CACCACGAGG AGCTGCAGGA ATATCTGTTA ATTGGAAACG ACCTAATGTT TTATTATCAG 1200

CGGCCATtGG ACGTTCACCT TGTAAAACGT GAATATCAAC AGCTGGTTGA TtGTCAGCAG 1260

CAGTTGAGAA TACTTGTGAT TTACwTGtTG GAATCGTAGT GTTACGATCa ATTAATTTTG 1320

TAAATAcGCC ACCCaTTGTT TCAATCCCTA ATGACAACGG TGTTACGTCT AATAAgACAA 1380

CGTCTTTAAC ATCACCAGTG ATTACACCAC CTTGGATTGC AGCACCCATT GCTACTcTTC 1440

GTCAGGGTTA ACTGATTTGT TTGGTTCTTT GTTTGTTTCT TTGCGAACAG CTTCAACTAC 1500

AGCTGGAATA CGTGTTGAAC CACCAACTAG GATAACTTCG TCGATTTCAG AAGGATTTAA 1560

TCCAGCATCT TTTAAGGCTT GACGTACTGG TACTTTTGTA CGTTCAACTA AATCACTTGT 1620

TAATTCATCA AATTTTGCAC GAGTTAAGTT CATTTCTAAG TGCAATGGTC CTGcTTCGCC 1680

AGCAGTAATA AATGGCAAGC TGATTTGTGT GCTTGTTACA CCTGATAAGT CTTTTTTCGC 1740

TTTTTCAGCA GCATCTTTCA AACGTTGTAA TGCCATTTTA TCGTTAGCTA AGTCAATGCC 1800

ATTTTCTTTT TTAAATTCTG CAACCATGTA GTCAATGATC TTATTATCAA AGTCATCCCC 1860

ACCTAGGTTG TTATCACCGG CTGTTGATAA TACATCGAAA ACGCCATCGC CTAATTCAAG 1920

GATTGACACG TCGAATGTAC CACCACCAAG GTCAAAGACA AGAATTTTTT CATCTTTGTC 1980

TGTTTTATCT AAGCCATAAG CTAAAGCTGC TGCTGTTGGT TCGTTAACGA TCCGTTCTAC 2040

TTCTAAACCA GCAATTTTAC CAGCATCTTT TGTTGCTTGA CGTTGCGCAT CGTTGAAATA 2100

TGCAGGAACT GTAATAACTG CTTTTTCAAC TTTTTCACCT AAATAGTCTT CAGCAAAACC 2160

TTTTAAATAT TGTAAGATCA TTGCAGAAAC TTCTTGTGGT GTATATGATT TTCCTTCAAC 2220

ATCAACTTTG TAGCCAGCTT CACCCATGTG ACGTTTAATA GATGAAATCG TGTTAGGATT 2280

TGTAACAGCT TGACGTTTTG CTACTTCGCC CACTTgAATT TCACCGTTTT TGAATGAAAC 2340

GACCGAAGgA GTTGTGCGGT TACCTTCTGG GTTTGCAATA ATTTTCgctT CGCCGCCTTC 2400

TAGTACTGCG ACAGCTGAGT TTGTTGTTCC tAAGTCAATA CCAATAATTT TACTCATAGT 2460

GAATATCTCC AATCTGTATT AGTTTTCTAT AATAATATAT TTTTAATTAA CGTTTGTTAT 2520

TGTGCAACAA TGACCATTGA TGGTCGTAAC ACACsATCTT GTAATTTGTA CCCTTTTTGT 2580

AGAACTTCAA CAATGGTATC TGCTGGGGTT TCTTCAGAAG CAGGGACTGT TTGAACGGCT 2640

TGATGTAAAT TTGGATCAAA GGTTCACCCA TTGCTGGGAT TTCTTCAATT CCTTCTTCTT 2700

TTAAAGCCAC AGTAATA 2717

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16870 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

GAATTTAGAA ATGAGGACTA CTGATGATTT TATTACAAGC AAATCAAGTT GCCCGGCATT 60

TTGGCTCGGA AACATTGTTT GAAAACATAC ATTTAGAAAT TGCAACAAAA AGTCGGATTG 120

CCTTAGTTGG TCGTAATGGT GCTGGAAAAT CGACTTTTTT AAAAATCATT GCAGGCATTG 180

ATGCTCCCGA TAGCGGAACC ATTGCCAAAA ATaAAACTGC TACGTTAGGT TATTTAGCTC 240

AAAATACCGG CTTAGAATCA GATAAAACCG TTTGGGAAGA AATGACAAAA GCCTTTGCTG 300

ACATCCTAGA AATGGAACAg CGTATGCGAG aATTAGAAAC TAAAATTAGT GAAATGGAGC 360

CAACCACTTC CGTTTATGAA GGAATTTTAA AAGAGTACGA TCAATTGCAA CATACGTTTT 420

CTGAAAAAAA TGGCTACGGC TATGAAAATG AAATTCGCTC AGTCCTTCAC GGCTTTGGCT 480

TTGATGAATC CTTTTACACG AAAGATATTC AAACCTTATC TGGTGGTCAA AAAACCCGGC 540

TTGCATTAGC GAGAATGCTT TTACAAAAAC CAGACATTTT AATTCTGGAC GAGCCTACAA 600

ACCACTTAGA TATCGAGACG CTTtCTTGGC TGGAATCTTA TTTGCCAAGT TATGCCGGCG 660

CCCTATTAAT TGTTTCCCAC GATCGtTATT TTTTAGATAA GGTAGTTAAT GAAGTTTATG 720

AACTGAGTCG CAAAAAAATG ACTCACTACA AAGGaAACTA TTCCAAATAC TTAGAGTTAA 780

AAGCAGAACA ATTAGCCAGT GAATGGAAAG CGTATGAAAA GCAACAAGAA GAAATCAATA 840

AGTTAGAAGA TTTCGTTGCC AAAAATCTGG TTCGTGCATC TACAACGAAA CGTGCACAAA 900

GTCGCCGAAA AGTATTAGAA AAAATGGACC GTTTAGACCG ACCTCAAGGA GATGAAAAAT 960

CGGCGCATTT TCTTTTCGAT AGTGAAAAAG TCTCGGGAAA TGTTGTTTTA CAAGTCGAAG 1020

ATGCCGCCAT TGGTTACGAC CAAGAACATA TTTTATCCGA ACCTATTCAC TTGGATATTC 1080

GTCGCAAAGA AGCCATTGCC TTAGTCGGAC CGAACGGAAT TGGTAAATCC ACTCTCTTGA 1140

AATCAATTAT TGACCGCATT CCTTTCATTA AAGGAAGTAA AACTTTTGGC ACCAATGTTT 1200

CTGTAGGTTA CTATGACCAA GAGCAAGCCA ATTTACATGG CAATAAAACG GTCTTAGCGG 1260

AATTATGGGA TGAACACCCA ACCACACCTG AAAAAGAGAT TCGAAGTATT TTAGGCGGCT 1320

TTCTCTTCAG TGGAGACGAT GTTGAAAAAA CGATTCCTTT ATTAAGCGGT GGCGAAAAAG 1380

CCCGTGTGGC ATTAGCAAAA CTAGCGATGG ATCGTGACAA TTTCTTGATT CTCGATGAGC 1440

CAACCAATCA CTTGGATATC GATAATAAAG AAGTTTTAGA AAATGCGCTG ATTGATTATG 1500

AAGGAACCAT CCTCTTCGTT TCCCATGACC GTTACTTTAT CAATCGAATT GCAACAAAAG 1560

TTGTTGAGCT TTCTGAAAAA GGCAGCAAAC TTTATTTAGG CGACTATGAT TATTATTTAG 1620

AAAAGAAACA aGAGGAAGAA GAAATCGCTG CCCTCTTAGC TAATGAAGAA GCGGCGAAAA 1680

AACCCGAACC AGTTACAGCC AAAAATACCT TTTATCAAAA CAAGGAGCAA CAAAAATTAC 1740

TCCGTACTTT GCAAAGAAAA ATAACACAAG TCGAAGAAAA TCTTGCTCAG TTAGATACGA 1800

CAATTGCACA ATTAGAAGCA CAAATGAGTC AACCAGACAT TCTAGAAAAT CATGTTGAAT 1860

TGCTGGCTTT AAACCAACAA TTAGATGAAG TTCGTCAGCA ACAAGATGAA CTACTTGAGC 1920

AATGGGAAAA TTTCAGTTTA GAATTAGAAG AAATGGAAAA TAACAATTAA CGGAGGGATT 1980

CACTATGGCT TCCAAAAAAA TCAGTTTAAC GAaGCCCAAA TATCAACAAA TCGCCGTTGA 2040

TGTTGCGGAA AAAATTGCAG AAGGCAAATT ACATGTTGGA GATAAAATCC ATGCCCGTTC 2100

CACCCTAGCT AACCAATACC AAGTGTCGCC AGAAACAGCT CGCAAAGCCA TTATCGTTTT 2160

AGTTGATTTA GAAATCGTTA AAGCCAAACA CGGCAGCGGC TTCTATGTGG CATCAAAAGA 2220

AAAAGCGCAG GACTTTGTGA CACAATATCA AGACGTTCAA ACGATTGCTG AAATTAAAGA 2280

AGAATTGCTA GATAGCGTAG cGAAACAAAA AGAAGAACTT ACTCATTTTT CTAGCATTCT 2340

AGATACTCTT GTTGAACAAA CAAAGCGTTT TGATTCTTTC AATCCAATGA ATCCCTATTC 2400

ATTAGTTTTA ACAGAAGAAG CTGCTTATCT TGAGGCGACG ATTAGCGAAA TGAATTTTTG 2460

GCAAAACACT TCGGCAACGA TTATCGCCAT CAAACACAAA GAAGAGCTTC TGGTCTCACC 2520

TGGTCCATAT GCAAAGATTT CGTTGAACGA CACACTTTAT TTTGTTGGAC ATGACGAATC 2580

AACCTTACAA CGTGTCCAAA ATTTCTTTTA TCCTTAAACA AAAAAGTGAC TACTTATCTT 2640

GCGGTAAATA GTCACTTTTT TTATTTATTA TTTCTATTTT TCATAAAAAC CCACTATAAA 2700

ACCTAGTTAA AATACTGTTT TTCTTCATTT TCAGAAAGAT TTTCATCTAA TTTGACAAAG 2760

TGACAACTTT ATAGTAGTCT GATGTAGTCA CTTTTTTTAA TGAGGAGGAA CAATGATTTG 2820

CCAAAAGTAA AAgTAAATCA CCTTACCAAG ATCTTTGGtA AGAAAACCAA ACCGGCACTT 2880

GAAATGATTC GTGCCaACAA AAGCAAAACA GAAATTTTAG AAAAAACGGG TGCTACAGTC 2940

GGTGTTTATG ATGTAAATTT CGAAGTTGAA GAAGGCGAAA TTTTCGTTAT TATGGGGTTA 3000

TCAGGAAGTG GGAAATCAAC ATTAATTCGT TTATTAAATC GTTTGATTGA ACCAACTTCA 3060

GGAAATATTT ATATTGATGG ACAAGATATT TCTTCTTTAG ATAAAGAAGG ATTACGTGAA 3120

GTACGTCGAA ATAAAATGAG CATGGTTTTC CAAAACTTTG GTCTTTTCCC CCATCGAACT 3180

ATCTTAGAAA ATACGGAATA TGGTTTAGAA ATTCGTGGTG TTCCTAAAGA AGAACGCCAA 3240

GCAAAAGCTG AAAAAGCCTT AGAAAATTCT AGCTTAATTG CTTTTAAAGA CCAATTACCA 3300

AGTCAATTAT CTGGTGGGAT GCAACAACGT GTCGGCCTTG CTCGTGCCTT GGCTAACGAT 3360

CCAGAAATTT TATTAATGGA CGAAGCGTTC TCTGCTCTTG ATCCGTTAAT TCGTCGTGAA 3420

ATGCAAGACG AACTATTAGA TTTACAAGAA AATGTCAAAA AAACCATTAT CTTTATCACA 3480

CATGATTTGA ATGAAGCTTT ACGAATCGGC GACCGGATTG CTTTGATGAA AGATGGTCAA 3540

ATTATGCAAA TTGGGACTGG AGAAGAAATT TTAACCAACC CAGCCAACGA ATATGTGCGA 3600

ACCTTCGTGG AAGATGTGGA TCGTTCAAAA GTCTTAACTG CCCAAAATAT TATGGTACCT 3660

GCTTTAACTA CTAACATTGA AATTGATGGC CCAACTGTTG CTTTAAAACG GATGCGTCAA 3720

GAAGAAGTGA GTATGTTACT TGCCGTGGAT AAAAAACGCC AACTAAAAGG CGTTGTACGC 3780

GCAGAAAAAG CTTTAGAAGC GCGTAAAAAT GGGACTTCTC TTGTAGAATG TGTGGATCCA 3840

GAAATTCAAA CAATCGACAA AGATATGTTA GTGAATGATA TTTTCCCACT GATTTATGAT 3900

GCACAAACGC CTCTTGCCGT GACTGATAAC GGCAAACTAT TAGGCGTGGT CATCCGTGGT 3960

AGCGTACTTG AAGCACTTGC AGAAACAGAG GTGAACGAAC ATGAATAAGT ATCAATTACC 4020

TGTCGCGTCT TGGGTTGAAA GTTTTACTGA TTGGTTAACT TCAACATTTG CTGGATTATT 4080

CAGCTTTTTA CAAACAATTG GTCAAAGTGT AATGGACAAC ATTACTGCCT TACTAACAGC 4140

CGTACCACCA CTTGTGCTGA TCGTGCTATT AACGATTGCT GCTTTCTTTA TTTCTAATAA 4200

AAAAATTGGC TTAAGTTTAT TCACGTTTAT TGGCTTAATG TTTATCTATA ATCAAAACTT 4260

ATGGAACGAT TTAATGAGTA CCGTTACGTT AGTTTTACTT TCTAGTGTTA TTTCCATTAT 4320

CATTGGTGTG CCACTAGGGA TTTTAATGGC GAAAAGCGAA AAAGCCAAGA GTATCATTAC 4380

ACCGATTTTG GACTTCATGC AAACAATGCC TGGTTTCGTT TACTTGATTC CAGCCGTTGC 4440

ATTCTTCGGA ATTGGTATGG TTCCTGGGGT CTTCGCCTCT GTTATCTTCG CTTTACCACC 4500

AACCGTACGC TTTACGAACT TAGGAATTCG ACAAGTGCCA AATGAACTTG TTGAAGCAGC 4560

TGATTCTTAC GGTAGTACTG GCTGGCAAAA ATTATTCAAA TTAGAATTAC CTTTAGCCAA 4620

AAACACGATT ATGGCTGGTG TTAACCAAAC AACCATGCTA GCCCTATCAA TGGTAGTTAT 4680

TGCTTCAATG ATTGGTGCGC CTGGATTAGG ACGTGGCGTC CTTTCCGCTT TACAACGTGC 4740

ACAAGTCGGG AACGGTTTCG TAAATGGTGT GGCTTTAGTT ATTTTAGCGA TTATCATTGA 4800

CCGTTTTACT CAACATTTAA ATAATAAAAA AGCAGCTCCT AAAGCAGCTG GTGCCACCTC 4860

TAAAAAGAAA AAATACGGGA TTATTGCAGC GGTTGTTGTA ATTGTCGCTG GCTTAGTTGG 4920

TGCTTCAATT TTTACAACAA CAAATGATAA GCAAATTTCC CTTTCTTATG TGGAATGGGA 4980

TACCGAAGTA GCCTCCACAC ATGTGGTTGC CGAAGTCTTA AAAGATATGG GGTATGATGT 5040

TAAAACGACC CCTTTAGATA ACGCAATCAT GTGGGAATCT GTAGCTAAAG GTGAAACAGA 5100

TGCCATGGTT GGTGCGTGGC TGCCAGGAAC TCACGCAGAG CAATATAAAC AATATAAAGA 5160

TAAATTAGAC GACCTTGGTG AAAACCTTAA AGGCGCGAAG TTAGGAATTG TTGTCCCTTC 5220

TTATATGGAT GTTGATTCTA TCGAAGATTT ATCTGATCAA GCTGGGAAAA AGATTACTGG 5280

AATTGAACCT GGGGCTGGTG TTGTTGCAGC AGCAGAGAAA ACTAAAGAAG CTTATCCTAA 5340

TTTGAAAGAT TGGTCTGTTG AAACTTCTTC ATCAGGTGCC ATGACTGTCG CCTTAGGACA 5400

AGCAATTAAA AACAACGAAG ACATCGTTAT TACTGGCTGG TCTCCTCACT GGATGTTTGC 5460

AAAATATGAT TTAAAATATT TAGCTGATCC AAAAGGCACT ATGGGCGGCG AAGAAGCGAT 5520

TCACACAATG GCTCGCCAAG GCCTAAAAGA AGATCAACCA GAAGCTTACA AAGTGTTAGA 5580

TAATTTCCAT TGGACCACTA AGGATATGGA ATCTGTCATG TTAGAAATTA ACGAAGGCAA 5640

AGACCCACAA GAAGCTGCAC GTGATTGGGT TGATTCACAT AAAGACCAAG TAGCAGAATG 5700

GnAAAAATAA TTCTTAACAA ATGACTTCCC CATCTATTAA AAGrGGGCGT TGAACAAATT 5760

GTTCAACGCC CTCTTTTTTC GCAAAAAAAA CCACTCAGTT GGGTAGACTG AGTGGAAAGG 5820

AGTTATTTTT ACTTGAGGAG TAAAAATGAA AAGTGTTTAT GTTGGGTGTT TTGTTGGTAT 5880

GCTTATATAT TAACGAGAAC CTGTGAATAA TTTGTGACGA AAACCTTGGA AATTTGCTAC 5940

GAATTACTTA AAAAAGTTTA AATGTGTCTC ACCCTAGTTA ATCGATACAG TGATTATTTC 6000

TTTTTAGCAT TTTGGATGGC TTTAAATAAT GTTGAAAATT GTCCTTTGAA AGACATCACA 6060

TCATCAAAGA TATCAACACC ATCGTTTAAC ATTTTGACGG TAACCCAGCC AAGCGCTTCC 6120

GTAATTGCTA CCGCAACTGA AGCATTGACC ATGCCACCAG TCACAGTTCC GACAACAGGA 6180

ATCACTTTTA GAATATTCCC GACTGCACTC TTACCTAAAC TAACCACTAC TAATTCTTTG 6240

GTTAAGCTTT TCCCTAAGCT TTCAGACCAA GACTGACCAA AAATTTTATG CAGCCGTGAC 6300

ATCATAGTTA ATTGAACAGG TACCAGTAAA AAAGCATCCG AAAACGGAAT TGGCGAAAAC 6360

CCAATAATTG CTGCTGACAA TGCAGCAAAA TGAATCGTTT GATGACTTTT CTTTTTCACG 6420

TCAGGTGAAA CACCAGCCAA AAATTCATCA AATACTTTGT CGAGTTGTGT TTGACTGTTC 6480

GTCACTAATT GTTCTGCTTT ACCTTTAGAA GTTGCGAAAG CATTCATTAT AGCCGTAGAC 6540

GTTTTTTGTG GAAATTTTTT CAAGACATCT TGAAAAAAAG AACGAAATTC TGGGTCTTCG 6600

ATTTCTTCAT TCAATTGCTT AGTTTCTTGT TCTGCTAATT CTTGTTTTTC AATCTCTTCA 6660

AAAAGTTCTT GTTTCTTTTT GTTTTTATGT TTTTTTCTTC TCATCTCATT AACTCCTTTA 6720

TCTATCAATA TGTAGAAACA AAGTATAGCA TAGCCCCTGA TTGATTCTCT TTTAGATGAT 6780

TTGAAAAGCC AGAACTTATC TGTTTTTTAA ACAAATAGAA ATGTTCCAGC TTTTCTCAAA 6840

AAAATCAGCT ATTTTGTTTG TTCTTCACCG CAATAAACCG TCAAGTGTTT AGGCAAAATC 6900

TTCAGCGTAA TTGGCAAAGC AGCTCCTTCA TCACCATCCA CATTAGTGGT TAATTCTTCT 6960

TGGTTTTCTA AGGAGATTGT TCCTTCTTTA AAGGTTAGGT AAACTAAATT ATCAGTGGAC 7020

TGGTCGACGC CTTTTAACAA ATCTGGGACT GCTTTTACTG CATCCCATAA TGATTGATcT 7080

TTTAAATAAA CTAAATGAAG TTTGCCATCA TCCACTTGGG CCTCTGGTAA TAATGTTTCA 7140

AACCCACCGA TTGAATTCGT TAAGCCAACT AAAACGGTAC TACTTTCAAT GGTTTGTTCT 7200

TTTTGATCCA AGCTTAAATG AAAAGGATAC GTTTGAGCAT TCGCTAAATG TTTGGCACCT 7260

GAAATAAAAT AAGCGAGTTT TCCTAACTTC GTTTTTTGTT CAACGTCCAC GTCGTTAATT 7320

GACTCTGGAA TCGTCCCAAT TGCAACGACA TTCATAAAGT AGTCGTCATT GATTTTCCCC 7380

ACATCTAAAG CACTGGTTTT TTCTAAATCT AATTGTTGAA TCGCTTCTTC AGGGTCCATC 7440

GGTAAATTCA AGGCTCGCGC TAAATCGTTC ACCGTTCCTA AAGGAAAAAA GCCAAATTTT 7500

GGGCGGTAAG CTTGTTCTGC CAACCCGCTA ATTCCTTCAT TAACGGTTCC ATCTCCCCCC 7560

ATTACGAAAA CACTGTCGAA GTGACTTTCC GCAGCTTCAC GGGCAAATTG TTCTGCATCG 7620

CCGCCTTTTT CTGTTTGTTT CACTACAACT TCATCAAATA ACTGCTTTAA TTTTTCTTCT 7680

GCTAACGTTT CAAATTCTTT GGCTTTTTCA CCACCAGAAC TGGGGTTAAC AATTAAAACT 7740

GCTTTTTTCA CTGGTACCAC TCCTCATTGA TATCTAAATC GTTTGATTCC TTACAATTAT 7800

TATAAAAGAT ATTTTTGATT CTGGCGAAAA ATATGCTGAC ACTTGCTTAA CAAACAAGCA 7860

TTCCTGCTCC TCACTACTTT TCAAAACGTT CTTTAAAGCA ATAGGCAAAA TAATAGAACA 7920

ATTTCCAAAA ATCTTGCCAA ACTTTTGGAT CCATCCCCGT CAATTCATGA ATTCGTTTCA 7980

AACGGTAATT CAATGTATTT CGATGGATAT GTAATACTTG TGAGGTTTCA TTGATGTTTA 8040

AGTTATGaTT GGCGAAAAAA ATAACGGTTT CATAGTATTC ATCATAAATC GTCTGGATGG 8100

TTTCAAAAAA ATCCAATTCA ATATGAAAAG GAAAATTAAA AATTCCTTGA ATGAAAAAAT 8160

CTTCTGCTTG GTAAAACTTA GCCTTTTCTA ATTGGAGAAA CGATAGAAGT AAATACGTGT 8220

TTTGTGCCTC TTGGACCATT TGTTGAATCG TGGCTTTTTC CGAACTTATC ACAAGGTATT 8280

CTGTTACTGG CTTTGTTATC TGAGGTGGTT CAGAAGTAAT CACCAGATAA CAATTTTCCC 8340

ATTCAAATAT TTCATGTCCA GGATAATAAC TGGATAATAA GCGCTTGTCC TTGGTCAGTA 8400

CACAATGATT CTGGTCATCT AATTGAAGGT TATATAGTTC CAACGCTTCA AGTTTTAAGG 8460

CTTCTGAGTA AGAAGAATCA ACATGCATTA AGTGCTGGAT AAACTGATTT TTTTGTTTCT 8520

TCTTTTTTTC GCGTTCTTGT TGTGTATTTA ATTCTTCTGT CAGTAAAAGA ACAATCGAGC 8580

GCACAAGCTT GGTAAATTTG CGCACCTCGT CTGGTTCTCC TGAAATGCCT ACGACACCTA 8640

AAATGTGATT ATTCAAAATA ATCGGTTCAT TGGTGCCTTT TTTCTCTAAA GAAGTATCTT 8700

TATAAATTTC AATATTTTTT CGTTCATTGA TGGCTCTTTC TGCACCACGA TGGCGCTCCC 8760

CAATTCGAGC AGAATCTCCA CTCGCAATAA TTTTTCCTCG TTCATCCATA ATATTGATGT 8820

TATAAGGAAT ATCTGCCATC AACTTTTCAA CAATTTTATT TGCTTGGATT TTGTCTAATT 8880

TTAACATTGC TATCGCCAAC TTTCTAAAAA GCGAGTaGCG TTTGGGACAT AAGCCGCAAT 8940

GACTTATGTC CCAAACgCTT AATTCTTCAG ATCAACGACT GAACATTTTT ATTTCAGCTT 9000

TCATTACTGA TGATACCATG GTTAGTTCTT AGTCACAAGG CACCAGCTGT AATGTACGAG 9060

CAATATTTTC AACGGTTCTT TCTAAATTAA CTGGTCCATT TTTCAATGCT GCTTCCAATG 9120

GTTCGGCTTT GGCCAAAATT CCAAAAATAG CCGTCATCCC TTCTTCATAG AGGACATCCA 9180

CATCTTTGCC GATATAGCCA GCACAAGCGA AAACCGGTTT TTGATACTTC TTGGCCAAAC 9240

GCGAGACGCC ATAAGGGGCT TTGCCGAATT TTGTTTGAAA ATCCATACCG CCTTCACCAG 9300

TAAAGACATA ATCAGCGGCT TTTATTTTTT CTTCTAATTG ATTCACTTCT ACGACAATAT 9360

CAATTCCTGG GCGTAACTGA GCTTTTGTGA ATGCTAGTAA CCCTGCACCC AAACCACCAG 9420

CAGCTCCTGC GCCTGGCATC ATCTCAACAT CAATGCCGAT CTCTTTTTTG ATGATCGCTG 9480

CATAATGGGC TAAGTTTTTA TCTAATTGAG CCACCATTTC TTCTGTCGCA CCTTTTTGCG 9540

GACCAAAGAC ATGAGACGCC CCTGTTGGCC CTGTTAGTGG GTTATTCACA TCACTTGCTA 9600

CTAGAACTTC GGTAGCAAAA ATGCGTTGAT CAAACTGTGT TAAATCGATC TGCGCCAGTT 9660

TATCTAATGC ACCACCGCCA CGTGTCAATT CTTGCCCTTC CTTGTCTAAC AAACGCGCAC 9720

CAAGTGCCTG AATCATACCT GCACCCCCAT CATTTGTCAC GCTACCACCA ATGCCGATAA 9780

TGATTTTTTT TGCCCCATGA TTGAGCGCAT GTTGAATCAT CTCTCCTGTA CCATAAGTAG 9840

AAGTGATCAA GGGATTGCGT TCTGCTGCTT TAACCAATTC AATTCCGTTG GCTTTCGCCA 9900

TTTCAATGAC CACCGTTTCT TGGTCACCTA ATAAGCCGTA ATAGGTTTCA ACGATTTGAT 9960

CAGGAAGCGG TCCTGCCACG CGCACTGGTA CTCTTGTACC TCCTGTTGCA TCCACTAAGG 10020

AATCCACCGT TCCTTCCCCG CCGTCTGCCA TCGGGACATG TTGAATGTCT GCATCTTTAA 10080

ATACAGCTAA AATCCCTTTT TCCATTGCCA GACAGGCTTC TTTTGCACTC ATACTTTCTT 10140

TAAATGAGTC TGGAGCCAAA ACAAATTTTT TCGTCATCTT TTCGCGCCTA CTTTCCAACT 10200

AGTTTAAAAT AAAGCCATAT AAAATGGTTG CAACAATCGT CATTACCCCA CCGACACACA 10260

TTTCATAAGG AATCACTCGA ATTCTTTCTT TAATCGTCAT ATTCATACTT TGTGCCGTAA 10320

TATGGAAATA ATTGCCTTGA GGAACACTAT CAATCACAGT TGCACCAGTA TGCACCATGA 10380

CTGCTGCTGA AAGTGAAGAT GTGCCCATAT CTAAAATCGC TGGTCCAAAA GAACCTGTCG 10440

CTAAAATAAC CGCTGTGGAT GTTGAACCAA CTGCTAAGCC CATTAAAATA CCAGAAATCG 10500

GTGCCAAGAA GGTTCCAGAA ATTCCCATAA CAGAAATTAA GTGGACAATT TGTGTTGATA 10560

AATCAGAAAC AGAAATTAAT CCAGCAATTG CACCTGCACC AATTAAAATT AAAACGGTAT 10620

CGGTTACTCT CTTCAAGCCA GAAGATGCAT AAAAGGCCAG TTGTTTTGTT TGCCCAATCG 10680

CAATTGTCCC AACAACTGAG GCGAACGGCA AAATATACAG AGCATCAATA TTCAAACTTT 10740

TGATGAAAGA AATATTTAAA ACGGCGCCAA TCGGATTAAG TAATAATAAG ACAATCGCTA 10800

ACACTGGCGC GATCAACGCC TTTTTCAAAG GTAATAATTC TTTTGTTTGT TCCTCTACAA 10860

GATCTGTTTC TTTGACCATT GATCCTTTGT TTCTCATGAT ATTTGCAAGA ACAACAGCTG 10920

TAATTAAGCC AAATACGGCT GGAATAAAGC CAGCAATCAT AACGTCACTT AGAGGTAAAT 10980

CAAAGCCATT CGCCGCTGCA ATAGTATTTG GATTAGGCGA AATAATATTG CCCGCTTTGC 11040

CGCCACCTGA TAAGGCAATC AATAATGATA ACTTACTGAT ACCCGTACTT TTCCCAACAG 11100

ATAGAGCAAT CGGGGCAACA ATCAACACTG CAACAGGAAT GAATACGCCC ACTGCTGTAA 11160

TTACCATTGT GGATAATGCC AAAGATAAAA TAGCTTTTTT CTCGCCAAGC TTTTGTACAA 11220

TGGCGTTGGC AATTGATTCT GCCGCACCAG ATTCCATCAT CACACCCGCC AACATGCCGG 11280

CTGCTAAGAC ACGGATTGCG GTTCCCATAA TACTTTGTGT TCCGTTTAAT AGAACAGCGA 11340

CTGTATCTGA CAGGGAAGCG CCTCCAATAA GACAACCAAC AATTGCACCA AAAAACAAGG 11400

AATACACAGG ATTAAATTTT TTCAAAATTA ATACTATCGC TACAACTAAT CCTAAAATTG 11460

CAGCAACCCA AGATATTGTA ATATCTGGCA TTCAAAATCA GCTCCTTAAA TTGTTTACGC 11520

TTTCATTCTA TCAAAGTCCT TTCTTTTTAA AATTGTAAAT TTAACAGGAT TTATTGTTCA 11580

AATGCACAAA AAAACCGTTC TTAAAGAACA GTTTTTTGAA ATCTTTTTCT ATCTTTCTCA 11640

ATAATGACCA CCAATCCAAC GaACAAAAGA ACACTCACAA TTAACAATAC TGCCGTTTTA 11700

TTATATCCAT GAAATAGCkT cGCCTCCACA TGCGTAAAGT ATTGAGGAAG GAATTACACC 11760

AATGAATATA GCGGCTATTA AATTTTTTAA TGGCAACTGC ATAGTATCTG CAGCAAAATT 11820

AATCACAGAA GAAGGAATAA TTGGAATCAT GTAGCCCACC ATAATCCCAA TTTTTGGATG 11880

CTTCATCCGT CGAATGGACT GGACCCAATG ATTGGTCTCC GTTTTTTTAT CTAAAAATTT 11940

TATATGATGC ATCAAAAAGA TTGCAACCAA ATTTCCCATA CTATTCCCTA AAACGTTAAT 12000

GAGTGTGCCG CTGAAGGGAC CAAAACTCAA TCCTGCAATC ACACCAATTA CTGAGGTCGG 12060

AATCCCTGGT ACTGCACACA TGATCGCGGT CAAAAAAACC AACAATAGCG CTGCTGAAAA 12120

ACCATGTGAA CGAATTCGCT CTAATAATTG TGTCTTACTG GCTTTAGGAT TCAAAAATAA 12180

TTGAATATCT GTTTGATATT CTAGATACAA ACGGTAACCT ACCAGCAGGA TTAAAATAAG 12240

CCCGCTACTG ATTAGTAGTA TTCGGATTGT TTTCCTCTTC ATTTTCTTTG CTCCTTTTAT 12300

GCTTTATACA AAGTGTAGCA CAGGATTCTT TGTGCCTCAA TTAGAAGCTT GTATGAATAA 12360

AAGCCGCCCC TAAAAGTTAG ACTAAAAATC TAACTTTTAG GGGCGATTCG AAAGTTACGA 12420

AGATACTTTT TTTATTTCCG ATACATCTTA AACTCCAAGA ACAGATCGTT ATAAATTCCT 12480

AAGTATTCTT GACCTAAGTC TTGGTAAACT TCTAAATGTT TGATAGTTTC ATCATCTGGA 12540

TAAAATTGTT TATCTTCGGC AACTTCTTTT GGTAATAGTT TTTTAGCTTC TTTATTTGGT 12600

GTGGAATAAC CAATATATTC TGCATTTTGT GCCGCATTTT CTGGTCGTAA CATAAAGTTC 12660

ATAAATGCAT AGGCACCCTC TTTATTTTTG GCTGTCTTAG GCATCACAAT GTTATCAAAC 12720

CAGAGATTAG AGCCTTCACT GGGAATCACA TAATGTAGAT GTTCATTGTT TTCTAGCATT 12780

TCAGCAGCTT CACCAGAGAA AGTTACAGCA ACTGCACTTT CTTCATTAGc CATGTACATT 12840

TTGATTTCAT CGGCAACAAT TGCTTTGACA TTGTTCGTTA ATCGGTTTAA CTTATCGGTA 12900

GCCTGACGTA ATTGTTGGTC GTTTTTACTG TTTAACGAAT AGCCTAAACT GTTCAAAGAT 12960

AATCCTAACA CTTCGCGAGC GCCATCAATC AGCATGACAT TATTTTTTAA TTCCGGGCGC 13020

CATAAATCAT CCCAATGTTG GATCTGACGG CCGTCAATAA ATTTATCATT ATAAATAATC 13080

CCCAACGTGC CCCAGAAGTA CGGAACGGAA AACTTATTTT TGGGATCAAA GGATTGATCT 13140

AAAAAGCGTG CATCAATGTT TTCTAAGCCT TTTAATTTGC TGTGATCAAG TGGTAAAAGC 13200

ATCTTCGCTT TCATCATTTT TTGAATCATA TATTCAGAAG GAATGGCAAT ATCATAGGCT 13260

GTGCCACCTT GCTGAATTTT TGTATACATA GCTTCATTAG AATCAAAGGT TTCGTAATTG 13320

ACTTTATAGC CTGTTTCTTT TTCAAATTTC TTAATCAAGG CCGGATCTAT ATAGTCCCCC 13380

CAATTGTAAA TGGTCAAGGT ATCTGCTCCT GCCATGCCAC TTGCTTTCTC CAATTGGCGC 13440

ACGCCAAAAA ACAAGATGAC AATAATAGCG ATAATTCCAA TAAAAAGTGA CTGTAATTTT 13500

TTCATTTTAA CTTGGCCACC TCACTTTGTT CCCGACGCAT TTTCTTTAAA CGTTTTGGCG 13560

TATTGTCTTT ACTGATAAAG TAATAACCGA CAACTAGAAT CATTGAGAAT AGGAATACTA 13620

ACGCACTTAA GGCATTGATT TCTAAACTAA TTCCCTGACG TGCTCGTGAA TAAATTTCAA 13680

CAGATAATGT TGAAAAACCA TTGCCTGTCA CAAAGAATGT TACGGCAAAA TCGTCCAACG 13740

AATAAGTAAA GGCCATAAAG TAGCCAGCAA TAATTCCAGG CGCTAAAAAA GGCAAAATAA 13800

TATTTTTAAT CACTTGAACA TTATTGGCCC CTAAATCCCG TGCCGCATCT ACCATAGAAT 13860

CGTTCATCTC TTGTAGTTTT GGTAAAACCA TTAAAACAAC AATTGGAATA CTAAAGGCAA 13920

TATGCGACAA TAATACACTT GTAAAACCTA AACTGATAAA CCCTACTGCC GTGAAGAAAA 13980

TTAAGAAACT GGCTCCAATA ATAACATCTG GTGAAACGAG TAAAATGTTA TTAAAGCTAA 14040

GCAACGTGTT CCGGTTTTTA CGACGTTTTG TATAATAAAT CCCCATCGCA CCAAATGTTC 14100

CAATGATGGT TGCTATTAAA GCCGATAAGA AAGCTAGCAA AAACGTATTC AGAACAATGG 14160

TTAACAAGCG TGTGTCTTCG AAAACGTTCC GATAATTATC TAACGTAAAA CCAGTAAAcT 14220

GATTCATCGT TTGCGTGTCG TTAAACGAAT AAAAAAtCAA ATAAAAGATT GGTGCGTACA 14280

GTAGCGCAAA GACAACGAAT AAATACACAT AAGACCATTT AAACTTTTTC TTTTGCATTA 14340

TTTGCCACGC CCTTTCTTCT TCTCACCCGT CAGCAACATC ACAATAAACA TGGCAACAAT 14400

TAAAATCACA CCTATTGTTG AACCCATGCC CCAGTTTTGC GTTACTAAGA AGTGTTCCTC 14460

AATCGCCGTT CCCAACGTAA TCACCCGATT GCCGCCAATT AAACGCGTTA ACATAAAGAG 14520

CGAGAGCGAA GGAATAAAGA CAGCTTGAAC GCCACTTTTC ACACCGTTTA GGGAGAGAGG 14580

AAAAATGACA CGGCGGAACG TCTCCGCACT GCTAGCTCCT AAGTCCCGAC TAGCACTAAT 14640

CAATGAAGGA TTCATCTCTT CTAAGGCATT AAAAATCGGC ATAATCATAA ATGGAATTTC 14700

AATATACGTT GCGACAAATA AAAAACTGAA ATCTGTAAAC AGAATCTGGT GCGTACCCAA 14760

ACCCATAAAA CTCATAAACT GATTGATACT CCCATGAATA CTGAAAATCC CAATAAAGGC 14820

ATAAGCTTTT AATAATAAAT TCACCCACGT TGGTAAAATA ACAAGCATTA ACCACAACTG 14880

TTTATGCTTG AGCTTTGTCA AAAAATAAGC CGTAGGATAA CTTACCAATA ACGTAAAAAT 14940

GGTAATCAGA AAAGCATACC ACACTGAATT AATGGTCATA CTTAAATACG TTCCTGATGT 15000

AAAATACGTT TGGTAATTCG CCAAGGTAAA CTGGCCATTC ATATCGAAAA ATGATTGATA 15060

AATAATCATT AGCACTGGCG CAATGACGAA TAAAAGTAAC CACATTACAT AGGGAATCGA 15120

ATAAATTCGA CGCATTGTTT TCATTTTATT TTCCCCTTTC GTTTACTCGT CGTAACTTTC 15180

TAAACGGGCA TCAAAGTCTT CTTCTGATTC ATTGAAGCGC ATCACGTGAA TATCTTCTGG 15240

CTCAAAAGAG AGCCCTACTT GAGCGCCTTC TTTGGCTTTT CGAGTGGAAT GCACCATCCA 15300

TTCGTTTCCT TGTTCATCAA AACAAATAAT TTCATAATGA ACGCCGCGGA ATAACTGCGT 15360

GTCCACCGTT ACGGTTAGCT TGCCTTTTTC AAGCGTGGTA ATCGTTAAGT CTTCTGGACG 15420

TAAAACGACT TCTACAGGTT CATTTTTCCG CATGCCGCCA TCGACACATT CAAACTTTTT 15480

GCCCACAAAT GAAACTAAAT TATCTTCCAG CATCACGCCA TCCACAATAT TACTTTCCCC 15540

CACAAAGTCA GCAACAAAGT GATTAATCGG TTCATCATAA ATATCGACTG GTGTGmCACT 15600

TTGCACGATA TGACCTTTAT TCATGACAAA AATTTCATCA CTCATCGCTA ATGCTTCTTC 15660

TTGATCaTGG GTCACAAAAA TAAAAGTAAT TCCTAAACGT TGTTGTAAAT CACGTAGTTC 15720

ATATTGCATC GCTGTGCGTA ATTTCAAATC TAAAGCGGAT AACGGCTCAT CTAAAAGCAA 15780

TACTTTTGGC TCATTGACAA TCGCTCGTGC AATTGCGACC CGCTGACGTT GGCCCCcAGa 15840

CATTTCACTA ATTTCaCGTG TTTCGTAGCC CGGTAATTGA ACCAAACGTA ATGCCTCTTT 15900

CACTTTTTTT TCAATTTCTG CTTTGGATAG TTTTTTGATT TTTAAACCAA ACGCAACATT 15960

ATCAAAAACA TTCATGTGAG GAAACAAAGC ATAATCCTGA AACACGGTAT TGACTTGGCG 16020

TTTATTGGCA GGTACATCAT TTATTCTTTT ACCATCAAAA TAAACATCGC CAGAAGTCAC 16080

GTCATTGAAT CCCGCAATAA TACGCAAAAT CGTTGTTTTG CCACATCCTG AAGGTCCTAA 16140

TAACGTATAA AATTTTCCTT CTTCAATTTC AAAGCTGACA TTTTTTAATA CTGGTTCATC 16200

GTCAAATTGT TTGACAACAT TTTCAAATGA AATCACATGG TTTCTCACTG GATAAACTCC 16260

TCCATTTCGT CCTTTATAAA TACGATTCTG TCGCCACAAT TAACGCACGA CTTTTTCCTC 16320

GACCATTGTT GCTAATTTGG TGAGGTTCTG TTGCTTGGTA ATACATGGTT TGACCTTTTT 16380

TAGCAATATA CGTGGCTTCC CCTAATGTCA AACGAATGGA GCCATCAATC ACATAAATAA 16440

ATGTTTCTGA TAGAGAAGGT TCAAACGTTT TGTATTCGCC ATTTTTATCA AAAGTAATAA 16500

TGACAGGTTC CaTTTCTTTT TCATTCGAGG CTGGAATTAA CCATTTTAGT TCATAGCCAT 16560

GTTCTTCATC ATAATAGAGG GTACTTTCTT CTTCATTGTA GACAATTTTC TGTTCAAGCG 16620

GTTGCTGACT AAAAAACTGC TCTGGCGTCA CGCCTAGTAC TTCTAAAATT GAGAAAAATG 16680

TTTCCATTGA TGGTGAGCTT AAATCCCGCT CCAATTGGGA AATATAGCCT TTGGATAAAT 16740

CCGTTCTTTC GCCTAATTCT TCTTGTGTTA AATTTTTCTG AATGCGTAAA TTTCTCAACT 16800

TCTCACCAAT TTCCATGGCA CACCTCGTTT CTATTTTCAG CTAATCTTTT CAACAAAAGT 16860

TTTGTAATAC 16870

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3905 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

GAATTATTTC AATTACCTGT TGACTATCAA GATGAACGAT TAACAACTGT CCAAGCAGAG 60

CGTTTCTTGG TGGAGCAAGC AGATGCTTCA AGAGCCAAAA GAAAAAAAGT GATTGATAAA 120

CTAGCGGCAG TAATGATTCT ACAAAATTAT TTAGATGCAC ATAGCCGATA ATTATCGAGC 180

TTTGAATACA TCAnGGGAGA GAAGAAGAAT GACAGAACAT AACCACGATC ATGATCACGA 240

GGACATGAAC ACATTACTTT AGTAGACGAA CAAGGAAATG AAACATTATA CGAGATCTTA 300

CTAACTGTTG ACGGTCAAGA AGAATTCGGT AAaAATTATG TTTTATTATA TCCAGCAGGA 360

ATTCCTGAGG ATGAAGATGT TGAATTGCAA GCTTACTCCT ATGTGGAAAA CGCAGAAGGT 420

ACAGAAGGCG ATTTACAGCA AATCGAAACA GACGCTGAAT GGGATATGAT TGAAGAAGTT 480

TTCAATACAT TCATGGCTGA AGAAGAAGAG TAAAACTGTA ACGAGAGAAG CGCTACTTCT 540

CCTATTGAGA AGTAGCTTTC TTAATAATAA TTGTTATAAT GTACATGTTG TTTCTATCAT 600

AAGATACTAT CACTATCATT GAGTGTGATG GATAAAGAAG GACAGCATTG TCTAGTTTTA 660

CAAGATGTTG TAAAAACGTA ACATCCTTGT AATGACGTTA ATTTAGTAAC TATCTAGCAT 720

AATTTATAAT TTGTGTTATA CTAAGTATGA GGTTTTAGTA ATACTGTTAA TTTkAkAaTA 780

TTAAATAAAA AAATGGCTCC ACCTACTGGC ATAGGTAGAG CCGTCAGAAA TTCAGTTAAA 840

CAAACACGTA ATTATTGGTG TGGTGCATTG CCACGAACTG AAATTCCAGT TAATAAAGCA 900

AGTAAATCAA ACATATATTT CACCTCCTTT ATTATGATAG CTTCATTATA ACATTATCTG 960

TAAAATGATG GTATGTTTAA TACAAAAAAA ATAGAAAGAT TGGAGTTTTG ACTGTGAGAG 1020

TAGCGGGAGA AAAAATCAAA ACAGCGCGTA AAAAGAAAAA ATTATCGCmA GCTGAATTGG 1080

CAAAAGGAaT TTGTACACAA GCmACAATTA GTAACATTGA AAACAAAAAT GTGTGTGACA 1140

GTTTAGACAT TTTCTCATCT GTATGTCTAC GTTTAGACTT ACAAGTAGAA GAATGTATCG 1200

AAGGTTCTAG TGAGAAAAAA CTCGAAAGCT TATTAAATAA AGTAGAAGAA TTATGCTTTT 1260

ACTTTAAGCA CGATGAAGCC TATGATTTGT TGAAAGACTA TCCAGATGAT ATCGAAAGTT 1320

CAAACAGAAT TTTAGAAACA AAATTTTTCT ATTATAAAGG AATTACGAGT TTATTAGGTA 1380

AGAAAAATAA TAGCGAACAT TGTTCTATTT ACATCGTGGT TCTGAAATTA GTCGTGATAT 1440

TAATATTTAT AATATTtTGA GTATGAATGC GATCGGTATT TTATATGAAT TAGAAGATGA 1500

TATTGAAAAA GCAAAAGTTT ATTACGATAA ATCTTTACAA TTGCTAAGCG AATTCAAATT 1560

AGATTATCCA TTAGAGCAAT GCCGTATTTA TTATAATACC GCAAAATTCT ATTCATTAAT 1620

TAAAGACTAT GCAAAGAGTA TTGAATTAAG TGATAAAGGA ATCGAAATTA ATCGGACGCA 1680

CAGTTCTATT TACTCTTTAG ATTGCTTATT ATACGAAAAA GCGTTTAATA AACAAATGCT 1740

AGGACTAGAT GCTGTTGAAG ATTATCGTAT TGCGTACTAC TTTACACGTT TCTTTGAAAA 1800

TAAAAAATTA CTTACTTATA TTGAAAAAGA TATGCAAGAG TTTAATATCA GTTTTAAATA 1860

ATAAATTATT GTTTCAGATT TtAAAAAGAA CAACCTAACT CAAGAGGAGT TAGGTTGTTC 1920

TTTTTTCTCT TTTTAATTTG TTTCTTTGAC AGCGTTACGG ACGCTTTCGG CGACGACTTT 1980

TGCGACTCCT TCTTGGAATG GATCGGGAAT AATATTAGTC GGCGTGAGTT CATTATCGGG 2040

AATCAATTTT GCAATACCTT TGGCAGCGGC AATTTGCATT TCAATCGTAA TCTTTTTCGC 2100

CCGAGCATCT AAGGCACCAC GGAAGATACC AGGGAAGGCT AACACATTAT TAATTTGATT 2160

CGGAAAGTCA CTACGACCAG TTCCGACAAT ATAAGCGCCT GCCGCTAAAG CTTCATCAGG 2220

AAAAATTTCA GGAACTGGGT TAGCCATTGC AAAAATAACT GGCTGCTCAT TCATCTGTTG 2280

AATCCACTCA GGTTTTAAAA CACCTGGTGC CGAAACACCA ACAAAGACAT CAGCGCCTTC 2340

AAGAGCAGTA GCTAAATCAC CGGTACGgTG TTcCCGATTG GTTAGcTTCG CAATTTCTGC 2400

GTGGTGTGGT GGCAGAGCTG TATCTGTTTC GCTCAGAaTC cCTGCACGGT CAACAATGAT 2460

AATATGTTTT ACGCCAGCAG CTAAAAATTT TCTTGTGAtC GAAAGACCAG CGGAGCCACC 2520

GCCGTTGATT ACCACATGAA TGTCCTCTAT TTTCTTGTTG ATTAATTTTA AGCTGTTGTA 2580

CAAAGCCGCT AGAACAACAA TCGCTGTCCC GTGTTGATCA TCATGGAAAA CGGGAATATC 2640

TAGTTCATCG ATTAAACGCT GTTCAATTTC AAAACAACGA GGAGCGCTGA TATCTTCCAA 2700

GTTAATCCCC CCAAAGGTTG GAGCCAAAAA TTTCACCGTT TGAATAATTT CTTCGGTATC 2760

TTGTGTATCT AAAACAAGGG GAATCGAATC CACACCAGCA AAGCGTTTAA ACAATGCAGC 2820

TTTTCCTTCC ATGACAGGCA TCGCTGCTTC AGGACCAATG TTACCTAAAC CTAAAACTGC 2880

TGAACCATCA CTAATAACGG CGACGGTGTT TTTCTTTGTC GTTAATTCGT AAGCGCGTTC 2940

TTTTTTTTCA TGTATGGCAG AAGAAACAGC CGCAACACCT GGTGTGTAGG CAATTCCTAA 3000

ATCTGCTTTT GTTTCAATCG GAACTTTAGG GCAGACTTCT AATTTCCCCC CAAATTTTTT 3060

TGCTTGCTCA ATCGCTAATT CTTTTACATC TTTTGTTGAC ATATCATTTT CCTCCTAAAT 3120

AATTCACTTA TAGTTAAAAA AGCATTCTAA AAATAGCAGT CATTGTTATT ACTGTAATGG 3180

CACCGCCTAA ACGTGTCGCA ACTTGTGCAA ACGGCATTAA GTTCATTCGA TTGGCTGTAC 3240

TTAAAATAGC CACATCGCCA GTCCCACCCA TACCACTCTG GCAAGCGGAG ACAATGGCAG 3300

CCTCAACGGG ATTCATATTC ATGAAGCGTG AAACGAAAAA TCCTGTAGCA ATAAywGTGA 3360

AAACGACACT AATTACCACA ACAAAATATT GCCAGCTAAG CATTCCTACG ACATCTTTTA 3420

AAGGAATATA TAACATACyT AAGCCGACCA TTAACGGGAA CGTAAAATTG CCAGAGATAA 3480

ATTTATAAAG CTGTTTTGAT CCGCGTTGGG TTTCTTTCGG TACTACGTTT AAGTATTTTA 3540

AAAAGGCTGC GACAACAATC ATCAAAACAG GACCAGGAAA GCCTGTTAAG TGTTGCAATA 3600

AACCGCCTGT AATAAATAAG GTACACGCAA TTAAAACACC AGCGCCCATT AATTTTACAT 3660

CAATTGGTGC TTTGTCTTCT TCTAAAGCAT CACTTAAATC ATCTGAATTG GTAATTTTAA 3720

TCAACTGACC TTGTCCAGAA AGTTCAGGGC GTTTTTCTCC TAAACGACTA AGTAAACCAG 3780

AACACATAAT AGCAAAAAAA TTACCAATAA TGGTTGCGGG AATTAATTGC CCCACTAATT 3840

GTTCACTTGG CAATCCAGTA ATGGCCACTA TAACCAAGGG ATAAAGGAAG AATGCCTTCA 3900

CCGAT 3905

(2) INFORMATION FOR SEQ ID NO: 100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32768 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

TAAAAAGCCC CTTAGAATCA TAGGTnTTTT TGATGGTTGT AATTGTTGAT GTTTTCTTnA 60

AGTTGTAAAG ATTTGATTTA ATAGTGATCG TGTCACCTAA TTGATAATCT TCACCTAATT 120

TAAACaGTTT AGAGCTTGTG GGAACCTCAC CACTTAGTGT TAAAATTCGT TTACGTTCTG 180

AAAGTTTGTT ATTCCCTCTA TTTTTCaAGA CTTCGAGGTA TTGTGAATCT GTCAAAGTGA 240

CCTCGTCAGT TGTCTTTTGT AAGTCTCGAG CATCTACGTA TAACTCTTTT CTTTGTAGAC 300

CTGTTTTCTC GTTACCCACC ACTACACTTT TACGTTCCGA GCCTTCACCT TCACCGAAAA 360

CAATAGCGGT TGAAGATTCA TCAAAGTTAT TGTTCTGATA CCCCGCTTTT GTTAGATTTT 420

CGTATTCGTC TGAAAACTCA ATCGTTCTCG AAACATCTCG ACCTTTAAAA ATAGATAACG 480

TATTACCAGG AACGCCTAAT TTAGTAGCTG TCTCTCTTAT CCCAAAATCG TAGGATGTGC 540

AAAGAGTTTC CACTTCTTCG GCTACGACAC CATAACTATT CTGATACTGG ATAGACGTAT 600

TCCCTAAATT AGCTCGGGAA TCCAATTTTA AATATTGAAT TTTTCTTTTT ACATCCGAAG 660

GgTTTACTAC TTCATTATTA AGATGTTCCC AAACAATTTG CTCTGGCGGT GCCGTTTTAT 720

TATAAATTCG GTAAACAATA CGATCTAATG ATTTCCCAAG AAGTGATTTC CCCGACACCT 780

TTATTTGTGC TGTTGCCTGA TCGTCTACAA CTACGGCATC AACATAGAAA TAGAGCCCTT 840

CCATATAAAT AACAGTGTCT GGAACAAATA TTTGTATATT TTCAGGTGTC AAACCAACAA 900

ACAGCTCAAA CGTTGAATAG GTATAGTAAT TTAATTTAAC AGTTAAACTT TTAAACCCAT 960

CAAATACTTT TTCGGAAATA AATTTCCATT GATTATCTCG ATCCTGAGTG AAAATTTCTA 1020

ACTCCATTAC ACACCACCTA CCAACGGCCG GAAGTCAAGT TGTACAATTA CGTTACCAAT 1080

ACCAGTAGTG GCTTTCACTT GAAAATAGTT ATCACCCTTT TCAAGCTGCA GAAAGGTAGA 1140

ATTAGGGTCT CGCAATGGCA TAGCGTTAGA TTCAACACCG TTTGGGTCCG TCATGATTGC 1200

TTCTTTCTGT CCACGAGTAG TTACCAGTTT AAACTTAGTA CCTGCTTCAA ATGAACCTTT 1260

AAATCGGAAA AATTCCTGAG TAATTACGTT ATACACTTCT GGATCGGTAG CTTCTGCACC 1320

TAAAGAAAAA TAGAACGTGC CTCCCACTGA TACATCACCA TCATTGGTTA CAGGAACAAT 1380

CTCACCACTT TTTAGTGTAG CAAACTCATA CCCTTTCGTT ATTTCTAACG GCCACATGTG 1440

TTTTTTTGTG GCAACGGCTA GGGGAATTAA CGTATTATAT TTACTCATAT CAGACCAATA 1500

CGGATCTAAA GCTAGAAAGG TACACGTAAA AGCCTGTGTA ATATTTTTAG CTGGATCTGG 1560

```
CAGGTCGGGA GCTTTTACTA CTAACACATC AATTTGAAAC TCACGATCAA ATACTCTGTA   1620
AATCAACGTA CCTGGTGTTT TGGGATTTAA CGTTTTTATC ATTTCATGCT GTAAATTGAA   1680
AAGATCCTTT GTTCCTCTAG CAATTATTTC ACCTTTTATC GTTAAATCGC GCTTATCTAA   1740
ACGCTCCGAA ACTTTAATAA CACCATCTAG TCCATATTGT TCTTCTGTTA TGATCTTATT   1800
TTCAACAGCT CCGAAGCCTG TTTTACTTTT AACGGTAAAA GGTGGTTGAA CACCGAAACT   1860
AATCAAAGAA CCTTCCGAGT TCTTGTATAT AAGTTCGTAC ATCCAACCTA CCTCCTATCT   1920
TCTCAGTGTT GCAAGGTCTT GCAACTTGTA TTTTGTTTCT CTTGCAATTT CTCTAGGCGT   1980
TAAAGGCTCG GGACTTGTAA TATATTGTGT AACATTGATA TCTCCGTCTT TTTGTTCAGC   2040
CATTACACCT CTCACTGCAG TTTCAACATA TCCTAACAGC GTGTCAATCG GTGCGACAGC   2100
CTCAGCCCCT GCTTCTCCAC CTATCATTGC GTTATTTCCG TTCATGCCAA ATATGGTAGG   2160
TTGCGTCATG ATACCGCCGT CTTTGTACCA TTCTATGCCC AAACTTGGTA TTTTACCTTT   2220
CAATAAATCG CCCGCAGACC AACCGGCCGG ACTAATAGAA AAGTGAGGCA GCGGAATTTT   2280
CGGCCAGCTA ATATTAAAGT TAAAGAAACC TTTAATCGCA TCTACTACAT TTTTCACTAG   2340
ATTTTTAGCC GCATTCATTG GCCCATCAAT AGCATTTTTT ACACCATTAA AGATATTAGA   2400
AACTGTTCCT GTAATACCAC CCCATACATT TGAAATAGTA TTTTTCACTG CGTTAACTAC   2460
ATTACTAATA GTATTTCTTA CACCATCCCA AATAGATGAA ACGGTATTTT TCATACCGTT   2520
AAATAAATTC TTAACGGTGT TCAcCGTTGC ACTAACGATA TTAGACACGG TCGATTTAAT   2580
ACTATTCCAC ACCGTGGATG CTGTATTTCT AATTGCGTTA AAAATGTTTG TAATTGAATT   2640
TTTTAAAGCA TTAAAAATAT TTATCACTGT GTTTTTAACA GAAGTAACGA TATTGGAAAT   2700
AGTCGTTTTA ATACTATTCC ATACACTAGA AGCCGTGTTT CTAATGGCAT TAAAAATGTT   2760
TGTTATCGTG TTTTTCATAG TATTAAATAT CGGTGTAACA AAGTTAAGGA TAGTCTGCAC   2820
AACATTAGAG ATTGTTGTTT TAATAGCATT CCAAACGTTA GTTGCTGTGT TTTTAATTGC   2880
GTTAAAGACA TTAGTAATCG TCGTTTTAAT AGCATTAAAT ATAGGAGTGA CAAAATTTTT   2940
AATAGCCGTA GAAACTGAAG TAACTGTATT TTTAATTCCG TTCCAAGCGA TTGTTACAGA   3000
GTAAACGACA GCATCCCATA TCTGTTGAAA GAAAGTTTTA AACGCATCAA AAATACCCTT   3060
AATAACTCCA ATGAAACCGT TAATTATTCC TGTGACGGTA TCACCTATTT TTTCCCATAC   3120
ACCGGTAGCA ATTGCTTTTA TGAGATCCCA TAGATATTGA AAATACGCTT TAAAACCATC   3180
AAAAATACCT TTCCAGATAT TAATTGTATT AACGATTATC GCCCCTACTA CAGCGTATAT   3240
TACATTCCAC ACTGTTTGTA GAAAATTCTT AATGCCGTTA AAAACGTTTT CTACTGTTTT   3300
TTTGAAACCA TCAAATGTTT TCTTGCCAGC ATCCGTGAAT TTCTTCCAAA TTTCTGAGAC   3360
TTTAGCGGTA ATATCATTCC ACGTTTTTAC TAAGCTACCT TTTAACCCAC TAGCTGTTTT   3420
AACGATTGAA TCCCACGTAT TTTTTATAAA ATCGGCTATA CCATTAAATA TCTCTACTGC   3480
TTTATCTTTA ACGGAGTTCC ATGTATCAAC TAGAAACTTA GTAAAAGATT TCCAAATCTT   3540
```

```
TTGTCCTGTT TCAGTTTGGG TGAAAAAATA AATTAACGCC GCTACTAACC CCACGATTGC 3600
CGAAATAATC AACACAAAAG GGTTGGCGTT TAAcACTGCG TTAAAAGCTG TTTGAATTGC 3660
AGTATAGGCT TTTGTGACAC CTGAAGCTAT TTTAGTAGCT GTTAATACAC CTTTCATGAT 3720
AATTTTATAG GTTGCTAAGG CGGCTCCTAC TCCAACTATA ATAGCTTTTA AAGAACTAAA 3780
AATAGCTTTA TTTTTAGTTA TAGCCTTTGT AAAATTAGAA AAAGCAGGGA TGACATTATT 3840
TACAATAACC TCACTCACAG CTCCTGATAT GTCCCAAATC ACCTCTACGA TTTTTTCAAT 3900
CGCAGGTAAT ACTACTGTAT TTATTGTTTT AAATGCAGAT TTAAAAATTT CTATCGCTTT 3960
AGGTACTACA GCTGTGAAAA TATCGGCAAA AGTAATTGAC CATGTACTAA ATACTCCTGC 4020
CAATTTCGTA AATATCATAG ACACTTGACT GCCCGTATCC TCAAAAATCC CTTTTATTTC 4080
AGGCAGTTCT TTTTGTACAA TCTGAGCTAC ATTTTTTATA GCATCACTAA TCCACTTAAA 4140
AGCCCACTGA AGATTCATCA CAAAACCTGA ATTTTGAATG CCGTTAAATG CATCCATAGC 4200
CGATTTTTGG AATCCCCCAA AGTCTGGTAT AACTGCAGCT ACCTTATCAC CTAAGCCACT 4260
TAAGCCTTTC GTAATGCTTT GAATAATCGA AACGACCGTT CCTAAAATAG GAGTGCCTAT 4320
TTTTGCTAAA AAGTCCTGCC ACGCTTGTTT TAAGTTACCC ATAACGTTTT CATATCCGTC 4380
GGCTTCGCGA GCAGCCTGTC CTACAGCTCC GGCTTGTTTA AGCATGTTGC TAGCATAATC 4440
TAAACGAGTA GCTTGTTTCG TCGCTTCATC TAAATTTTGC CAATCTTTTG TAGAACCTAC 4500
TACTCCCTGA GAAATAGCAA ACTGCGCCAT TTGAGTATCA TTTGCAAAAA TCCCTATCGC 4560
TTCACCGGCT TCGTAGTTTC CTTTTAGAAA CGAAGTAAGA GAACTTGAAG CATCTTCCAG 4620
TGAAACATCA TAGAAAGCTG CAGCATCTGC CGCTAAGGTA GTTGCTTGCT CTGATTTCTT 4680
CATTGCGGCT TCTGTATCTA GCCCTAAGCC TTTAAACATT GAAGTAATTC GAGAAAATGT 4740
CGGCTTTATA CGGTTAGGTA GGATATTCAT TTTTTTACCC ATGCCATCTA CCATTTTTTG 4800
TGCTAATGGT TCTAACTTAC CAAAAACCTG GGTAAATTGT GCATTCAAAG CTTGTGCACT 4860
TGCCGCGGCT TCTACGCTTA GCTTCCCAAA ATTAACGATT TTTTCTACGG CAAAAGCAGC 4920
ACCAATTGCT AGCGCTGCTT TTTTAAAGAA CCCCACTAAT TTACTTGTTG TTGATTGTCC 4980
CTTAGATTCC GTCTCGTCCA AAGCCTTATT GGCTTCACCA TTATCGACAC CGATTGTACC 5040
AAAAAGCTTG AAAATTTCGC CCATTAAGAC ACCTCACCTT CTTTTCGAGG TTTAACAAAT 5100
TGACTAGCAA AATCTAATGC TTTTTGTTGT TCTTCTTTTG TGGTAGATTT CGCTTTATTT 5160
TTACGTAAAG GTCTGTATTT TTGTTGGTCT TTAAATTCTT TTAAAGACTG CGTCATCTGT 5220
GTGTGTAACC ACTGATTGTA GATTTTTTCT TCTTGTTCCA CTTCTTGAAC ATAAAAAATA 5280
AATTCCATTA AGTCACACAA ACGATAGGTA CGAAGTAACG ATATCGGATC TCCGTACCTT 5340
TTGAACAACA AATCTTTAAT TCGGTGTTCG CCGTCTTTTA TTTCAATAAG CTGGCGATAG 5400
AAGAGAAAAA ATCAGCTAGT TCTGGTTTCT TAAAGAAATC GACTAATAAA GTCGTATACT 5460
CCTTTAATCC AAGAGCTGAA ATTTCTTGTA TAGAAACGCC TGTTAACTCT GCTAAAAGTC 5520
```

CGTTAATATC ACTCTTAATA TCTTTTAAGT TAAGCATGAC TTTTTGCAAC ATCGctGCGC 5580
CCATTTCAAT GCCACGCTTA GTAGCTTCTT TGTCAGCTTT TTCTTTGGCT GCATCAATTT 5640
TTTTCTGCTC TGCTTTTGTT GGCTCTTTTG ATTTGTGGTC CAACGGTACA ACTTTTTCAG 5700
CAGATTCAAT ATTTTTTTCA AATAAATTGA TAAACTCGTC TTTCACATCA AGCTTTCCTA 5760
TGATGGCAAG TAATGTAAAT AAGTCATCAC CTTTTAGTTC ACGCATTTCT AAAGTCATTT 5820
TGTTATTCCT CCATGTAAAA AAGAACGACT GATAAGCCGT TCTTTAATTA TTTCAGTTCG 5880
TCTAAATTTG TTATTTCGAA ATATTTGATT CCTCAATAAT CTACAAGCTC TCAGTTAGCT 5940
CTATAGGATT TTTCGATGTA CTTGGGTAAT AAATACGGAA TGGGAATTCG TCATTTATTA 6000
GCTGTTCATA AGAAGCGTTA GCTGTGATTT CCTGTTCAAT TACTACCTCT TTGCCATCCT 6060
CTGTTTTAAT TTCAAGTCCG CTTTTCACTA ATCCGTTATC TAACGCAACG ATGATAGGCT 6120
GTTTCGTTCC ATTATGAATA CCCACAACGG CCATGTTAGG GATATAGTCC CCTTCTTCTA 6180
AATAGCGTTT AGGTTTAATA ATTTTATAAC CAGCAGGTGC TTCATCTGTA GTAGCATCTG 6240
TCATCGTAGC ATTTAATGAT CGGCGTAAAT TTTCTGCTGT TAATTCAATC AAGTTGGCTT 6300
TCATCGTAGC TGTGGCTGAT TCAAGCACAT TCAAACCTAC TACGTCCATA ATATAAGCAC 6360
CGTCTACTTC CACTTTACGG TAAGAAAGTT CTGTTTTGAT TTCTACACCA CCAGAAGTTG 6420
CCCCCATGGG AATACCGGTA AACTCCTCTT TCTCTTTATC GTATTTAAAA TCAGTAAATA 6480
CTGTCGCAGA ATCAATGACA AAATTGTCTG CGGTTGTTTT TGTATAACCT GTTTTAGGTA 6540
GTGCCATATT CTTTTTTCCT CCAATCGACC GCTACATAAA AGCGGACATT TCTTCTTTTT 6600
AACAGCTCTT CTTTTGTTGG TATTTTATTT GATCCTTGAA AACTAAAAAT GAGATTTAGT 6660
TCTGGAGTTA GTACCCTTTT GTAACAAAGA GCTGTTTTAA ACTTGTCTTC TAGAACAAGC 6720
AAGTTTAAAA ATGAATTATT TTTGTCAAAA ATATCAATAT CCAGATAAAA ACCATCTTGA 6780
TTTCTACGAA TAGGCTCGGA ATCAAAATCA AAGGTGGCAT ATGGATAGAT GACCTCTTTT 6840
CGAGGATTCG TTTCTAAAAA TGTTTCTGGG TGTATCTCTT GTAAGATTCT TACCAGCTCT 6900
CGTATAAACT CTTCCATCAC TTATCCCCCT TTAAATGTAG CTCCGTATTC TTTCCCTAGA 6960
ATTTCGATGA CTTTTTCTTT ATTCTGCCGA AACGCGTTTC TTAAAAACTT TTGTGGTTTC 7020
TGACCTCTTG TAAAATGCCA TTCTCCATCT GGTCCCTCAA AGAACCATCC ACCTTTACGT 7080
CCTGAGCCAT TTTCAGCAAA TTCTCCGGTT CCGAATTCAA CATAAATGGC GTATTGTTCT 7140
GGCGAGCCGA CAATGCCGAC CACTTTTCCT CCAACAGCCT TAAGTTGATA GTCTATCCGG 7200
TCCCGCAACT GTCCTGTTTC AACTGGCGCA CCAGACTTTG CAGCAGCTTT GATAATCATG 7260
AGGACTTTAG TCATCCCCCG TTCTGCCGTC GCGTCTAACT CTTTTTTTAC CTTGTCTTTG 7320
TAGCTGATAA ATTTAAAGTC ATTCTTCGCC ACTTAAAACA CCTTCAAACT TCAAATAAAT 7380
TTCATTGTGA TGCTTTATGC CAACAGGATT ATCGGAATAC GTAATATCGT AATAACGATT 7440
TGTTTGATCA ACCACTCGCA TATTGTCTGT TATCCCCTCA ATAAATTCGG GAATAATAAG 7500

AATGTGTGTA GATTCTTCTA CAAATGCGTT CTGTTTCGTT GTTTCATCTG TTCCTGTCAC 7560

CAAGTCGATA TATCCAGAAA CTTCTTTAAA CAAGCCCCAA TCGTCAATAA GCCCGCCTAA 7620

TCCGTCAGGC TTTTGGCCTT TGACCTCTTG TAGAAAAAAG GTTTGAGGTG AGTACATTAT 7680

CCCCACCTCA TTTTTTTATA TTTATTTAAA AAACTGAACT TTGCTGCAGG AAAGCCCTCG 7740

ATATTAnTCA CTAGCGTTCA CATCGTAATA CGTGATAGAC ATTCTAGCGA TAGATTCTGA 7800

CTTAATGCCT AACTTTGCGC CCATCGATTT TTTAAATCGT AAAAGTTCTT CAATACCCAT 7860

TTTAATATCT GCTGGATATT CTACTTTTGT AATAAAAGCA CCAGAAAAAG AGCCATCAAA 7920

AAAAGGCTCT TCAACATCTG TTTCCAGTGT CTTTTCTGTG ATAGCTTCAA CAGTTGTTAG 7980

GCAATCGTTT ACTTTAGAAT TACTAATTTG AATCGTATCA CCTACTCGTA AACCAAGAGG 8040

GTTATCTTTT AACACAATAG AATTACCCTT AAAAACAACA TCCCTAAAGC GTATATGAAC 8100

GTTTTGAAAA TTATTATTGG TTAGTTGCCT AACTGTCTGT TCGAACGCGT CTAAGTCACT 8160

CTGAGTAACC GATGGATCAA TTTTTTTTGC TTCTTCAATG CTGATAATCA AGTTAGCCAC 8220

CTCCTAAGAT GCAGGTGTAC CTGTTACTGC AATAGAGGTT GTAAACGCAC CGGAAGTAAA 8280

TTCAAATGTC GCTTTACCTT CTGCAATAAT TGAGCCATCA AATCCACCGT TTTTATTTTT 8340

AGTGATAGTA GCAACACTCT CATTACTAGA TATCACTGTA ATTGCTGAAA TTACAGCTTG 8400

TGCGTTTGTG GCATCTACTG GATTAGCTGC AATAGTAAAT GTCTTAGTTG CGCCTACTGC 8460

ACCCGTCCAT GTTTTCTGAC TAGGCGTGAT ACCCGTTGCT GGATTAGTAG GCGTTACGCT 8520

TTTTTTACTG CTTTTAAAAT AGCTACTTTA TTTTTTTCAG GGATGTGTTT ACCATATTTT 8580

CCAGCTCCCT GGATTGCTAC TCCGTTAAAA TCTTCGGAGT CAATAGCACG TGTTACAGAA 8640

ATACCAACGC CCGCAATACC TACGCTATCC GCGGAAAATT GAGCCATTTC CCCAGTTTGT 8700

AATTTGCTTT CTGGAATTTC AACCAATACA AAACCTTTAA ATTTATATAT TGTTTGTTCA 8760

TCAATGTTAG CACTTGAATT TTTAGTTGTT GTTGCTAAAC CATTGTCTAC TAAAAAATCA 8820

TAAACATCAG GGTGAACATA AGCAACCCAA GCAAGAGACT TAGAAACTAA ATTATTCACA 8880

AAAGTCTTAT GCGCCGTTGA GAACAGCTTA GTTACTCCTT CACTGGTTAA TTCAAACCGA 8940

ATCGTTTCAG AGGCTGCATC AGATAATGCT TTACCTAATA AACCGTCGAT ATACTCAGCC 9000

CAAGCAATCG CTTGTTCCTC TAATCTTTCA GCTACTACTG CGTCAGCGTC aTCGTTTACT 9060

GTGATATTAT CCACTCCTTC GTGAATCGCT AACGGAGATT CATAAGGAAC CGTTGTATCA 9120

ATAGATTTAA TTTCTTTACG TGGCCCAAAA CGATTAGAGT TGCCAGTTCC TGTACCAAAC 9180

GCTACGTTTT CACCAGTGTT GTAGGGTTGC ATTACCACAG GTGTATCTGA AGTTTTTAGC 9240

ATCAAAAATG TGTCGCTATC TTTTACAGCA TCCGACGTCT GCAACTCTCC TCCAAACGCA 9300

CGTAAAAAAT AAGATTTTTT CTCTGTAATT TTGGCTAATA AGCCTGCATA TTGTTTCGTG 9360

TAAAATTTTG TAGCCATAAA AAATTCCTCC AATTACTCGT ATTTTGATAA GATTGCTTCA 9420

AATGGATCTT TTTCGTTTGG TTCCTTCCCT TTTGGCAGTT TGTTATCAAT TGGTTGATAG 9480

CCtGCTTTAT TTTCAGTCTT TTTTTTCGAC TTATCATCGA GTGTATCCTT CTTTTCAAAG 9540

TAATCAGGGA TGGAAGCTTT CAAGTCTTTC ACTTTACTTT CTAAGTCCTT CACATTCCCA 9600

TCTTTGTCAA GTTCTAACTC ACCTAATTTA AATAAAGCAT AATCAATATC CTTAGCCCCT 9660

GCACTTTGTA ATGCAGCAGA TACTTGGCTA TTGATTTTCA AATCTTTGTT TTCCTTTTCC 9720

AACGCGTCTG CCTTTTCTTT TAAATCAGCT AGTTCTTTTT GAACATCCGG ATTGTCCTTT 9780

GTTTTAGCTT CTAGCGATTT TAAGGTTTTA TTTGCTTCAG CTAACTGGCT TGAAAGATTA 9840

TTATATTGGT CTTTTGGTAC TGCATTTACA GGAAATTCTT TGTCAATTTC TTTATTAGCA 9900

GCCTCCATAT CCAACGTACC ATCTTCTTTC ATATGTTTTG CTAAAATTTG TTTGATCCAT 9960

TCCATCGTTT TCTACCTCCA TAGCATTTAT ATAGCGGTCG CTGCCGCTTA GAGTGTCAGA 10020

ATATACCGTC TGCTCGGTAG CGGTGCCCTT TTAACGTCAT GGCTCCTGGA CAAAATAAAA 10080

AGACTTATCA ATAAGCAAGT CTTAACTCCT CATATCATCT AATACTTGTT GTTTTGCTAT 10140

ATCTAGCATG CCTAGCAGCT GCAAACTTCC ATTTTGACTG TAATAAATAG TGACCTCTCC 10200

ATTTGGCCGC TGTACAACAA TTGTGGCCTG TTCAAAATCA GGCTCTTGGC TTGATACCCA 10260

TTCGTAAAAC TCTTCGTAGT CTACATCTAA ACAAGCGTAA GGATCTTCCA TCTTATGGCC 10320

TCCCTTAGAT ATAAAAATAG CACTCAGCCG CTTAATGGGT GAGTGCTACT ATTCGTTATA 10380

ATTATCTGCA TCCGCTTGCC ATCTATCTAG CAAAGTGGGT TGATTAAGTT TATGTTGTTG 10440

ATAATATTCC TGTCTCAACC GCTCACGTTC CTCAGGCGTT TTCGCTGTGT CTAGCAGCTT 10500

ATATAACTCA TCAAATTCCT TATCAGTCAT CTCTAAAGCG cTACGCCTAT AATCCACATC 10560

TACCTTACCC AAGAATTGTC ACCTCCAATA TATTATCGAC TATCTCCACT GTATACTTCA 10620

AGTTTCTGTT CAGTAGGAAT TCACTTTCCT TTTTCATCGA ACTGTTCGTT CCTATATAAG 10680

CGCCTTGCGT TCCTTTAGGA ATCAGGAATT TAACTACTCT ACCATCCTTA TTTGCACTGG 10740

TAAAATTTAA AAAATCCTCG GCAACTTTTT TGTTGATACT AGTAGACTTG AATTCTTTGA 10800

ATACATTTCC GTTCAGAATA GCATCGTATT CATTTGCAGA TACACCTCTA TAAGTTATTA 10860

TATCATGTTT CAATTTAAAT TTACTAATAG CATCATCTAA ATTTTGAATA ATATTGACAA 10920

GATAAGGTTT TTCTCTTGGA TTATATTTTT CATTTCGGAG TATCTTATTC ATTTTAGCAA 10980

ATGCGTCTGA TGTATAATCT TTCATAGCAT TAACTTGTTC AGGCTTCAAT CCAGCAACCC 11040

ATTTATCACT ATCATCTTGC AGATCAGATT GACCGAAGAA GTCCAGCTCA GTAAAATTCG 11100

GCTTTGGTTT CGATTCATTT TGTCTGATTC TATCTGCATA CCACTCATCA TAGCTTTTAA 11160

ACTCCGACAT CTCTTTAGAT TCATTATCTT TTCTAAGTTC AGGGCTGATA CCGTCGACAA 11220

CTTCAATTGT TGTACATCGG CAATTCACAT CTTCTGATGC TACCCCAAAC ATCCGCGGAC 11280

CTTTTGCCGA ATGCCCTCTA ATGGTAAACT CTTCGTCTAC ATCAACCTCT TTACCGTCTA 11340

ACTCTTGGTG GGTTCGGCGT GTATGTTTAT CGATCGTGGC AAGCCATTTC TTTTTAATAT 11400

TGATGCCCAG CTCTTTTGCT TCTTCATATC CTTTTTGGGT AGTGACAGAC TGAGTACGTC 11460

CTGCTTCTGT TCTTGCAATA CGTAATGCTT GTTTGTAGCT AGCTTCTGTT TCCTCATTAA 11520

TCCATCTAGC TATTTCAGCA TAACTTTTAC CCTCGAATAA GCCCGTTATG ATATTGTTAG 11580

TCACATTTTG GGCTAATTCA TCACGGTACT TGTATAAACG CTTTGAGAGC CTCTTACCCG 11640

CTACAGGTGC ATTGACGAGA TTCATAATAT AATCATGGTT AATTAACGGC ATACTTAGTG 11700

CTATATTCTG TGACTGCTCT AGCGTATACC ACAGTCCATA ATATCCCTGT TCAGCTTGTT 11760

TTGCAGAATA GCCTTTGATA GTTTTTTCAA CTTTCGGTGA GTTTAACTGA AGAATGGCAT 11820

TAATTTCATC AGCAACACTA AAAAGTCTCT CTACTTCCAA ACGGGTGGAA AAAGAAAGAC 11880

TTTCTGCATT TTCTGTATAA ACTTTTAGCC GCTTTTTGAT GTCGATTAAT GATTGACGAT 11940

AAACATTAAA TAGCTGATTA TCCGTTTCCT GGTAATTGGC TTTCTGTAGA GCCTGTAGCT 12000

CCTTCTGCCA TTTGTTGAGT TGGGACATTT GGGGTCACCG CCGTTTCTTC TTCATCGGTG 12060

TTGTCCGTCA AACCTTGCGT ATATTCCGCC TCTTCTAATG CCTCTTTCAC GTCTTCCCAG 12120

TCAAGGTCAA ACTGTTCGCA GATAAGGCGT AAAACGTTGT CGTCGTCAAG GCGAGGTGCT 12180

ACTTGCAAGA TAGATTCAAG GATAATTTTT CTTGTCTCTG CTTCGGTTTT CTCATTGTTA 12240

ACAATATCTG TTTCATTCAC CATTACTTCT CTAGTAAACG TAAATGAAAC TTCTGTTGGG 12300

TCAAATGCTT TAGTATAACG ACGATTAATG TCATCAATAA CAAGCTTATT CATCCATTCT 12360

AGTAAAGCTC GTAGTCTGGC TTCAGTCTTG TTGGCTTTCA TGTTTAATAG TGTATAACGT 12420

GCTTTTATCA CTATGTTGGT GATATTCCCG TCGCCTACTT GAGTTGAATC AAAGGCCATT 12480

CCAAACTTAT AGATATTTTC TTTGTCAATC TCCATTTTGG TTTTACGCCC TTCAGTTGGA 12540

ATCGTTACTG TTTTAACATC TAAGCCCCCA TCTGAACCAG TACCAACAAC CTTTTTAGAT 12600

TTAACATTTT GTCTTAATTT TGAAAGATCG TCACCTTGGA AACCTGATAC AACGTAAATG 12660

GCTTCCGCAA AATCTTGTAA ATTGTTTGAT AAAAAACAAT TCATTAAGTC GTAATCATCT 12720

ATCAGTGCCT TAATAGACTT TAAATCCGTC GTTTCCTGTT TATTATTTGA TAACCGATAA 12780

AAAGGTATTT GTCCGTAGCT TCGTTGCAAT AAGCTTTCAT TATCGCTATC AACTGCTAAA 12840

ACATGTGGCC TTGGATTAAT TGGTTCAGCT TCATCTAATT CATAATCTTT GTTATCCTCA 12900

GCTACGAAAA AATAAACGTT CTGGTCAGTC CACACTTCCG CGTGATGGAT ATCGACTGTC 12960

TCGCCGTCTT TCTCGATTTC AGTGATATAA TGACGGCAAA TACGTTGTAA TTCATTGTAT 13020

TCATTGTATA CACCAAACAC GTTTAAACTA TCAGCCACTT GAAAACACAA TCGATCTTCT 13080

GCATTGGTTC TTGCATAAAC ATACTCAAAA CCTTTTTGGC TTGAACCTTC CACAAGCTCT 13140

TGTAGCACTA CTTGAAATTC AGAATTGTAA TACTCTGCTA AATACTCTTT AAGTTCTTCG 13200

TTTTCTGTTT CGTATTCAAC AGGATTAGAT AAAAGGTATT GAGTTTTTTG GTCAACAATT 13260

TCTGGGAAAA AGCCGTGCGG GATTCGTACA TTCGAGGCGT ATTTGTCTTC TCGCAGTACC 13320

CCCTCATCGT CTACGTAAAA GATACGATTG TTCATGATAT CATTTTCATG GTTATAGTAC 13380

CGAATACCTG TCTCTGCCTC TCTTTTTGAG GTAGATTTTC GATCTTTGTC AATCGCCGCC 13440

TTTAAAGCAC TAGCGATAAT CTTCACGTCT TCACTAAGTA AAGCTTCCAT TTACAGCCTC 13500

CTTTCTAGTA TAGCCATTTA TTTGTGTTCT TAATTTCTCT AAGCAGACTA GCTGCACTGT 13560

CTGGTGCGTC GTCGTGCTCT GCGTTCTCTG TATAGTCTAG TATCTCGGCT ATATATTCTT 13620

TGTCTGTGTC TTCAAGCCAA ATTATTCGGC TCCAATACTT CCGCAAATAA GAAGATATCT 13680

TGATAAACTT ATTTGTTTTT TCGTGATACT TTTGCACGTA TTGACTTCGT TCGATTAGGT 13740

GTTTAGCTAA ATAACCTTTA TCGCCGTTCG TTTCAGTGTA AAATGTTCCA GCTTGGTAGT 13800

ATTGATGTAG TTGCAATATT TCAGGTAGGC AATCATCAAC GTGTTTCTGC CATTTCTTGC 13860

CAAACCCGAT AATGGTTCCG TCCTTTTGCT CTTTAAAAAT GGTAAATGCT GTGCTATCAC 13920

CACCTCCGTA TGCCGCGTCG ATGTGTGCTA CACCGTTATA AATAAGGTTA GTGTTATCGA 13980

TATAGGTCGG TGCAGTAAAT AGCGATTCAC TATCAGCGAT GTGTTTTAAC TCGTAGTTCG 14040

CTGCAAAGAG CGACGGTGTC ATAGACTGCT GTAACGCTTT ACGTTGTTCT TTGTCAATTA 14100

ATCCTGTTTC GTAACAATCG AATTTCTTGA CGTTAGGCAT TTTAGAAATT GCATCTTCCT 14160

TATGCCAAGG CGTGCCTGTG TTAATAAAAC GGCCCTCACG ATTCTTCACA TTCTGCAATT 14220

CTTGATACTG CAGCTTTGTT TTCTCTCGTT CAGCACGACT TACTCGGTCT TTAATGTTAA 14280

CAATATCATC AGTGATAACG ATATCCGCGT GTTTACCTGT TAATGAAGCA TAAATCCCCA 14340

TACCAAGTAA TTGAGATGTT CCCCGGGTAG ATGTTTTTAA GTTGGTATCT ATTTCTGTTG 14400

TAGTCTCTTT CAAAAGCACC AATTCAACAC CATATAATGC AAATACAAGC GTTTTAAAGT 14460

ATTTGCTAGA TAAAACCTTA GCCACTTGTA AAATAATCTC TACGACGTCT GTATCGGTCT 14520

TACGTAAGAA GATAATATTT TTGTTGGGAA AAAGAACCAT CAACAATGCA ATAGCAATCG 14580

CCAATGTGGT TGTTTTAAAT GAGCCACGAT GAGCCAATAG CGTTTGATCG TCTTTCTCGA 14640

ACAAAAAAGA CTTTAGCCAA TCATTGTGCA GTTCTGCTAA GTCGGTAAAC CCAACTAAAT 14700

TTCCGAAAAT TACCGGGTTC GTTTTGATTA AGTTCAAGTA TTTTCGCTTT TTAGGGCTCA 14760

TGTAGAATCA TCCTCGAAAA ATTGCTCGAT TTCGACAGCT GCATCGGAGA TATTAAGCAC 14820

GCCTGAAACT TCAGTTTCTT TTCGATCTCG CCATTCGTCC GGCTTTCTAT TTTTCAACCA 14880

GAAAATAGCT GCAGTTGGAT TAGGAGCTAC TTGTTTTTTA ACCTTTTTTG TAATTTCCAT 14940

ACCGTTCTCT GTTAGTTCTT CTGTAACTTC GGTATATTCG TAGCCTGTAG CACTCTTAAA 15000

TAAAGCATTT TCAACTTGTC TATCAACGAC TTCTTTTCCC TTTTTTAAGG ATGCCGAAAG 15060

TGCCGGAAAT TTTTTTACCC ATTCTCTGAA AGTTGAATAA GCTACCCCGA TATTCTGGGC 15120

TATCTGCTTA TCGATGAGGC CATCTCGTGC CCATCCTTCG ATTTTGATTA ACCCTTCCTC 15180

GGTTAGCCAC TCTGTGTAcT TCGCCATGAC CTCACCGCCT TTCTGTATCA AAAATAGACA 15240

CCTTCCCATT TTCTTGGGAA GGTGCCTTTT CTGCGTAATT TTACTGATAA CAGTATATCA 15300

CGATTTATCA ATGGGAATTG TCACATCATT GTCACTTTTT TCGATTTTCT TCAAAATCTG 15360

GTTATGTTTA TTTCTGATAT AGCCATAGCT ATGCCCAAAT TCTCCCGCAA TTTGTTTTAA 15420

CGATAGCCCT TCGATATACT TTGCTACTAA GATTTTTTGA TCTAATCCTT TGAAACTATA 15480

AACTAGCTTT CGTAAATCGT AAATTGaGTT CATTTTCCAC GCTAAACGTT TTTCTAAGTC 15540

TTCGATAATA TCTTCCAACT TAGAAGCTTT GGAATCTTTT GTTAGTGTGT ATCTTCCTAA 15600

ATCCTCTGGG TCACACCATC TTTCTAACTC TTTTTTATAC GTTTCTAGCT CCCAGTCTAA 15660

GTAATAGATT TCTTGTTCTA ATTTTTGATA GCTGACTAAC CATTCGTACA ATGTACTTTC 15720

ACCTACCTTC TTGTTCATCG CTTTCACTAT TAACCGTAAC TTGTACATTT TTATACAGAA 15780

TTAGAGTTAA TAATGCGGAT AGCAATATAA ATCTAACATC ATACGGTATC TTAGCTAATA 15840

TCGAAATAAT CCACTCCGAC ACTTTGTATA TCATCATAGC TAAACCTAAA CATAACCAGA 15900

ArGATATTAA TTTATCCAAG TTTATCCCCC CTATACTCCT CAATTTTGGC TTTCACTGCA 15960

GCCATTAGCG CTGATTGCCC TTGTTCTTTT GCTTGTAAAG CTTTGATTAC TTGTTCGTCA 16020

ATAGTTCCTT TCGTCACTAA GTGATGAATA ATAACAGGCT GTGTTTGTCC TTGGCGGTCT 16080

AATCTGGCAT TGGCTTGTTG ATAAAACTCC AAAGACCATG TAAGCCCAAA CCACACGATA 16140

ATGTGCCCGC CTTTTTGTAA ATTTAGTCCG TGCCCTGCTG ATTGTGGATG GGCTAAAAGT 16200

AAAGAAATTT TTCCCTCGTT CCATTTCTCA ATATCCCCGT CAGACACATT TAAAGCCTTT 16260

GCTTGTTTAA ATCGTGCTTG GATTCTTTCT AGGTCATGTT GGTATTGATA AAAGACTAAA 16320

ACCGATTGCC CTTGTGCGTC CTCTATAACG CGTTCTAGCG CATTTAACTT TTCTTGGTGT 16380

ATTTCCCTTC CGTCGCCGTT TTCGTCGTAT ACAGCGCCGT TAGACAGCTG TAAAAGCTTG 16440

TTTGATAGTG TAGCCGCATT ACTAGCTACA ACGTCTGTTC CTTCGAGTTC TAACACGTAT 16500

TCCCGTTCTA GCTCTTTGTA CTGTTTCCAG CTTGTCGGGT TCAAGTCTAG CTCGATAATA 16560

TTTTCCGTTC TCGGTGGCAG TCGCAAATAA TCTTTTGCTT TCATACTCAC GCATATATCG 16620

CTTATTTTGT TGTAAATCGC TTCTTCTGCT CCTGGGATTA ATTGCCAAGA ATACACGATA 16680

TGCCCGTTTT TTTGTGCGGG GACAAAATAT TTATTACGGT ATTGTGTGAT AGTCTTTCCT 16740

AGCCGTTCGC CTTGGTCCAA TAAATACATC TGCGCCCACA AATCCATCAA GCTGTTTGGG 16800

GAAGGTGTGC CTGTTAGCCC GATAACACGT TCCATCTTCG GCCGTACTTT TCGTAATGCT 16860

TTAAATCGTT TGGCACTACT TGACTTAAAA CTGGATAACT CGTCTATAAT CACGGTTTTA 16920

AAGGGCCAGT TTCGTTGATA GTAATTGACT AACCACTCCA CATTTTCACG ATTGATTAAA 16980

TAGACATCTG CCTTTTTGAA TAAAGCTTCT TCACGTTGCT TTGGGTTTCC TAACACTTTT 17040

GAAAAAGTGA GGTGCTGCAG ATGGTCCCAT TTTTCAATTT CATCGGTCCA AGTTTTCTCG 17100

GCAACTGATA ATGGGGCAAT TACTAACACA TTTTCAATAA TTTCAAACGT GTGTAACAAC 17160

TCGTCAATAG CTGTCAAACT GGATAAGGTT TTCCCTAGTC CCATATCTAA CATCAAAGCA 17220

CAGTAAGGAT GATCAAGAAT AAAATTCTTA GAGTATTCCT GATAGGGATG TAACGTTGCT 17280

TTCAAACTCA ATCACTCCTT TTTATGGGAC ATATCAGGCA TAACTCTTGC TATACCGCTT 17340

TGCATTGCGT AGATTAACAT ATCCACTCCT TGCTTGCTGT CAATGACAAA CACTGGGATA 17400

AGCTGCTCTT TGaATTTCTC TATAATTTTT AACTGGTCTT TTCTTGGTTT ACCTCCAGGC 17460

CTCTTTAATT CAACGAAAAA CACGTTGCCT TGATATAAAA TAATCCTGTC aGGAACACCC 17520

CTAGTTCCTG GGGAAGTGAA CTTATAACAT AAAgCTCCtA TGCGTTTTAT CTGCCTAACT 17580

AGATACTTTT CAATGTCGTT CTCAATTTGC ATTACTTCGG CCTCCCATCT GATAAGATAT 17640

TGGTTGATGT GTTCCATCAG ATCTATTTTC TTCAACCTCC GTTTAAAAAT ACCGTGTCAC 17700

ATGTAACACA AAAATTCTTC AACTTATATA TATTCCCCAT GTACATATAT ACATATACGC 17760

CCTATATATG TATATTAATT ATTTCTATAT ATGAATAGAA AATAGTGTTA CAGGTGTTAC 17820

GATTGTnCTT CATTCTTACT CTCCCAAGAT GTAATAGCGT AACACTCCCT GTAACACCAA 17880

CGGGAAATAA GGTGTTACAG GTGTTACAGT GAATATTTTA GGTGTAACAC TCGTAGCTTT 17940

TTTAGGTGTT ACACCTAACC CCGCGTCATT ACTACCTTTT TGTTTTAAGG TGTAACACGT 18000

AAATAGGCTA CTTGTACTCC ATACCCAGGG CCAAATCTTA AACGTCCTTT ATTACCTTTG 18060

CTGTTTTTCT CCCAACCAGA TAAATGGCTT AAAACTTGAC GAATTTCTGC AGCTTTTGCA 18120

GGATGTATGT TTTTAGAATC TCCGTTATAC AATTCATTCC AAACCTCTGC AATACATACT 18180

TTGTTTCTAA CAATCTCGCC TTCTTCCTGA ATATCCGTCC CCCACCCCTG AATATACTCT 18240

CGTCTTTCTT GCTTACTTCG TTTGTACCAA TCTTCCGTTA TTGGGATTTC CAAATATTCA 18300

AGGATTTCGC CTTCCATGCT AGAGGTTTCC GTGTGCATTT CTTGGGCTTC TAGTGCTAGT 18360

TTTTCTTGTT CGTCTGTTAG ATACAAAGGT TCTCCTGCTT GCCACAACTC AACGGCTTCT 18420

GcCCATATCT GATTGCGCGT TTCGTCGGTC ATTTCCCACA CTTTATTTTT AATCGGTTGA 18480

ATGCCGACAT CTACTGGCCA AAAACGCCGA TTTCCTGTTT TATCTCTTAG AAACTCGTTG 18540

TCGTTCGTTG TTCCCCAAAA GACACAACGG CGTTTAAAAT AGGACTTATG CCGACCGTAA 18600

GCAACGCGGA AAATGTCTTC TTGCTTGCTA ATAAAATGCT TGGTGGCTTC AATGTCTGCT 18660

TTTTTTGTCG CTGACAATTC GCCCATTTCC ATGATCCAAA CACCTTGCAA CGCTTCGTAA 18720

GAGTCTTTCC CCGTCACACC TTCTAAGCTG TTAGAAAACC AGTCTCCCGC CAATTTTGCA 18780

GGCAATAGCG TTTTTCCTAT TCCTTGCGGA CCGCTAGTCA CTAGCATGTA GTCAAACTTA 18840

ATCCCTGGCA CAAAAATACG CCCTACGGCA GCTGTAAGAA ATTTTTTCGT AACTACTCGA 18900

ATGAACGGTG TATCTTCTGC GCCTAAGTAG TCTATTAGTA AAGTTTCTAC GCGAGGCACT 18960

CCGTCCCAGT GCAAGCTCTC TAGGTATTCT TTCACTGGAT CATATGAGTT TCTTTCGATT 19020

TCTTGGACTA GCGCATCATC GATTTTGCCA CGGTTCACAA TACCGTAAAT TTTTTCGATA 19080

TATACCCGTA ACCCCGCATC GTCGCTGTCC TTCCACATCT TGTCTCGATC TAACTTCCGC 19140

CAAGGCAAGT TGTCTTTCAC CTCAATACGA TTAGAAAAAC TGTTCATAAA AATTTTCTTT 19200

TTGAGGTTCG GGTCATTCAG CATAATGACT TCTAAATTTT TCGCGGAAGC CATAATCTGA 19260

CCGTATTCGT CAATTTCGAG GTCCAATTTT GTAAACCAGT CTTTGTCGTC CTCTTCCAAT 19320

TCGTCTAGCT CGCCGTCAAA GTCTTCTAGT GCTTGTGACA GCTTTTCACT CTGAATCAAG 19380

GTCTTAACTT GCTTGTCTTC CATTGCGAAC TCTCGCATCG CTTTAAACGA TGGATAACGG 19440

TCAATCCGCG TCGTAGGTTT GACGTCTTCA TCTAAATCGC CAAATAGATG GATACGGACA 19500

AGGTCAAACG CGTTCACCAG TTGATCGCCT ACAGGGTCCG TACCGTGATG AGAATAAGCG 19560

AACTTATCAT CATAAATCAC TAGCCCACCA CTAGTAGAAC CTTCTAAAAA CGTCCAACGG 19620

TCCTCTCTTC CTGTGGACCC GTAGATGTCA GGCAAAAAGG TTTCAATCGC TGAAATGATG 19680

TCATACGTGC GACAAAATGC CCCGACAATC CCTTTTTTCT CTAATGGGTC GCCTGCTTTT 19740

TTTGCTTGTC TTTCGCGGAT AGAATGCCCA CGGGAACTTT CTGGCCAAAA ACTTGCATCT 19800

CGCCAGTCTT CGTATTGGCT TAGAATTTCG TCAGGATCGA CCCACGGCAA ATCGATGTTA 19860

TCGGTGAAAT ATTCCCCGTC TATCGAATGA CTGGGCCAAT ACATTAAACG TTCTGCCTGA 19920

TAGGTCGTAT CATCGAAGTT ATCCATGCCG AAAAACTCCG CAAGCTTTCT CGCTAGGGGC 19980

TCATACTCTT CGGCGGTCAC TGGGCGACTT AACGGAATAA TCAAGCGATA ACGTGGACCT 20040

TTTACTAAAT GGCTGTGTGT CGTGTAGACA GCTGCAGCGT GGTCAAATAA CAGCTGTACA 20100

TCGTCCCAAA AGTCTAACGT CGTGCTGTCG GCATCTAACG TAACAAGGcT ACGCTGTTGC 20160

GTGTTTCCTC TTTTCCGTCG TCCTTCTTTT AACCAACCAC CAATAAACGC CCCAACGTCT 20220

TTAACTTCGC CCTGCTTAGA CTTTGGCATT TTTTTGTAGT CTTCGACAGT CTCTTGTGTC 20280

ACTGTCGGCG TTTTCAATCG TTGGGTAAAA TCCGACCAAG ACATCTGCCG ATTTTTCCAT 20340

TTTTTCTCGG TTTTTGAAGC ACCCACTGCA AGATGGATTT CACCATCATA GGCTAATTTG 20400

ATGTTTTTTT CTGGCTGTTC CATCTTTCCG CTTCCTTTCT TTTATTGTTT AAGGATGAAA 20460

CCTAAACAGT CAAGCCATCC TTCTTGATAG TCACACGTAG AGGATTTATG ATCATAATTA 20520

AACGCAGCTC TCATTGCTTT AAGATATTTA CGCTTACCAA CAACGTAGCC CAACAAAAAG 20580

CAAAGTACTA AGTTTATTAG TATAAGTATG TTATTTCCCA TTTATTAACA ATCCTTTCCA 20640

TATTCGTACT CGTAAATTAG ATTAGTTAAC CAGCTATAAT CTAGCTCTTG CTCTTCGTCA 20700

AACTCCTGTT CGTTGACATA AGCAGCAAAA TCTACGATAT TTTCGATATT TTCCTCATCA 20760

TAAAAACCCC AACTCCAAAT GAATGCTTTA ACACGAACGA GTTCCCCAAT ATCTTTTTTA 20820

AAAACAAAAA CGGGCTCATT GACCATTTCA ACACTCATAC AAAATTCACA TGTCCCCACT 20880

TCTTCTTCGT GAGGATTAGT TTCAACATCA ACTAGTTTTA GTTTCATTTT CTATCGCTCC 20940

TTTTTCTAAT TTTTAGTGAT TTTCAATTCC CAATCATCGC TATCAAAAAA TACAAATTCT 21000

AATTCCTCAA AATCGTTTGT ATTTTTTATC AAATCTTTAA ATTTCTTATT TGAACAAATA 21060

TTTCCAAAGC GTTTTAAAAA GCGGTGCTTA AAATTTTCAT AATCACTGTA ACTAAATTCA 21120

AAAACTTTCC CAGTTGAAAA AACTCCACCA GTTGTCAGTT CTAACTTTAC ATAATTTAGT 21180

GCCATTATCT GAACCTCCTA ATCTTTCATG TAATATTTCG TTTCAAATCC TTCGGCGTTT 21240

AACGGCAAGC CTTCCGCCCA ATCAGGAACT ACAGACATGA TTTTGTTCAT TTCTTCAATA 21300

GATTTTTCCC CTTCTGGCAC TTCGGCTACT GCCTCATCGT GCACATGGAA AACAATAGGA 21360

TAGCCCTCAC GTTCAAGCCT CTGCATCGCT TCTGCTAAAA CGTCTCTTGC CGTTGCTTGG 21420

ACAATGTTTT CCACCAATTT ACCGCCGTAG GTTTGCTGTT CCGTGAAAGC GACTTTATCG 21480

CCTTGCCCTT CATAAAAAAT AGCAGGACCG TAGTCACCTT CTTTTAAATG CGCTTTTGCA 21540

TACGCTAGTT TTCGTCCGCT TGGTAGCTGA ATAAACAAGA ATCCTGCTTT TTTATAAAAT 21600

TTCAACCCAC GAGGGCCTTT TTTAATGCCG CCGTTTTGTA AACAGTCAAT CACAGCTTTT 21660

TGGGTTTCAT GCCAAAAGTT CTTAATGCGT TTATTGGCTG TACGCCAACG GTCCACAATG 21720

TCTTGTAATT CGTGTTCTTC AATGCCGTTT TCTAATGCAC CCATTGCTTT AAGCGCCCCT 21780

GGTCCGCCTT GGTAACCAAG CGCCAATGTC GCCACTTTTC CACGTTGGCG CATGTCTTTA 21840

CcTTCGTGGC TTTTCCAGTC GTAGTCCGTC ACTtCGCCTA AATGGAACAT CTGCGCCGCT 21900

GTTGCTTCGT AAATTTTGCC GTGTGTGCGG AATACTTCTA ATACCCAATC TTGTTTGGCA 21960

TACCAAGCAA TCACTCGGGC CTCAATGGCT GAAAAGTCAG ACACGATGAA CCGATGCCCT 22020

TCTTTGGCAA CTAAACCTGT TCGGATAAGT TGTTTCAATG TGTCTGGCAC ATCTTCATAC 22080

ATTAATTCGA TGCCTTCAAC ATCTTTTGCT TTCACAAGCT GACGGGCAAA ATCAATTTCA 22140

CTTAAATAGT TTCTAGGCAA GTTCTGTACT TGTAATAATC GCCCTGCCCA CCGTCCTGTG 22200

CGGTTGGCAC CGTAAAATTG TAAAATGCCA TGAATGCGAT TATCTGAACA ACGTGCATTA 22260

TCCATCATCA GATATTTTTT CGTGCTTGAA TTAGACAAAC TAAGGCGCAA TTTTAATACC 22320

TCCGCCACAT TCTCGGGCAA GTTTCCTAAT GCTAGAGCCT TTAACACAAC CTCTTTTCCT 22380

AACTTCTCGA AAGGCGTTCC TTGTTCTTCT AGCCACTTTT TCAGTTGCGC TAAGCTGTTA 22440

GGATTTTCCA ATCCTGTTAA AGCCTTCATT TCGTTTAATC CCGCTTCGCT TAAATCGGCC 22500

ATAATATCAA TAGCAGCCGT TGCCAATTCG TGGTCAATTT CTGCCCCTCT GTCGTTAATC 22560

CGCTGATCCA AGGTGTAAAG CTTCCACTCA CTTTCGGGGA CAGGGAAGCG ATTCAACTTA 22620

TTCGCGATAg CCATTTCTAC ATTGACGTCT TGAATACAAT ACTCTATAAA CGTTTGCCAT 22680

TTTTCGGGAG CGTGTTCAGG TAAATTCCTT GTACGCATGC CGTTTTTCTT GGTCGGCTTA 22740

CACGGTTTAG AAAAGAAGTT AATCAGTTGC GTACCTCTTG TGTCTTTTTG CTGTTCGATG 22800

TTTAAATATT TCGCACATTG CCCTAAAGAA GCTGGCAACC CTAATTCGTT TGCGTGAACC 22860

ATCGTACAGT GCCATTGTGC AGGGTCTAAG TAGTACGGAA TACCTAAATA CTTAGACAAA 22920

CAAACGCGCT CAAATTGGGC ATTAAATGCA ATTTTAAGCA CGTTCTTATC TGTTAGCGCA 22980

GCCACTATCT CGTCAGGTAT TTCGCTAATT GTTAAGTCCT CACATTCGAC AGGCTGACCA 23040

TCAACCGAAT AAGCAAAAAG TAAAATTTCA AAGTTGGGGC TATCAGCGTA TTTATAAACC 23100

CCAACTTTGG TTAAGTCTTC ATCCGAATAC GTTTCAATAT CAATATTTAA CGTTTTCATT 23160

TGCTATCAGC CGTTTCAGAT GGTAACCGCA ATTTACTTTT ATCAACGGCA ATCATCGGAA 23220

GCTGTGCTTC TTCCTCTTTA AAAACATCAA CGTTGATCCC TAGCTCTTCA AGATAAGCCA 23280

CTCCTTGTTT AGCAGGCACT CTTTCTACAT TTGCAATAAT CGCATTGTAC GTTTCTCTTA 23340

```
CCTCTTGGAT TTCTTGATCG TAAACAGCTT CTACTTTCGC AAGCCCTTTT ATTGAGTTAT   23400
ACTTAAGTTT TTCAAGAAAC CCTTTCTTAT AGGTTTCAAA ATCTCTAGCA TCCGTATCGC   23460
GATAAAAGCT CCATATTTCC AAACTGTTTT CTTCTAACAA ATCGACAAGT TCATTTTTTA   23520
CGTTTTGGCG AATTCTAAGT GAGCTGAAAT ATCTATGCAA TTTCTCAAAT TCAGGATCAA   23580
TTTTCAGCCA CTCTGATTCA ATATGTGCAC GAATAGCCTt CGTCTTTTTC GCTCTTATTT   23640
CTTTTTCCAC TGAATTTTTT AAAGTAATCA CATGGTTTCT TAAACGTGCT TTAGTTGGTT   23700
TCATTTTTTC CTACCCCCTG ATCTTCTACT GGTGTATATT CTTGTTTGAC ATAAAAATTA   23760
GTATCTCCGA ATTCTTTGAA TTCCCAACGG TGCAATCTGA CGGATTCTAA ACGGCTACCG   23820
TTTTCAAAAT ATATAGAATC CAGTTTTCTC TCTTCAATTA GCTCAAAAAC ATCTTGTAAT   23880
TCAATAGGCA AATAAATTGG TCGTTTACCC ATTTCAATTT CTCCTTTCAA AAGAAAAGGG   23940
GCATTTTAGC CCCTTATTCA ATTTTTAGCT AAACATATCG TCTTCTTCGT CTTCCCATTC   24000
TAGGTCACCA AAATCGGATT CAGCATTGGC ACGACCGCCT AAGAAGTCCC CTTTACATAA   24060
CGTTAAAATG TTGTTTAGCC CAGCAGTAAC CCCTTTGTTT CCTGcAGTAC TGTAAGCATA   24120
AAAGTTAATA GATGCAATCG CATAAACACC TGAATAGACT TCGTCTGGAT CGTCTGTTTT   24180
TACAAGCACA CCATCTTCAC GTTTTACGAC TTGTGGTTTC GTTTTACTTG ATACGTTGAT   24240
AAACATTGCG TTTTCAAATT CTGGGCGTTC TTCGGTGTCC ATTTCTTCAT CGCCATCGCG   24300
TAGtAgTTTT CAAGCGCTCA AATTTAACGC CTTTTAATTT ATCCCCTTTT GCGCCTTCAT   24360
AGGCTGTTTT GATTGCTTCT TTCATCGCTT TAAGGGTTTC CTTATCGTCT TTTGGAATAA   24420
TCAACATGCA AGAATATTTT TTCTCTTGCC CTTCTTCCAT CGCATGAGGT TCTAACACAT   24480
GCACAAAACT TAATCTCACT TGATTTGTAA TTACTTTCGT TCCAGTTACT TTTGCCATTT   24540
TAAATTCCTA CTTTCTTATA TTTTTTAGTT AAATAATTCG TCTTCTTCTA AATCATCTAA   24600
CGAATTAGAA GCTTGTTCTG TAAAATACTG TTTAATGTCC GAGATGATTT GTTCTTTCAT   24660
CTCTAAAATA GATTCTTTAT AGTTTTTTAG GTTCATTGTC CACGTTTCAT CTACAAGCAA   24720
TGAGTTAGTA ACTAACGACG ATAACGCCGC TTCAATCGTG TTAAAATATC CTTTAAATAA   24780
CACTTTATTC TCACTGAATA CATAAAGCCC AACATTGCGT GTATCTTCTC TTGTGATGTA   24840
ATAACCTTTA CCGCTAAAAA TCACAGGCTT TTCTTTACAT GATTTTTGAG GCATTTTATC   24900
CCCCACCTTC TACTTAAATA CATCCAAAGT ACCTTCTTGA CGTTCTAAGA TAACCACCGT   24960
ATCTGTCACC TTAATATTTA CAGGGCGATT TCCTGTGCTG TTTTCAACGG TAACAATAAC   25020
CGCATTTAAA TCATCAAAAG GCTGTATGAC AGTAAATTCA TCCTCAGaTA CGTTAATTCC   25080
TGAATGCTCC GCTTTCTCCA TTCCCTTTTT TAGAGTTGAA GCTATTGCCT CTTGTGTGTC   25140
AATTTTTTTC ATAGCCATTA CTCCACCCCT TCAAAATCAT TAAGCGCATT TTCTACGCTG   25200
TTTAACGCTG GTCGCTTGTC ACTTkCAGGA ACAAGAACAG GCTTTCCTTC TGGTTTGATA   25260
ATGAAGTCTG CAGCTAATTC CGCAAATTTT TTCTTGCCGA CAACTTTTTC TAATTGCCCA   25320
```

ATGGCTTTCA GTTCCTGCGG TTTTAAAATG TCTTCATCTT CAAATCCTTC GGCTTCTAAA 25380

ACCATCAGTA GCCCTTCATT ATCTGAAATT TTTCGATTAC TtCGCCCAGC AACTACTTTC 25440

CAACCAGGGA ACTCTTCGCC TTCGTCCCTT GCTTTCTGTA AAGCGTAAGT TTCCACATGT 25500

TCCAGCCATT TTTTTATCTC TGGCGCACGT TCTAAAATTT CAGCAATTTC CTCATTGGTT 25560

AAAAGCGGTG CTTCTTTCAG TTCGTGCTTA TCAATCAGCT GGAAATTTCT CTCCGCACGT 25620

GGCCGTAGTT GGGCACGAAC TTTTGAAAAC TTCACAACGT CGTCTGTAAT CGTCCATTCT 25680

CCTGTACCTT CCCAAGCCTG CACTGCTCGA GGCGCCACGT AATTATCAGC CCAGTATAAT 25740

AATTCTTCTT TTTCAATTTC AAAGGTCGAA ATATTGTCTA AACGAGGTTG AATAATCGTC 25800

ATTCGCACGG TTTCAAATTC GTAAATAATG TCGTACTTAT CCACCGCACC AAGCGCATAT 25860

AACATCAGTT GAGGGTTTAA ATACGCATCA ACAGGAACAC CTTTGCCGTA TTTCAAGTCG 25920

ATAATTTCAA TCGTCTTATC TGATAAGACA ACCACGTCCG AAGTTCCAAA TCCTTCTGGG 25980

ACCCATTTTG AAAAATCTAC TTTTTGTTCT AATTCGACTG TGGCATTTTC GTATTGATTG 26040

ACACGTTCTT CTACTAAATC ACAATAAGCG GTGACGTATT CTTTCATGGA TTCATCACAA 26100

TACGGATGGT CTTGtTCAAA AAATTTCAAT CGTGAGTTTA CCGCTCTAGC GGTTAATAGC 26160

TTAAACCGct TCGCCAAATA CAACTCTGCT AATTCGTGTG CAGTTGTTCC CTCCTCCGCA 26220

TAAGAGcTGC CGCGATCTTT TACTTTTTCC TCTAGTCGTG CCAAAGGTGG GCATATTAGC 26280

CACCGATGCG CACTACTCGC CCCTAATAAG GCATGATTTC CTACTGGCAT GCTTATTCAC 26340

CTACCAACGT TTCTAAGTCA GTGATAAATT TACCGTAATG CTCTTCTTTC AAGTCTGAAA 26400

GTTTTTCTGC GTTGTAACGA CTAAAGCAGG TCTTTATCAT AGGTCGCTTG CCTTTATCCA 26460

TAGCCTTTCT CATCGCTTCT TGCACATCTG CTTTTGTTGC CCCTAGGTGT AAATCGGCTG 26520

TTGCTGATAC TTCTTCTGTT TCTGAATCAG TAGGCGCGAC TTTCTCCCCA GTTTTCGATG 26580

GTTCGGTGTC TGAGGTTGAG GCCGTAGAAA CTTCTTCTTT CTTCTTCGCA GTTGCTTTTT 26640

TCTTTTTTGC AGGTtCTTTC TTCGCTGCTT TTTCTACTTC TTCTGTAGCT TTTTTATCCG 26700

CTGTATTTAA AGATTGCGTG ACTTCTACAA CGTTGTCTGT TGCTAACTGC GATAAAATTT 26760

CTTTTAACTC TTGTGCTGAA TCTGCTTCAA TATTTAAATA GATGTTTGCC ATTTTTTATT 26820

CCTCCAGTTA TTTTAATTGG TTAATGTCTT CATCTGTTAC ATTGAAAATC GCTTTAAAAT 26880

TTGCGTGATA ATGCAACGGT GGGTCTACAT TACGTTTTTC ATAAACTTTT ATCAGCTTCA 26940

CAGGAAGTAG CATTCTTTTT GCTAATTCTT CTTGTGACAT GCCCGCTTTT CGTCTAATTT 27000

TCTTGTACGG CGATAACATG AGTTGCTACA CTTCCTATCT GGTTTAGCTC TAGGGTTAAA 27060

CATTCCCATG CTTCAAATGC TTGGATTCCC TGTTTTCTAA GGCAATTAAC AAGCGTTGTT 27120

TTTCCTGTAG GCCCcTGCTG ACCTTTCACA ATGATTGGTT TTCCCCATTT ATATGCTTTT 27180

AATATCTGCT CTTTTTGCTT CGATGTTAAA AAGTCAGGAA GTACTATCTT TTTATCGCTC 27240

ATCAACTTTC ACCGCCAATT CCCGATAAAA TCCAGGTACT TCTTCTTTTA GTTCTGTTTC 27300

TGTGAAACGT TGTTTTAGCT TGTCTGATTT TGCTGTACAA ACAAAACGAT TCTTTTCAGT 27360

GTTGTAAGAT AAATATTGCT GTTTTCCCGT AGCATTTATC GCGTATGGCA AAGGAATTTC 27420

ATACATTGGT TCTTTTTTCA CTGTGTAACC ATCAAGAATC GCACGTGTAT GCAGTTCTCG 27480

ATTTTCTTCT ACTTCTTTCG TCCACTCAAG GCCTTTCCCG GTAAGCAACT CTTCATCTGT 27540

AATAGGATCT TCTAACCAAT GATCCCATCC CATCTTATTA ATTTGCCAAA TACACCAACT 27600

CTTATAATCA ATATGTGCTT CTTTACTTTT CTGTACCCAT TTATCAAATT CTTTTGAGAC 27660

TTCAATTGAT TTTTTAGGTG AAGTTGCTGT CCATTCAATA AAATCAGTAA TTCCTAATAT 27720

ATGCTCTAGA AAAGGAGCAC TAATAATTGA CAATTCTCTA TTCGCTTTAT TTACTTCAAT 27780

AACTGTTTTT TCTTTATGCC ATTTCCATAA ATCGGTTTTT GTAGGCTCAA GTCCTGTAGC 27840

CCAGACATAC CCTTGTTCTT CCAAAAATGC CATTAAGTTA TCATATGCTT CTTGTGTTTT 27900

TACGTGATAA TAGTTTCTCA TTTGCGTTTC TCCTTTTTCG TGCTATTATT TACTTGTATA 27960

TTTTTTGTTT AGTAGCTTAC TTCGTTGGCG GACGAGGTAG GCTCTTTTTG TTTCAATACA 28020

GCAATCATCA CCTAACACCC CACTTCGTTG AGCATTTTTT GATAAGCAAC TTCGTAACGT 28080

TCTAACTCTG CTTGAAAATG CTTTAGCGTG CGGAGGTCCT GCATTGTCGG ATGCTTTGCA 28140

CTTTCGTGAC GTACTGCATC TCTTAACGTT TCAATCTTTT CTCGTACTGC TTCACGTACT 28200

AAAAAAGCTT CATTAGCTGT TAACATTTAT TTCACCTCAC GATCTTCAAG CGCTAAATCA 28260

TAAATTAAAA GCCAAATGAT AAAAACCGCT ATATAAATGT TTTGGATTAA TGGTCCAATA 28320

TTGCCACCTA CCAATAGACC TAGTCCGAAT ACAACCAACA GTGCCGCTAT ACGTCTTAAT 28380

TGATATATTT TTTTCACTGT AATCATCCTT TCTTTCCAAT GTAAACATCT GCAAATCTTT 28440

CATCGTACAC CTCATAAATT GAGATATTTT TTTCTGGATC TGTGCAGCAA TAACGTAAAT 28500

TAATCtTTCT TTCAAATTCT TTAAATTGAG CGGCATCTAA AAGTTTGTAG TCAATTGATC 28560

CATAAAATGT TTCATAAAAT TCCACGCTAG CCTTTTGCAA TATTTTGTGT AATTTTTTCA 28620

TTGCTTTCAC CTCCCTAAAA CCATCACAGA ACCGTTAAAA TTTCGGATtG CTTTTGCTTC 28680

TGGCAAACGT TCATCGCCGC CTACAATTTC ATTTGTTCGA TGGTTATACA CTCGTTTTTG 28740

CCCTTTAAAC AACACAACTG AAATGTTGTC GTATGTGCTA CAAGTGTCTT GCACTTTTAA 28800

ATCTTCGCCT TTGTAGCAAA TAATCATTTT TGTAAAACCT CCTACGGAAT TAATCTGTGC 28860

ATGCGGTTCT TATCTTCTAA ACGACTTAAA AAATCTAATT CGTTTTGAAT TTTTCCGCTT 28920

CTCTGTCTGT TAATAAATCA GCGCTTCTAA GAGCTGCCCT ATCTGATTGT TTTTGCTTGC 28980

GATCTTTTTG AATTTGGCAT AATATCCACG CTTCTTGTTC TGTCGTATAA GCCACGTACT 29040

ATCCGCCCTT TCTTTTCTTG TAACGGTTTT TATCCATCCA GCGAACAAAC TCATCAAACA 29100

CATCTAACTC AATTAATACA ATCTTATGTG TTGGATGAAT TACACCAGAC GAAAATTGCG 29160

GCGTGTCCTG CATTAAAGCC AGCCATTTCT CTAAACTAGA TTTAGTGATT GAGTACATTT 29220

TTTGTACTAC GGCTTTATTT CCATATTTAA TTGACGCACG ATCTGGATTC TGTTCGACAG 29280

CTTCCAGATC AACCTTAACC GGTTTTGGCA TACCCCTTAC CCCCTCTCAT GCTTTAGCTC 29340

GTTTAGTTCA ACTAGCACAG CCACATAATC GGTAACAAAA GAATAACGGT CCTGATGTAG 29400

TTTTTCTGTC CATCGATAAG ACACATCTAA CTGCAGTAGT TGATGAAGTA AAAAACTTTT 29460

TTCACGAATT GATGCATCTA GATTTTCTTG TGCTTTCCAA TAACGATTTG TTCCATCGCA 29520

TAATGAGCTT GATGAAATTT TTCTCTGCA TCGATTAATT CTGTTGTTCG TTTTAAAAAC 29580

TTCTTTAAAT AGATCTTTGA CAATTAAATG CAGTCTTGAT TTAACAATGT CTTCTGCCGT 29640

TAGTGCGGGC ATTTCGTGTG CATAGTTCAT TTTCTAAGCT CCTTTTGTTA AAAAAATAGT 29700

AATTACCACT TGCGTAAATA AGAATTACGT TTTATAATAA AAAAAGAATT AAGAATTACG 29760

AAAATCGTAT TCCCACAAAG TTTCAAAATC TACATTAAGT AATTTTGCTA ACATAATTGA 29820

TAGTTTGAAA CTCGGGCAGG AACGGCCATT CTCTATATTA CGAATAGTTG TTTCAGTAAC 29880

ACCaGTTAAC TCAGCTAAAC GTTTTTGAGT TAATGACTTT TTGATTCTCG CTGCTCGCAA 29940

AGACCCTCTT GTACAATTCA TATGAACATC TCCTCTCATT TAATTTACGT AATTGCCAAT 30000

TACATAAATG AGTATACAAT GTAATTAAGr ATTACGCAAG AGGTTTATTT GCTATTTTTA 30060

CAAAAATAAC ATAAAGGAGC TTTAAAATGA CCTCATTTGC GCTGAGATTA AAGGAATTAA 30120

GAAAAAATAA AAAAATGTCA CAGCAAGAAC TTGCAACTTT TTTAGGTATG AAAAGAGAAA 30180

ATATTTCTAA CTATGAGCGA GGGGTAATTA CCAATGTATC TAGTGATACA TTAGAAAAGA 30240

TTTCAGAACT TTTTGATGTA TCTATAGACT ATTTGCTAGG AAAATCAGAT ACATCTAATC 30300

AGAAAAATGA ATCGATGGCT AAACAAGTAA TGATGCGCAT GGATACAGAT GGACTTTCTT 30360

CAAAGCAAGT AAAAGAACTA GAGGATGAGA TGGAACGTTT CTTTTCATGG AGAATAGAAG 30420

AAATAAAAAA AGAAAATTTA AGTGAATAAT TACAAGGTGG TTTTCAGGAT GACAGTAGAA 30480

GTAGACGAGT ACCTTAACTT TTGCGGTACC ATTAACGAAT TTATTTCAGC ACACATGTTA 30540

TGTTTGGGAA TGAGTGTTAA TAATTATGAG CATAGATATA TTTGGGATGA AATATTAACA 30600

TCTAAATCAA TTAAAATAAG GCCCTTTCCA TTTGAAAAAA CAGCACGTAG ATCTATATCA 30660

GGTATGATTA TTAAAGATGA TTACGAGACT ACGCTAGCTT ATAATTCTAA CATGGGCGAA 30720

AAAAGAAAAA ATTTCACGAT ATCTCATGAA CTTATTCACG CCATGTATCA TTTAGATAGC 30780

GAAAATAAGG TATTCACCGA CACAAAAGAT ACTTTATCCT ATTCACTGGC AGATATCTTA 30840

CCAGAATTTC AAGCGAACAT TGGAGCATCT TCTATACTAC TACCCGAACC TGTTCTTATT 30900

AATGAGCTGA AAAAAGGAAC TCCTCCTTAT TTTATCTCAA ACCGTTACGG TATATCTGAA 30960

CAAGCTATTT TTATGAGATT ACTTCAGCAA ATGCAGGCTA GTTTTGAAGC ATCTTACGTT 31020

GCCGCCTATG ATACGGCCAA TAAGATAATG AACGGCAACT CTAAAAATTT GGCAATTGAA 31080

TTAGGACGAA ACTTAGAAAG AAAAATTTTA TACAGCAACC CTTTTTATGA AGCAATTACG 31140

CTATAATGAA TTTATATTAA ATCGAAAGAA GGATGATAGA AATGAAAAAA ATGATTAAAT 31200

TTGCAGGCAT TGCTCTTATT TTTGCAGCTC TTCTCTCTGC CTGTAGCAAC GCAAAAAAATA 31260

ATACACAAAA GAAAGCCGAA ACTGCTGCCC AGTCAAGCAC TATTGAAGCT TCAGACAGTA 31320

ACGAAAACGA GCCTAATACA GAAAACATAA CCCAAGCAGT TAAACAGTTA GAAGAAAAAT 31380

TTAACTCTGA CGAGAAATTA GTAAAAATAG ATGTTAAAAA TAATGTTAAA GATGACACAT 31440

CAGATAACCC TCACGCTGTC ATTACGGTTA AGGTAATTAA TGATGAAGCA AAAAAAAATA 31500

TGGAAGAAAT GCAGACTGCG ATAGATTCCA ACTCAGGTAC AGAGGCACAA AAGACTGCCA 31560

TATACGGAAT TCAATTAAAT GTTGAAGAAG TAGCCAAAAC ATTAGAAAAT GATAACGATG 31620

TTATTTCTTT CATCACACCT TACACGAATG GGAACGACAG AACCATAGCA AAATCAACTA 31680

AAAATGAAAA TATTATTCCG TTAGTAAAAT AAAAAAGTGC ATCCCCCGCC GGCAAGCTAG 31740

TGGATGCACC TTATCAGAAA ATAAACCAAA GGCTTATTTA ATGCTATTAT ATCAAAAAAT 31800

AAGCCTCATT ACGAGGTTTT TATACACCTC CAAAAGAACG TAAGTTCCCA TAAAAGGAGA 31860

GTTGAATATG TGGGTAGAAA AGAAAAAGGA TAAAAAAGGA AATATCACAT ACCAATATCG 31920

CGAAAGATAC ACCGACCCTC GGACCGGTAA AACACAAAAA GTCACCACGA CTTTAGCAAA 31980

AAATACAAAA CAAGCTGAGA aTGCAGCACG TCGTGAGTTA GAAATAAAAA TCCAAAAGAT 32040

ACTGACCGAC TACGAAGAAG AGAAACGTCT AGAAAATTTA ATCAAAAACG AGGGCGTAAT 32100

ACCAAATGAT AAAACGTTAG AAGATATGAT AAAAGAATGG TTTGATTATA TTCAGGATCC 32160

AAAATCAAAA GAACGTAAAA AAGGTTCCAC GCTCTCGGGA TATAGCTATG GCATTGATTC 32220

ACTCTTAAAT GAGTTCTTAG TAATAGAAAA AGATAAACGA ATCCAGGATT TAACTTTTGA 32280

TGATATACAA GCATTTTATG ACCGTATGAT TTTTGACTTT GAATATCAAA AGTCCTATAT 32340

AAAAAGATTC CGTGCGATTA TCCGTGGTGC ATTTACTCTC GCCTATAAGA AAAGATACAT 32400

TAAAAATCTT GATGCCATTA ATCTTTCTAA AATAGTAACG CCAGCAAAAA CGGTGGaTGA 32460

TCTTATTAAT GAGCAAGTAC CTAGATACTT AGAAAAAGCC GAGGCTGAAC AAATATTTCA 32520

AGTTTTAAAT GGATACAACC CACTTTATAC GCAAATACTC GAACTACAAT TTCACACTGG 32580

GATGCGTTTT GGTGAATTGA TTGCTTTAGA ATCAAAAAAC TTTCACGACG GATTTTTGTT 32640

AGATATTCAT GGAACTTACG ATCATGCAAC ACGTAGTCAC ACCCGAGGAC AAAAAGTCCC 32700

GCCAAAAACA AGAAAATCAT TTCGTACTAT TGAGTTAGGT GATGCTGCAG TACAGATTAT 32760

AAAAGACC 32768

(2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7964 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

AGTGAGCGGG TCACTCTAGA GAAAAGGAGT AGAAAAAAAT GGCGGATTTA AAACGACTGT 60

TCATTGGGAA ACCATTGAAA TCAGCCGAAA ATGATGAACA TAAATTAAGC AGATTTGCAG 120

CGCTTGCGTT GTTATCTTCT GATGCACTTT CATCCATTGC GTATGGTACA GAACAAATAG 180

TGGTGGTTCT GGTTACTTTG TCGGCGGCTG CGATTTGGTA TTCATTGCCA ATTGCTGCAT 240

TTGTTATTAT TTTACTTATT TCTTTAACAT TGTCGTACCG ACAAATTATT CATGCGTATC 300

CGCATGGCGG CGGCGCATAC GTTGTTAGTA GTGAAAACCT CGGCAGAAAT GCAGGTTTAT 360

TGGCTGGTGG TTCGTTGTTA GTCGATTATA TGCTAACAGT AGCCGTTTCA GTTTCAGCGG 420

GTGCAGAGGC GATTATATCG GCGATTCCAG CGTTATAtGG TCATCAAGTA GCAATTTCTA 480

TTGGGATTGT GATATTGATT ACCTTAATGA ACTTGAGGGG CCTGAGAGAA TCAGCTTCGT 540

TCTTAATGTT GCCAGTGTAT AGTTTTATCG CAATTATTAC CTTACTGATT GTTGTTGGGC 600

TATTTAAAAT TGTGACTGGC GCACAGCCAT TGAATGCGAC AgcTTTACCA GGAGCCGTGG 660

TTCCAGGAAT TTCTATTGCC TTGGTTTTAC GGGCTTTTTC ATCTGGTTCT TCTTCCTTAA 720

CAGGGGTCGA AGCAATCAGT AATGCAGTGC CATTCTTTAA GAAACCACGA GCAAAAAATG 780

CTGCTGGGAC GTTAGCGTTA ATGGCAACGA TTTTAGGGTT CTTCTTCGTG GGGATTACGT 840

TTATTAACTA TTGGTATGGG ATTGTGCCAA AAGAAGAAGT GACAGTTCTT TCACAAATTG 900

GGAAAGCGGT CTTTGGTCAA AATATTTTAT ATTATTTATT ACAGTTTGCG ACAGCCTTAA 960

TTTTAGCGGT TGCCGCTAAC ACAGGATTCT CAGCCTTTCC AGTTTTGGCT TATAACTTAG 1020

CCAAAGATAA GTTTATGCCG CATATGTATA TGGATCGTGG CGACCGTTTG GGCTATTCAA 1080

ATGGTATTTT AACTTTAGCT GCTGGTTCCA TTGTTTTATT ATTGATTTTC CAAGGTTCCA 1140

CAGAACGTTT AATCCCATTG TATTCCATTG GTGTGTTTAT TCCGTTTGCT TTGTCTCAAT 1200

cGGGGATGGT TGTAAAaTGG CGAAAAGAAA CTAAAAATTG GTTGCCAAAA TCAATTGCTA 1260

ATATTGTCGG CGCCTTTATC TCTTTCGCGA TTATCGCCAT TTTGTTTATT TATCGATTAG 1320

GGGACATCTG GCCGTTCTTT ATTATTATGC CAGTGTTGAT TTATGCATTT TATCGGGTCA 1380

ATACTCACTA TAAAAATGTG GCGGAACAGC TACGTTTAGA AGATGGGGCA CAGTTACATG 1440

AGTTTGACGG GAATACAGTT ATTGTTTTAG TAGGGAACGT AACCAAGGCA AATGTCGGCG 1500

CCTTGAATTA TGCGCGTTCC ATTGGCGATT ATGTCGTTGC CATGCATGTC TCAATGGATG 1560

AAAATGTTGA GAAGGAAAAA GAAATTCAGG AAGAGTTTAA AAAACACTTC CCGGACGTTC 1620

GTCTATCAAT TGTCCATTCT TCGTATCGTT CACTGCAAAA TCCTATCTTA CGTTATGTTG 1680

ATTTGGTTAG TAAAAATGCC ACGAAGCATA ACTACAGTAC AACGGTTTTA GTGCCACAAT 1740

TTGTACCAAA TAAACGTTGG CAAAATATTT TGCATAACCA AACAAGTTTA CGCTTGCGGA 1800

TTCGTCTAGC GTGGCGGGAA AATATCATCG TGGCTACCTA TAGTTATCAC TTGAAAAAAT 1860

AAATAAAAAA ATCTGCCATC TTTTGATGGC AGATTTTTTT ATTGACGTAA TTGTTTTTGA 1920

TAAATTCTAG CGCAAGACTG CTCTTCGGTG GTTACCATTG TCAAAAATTC CCAGTCATAT 1980

CTTTCATAAA AAGTAGTATG GTCTGTGAGC AAATAAAGCT TTTTAAATCC AGAATTAGCT 2040

```
AGAAAATCAC CAACTGTTGT TAAAAGAATT CCTGCTAAGC CTTGGTGACG AAATGTTGGT   2100
TCGACATACA ATGCACAAAG GTTTGGCGTA AGGTCTGCAC GATTGTGGAA GTCATTATCA   2160
ATTACGCCCA CGCCTGCAAC GATGTCCTGC TGTTTATTCA AGCAGACCCA CCAATGTGGT   2220
ACGTGGTTTT CGCTAGTTAG CATTTCCTGC ATACTTGCTT GATACGCTTC AGCGGGGATT   2280
TGCCACTTTT GGCTAAACCA TTGAGCAGCT GACGGAATTA AGTGTGGTTG TTGGTTTAAA   2340
TCGATAATAG TATAGTTAGA CATTATTGAT CTCCTAGATA AAAAAACAGA TCCCATTGAA   2400
AAGAAACAAT GTTGGATCTG CTCTGTATTT TTTATTAATT GGAATGTTTG CTAGCTGGGT   2460
CTAAACGCCG TTCAATAATC CCTAGAACCC AATCTGTGAT AATCGCCATA AAGGCCGTTG   2520
GTAAGGCACC GACTAAAATG ATAGAAGCAC CATCTGTGGC ATTGGTTCCA CGAATAATAA   2580
TATCCCCTAA ACCACCAGCA CCAACGAAAG CACCAATCGC AGTAATCCCG ATTGCGACGA   2640
CTAACGCATT TCGAATACCT GCCATGATTA CTGAAACAGA TAAAGGTAAC TCGACCATAT   2700
AAAGAAGTTG TACTTTTGTC ATACCCATTC CTTTACCAAC ATCTAAAATA TTGCGATCCA   2760
CTTGAATCAT GCCAGTGTAC GTATTCTTGA TGATTGGTAG GAGAGAATAA AGGAAGACCG   2820
TTACAACCAC CGTGTTGACG CCTAAGCCCA TTCCCAACAT TAAAATCGAA AGCATTGCCA   2880
AAGCTGGTAC TGTTTGGATA ATGTTCGCTA AACGGATAAT CCAGTTGGCT AAGCTACGTT   2940
TACGTGAAAT CATAATCCCA ATAGGaATCC CGaCAATTGC AGCAAAGATT ACGCCATAAA   3000
TCGAAATTaA AAAGTGACGT AAAAATTGTT GaAAGACGTA GCCGCCATTT tGTTGGaAAT   3060
AGTAGATAAA TTGTTGGAAT AAATcATATT ATTCAtCTGk TACTCCCCCT CAAAATAGTT   3120
ATTTtCTTTT AAGAAaCGTT CGGCTACGAC AGCAGGTTCA ATTAAATCAT TATCTGCTTG   3180
ATAGTTCAAT TCTTGCATTT TTTTGGTAGA AATTTTACCA TCTAAGCGAT TTAAGACATC   3240
TTTAATTTCC GGCGTTTCTT TTAATAATTT GTCACTAACC ACAGGAGCGG CATCATAAGG   3300
CGGGAAGAAG CGCTTGTCAT CTTTTAACAT GACTAAATCG TAACTACCAA TCCGGCCATC   3360
TGTTGAGTAA CCTAAGATGA CGTCCATTTT TCCAGCGTTG ACTGCATCGT ACACCAAGCC   3420
AATTTGCATT GGTAAAATAG ATTTAAATTG GAAGTTGTAT TCCTTTTTAA AGCCTTCATA   3480
ACCGTCGCCT TCACGTGTGA TCCACGCAGT ATCGACACCA GCAACTAATT TGTCTGCGAC   3540
TTTCCCTAAA TCACTGACTG TTTTTAAATT GTATTTCTCA GCCGTTTCTT TTGTAACTAA   3600
GAATACATAA GTATTTTCAA AACCATATGA AGGGAACCAA GTTTGTTGAT AACGTTTTTC   3660
AAATTCTGAT TGAACTAGAT TGAACGCTTT TTTAGGGTCT TTTTCCGCTG GTAAGCCAAG   3720
CGTAGTAGTC ACGTCAGTAC CAGTATAACG AGCTGCAGAA ATTGAGGCAT CTCCGTTTAA   3780
CATTGCTTGG TGATTGATGG TAGTAGTGGC TAAGTTGTTA ATCACATTTG TTTTTTTATC   3840
TGTGTAATGC TCCACCATAT CCGCTACAAT GCTCCCTAAA ATCTGCGCTT CTGAAGTAAT   3900
CCCACCTGTA ATGGCAACAG TGTTATCATC AGAAGTTGAA GCGAGACCAG GTAAAGAACA   3960
ACCTGCCAAT AAAAGTAAAC TGCTTAAGAA GAGTAAGCTT AACTTGAGTT TCTTTTTCAT   4020
```

CATTATTCGC CCTCCCTTAA AGCTTTTGGC GTTAAACGAT TTTCTAATAA TCCAAGAAGA 4080

ACATCTGCTA AAAGCGCTAA AATAGTTACC GGAATAGTCC CGCCAAAGAT TAATTCTGGC 4140

TGATAGTTAT TTAATCCACT GAAAATCATA TCACCAAGAC CACCTGCACC AATATAAGAA 4200

GCCAGTGTAG CCCAAGCAAT GACATAGACA GCAGCTAAAC GAATGCCGGC CATAATTGTT 4260

GGCATGGCAA TGGGTAATTC AACACTAAAA ATAGACTGTA AATTGGTCAT GCCCATACCT 4320

TTAGCAACAT CTCGGTAATT TTCATCGACA TTTTTCATTC CGATATACGT GTTTCTTAAA 4380

ATAGGTAGAA GGGAATAGAT GAATAGCGCG ATAATAGCGG GTGTTTTGCC GACACCAAAA 4440

ATTGGAATCA TCAAAGCTAA TAAAGCCAAC GAAGGAACCG TTTGTAAGGC ACTTGTTAAA 4500

CCAATGACAA TATTGGCTAC TTTGGGCAAA CGTGTTAACC AAACGCCAAC GGGCACAGCG 4560

ATAGCGACCC CTAATAAAAG TGCAAAGAAT GAAATATAAA TATGTTCAAT ACTTTTTGAA 4620

ATCATTTCTG ATCCATATTC TTGAAAGAAG CTACTCATTG yTaCGCCTCC TCTTTATCAG 4680

TTTCGGACTC GCTTTGTTTT GCTTCTACAG CTTCACTAAT CGTTGTTTCA TCGCCCCAGA 4740

TAACATCATA AACAATGTCA ACAAGGGAAG CACGTGTTAA GATGCCCACC ACTCGTTTTT 4800

GTTCATCAAC GACTGGCACA TATTTTAAGC CGCGTTTTAA GATTCTTTGT AAAGCATCAC 4860

GTAAAAGAGC GGTTTTTTGA ACAAAGAAGA CATCTTTATT CAAGATATCA CCGACACTAG 4920

AAGCTTTGCC ACGTTGCTGA TCTAATGTTT CAACATCGAT AAAGCCTTTA AGCACATGGG 4980

AATTATCAAC AACTAATAAG GTATCAACCC GTTTTTCACG CATCAATTTA ATCGCTTCTT 5040

GTAAAGATTT TTCCGGCGTG ATAGTAATAG CTGAATTTAA CATGACTTCA TCAACAGTGG 5100

TGAAATCAGG TTTGGCTTGT AATAGGCGAT CTTCTCCGAT GAGTTCTTCT ACAAATTCAT 5160

TGGCTGGATG ACGTAAGATA TTATCTGGTG TGTCAAATTG AATTACTTTG CCTTCGCTCA 5220

TAATGGCAAT TTTATTTGCT AATTTCAACG CTTCATCCAT ATCATGGGTA ACAAAGACAA 5280

TAGTTTTTCC TAAACGTTCT TGCAAATCTT TCACTAAATC TTGTAATGAA TCACGGGTGA 5340

TAGGATCTAA AGCACCAAAA GGTTCATCCA TTAGAATAAT ATCTTGATTG GCAGCTAAAG 5400

CCCGCACAAC GCCGATTCTT TGTTGTTGTC CGCCAGAAAG TTCATTGGGA TAGCGGTCCA 5460

ACATTTCTCG TGGTAACTCA ACCAAATCAA TCATTTTTTC GGCTATCTTA TTGCGTTCTT 5520

CCACAGGTAC TTTTAATAAT TTAGGCACTA AAACGATATT TTCACGAATG GTCATGTGAG 5580

GCATCAAGCC GATATTTTGA ATGACATAGC CAATTTGGCG ACGTAGCTCT ACAGGATTGA 5640

TTTTTTGAAT ATCTkGGCCa TCaATTAAAA TTTTTCCTTT TGAAGGGTCG GTCATGCGGT 5700

TAAGCATCCG CATTGATGTT GTTTTTCCAC TACCACTTGT TCCGATGAAG CAGATAAATT 5760

CACCTTTATC GAAAGACAAA TTAATGTCAT CAACGGCGAC TTTTCCGCCT TTATAAAATT 5820

TTGATACATG CTGAAATTCT ATCAACTTAC TCACCTCAAC TAAATATTTT GAAAATGTAT 5880

CTTATTACTA GTATGCTATA AAAATGCGAT TTCGACAAAT GTAATCTATT TTTAATGGGA 5940

AAAAGAAGGC GGTAAAATTT TTTTAATTCG TTCTAAAAAA AAGAAAAGAT AACTATTGTT 6000

TAGTAAAAGT TATCTGAACG AAACAAGATG TTAAAAAGAG TATTCTGAGA AGGAAGAAAA 6060

GCCGCTATAA CAAAGAAAAG TGTCCTTTGT AGCAAGTTGA CGGGAGCTGA GTTTTTCATT 6120

AGAATGGAAA GAGAACAAGG GAATAAAATA AAAAAAGGCG CTTCATGATT TAGCGGCAAC 6180

TAAATCACGA AGCCTTGAAT AGTCACCTGT GCAAGTGGAC CATCAAGTTG CCCGTTCAAT 6240

CAATGGAAAC CATTGATCAC TTTTATTTTA CTGAATTTTT GAAAAAATTA CCAGTTTTTT 6300

TAAGAAGTAA GAAGTGGAAG AACGAAGGAT GCTTTTTATG GAATGTGCAA TGACGTTTAG 6360

AAAAATCTGA AGGTCATTTT TTATTCTCTT GTTCAAACAT TCAAACATAG GGGAAGAGAG 6420

ACGGACGTAA TTTTGTCACC GATTATTACG TCCGTCTTCC GTTGTTTTCT AAAGCGTTGG 6480

TTACATGTCT TTGTAGAGGT TGGCTTCAAT TGCAGTCCGT TTTGGTTCTA AAAAAGGCGG 6540

TAAAGCCAAC GTTTCCCCTA AATGTTCCAG AGGCTCATCC AATGTAAAAC CAGGAGTCAA 6600

AGTAGCAAAT TCAACCCGAT TCCCGCCTGG TTCACGAATG TATAAACTTT TAAAATATCC 6660

ACGACTTACA TACATTTCAA TTTCCCAGCC TTGTGCTAAA GCGCGCGTGT GTAGTTCTTG 6720

TAATGCTTCA TCATCAGCCA CAGAAAAAGC AATATGATCC ATACTGCCAC GCCCCATTCG 6780

GCTTGTTTTA TCCGACGTTG TTTGATGAAC AGAAAAGCTT TCCGTCTCAC TTAAAATAAT 6840

CGTATTGCCA ACTGCTTGCC AACCAAGAAA TGTTTCAAAA AAGGCCAAGG TTTTTTCAGA 6900

ATCTGGCACA TGAAATTCAG TAGAGGCCAA ACCGATAATT TGAACCGCTG CTGGAATGTC 6960

GTTAGAAACA ACGAGTGTTT CTTTTAATGG ATTTTCAGGA ACTTCTGTCA GGATAATATG 7020

CGTATCATCT GGATCACTGA AATGTAGGCT CTTTTCTTTT TTGTTAAATG GAATTTTGTT 7080

TTCTGTTAAA CGTTGTTCCC AAAAGGAAAG GCTGTTTTGG GGAATTTTTA AGCCAATCGT 7140

GGCTAAATAA TTTTTTTCTT CATAGCGGTT GCCTAATAAG GGCACGATAA AAAAAGTTAC 7200

TAAGCTTCCC GGCGTGCCTT GATAATCGCC ATAAAAATAA TGAGGCATGC GATTGTTGTC 7260

TTGATTTACG GTATTTTTGA CAAAACGTAA GCCTAACAAT TTTGTATAGA ATGATTGGTT 7320

TTCTTGACCA TGACGCGCTA ATAGTGAAAC ATGATGAAAT TGAGCATGCA TTTTGAGTTC 7380

CTCCTCTAAA TAACTATGTA TACTTACTTA AAGTAAGTAT AGTGCTTCTG ACATTTTTTG 7440

ACAACTATTT AGTTTGATTC AGAAGGCTGA CATAGCCAGA AACmAGAAAA AGTGTTAGAA 7500

TAACCAAGAT wAAAACTAGT AGAAAGAGTG ACGGAAATGC TAAGTACaGA AGATTTTGAT 7560

TTTGATTTAC CAGAAGAATT AATTGCTCAA ACCCCACTAA AAGATmGCGC TAGTTCACGA 7620

TTATTAGTTG TAAATAAAGA AaCAGGGGAC ATGGAGGATA AACATTtCCA TGATATTCTT 7680

GATGAATTAC AGCCTGGGGA TGCATTAGTC ATGAATAATA CACGGGTACT GCCGGCCCGT 7740

TTATATGGAG AAAAACCTGA AACAGGCGGT CATTTGGAAG TTTTACTTTT AACAAATACA 7800

GAAGGCGATA CATGGGAAAC CTTGATTAAA CCCGCGAAAC GTGCAAAAGT CGGCACGGAA 7860

ATTCAATTTG GCGATGGTCG CCTAAAAGCG GTGGTTAAAG AAGAGTTGGA ACATGGTGGT 7920

CGAATCATTG AATTTAAATA TGATGGTATT TTCTTAGAAA TTTT 7964

(2) INFORMATION FOR SEQ ID NO: 102:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7215 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

AGAAGTGAAA AGAGCTTTTG ACTTACGCTn nTATTATGAG GAACGAACCT TTCAGGCTGG 60

GAAAATAAGG AGGTATCAAA AGAATGAATT TGTTCAAaAA GArACAATCA AAGArACAGA 120

CGTTATCTAG TAACGTTGAA AAAGTAGATA AAAAAGTAAA AAAACGTGTA CAGCAACAAG 180

CAAATCCTAC CGTACAGGAG AGTATTTTTT ATACGAGTCA GTTTGAAGAA GnGCTCATGC 240

ACATTGTAGA GGATGAGTTT TCTAAATGCT ATAAACTtGG AGAaGTGGAC TACGAAATTG 300

CTGTAGAAGA AGaTCmATTA AACaCGGTAA TGGGGTATGC AGAAGGGTTA AATACACTAG 360

ACAAAAAATC TCGGTATCaG CTGCTGGTTG TTAATAAAAG AGAAGAAGAG TCGTTGATGG 420

AAGACGTTTT ATTGGCGTAT CaGGGAGATC ACTTAGATAA cTACCGTCAA GAAATCAATG 480

AAATTATAAC GAAGCAGTAT GAACAAGACG AACGAAACTT TAAGATTGAA AAATATGCTA 540

TTTTCACCAC AAAAAGTTCG TCCATGAAAC AAGCTAACAA AGAGTTAGAA GGAATCATGC 600

AGAACTTTGC CAATCGTTTT GAAGCCAACC AAATTAAATT AGAGGTAACC CCAATGACTG 660

GCTTAGAGCG ATTGTCTTTG TTCAGCTATT TTTTGCGCCA CGGACAATAT TTTGGCACAA 720

GCTATCAAGA TATTGCCGTA TCAGGGCTAA CGTCTAAAGC GTTTATTGTT CCAAGTAAAA 780

TCAAGTTTCC AAACAATAAA GCTTATTTTC GGTTAGGAGA AAACTACGCA GCTGTTTTAA 840

CCATTCAACA ATACCCAAAA TACCTAGAAG ATAAGTTAAT CAAAGAATTA TGCGCTACAG 900

GAAGAGAGTT AGCGATTAGC ATTCATGCCA GACCTTATGA CATGATGGAA GCCAGAAAGG 960

CAATACAGGG AACCAAGACG TTAAATGATG TTGCCATTCA AAAGCAACAG AAGCAAAACT 1020

TCAGAAGTGG TGTTTCAGAG GATGCGATTT CAGGAGAAGC AAAGGAAATC CAAGAAACAA 1080

CGAACGCGTT AATGGATGAA ATAAAAAATA ATGGCCAAAA GTTATTTAAT GGCCTATTTT 1140

CAGTTTTCCT AATTGCAGAA ACAGAAAAGG AACTTTTAGA AGCTGTTGAA TCGGTGAAAA 1200

ACGTAGGAGC AACCTGGCAA GTTGTGTTTG ATACGGTAGA CAATTACAAA GAAGAAGCAC 1260

TCAATACTAT CTTACCGATT GGAAAACCTT ACTTAGACGT TGAAATGAGT TATATGCGtG 1320

ATCTAACmAC mACmAATATt GCTACACAaG TACCTTTtAC CAATGTGGrA TTACAAgCCC 1380

AcAGGTCAGT TTtATGGCCG GAaTCmGTTG ACGAAGAaCA TGATtACTAT tGrCCGAAAA 1440

AAaGwTTAAT TwCACCAgTG GGCTTATTTc GGAcGcyGGT tCTGGTAAAG GAtGGCGACC 1500

AAATGGGAAa TTCTTtCAaC GCGTTtAAAA TtTCCGAAGa TCGCTTTTTA ATTGTTGATC 1560

CAGAAAGTGA ATACTTACCG ATTGGGAAAG AATTGGATGC AGAAATACTC GATATCTCaA 1620

CTGGTACGAA AAACCACTTG AATATTTTAG GAATGGTAGA CAAGCAGTTA TTGGATGAAG 1680

AGGACCAGGA ACTTGATTTA GTTAAAGAAA AATCTAATTT GTTATCCACT ATGTTTGAAT 1740

CGCTACTAAA AAGCTATACG GACGTGGATG CGGGACTTGT TGATCGTGTG ACGTTGTTGA 1800

CGTATGAACG CTTTGAAAAC AAAGAGACAG AGCCAACATT AGTCGATTGG TTTAGAATCT 1860

TAAAAGAACA AACGGATGAA GGTGCCAAAG AATTAGCGAA TAAAGTGGAA ACCTATTGTG 1920

TGGGTTCACA AGATATCTTT GCGCACCAAA CGAATATCGA CTTAACGGCA CCCTTTATTG 1980

TCTTTAATAC AAAAAAATTA GATGATCGCT TGaAAAAGTT TGCCATGAAA GTGATTTTAG 2040

ATCAGTTGTG GAAACAAGTA GTATCCGGAC AAGGAAAAGT AACGTCaCGT TTGTATTTTG 2100

ATGAATTACA AGTCAATTTC ACAACGGaAG AAGAAGCAGA ATGGTTCTTA AATTTATGGG 2160

CGCGTATTCG TAAGTATGGC GCCGTGACAA CAGGTATCAC GCAAAACCCT ACGACGTTAT 2220

TAGACAGTAA AGCGGGGCAA AAAATGATTG CCAATAGTGA ATTTTTGATT TTATTACGGC 2280

AAAAaCCAAT TGATTCACAA CGATTAGTAG rAATTTTAAG TCTGACACCT GCATTgAAGA 2340

AGTATaTCmA TGATCGAGCC CCACAAGGrA CGGGGcTAAT TtCTGCAGGT GGGGTAGTTG 2400

TTCCGTTTGA AAATCCTATT CCAAAAGAAA CACATCTATT TGAAATTATG AATACCGACG 2460

CTTAGGATGT GAGGACATGG TAGAAAGTAA GCGTACTTGG AAAGAAAAAG TACATGTTGA 2520

ATCTCAAGTA GAAGAGCGTC TTAGAAAAAA TTTGACCTTA AAATCAAGAG AAGCGACGAA 2580

tTGCAAAAAA AAGCATTAAA GACGACTAAA AAGGAGTTTA AACTTTCTAA GAAAAATTAC 2640

AAGCAATCAC TAAAAAAACA TAAACTTACC TTGAAAACAG GGAAGGAACC ATTAAGTATT 2700

AACCATTTAC TGGATAAAAA GAAATTTGTA GAGCAAAAGA AACGCAAAAA AGCACAGAAA 2760

ATTGCTTACA AACGAACGAA AAAAGTAGAT GAAACGCGTG TCGTTAACCA AGTGAAAAGG 2820

GAAACCAAGG AAGGGTTGAA AAGAGAAGCG ACGCAAAAAG TACGAACCAC GTTAACGCAG 2880

GAAGATACAC TGAATGAAGC AATGACCCTT TATGAAAAAA CACAACAAGC TAAATTTAAC 2940

ATGCGTACAG CTTTGAAAAC AGGAAAAACG GTTAAGAATC TTTCGGTGAA AACTGCAAAA 3000

GATACGTACG GATTGGGGAA TCGCTTATTT AACTTTTCAA GAGGACGTGG GTTTCAGCGC 3060

ACACCAAAAG ACTTTACGTT AAAAAAACAA CTTATGAAAC AACTACGGAA TCGCGCTATG 3120

CGCTTTAAAG CAGCAAAAGA AGCGAAGAAA GCCGAACAAG GTATTGGCTT GATTCGTTCT 3180

TTTTTTAATG GTCAGAAGAC ATTAGGGAAA GTCGCTGCAT TGATTCTAAA AAATCCAATT 3240

TCTTGGGTTG TTTTATTGGT TCTTTTTCTT GTTTTTTTAT TGAGTGGCGT TGCAAGTAGT 3300

ACACAAAAAC CAGCGATTGT TCAAGAAGAG GAAGATTTAA CCGCTTCATG GACGTATTTT 3360

ACTAAATTAG ATGCGCAACA TACAGACGAT AACAACCTGT TTTATAGCAA TATCGATGAT 3420

GTTCTTTTCT ACATGAATTA TCGGTACGAT GATTTTAAAT TATTAGATAT GGATTCCACT 3480

GGTACAAAGA ATTTTGAAAC GATTCTATCT GAACTTTGGA CGGCACTAAA CGGAAAGAAA 3540

CCTGATTATC AGTTGAAAAC AATGCAAAGT TTAGAGACAG ATAAGAAATC GAGTTATTTT 3600

ATCGAAaGAGG AACAAGCAAA GCATTATCAA GAGaTAAAAA AAGAGTTAGG CTACCAAACA 3660

```
TTGGATGATC TCCTTTCTTT TCCAGTAAAA ACTGACGCTT TAATTGTCaA TaAGCGTTAT   3720
GGCTATGACA AtCmAAAGAG aAGCTCACCT TATATAAAGg CATTGATGTA TTGATTGAAG   3780
ATAATCAGCC GTTTCATTCA CCAATGAaCG GCCAAaTCGT TTCAGTGCCT GATACaGAAA   3840
CGCTGGTTAT CGAAAAaGAG AAAGTAGCTA GGTTAACGAT TCGTGGTGTG AaTACCCTTC   3900
GACTAACAAA AGGTATrGAT GTTGAAGAAG GAACATTTCT TGGGAATACA AAGAATTCAA   3960
CAGTTACGTT CCAATATGAA AAATATAAAA AAGAAACGAA AGACTGGTTT TTTGTAAATC   4020
CAGCCTTTTA TTTTCCACGT GTAACCTACA CGCAAACGAC TCTTTTAGGA AGTGCAGAAT   4080
TTTCTCCTGG AGCAAGTGTC GAAAAAAGAG CGCAAGCTGT TTATGATTAT TTATCAAAAA   4140
AGGGATACAC CAAAGAGGGA ATTAGCGCGA TTTTAGGCAA TTTTTCGGTA GAAAGTGGCA   4200
TTAATCCAAA ACGTGCAGAA GGGGATTATT TGAATCCGCC TGTAGGTGCG CATGGGAATT   4260
CATGGGACGA ACCGAGTTGG CTTGCTATGG GAGGACCGCA AATTTATGGA GGACGTTTTC   4320
CAAATATTCT CCATCGAGGA CTAGGTTTAG GTCAGTGGAC GGATACAGCT GATGGTGGTC   4380
GTAGGCATAC TTTGTTATTA GACTATGCCA AAGGTAAAAA TAAAAAATGG TACGATCTGC   4440
AATTGCAACT GGACTTTATT TTTGACGGCG ATGCACCAGG TTCACGAACA GCAGCAGACA   4500
ATGTAGCAAG AAGTAAGGTG GCTGCAACTA TTCCAGAACT TACTACTTAT TATCTAACTG   4560
TTTGGGAAGG AAATCCCGGA GACAAGTTAG GAGAACGAAT TCAAGCTGCA CAAAATTGGT   4620
TTACATTCTT TTCTCGATCG GGTACACCAA TCGGTGGTTC TGGGAGTGAG GTCTTTGCGC   4680
AGTATAAAGA AAAAATGCAA CCATTACCCA CAGACAGAGA AmCAAAaGCA GGACaAGGCT   4740
GGCCAGGCmA TGCGTATGCG CCAGGGAATT GTACTTGGTA TGTGTTTAAT CGTTTgCACA   4800
GTTAGGAAAA AGCATTTACC CTACTATGGG AAACGCCAAT CAATGGGTGC ATAACTACAG   4860
TCmAACGCCC GGAGCsACAC TTGAATCGGC CCCTAAAAAA GGGGATGCTG TTATTTTTAC   4920
AAATGGTGTA GCTGGTTCTA GTACGCAATA TGGACACGTG GCTTATGTGG AGCATGTAAA   4980
CTCTGATGGT AGTTTTGTCA TTTCTGAAAT GAATGTCAGT GGTGAGTATT CTATGAATTG   5040
GCGTGTCTTA AAAAAAGAAG CAGGCGAATA CTTTATGCGC CTAAACTAAA AAGAAAGGAA   5100
CCAATAAGAA CAATGTATAA TATGGAAAAA TTAAGTAGAT ACAAGAAAAT TGGCTTTATT   5160
GGTGGCAGCT TACTGCTTAT TGTGCTTATT TTTTTGATTG GGCTTAGTGT TGGGCAAAGA   5220
AAACAAGTAA ACACCAATGA AAAACAAGTA AAAGTCGAGA AAAAAGAGGA ATTAACAACA   5280
AGTACGGTAA AAAAATTTTT GATTGCGTAT TATACAAAAA AAGATCTTGG TGAAAATAGA   5340
AATCGCTATG AACCTTTGGT GACAAGCGCC ATGTATAACG AATTAGTGAA TGTAGAAAAA   5400
CAGCCTGTTA ATCAAGCGTA TAAAGGCTAT GTAGTCAATC AAGTACTTGA TACGTATAAA   5460
ATCTATATTG ATACTGAAAA CAATGAAGTC ATTGTGGACG TCACGTATAA GAATACACAA   5520
CGAACCAAAC GAAATAATGA TGAAGGAGCG TTGAAAAATC AGAGCAATCA AGAAGCGATG   5580
AAATTAACCT TTGTAAAACA gGCGCTAATT TTTTAGTGGA TAAAtGGCAC CAGTAACACT   5640
```

TACCAATGAA TTmCAAGAGG AGCCAAATTC TTATAATACa CATGTAGTAA CGACAGAAGA 5700

ATCAACAAAG GAGAGTGCAA ACAGTGGGGA AAAGTAAAGT AAGTGAATTA GAAAAAAAGA 5760

AGGATCGAGT GGAGAAAGAG CTTAATAAAG CCAATGAACG TCTCGATACA TTGAAACAGT 5820

CGATTGCTAA AAAAGAGAAT GAATTGAAAC AAATAGAAGC GGAAATTGTC AGTGCTCTAT 5880

TAACAGAAAA TAATTTATCT TTTAGTGAAT TGACTGCGTT GTTGAGTGAA AAAGAaTCAG 5940

GTGGTGAAGC mAGTGTTTAT AGGGAaTAAA ACAACCTATC GCTATGAACT TCTTCATTCA 6000

AAAAATATTA AAGGCGAGGC GATCGACTCT ACCGCTACAC TTGTTCGTTA TGTGATTGGC 6060

GTAAmCAaCG ATCTTATTAA GAAGGAAGAA GAAGGATGGC TTTATGTCGA AAAAATAAAA 6120

GAAGGAAAAC AaGAGGACCA AATCCTTTAT GGCAAGCGTA TAGATTTACC AATTTTAGAA 6180

GACAATCCTT ATTTTCATGA TCTTTTAGAT CCGTTTGATA CAAAAAAAGT CATTCCTTTT 6240

GTAGAGCTAT TGCCTTCAGC TGTAGAAAAT CAGGTAAAGC CAGAAGAAAA AATTAAAGAA 6300

CCACAGGATA TTGTAGAGTT AGAGAAAAAT AACAAAACCA CTATTACTAC TGAAGAAAAG 6360

GAAATCACTG AATTAAAAGA GCAACCAGAT GAAATGGCCG ATAAAGAAAC ACAGAAGAAA 6420

ATTCTGGAAC AAAAGATTCA GTTTAAAAAA GACGAGTTAG CCACATTGAC CGAGGAGATT 6480

GCGACTAAGC AGGAAGAAAT CAATGACTTA GAACAAAAAT TAAGTACATT TTCAAAGAAG 6540

CCAGAGGGAC TTCATTTATA AAAAAGAGTT TTTAATCAGC TAGTAAAAAA GCyTCTTCTA 6600

GCTAAGAAAy AAGTATTATC ACTTtmCCCT TACAGTTGTT TATGGCCTAT CAmAAAAGTA 6660

GAGAAATAAA AAAGACTaGC AAAAGTTGCT AGTATTTTTT TTGTAAAGGA GGAGAATTTT 6720

GGTGGAACAA AGAGAAATTG TACACCGCTA TTATGTTTTA GGTACTAGTA CTTTAGATAA 6780

AGTATTAAAA GTGCTGCTTA ATTTATCCAG TGATGGTGTA TTGGCGCTAA AAGACAAGTT 6840

TCTTCCTTTA AAAGGGGAAA ATAATTTGTA TAAGCTAATG AATCGTGAGG ATCCACTGAC 6900

TTCGGCGCAA CTTCATGAAA AAGTGAATTT AAATAAGTTA AAAGAACAGT TGGAAGCCCA 6960

AGGGCTTCCT TTTGCGTTTA AAGAAACCAA AGAAGGGACT AACTTCTATT TTCGAGTAAA 7020

AGATACGGAA CTCGCAAAGA AAGCCTTAGA ACGCGTATTA ACGGATATTA AGAAGAACCC 7080

ACAAATGATT TTGCGTAACC AAATACCATG ACTTTTGATG AACGGTTAGC CTACACTGCA 7140

GCCAATAAAA AATACGTGGG tAAAAtCGaT CAAACCAAAA GCcAtCcCcA AAGGGAAGGA 7200

AACTGTTGAT TGCnG 7215

(2) INFORMATION FOR SEQ ID NO: 103:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5504 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

TGTTCGCCTA AGAGATTCAC CATAACCACT GATTGAAGTA ATTGGGGTAA CCCAATGCCC 60

CAATTGCTAA TCCCTCGAAT ATGCGCATCA AATTGACTAA ATGAGCAAGC TTCAATAGAG 120

TAATGACCTG AGTTATGAGG CCGTGGCGCT AATTCATTGA CGTACAACTC TTCTGAATCC 180

GTTAAAAACA TTTCGATAGC TAAAACCCCA GATAGATTTA ATTCCTCAGC AATGTATTGC 240

GCCAATTCCG CTGCTTTTTG TTGCACATTC TCAGAAATTC GTGCTGGCAC AATTGATTCA 300

TGCAAGATAT TCTTTCGATG AATGTTTTCA GATACAGGAA AGCTTTGGTA ACCATGCTTA 360

TTCCCTGCAA CGATGATTGA AAGTTCTTTA GTGAAAGGCA CCCACTTTTC TAATTCACAA 420

GTGGCGTTAG cTAATAATTT AGTCGCTTGG GAGAAATCCT CTTCGCATTT TAAGACGACT 480

TGTCCTTTTC CATCGTATCC ACCTTGAATA GTTTTTAAGA CACATGGAAA ACCAATCTTT 540

TCTATTGCTT GTGGCAGCTC TTTCTGGCTT TTTACTTCCG CAAAAGGGGC AATCTTACAG 600

CCAATCGATT GAAGAAACGT TTTTTCAAGC AATCGATTTT GAGTAATTTT TAACAATTCT 660

GTTCCTTGTG GCAACTTTTC TAGTGGAATG GCCTTTTCTA GGGAGGCCAC TTCAATATTT 720

TCAAATTCAT AGGTGATCAC ATCGCTTAAC TCTGCTAGTT CAGTTAAAGC AGATATCTGA 780

TCAAAGGCTG CCACAATTTG ACGGTCAGCT ACTTGACCAG CTGGACTATT GAGCGTTGGG 840

TCCAACACAA TCATTTGGAA ACCCATTTTT TTACCACTCA TGACTAGCAT TCGCCCTAAC 900

TGTCCCCCAC CTATAATGCC AATAGTTGCT CCTACTTTTA AAGGTTTATT CAAGTTGTTC 960

CCCACTTTCC ATAACCGCTT TTTCCAATGA TTGTCGTCTT GCTTGGATTC TTTCTGCTAA 1020

CTGTAAATCA TCAATACTTA GTATTTCTGC AGCTAATAAC CCAGCATTGG TAGCACCTGC 1080

TGAGCCAATC GCAGTTGTTG CTACAGGAAC TCCCCCTGGC ATCTGCACAA TAGACAACAG 1140

TGAATCAATA CCGTTTAACG TTCGTGTTTT AACTGGAACA CCAATGACAG GTAAGGTTGT 1200

CTTTGCAGCT ACCATCCCTG GTAAATGGGC CGCTCCACCA GCGCCAGcgA nTAATAACCT 1260

TCAATCCTTT TGTTCGTGAT TCCTCTGCAA AACGAAACAT CAAGTCAGGG GTCCTATGTG 1320

CTGATATCAC TCGTTTTTCA TAAGGAATTT TCAACTCTTC CAACACAAGA CAAGCAGCTT 1380

TCATCGTTGC CCaGTCTGAG GTACTTCCCa TAATCACAGC GACTAATGGT TtCATGAATG 1440

TAATAACTCC TTCCGCCTTT TTTCACTGTT ATTCTAACAA TGAGAAAAAG AGATGTCAAr 1500

GAAAAaCCGA AATATTATAG ATTAAAATTA TATAATGTTC GTGTTTATCA AAAAAAGAGT 1560

CATCGGAAAG ATgcTTTTTT GCtaTTTTTA GCGTTGCTTC CTCTTTAAGA AGAAACGTCG 1620

CCTGTATTAA TAACTGGTTG TGTGCCCTTA kCAGTAATGA TTGAGACTAG TAAATTATTA 1680

ATAAGTTGTA CTTTTCGTTC ATCGGTAAAA TTAATTTCTT GCCCTTCTTC AATCTGTTCT 1740

AAGGCCATTT GTGTCATGCT TACCGCACCT TCAACAATTG TTTGTCGAGC CGCTAAGATT 1800

GCTTTTGCTT GCTGTCTTTG CAACATGGAG CTAGCTATTT CAGTTGCATA AGCTAAATGA 1860

TTCaACCGTG TTTCGATAAC CTCTACACCA GCCACTGCTA AGCGTTCTTG CAATTCTTTG 1920

GTTAACTCCT CGGAAATTTG TTCGGTATTT CCTCGCAGAG TTACATCATT ATCTGAAAAA 1980

GTGTCATAAG GATATTGTGT TGCTACATGC CGGATTGCCG TTTCACTTTG AATTTCCACA 2040

AAGTCTTGaT AATAATCGAC ATTAAATAAC GCCTTAGCTG tATCGACyAC CCGAAAAACG 2100

ATGACCGCAG AGATTTCGAT TGGATTCCCG TCAGAGTCAT TGACTTTTAA TAGAGAACTA 2160

TTGAAATTTC GCACTTTCAA AGAAATGTTC ATTTTTTGCG TAAATGGAAT GGTAATAAAC 2220

AACCCATTTT CTTTAATCGT ACCTAAGTAG CGACCAAAAA ATAAAATAGC TTTCGCCTGG 2280

TTGGGACTAA CAATGGTTAA GGAACTTAAA AATAAAATAG CCCCAACTAA CAAAATAATA 2340

CCTAAAACAA CAAGAACACC ATTGGTATGA CTCGCTCCAT AAATAAATAA ACCAATTCCC 2400

GCAATTAAAA TAACTAATAA GCTCGCAATT CCTACGTAGC CACTCATATA AAATGTTTTT 2460

TTCTCTTCCA TCTGTTTCAC TCTCCTAACT CATTTATCTG GTCCTTATTA TACACCTCAT 2520

TGCCTTCAAT AGAGTAACTT TTAACATTAA ATAAACTGAT TTTTTCGGGG CTCTTGTAGT 2580

ATACTGAGAA CAAAGAAATC CCAAGAAAAG GAGTGATAGA ATGCGCTATT TAAGCAAGGA 2640

CATTTTAGAG GAAGTCATCA CTCAACGCCC CTCTGACTCT TATAAAAGTA ATTTTGGGCG 2700

TGTTGTATTA ATCGGAGGAA ACCGACAATA TGGCGGAGCa TCATCATGAG TACCGAAGCG 2760

TGTATCAATA GTGGCGCTGG TCTCACCACT GTGATTACTG ATGTTAAAAA TCACGGGCCT 2820

TTACATGCAA GATGCCCAGA AGCGATGGTC GTGGGCTTTG AAGAAACAGT CCTTCTGACG 2880

AACGTCGTAG AGCAAGCGGA TGTTATTTTA ATTGGTCCAG GTTTAGGCTT AGATGCTACT 2940

GCACAACAAA TATTAAAGAT GGTACTCGCC CAACATCAAA AACAACAATG GTTAATTATC 3000

GATGGCTCAG CAATTACCCT GTTTAGCCAA GGCAATTTTT CCCTTACCTA TCCTGAAAAA 3060

GTTGTTTTCA CACCTCATCA AATGGAATGG CAACGTTTGA GTCACTTACC TATCGAGCAA 3120

CAAACACTGG CAAATAACCA GCGCCAACAA GCAAAATTAG GCAGTACGAT TGTTTTAAAA 3180

AGTCATCGCA CCACAATTTT CCACGCAGGA GAACCTTTTC AAAATACAGG TGGCAACCCT 3240

GGGATGGCTA CTGGCGGAAC TGGGGATACC TTGGCTGGCA TCATTGCTGG TTTTTTAGCT 3300

CAGTTTAAGC CGACAATTGA AACCATTGCT GGCGCCGTCT ACTTACACAG TCTCATTGGA 3360

GATGATTTGG CCAAAACTGA CTATGTTGTT TTACCAACGA AGATTAGTCA AGCCTTGCCG 3420

ACATATATGA AAAAATATGC CCAGCCCCAC ACTGCACCTG ATTCTGAACT ATTAGAACAG 3480

AAGCGTTCGA GATAAAACCC GAAAAGGCTA AATGGTGACA AATTTTGGGT TAAAAAGCAG 3540

GATTTAAGCC TGGGCCGAAA ATCAAAAGCG ATTTTCAGCC CAGGCTCTCT TTTTATTTTT 3600

CCTTATAAAC ATTTTTTAAC GTTAATGGAG CGCCAAAGGT ATGAAAATTC AAGCCTTTGA 3660

CTTTATCATT TTGCAGGACT GTTTGTTCAT TTtGGCTAAG TGGAATCATG CCAGCAGTCG 3720

TTTCAATCAC TTCTTTTTCA GCTGCAATCA GTGTCGCCCA TCTTTTTTCT GGCTCTAAGG 3780

CATAGGTTGT GGCTGCTTCA TCTAATAATT TGTCATAGAC AGGGTTCTGA TAATTGCGAT 3840

CATTGCCCTT GTATAAAGTC ATCAGGGTAG AAATAGGGTC TTGATAGTCT GGTGTCCAGT 3900

AAATTAAGAA TAAATCATAG TCACTTTCTC GCCCAAAGTT CAATGCAGCT TCAGTCGGCA 3960

ATGCGGTTAG CTCTATTGTC AAACCAGGAA ACAATTCTTG TAGCGAGCCT TGCAAACTTT 4020

CACCAATTTT TTTGTAAGAA CCATCATCTG TTACCATCAA TTCAACGTTA ACCTTTTCTC 4080

CTAATTCTGC TTGTGCTTTT TTCCAATAAC TTTGCGCTTT TtCTTTGTTA TAAACCATTA 4140

AATTACCTGC TTCTTGACGA AAATCGAGAC CCGTTGTGGG ATTCGCCACA AAGCCTTCCG 4200

TAATCGCACC ATGTAGCGCT TTAGAACCAT CTGCAATAAT ATTATtGACT AGATTTTCTT 4260

TATCTATTCC TAAAGCCAAT GCTTTACGCA GGTTTTCGTT TGCAAGCGGC GTkGCTTGCC 4320

CTTyCCGTTT TTGATTTAAG CGCAAATAGT TCATTGTCGC TGTTGGATAC GAATGATACA 4380

ACGTATTATT TTTATTCTGT TGCGCCAGTT CACCACTTAG TGTAGCCACA TCTAATTGTC 4440

CATCTTCAAA AAGATTCAAT GCGGTAGATG TTTCTTTGAT AACTGTATAA TTGATAATGT 4500

CTGAGCGCAC GTTCTGGTGA TCCCAATAGC GATTATTTTT TGCTAGTTGC CAGTCCATCT 4560

TTGTTTGCTG CCAATCTTTT ACCACGAACG GACCATTATA GACGACTTTA TCACtAGCAG 4620

TTCCATAGTC CGCACCAAAT TGTTCGACTA CTTTTkGATT TtGCGGGaAA AATGTCGGAA 4680

AaGCTAACAA GGACGTAAAG TACGGTTTTG GCTCTTTGAG CGTCACCTTT AATGTATAAT 4740

CATCCACAGC TGTGACACCT AGTTCATTGG GTGCTAATTT CCCCGCTGAG ATTTCTGCAC 4800

CATTTTGAAT TGTTTCAACG ATGAGGAAGC TATAAACAAA GCCGTTTTTA GGATCAATCA 4860

TTTTTTTCCA AGCATATTCA AAATCATGTG CTGTGACAGG ATCATCGTTA CTCCAAACCG 4920

CTTCTTTTCT CAAAGAAATC GTGTAGGtTT TTCCATCTTC TGaAATCATC GGCAATGCTT 4980

TTGCTACGGC TGGTACCAAT TGGTCTTGTT CATCTAAACT ATATAACCCT TCAAACGCTG 5040

CAGTTTGGAC AATAGCATCT GGAAAATCCA ATAATACAGA GGTGTTGAGC GTTGTTAATT 5100

CCGAGGGTGA CATCAGTTGA ACGGCTTCTT TGGTATTTTT CTGTTTCTTT TTATTTTCAT 5160

TTTTATTGGT ACACCCAGCT AAAAATCCTA CCAATAAAAT ACAAACCATT GTGATTTTTA 5220

AATACTTTTT CATCTTCTTT CTCCTATAAC TCTTCATTTT TTTATGCTCT TACTTACAGA 5280

AAGACAGCTC CCCTTTTATT TTGTCCTTTC CACAAGCATC CAGACAACGA CCGCTGGCTC 5340

CAACCCATTA TTATAATTAA TGTGCCCCTC ACCAGCTTCA ATTAATTGGG CTTGTCCAGC 5400

TTTTCCAACT AAGTCTTGAC CATCTTCTAA AATCGTATGT GTcGCCAGAA ATCACATAAG 5460

AATATTCATC CTGTTCATGG CGAGCATATC CAGCTAAAGG CCGT 5504

(2) INFORMATION FOR SEQ ID NO: 104:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6433 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

GTTGCGACTA AACTATAGCC AATTGCTGAA AAACGCTTAG CAATTTCTAA CGCCTCTTCT 60

```
TTTGTTTCAT CAGCAATGGT GAAAAGTACA GCCCCATAAC TTGGCAAATG CAAGCCAGAT    120
GCTTCAAAGG CTTTATACAA GGCTTTTTCT AAATAATAAT CACTACCCAT AACTTCCCCA    180
GTTGATTTCA TTTCAGGCCC TAAGTACGTA TCTACTTTTT GTAGTTTCGT AAAGGAGAAT    240
ACAGGGGCTT TCACATGAAC TTGTTTACTT TCAGGATATA ATCCATCTTG ATATCCTAGG    300
TCAGTCAGTT TCTCACCAAG AATCGCTTTT GTAGCAACTT GGGCCATCGG AATCCCTGTG    360
ATTTTACTTA AGAAAGGAAC TGTTCGACTC GCACGAGGAT TCACTTCAAT CACATAAACA    420
CGATTTTCAT GAATAACAAA TTGAATATTC ATCATACCTA CACAGTTTAA ACCAAGTGCC    480
AGTTTTTTTG TATAATCGGC AATTGTTGCT TGAATTTCTT GAGAAAGATA CTGCGGTGGA    540
TAAACCGCCA TTGAATCACC AGAATGCACG CCAGCTCGTT CAATATGTTC CATAATCCCT    600
GGAATTAGTA CTGTTTCACC ATCACAAATG GCATCAACTT CACATTCTTG TCCCAATAAA    660
TAGCTATCAA CCAAGACAGG ATGTTCTGGA GAAGCTTTGA CTGCATGTCG CATATAATCT    720
TCCAAATCCT TTTGATTTTC GACAATTTCC ATTGCACGAC CACCTAAAAC ATAACTTGGA    780
CGAACTAAAA CGGGATAGCC GATACGATCA GCAATCACGA CTGCTTCTTC TGCACTTGTG    840
GCCGTATCTC CTGGTGGTTG AGGAATCGCT AACTCTTGCA AAGCTTGTTC AAATAAATCA    900
CGATTTTCCG CACGATCTAA ATCTTCAATC GTTGTTCCAA GAATTTTAAC GCCTTGTTTT    960
GTGAGCGGCT CAGCTAAATT AATCGCTGTT TGTCCACCAA ATTGGACGAT CACGCCAATC   1020
GGATTTTCTA AATCAATCAC ATTCATGACG TCTTCTAAAG TTAATGGCTC AAAGTATAAT   1080
TTATCAGAAA CAGAGAAATC TGTTGAAACT GTTTCAGGAT TACTATTCAT AATAATTGCT   1140
TCATAACCAG CCGCTTGGAT CGCTTTGACA GAATGAACCG TTGCATAATC AAACTCGACC   1200
CCTTGACCAA TGCGAATGGG ACCAGAACCT AATACTAGTA CAGAAGGCTT TTTCGAAACA   1260
TTGCTTTCGT TTTCCACTTC ATACGTGCTA TAAAAATAAG GCGTGTGTGA TTCAAACTCT   1320
GCAGCACACG TATCCACCAT TTTATAGACG GGCACAATTT GATTGGCTCT GCGGAAGTCA   1380
GCGATGGCTT GTTCTGTTTG TCCCCAAAGT GCTGCGATTT TCCGATCAGA AAAACCATTT   1440
TGTTTGGCTT CTTTCAAAAC AGCTACGTTG TCCACGTGAC TTTCAAGGGC TGTTTCAATT   1500
TCGATAATGT GTAAAAGTTT ATCCAAGAAG AATAAATCAA TTTTCGTTAA ACTTTGTAAT   1560
TCTTCGATGC TATATCCTCT GCGAATGGCT TCTGATAAGT AGAACAATCG ATCATCTTGC   1620
GCATGGACCA TTTTCTTCGT TAATTCCAAA TCACTGACAT GACTTAGTTC TGCTAACTCG   1680
TTGTGATAAG CACCAATTTC TAAAGAACGA ACTGCTTTTA ATAAGGATTC TTCAATATTG   1740
CGGCCAATTG CCATGACTTC CCCGGTTGCT TTCATTTGCG TTCCTAGTTC TCGAGCACCT   1800
TTTTCAAATT TGTCAAAAGG CCATCGAGGA ATTTTGGAAA CAACATAGTC CAAAGCTGGT   1860
TCAAATTCGG CATAAGTGGT TCCAGTAACT GGATTTTTCA TTTCATCTAA TGTTAAACCA   1920
ACAGCAATTT TTGCAGCTAA CTTCGCAATG GGATACCCTG TGGCTTTACT AGCAAGTGCA   1980
GATGAGCGAG AAACACGAGG ATTCACTTCA ATcACaTAAT AwTTAAAGcT ATGCGGGtCC   2040
```

```
AArGyTAATk GAACwTTACa GCCACCTTCA ATTTTTAATG CACGAATaAT TTTCAArGAC      2100
GCATCTCGCA ACATTTGGTA TTCATAATCT GATAAGGTTT GACTCGGAGC AAAAACAATC      2160
GAATCCCCAG TGTGAATGCC GACAGGATCA AAGTTTTCCA TGTTACAAAC CACAATCGCA      2220
TTATCAGCCG AATCACGCAT CACTTCATAT TCAATTTCTT TAAAACCAGC AATACTTTTT      2280
TCAATTAAAC ATTGTGTAAC TGGTGATAAT TTCAGACCGT TTTCAGCAAT TTGGCGGAGT      2340
TCTTCTTCCG TATCACACAT GCCGCCGCcT GTTCCACCTA AAGTAAATGC GGGACGAACG      2400
ATAATTGGAT AGCCAATTCG TTTGGCAAAA GCCACGGCTT GTTCCACCGT ATTGACAATT      2460
TCACTTTCTG GAATCGGTTG TTCTAATTCT TCCATCAATT GTTTAAATAA ATCACGGTCT      2520
TCCGCTTGGT CAATCGCACT GAGCTTTGTT CCTAGCAATT CTACATTTAG CTCATCTAAA      2580
ATTCCTGATT CCGATAATTC CATCGCCATA TTTAATCCAG TTTGACCACC TAATGTTGGT      2640
AGAAGAGCAT CAGGACGTTC TTTTCGCAAA ATTCTTGAAA CAAATTCTAA AGTAATCGGT      2700
TCGATATAAA CATGATCCGC GATTTCCTTA TCGGTCATGA TAGTCGCTGG ATTTGAGTTT      2760
ACTAAGACAA CTTCATACCC TTCTTCTTTT AGCGCTAAAC AGGCTTGTGT GCCTGCATAG      2820
TCAAATTCTG CTGCTTGTCC AATAATGATT GGTCCTGAAC CAATCACCAT GATTTTCTTG      2880
ATGTCCGTTC GTTTTGGCAT TAGTTTTGCT CCTTCCATGC ATCCATTAAT TCCATAAATT      2940
CGTCAAATAA ATGAAGACCA TCGTGCGGTC CTGGAGCCGC ATCTGGGTGA TATTGCACAG      3000
TAAAAGCAGG ATAATCTCGA TGACGCACGC CTTCAACCGT TCCATCATTG ACTTCCACGT      3060
GGGTCACAAG TAATTTTTCT GGATCAATTG TTGCTTCATC TACGGCATAT CCGTGGTTTT      3120
GTGAAGTGAA ATCAATGCGT CCTGTTGCAA TTTCTCGTAC AGGATGATTC AAACCGCGAT      3180
GTCCAAACTT CATCTTATAC GTATCCGCAC CATTCGCTAA TGAAAATAAC TGATGACCTA      3240
AGCAGATGCC GAAAATCGGA ACTTTTCCTT GGATTGCTTG AATCATTTCA ATAGCTTCAG      3300
GCACATCTTT GGGATCTCCA GGCCCATTTG TCAACATGAC ACCATCGGGA CTTAATTCTA      3360
AAATCTCTTC GGCCGTTGTA TTATAAGGCA ACACGGTTAA GTTGCATTGC CGTTTTGATA      3420
ATTCTCGCAA AATACTATGT TTCAACCCAA AATCAACGAC GACCACATTA CGCCCAATAC      3480
CTGGACTCGG ATAAGGCTTA GTGGTGGAGA CTTGTGCCAC TTGATTTTTC GGCATGACTG      3540
TCGCTTTTAA TTGATCAAAC GCATGCGGCA AATCATCCAC GGCATCAATG ATGCTCCCTT      3600
TCAkTGTTCC AGCTGATCGT AACTTACGAG TCAGCGCACG CGTATCAATT CCTGAAATTC      3660
CCGGAATCCC TTTGCGCTTT AAAAATTCAT CCAAGGTCAT TTGTTGACGC CAATTTGAAG      3720
CAACCCGCGC ATGCTCTTTA ACAATGACAC CTTTACAGGT TGGTGCaATG GATTCATaAT      3780
CaTCTCGGTT GACCCCGTAA TTCCCaACTA ATGGATACGT AAAAGTAATC ATTTGTCCAT      3840
TAAAACTTTG GTCAGTAATT GCTTCTTGGT AACCTGTCAT GCCAGTAGTA AAAACAACTT      3900
CTCCTACTAC ATTGCCtTCT GCaCCAAACG CTTTTCCTTC AAATACTGTT CCGTCTTCTA      3960
AAATCAATAA CCTTTTCAAA CCTACTGCGC CTCCTCCGAC CAAGCTAATG CTCCGTCAAC      4020
```

```
AAATGTCATA AGTGTCTTTC CTTTGACTGT CCACCCTGTA AATGGTGTAT TGACAGCCAT     4080
TGATTCAAAT GCTTCTGCAT CAATTGGTGC AGCTGTTGCT AAGTCAAAAA CTGCAATATC     4140
TGCTGGTGCA CCAATTsTTA GTGTTCCTGC GTTTAATCCA AAAATTTCTG CCGGTTTCAC     4200
TGCCATCCAA TCAATCACTT GTTCTAATGT GAAAATGCCT GTTTCGACAA AGTTCGTATA     4260
AATCAATTGG AAGGCCGTTT CACTACCGAC AATACCAAAA GGTGCGTTTA AAAATGATTG     4320
TTGTTTTTCT TCTAAACCAT GGGGCGCATG GTCGGTAGCA ATACAGTCAA TCGTGCCATC     4380
TAACAAGCCG TCAATCAATG CTTGTCGATC TGCTAAACCA CGTAATGGCG GGTTCATTTT     4440
CCAAAACCCT TCGTCACCAG GAATATCCTC ATCGACTAAA ATTAAATGGT GTGGCGAAAC     4500
TTCGGCTGTT ACGTGAATCC CTGCTTTTTT AGCATCTCGA ATAACGCGCA CGCTTTCTTC     4560
TGTTGAAACA TGACAAACAT GATAATGAAC GCCTGTTTCT TTCGCTAAAG TAATATCCCG     4620
AGCAATTTGT GAAGCTTcCG TTGCACTTAA AATCCCTGGT AGGCCTAATT TCTTAGAAAC     4680
TTCTCCTTCA TGCATCACAC CACCAAATAA TAAAGATTCA TCTTCCGTGT GCGCAACAAG     4740
TGCCATATTT AATGCAGCCG CTTCTTTCAT GGCTAAGTAC ATCGTTCCTG CCGTTTGGAC     4800
ACCAACACCA TCATTCGTAA ATGCAAAGGC GCCCGCTTCT TTTAAGGCCT TCTGATTGGT     4860
TAAAACTTCA CTCCGTAGCT CTTCTGTAAT GGGGGCATAT TGTAAAACTT TCACGACTGC     4920
ATCTTTTTGA ATTAAGTCGT AAACCTCACT TAATTTTTCA GCCGTATCTG GTACTGGATT     4980
CAAATTGGGC ATTGCGCAAA CCGTAGTAAA ACCACCGCGA GCAGCAGCTT TACTACCCGT     5040
TTTAATCGTT TCTTTATACG TAAAGCCTGG TTCTCTAAAG TGAACATGCA CATCCACCAA     5100
GCCTGGCGTG aTTAGTTGCC CGTTCGCATC AAACACTTGC TCAAAAtCTG CTTCGTCAAA     5160
CGATTCCCCG ATTGCGTGAA TCACACCATT TTCTAGCCAA AGTGCTGCCT CAATCAATTG     5220
ATTCTCTTTT TTGATGATTT TTCCATTCTT AATTAACGTC TTCATTTCTC CACCTAGGCC     5280
TTTCCATGTA AAATTGCTTC CAAAATTGCC ATGCGCATAA AGACTCCGTT ACTCATTTGC     5340
GCGACGATTC TTGATTGCAG ACTTTCCACT AATTCATCTG CCAGTTCAAC ATCCCGATTG     5400
ACTGGCGCTG GATGCATAAT AATTGCGTGT TTCTGTAAAC GAGTTGCCCG TTCGTTGGTT     5460
AAACCATATT CCAAATGATA GCCTTCTTTT GAAAAACTTT CCTTACCATC ATGACGTTCG     5520
TGCTGCACGC GAAcAACATC ATCACATCGA CTTTTTCAAC AATTTCATCT AGCGGCACAT     5580
ATTGTCCATA GACATCAAAC TGATGATCGT ACCACTCTTC TGGACCTGAA AAATAAATTT     5640
CAGCGCCTAA ACGGTTTAAT AACTGCATAT TTGATTTAGC AACGCGAGAA TGTGTAATAT     5700
CTCCCACAAT CGCCACTTTC AAGCCTTCAA ATCCACCAAA TTCTTCATAA ATAGTCATCA     5760
GATCCAATAA ACATTGTGTA GGGTGTTGTC CACTCCCGTC CCCACCGTTA ATAATAGAAC     5820
ATTGAATCGT TTTACTTTGA ATCAATTCAT CGTAATAATT TTCTTTACCA TGACGAATGA     5880
CTGCTACATC TACACCAATT GCCGACATCG TTAGTACTGT GTCGTATAAC GTTTCTCCTT     5940
TTTGAACAGA ACTTCTACTT GCTTCAAACT CAATCACTTC TAAGCCTAAT TTTTTCTCAG     6000
```

CTACCTCAAA ACTTTTATGT GTTCGAGTAC TATTTTCAAA AAACAAGTTT GTTGCAAAAT 6060

ATTGACGTTC TTCTGGGTGC CATTTTGCCC CCTGTTTAAA TTCTCCTGCG CGACGGATTA 6120

ACCCCATGAC TTCACGGTCG GTTAAAGCCT CAGCAGTTAA AAGATGTTTT AAACTAATTC 6180

GTTCAGATGT AATAATCATT CGTCCAGTCC TCCTACTGTT CTTCGCTACG TGCGTTTTCC 6240

GGTAAAATTA AATTCAAAAC AATTCCTAAG ACAGTTGCTA ATGCCATTGA AGACAATTCA 6300

AAAGTTCCTA CTTTGAACAC TAAGCCACCA ATGCCGATGA CTAAAATTAC TGAAGCAATC 6360

AATAAATTCT TTTTCTTATC AAAATTAATC TTATTATCGA TTAAGATTTT CAACCACTGG 6420

GGnAACAAAC CGA 6433

(2) INFORMATION FOR SEQ ID NO: 105:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5855 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

ATAGAAACCG CCTCTCTCTC GACACGCAAC TTGCGTGTTA TCTGTACGAT TTGTTTACAT 60

GTTGCATTGT AAACAGTTTT TAGGCAATCG TCAAGCAGAA AATAATTAAC gTTGCAtGAA 120

GAAACtAATA ATTGCAAAAA TAGCAATGTG AGCTGGATAG AAAATATAAA AGAAATTCCG 180

CATGCCACGG CCTTTTTCAC CATTATAAAG AACAATTGGA ATCGCTGCTA AAATCATCAT 240

CCATTGGTGG TTCACCCCGA ATAGTTCTTG GAAATTATAG CTGGCTGCTG AAATCAACGC 300

AACAACGCCA ATTGCTGAAA TTTGTAACCA CTTTTTGCCA TGGAATAAAT AAAAAGCTAC 360

GGCCAAGAGA ACAAATAAAA ACCCACCTTC TGTCATCATT AGTGATGGCA CAATCAGATT 420

GAAAATTTGG AAACCAATTA AGATAGCAGG GTTGCTTAAT ATGTCTGGAG AACTAAAAAC 480

AACGAGAGGT AGGGCACTGA TTAATAAAGG CACAATAAAC CAAAGAAGCC CTTTACCAAT 540

TTGTTTTTGT TTAAAATAGT CCATAGACTG CATGTAGACC GTACCTAAAA ATAACGTCCC 600

AAAAATATTA TTGATAATTA AGCCACCTGT TGAAAAATAC GCATTTAAAA AGCCGTTCAG 660

AATCCCCATA ATCCAGAAGC CTAACAATAA CCGGAACATG TATTTTCGCC GATTGCTTGT 720

GTGAATGAAG CCTTCAGAAC TTTCAAATAA GAAAATAGGT GCAGCAATTC GTCCAATCCA 780

TCCGAACCAA CCAGGCACGC CTAAAAAACT AAAAAATTGC AATAGATGGT CAAAGACCAT 840

TGAAATAACC CCAATGACTT TCAAATGAAA GCCAGTTAAT GTTTTACCCA TAGTATAAAC 900

GCCTCTCTTT CATAAAGAAC AATATAACAA AGTCCAGAAA AACCAACCAT TTATTTGCCT 960

AACAGTTATC TTACAAAATT GTTTGATTTC TTCAGCTTTT GATGAAAGAA AAAAGTGCCA 1020

AGACTAGTTT TTATAATTAT TCTATGCTAT AGTAAAAGTA CATTTTAATA AAAAACATTT 1080

GGGGTGCTGT TATGGCTGAG ATGATACCCA TTGAACCTGA TGCAGTTAGT ACTGTCGCAG 1140

GGAAATGCCG ATTTACTTTG ATTTTTCACG CACTCATTCT GTCCTAGGAT GAGTGTGTTT 1200

TCTTTTATCA GGGTTTATTT AAGTCACTCC TTTTGCTTTT TGTTAAGAGA AAGGAGTTTT 1260

TTTATTGAAA AAATGGTCAT TGCAACAAGT CGTCTTACTC GCTTTTTTAG CCTTTTTGTT 1320

TGGCGGTGTT TTTATGGGGG CAGGATTCCT CTATGCGTTA TTGAAtGCTG CGCTGTTACC 1380

ATTAGGACTG AGTCCATTTG CCAATGAGCT TTTATTTGGA ATGTGGACCA TGGTGGCACC 1440

AATTGCAGCG ATGCTGATTC CACGAGCGGG AAGTGCTGTT TTGGCGGAAG TTTTAGCCGC 1500

TTTGGCAGAA ATGCTTTATG GTTCGTATTT CGGTCCTAGT GTATTAATTT CTGGTGTGAT 1560

TCAAGGTCTA GGTAGTGAAA GTGGCTTTTT GGTGACACGC TACAAACGGT ACGATACACT 1620

AACGTTATTT TATAGTGCGA TTGGGACGAC TATTTTTAGT TATGTGTATG AATACTTTAA 1680

GTTTGGTTAT GGCAATTACG GTCTGGGGAT GAACATTGCT TTAATTAGCG TTCGGTTCGT 1740

TTCTATTTGT TTCTTTGGTA TTTTCTTAAC GAAAGTTATT TTACGAATGT ATCAGTCTGC 1800

GCAAGGGTTA GCTGTGAAGG CAAAATGAAT AGTATCCAAT TAAAAGCTGT GTCCTTTTCA 1860

CGAGAACAAC CCTTGTTTGA AAAGACAAAT TTAGCAATTC CTCAAGGAAC GTTTTCCTTA 1920

TTAATTGGCG ATAGTGGTAG TGGAAAATCA ACGCTCTTAC GATTGATTGC CGGATTTGCT 1980

CCGCTTGATT ACCAAGGAGA AATTTTGATT GAAGGAATGG AACGTCGACA ATTATCAACT 2040

CGTGAAAAGG CGCAAAAAAT TGGCATGCTG TTTCAAAACC CCAGTCAACA ATTTACAATG 2100

AAAACCTTGG AACGAGAACT GATTTTTGCT TTAGAAAATT TAGGAATCCC GCCTGAAGAA 2160

ATGAACCGGA AAATCCAAAC AGCGCTTCAA CTTGTTCAGA CTCAAACCTT ATTCACACGA 2220

GAATTAGCGA CGTTATCTGG TGGTGAAAAA CAAAAAGCGG CGTTGACAGT TCTGCTAGCA 2280

ATGAACCCAG ATATTCTTTT ATTAGATGAA CCTTTTGCAA GTATTGATCC AACCTCTCGA 2340

AAACAATTTA TTCAAATATT AGCAAGACTC CATCAAGCAG GCAAAACGAT TTTGGTGTGC 2400

GACCACGATT TCAGCGATTA CGCTGATGTT GTGGACCAAG TCGTCACGTT GAAAAATGGC 2460

CAGTTTGAAA AGCAACCGTT GACGTTTATC AAAACAAAAC CACAGACATT TCAGCTTACA 2520

ACTTCTGTGG TAAAACAACC AATGCTCCAA TTGAAAAATT TTCGTTTGTC TCAGGGAAAA 2580

CGAGTGCTAC TTGAAGAAAA AGAaGCGCTT TTGTTTAAAG GAATTACCAC GTTGACGGGG 2640

CCAAATGGGG CGGGAAAATC CACCTTATTA AAAGCCATCG TTCAGAGACA AAAATATCAA 2700

GGGAAAATGT TCTTGGCAGG CCGCCGCTTG CGGGCTTCAA AAAAATTGTA TCAACACATG 2760

ACATTAGCGG TGCAACAAGC CAATCGTCAG TTTGTCACGT TGACTTTACG AGAAGAGCTT 2820

TTGTTTGGTC AGAACATGAC AGCAGAAAAA AGAAGAAAAC AAGAGGAAGC CTTAACTTTT 2880

TTAGGTTTAA AAGAAAAACT AGAGCACAGT GTTTTTCAAC TAAGTGAAGG GCAAAAAAAG 2940

ATGGTGCAAT TAATTAGTTT GCTGAGCTTA GATTTGGACT GTTTACTTTT AGATGAACCT 3000

TTTGCTGGAT TAGATGAGCG AGCATGCAAT TATTTTGTTG AATGGATCAA AGAAAAAAGT 3060

GCGCAACAAG ATTTTCTAAT CGTCACGCAT CGGCTAGAAC CTCTCTCTGG CGTGAGCAAT 3120

TACCGAATAG AATTGCTGCA GCAACAACTG ATTATCTGGC AGGAGGGAAC CACATGCAAA 3180

TAAAGCAAAC AAATGCAGCA ATTTATGCTT CTTTGATTCT TATTTTGACC TTTGAATTAT 3240

CTTTTTCTCA GTCAATTTTG GCAAATTTGG CGGTTTTTAT GGGTTGTGTC ATTTTTTTAA 3300

TAGGCCAGCG AAAATCCCGC TTACTGCTTT GGTTATTTTT CTTGCCACTT TTACCAGCAA 3360

TCGGGACGTT TTGGTCTATT TATTTGCATG GTACGAGTAG TCAACAAGCC TGGCTTTTGT 3420

TTAGCCGAAC GTATGCCTTT GCCGGTTTAG GACTAGCTTT TGCTGTAGGT GTTGATTTTG 3480

AAGAATTATT GTTGTTATTG GAACAAAAGG GGCTAGCACC GAACTTTGTT TATGGAATTT 3540

TAGTGGTGGT TCACGCCTTA CCGGAAGTGA AAAGAGAAAT TAATGATTTA AAAGAAGCAA 3600

GCCTACTCCG TGGCAAAACG TTTCATTTTT GGTCGCCGAT GCTTTATGTT AAAACACTGC 3660

TCGTCGCTGT TTCTTGGCGA GATAAATATA CTGAAGCAAT GTCTGCCCAT GGCTATGAAG 3720

AAGGAGCCAC CAGAAGTCGA AAAGAGCATT TTGTGTcAGC AAAAGTCAGT CTGGGGCTTG 3780

CTATTTTGAT CATCATTTTT ACAAATTTGT TTATTTTTCT AACGTAAAGG AGAAACCTAT 3840

GACATTTACA GAACAAGCCA AAGATCaAGC GCAAGCTTCT TGGCAAGGGA GTTTTCAGCA 3900

TCCTTTTATC ACAGAATTAC ATGAAGGAAC CTTGAGTCCA ACGATTTTTC GCTACTATTT 3960

AATTCAAGAT CATTACTATT TAAAACACTT TAGTCAACTT TATCGCTTGA TTGCGGCACA 4020

AACGCAACAA CCTCGATTGA AAAAATTACT CTTAACAAAT GCGGAAAATC TCGCGTTGGG 4080

CGAATTAGCT ATTCGTGAAA CTTTTTTTGA AGAACTAGCG ATTACGGAAG AAGAAGTTGC 4140

AGCGACACCG ATTGCACCAA CCGCGTATCA CTATGTTTCT CATATGTATC GCCAACTTAT 4200

TGAAGGAACG CCTAAGACGG CCGCAGCAAG TATGTTGCCT TGTTCTTGGC TCTATCAAGA 4260

AATTGGTGCA CAGTTAGTCA AACAACACTC ACCAGAGCCA CTGTACCAAC GTTGGATTGA 4320

AACCTATGCT GGGGAGGAAG CGTACCAACA CGTCCAAGAG GAACGCCAAC TGTTAGACCA 4380

ATTATATGAA GAAAGTTCAC CACAAGAACA AGCAGCTATG ATCaCCGCTT TTGTTATCAG 4440

CaGTGAGATG GAATACGCTT TTTGGGAAAT GGCCTATACA CACGAAACCT GGATTGGTTA 4500

ATACAAAAGC AAGTCGTTTA GTGCTCGTAA CTAACGACTT GTTTTTGTAT TAGAAAGCAG 4560

TAATTGTCAG TTGCTTTTTT GACAGGGCGT GATAAGATAT TCAGTGACGC AAAATCTGTT 4620

GTAACTCGCT CAGTAATAAC AGAAACGCCT TGCTAAGAAA GGACTGTAGA AGCCTCATGA 4680

ATTATCCAGA ACCGATTGCA AAATTGATTG AAAGTTATAT GAAATTACCA GGAATTGGTC 4740

AAAAAACAGC GACACGGCTT GCTTTTTATA CAATTGATAT GAAAGAAGAA GATGCTAATG 4800

CTTTTGCAAA AGCGCTAATT AGTgTCAAGA GAGaTCtCCA TTTCTGTAGT aTTTGCGGAA 4860

ATATTACTGA AGAAGATCCT TGTGAAATTT GCCAAGATAA AAATCGTGAC CGTAGTATTA 4920

TTTTAGTAGT GGAAGAACCA AAAGACGTTA TGTCAATGGA AAAAATGCGT GAATATCAAG 4980

GCTTGTATCA CGTACTCCAC GGGGTGCTTT CACCGATGGA AGGAACAGGA CCAGAAGATA 5040

TTAATATTGC CTCTTTGATC AAACGATTGC ACGATGATGA AGTGAAAGAG GTCATTATAG 5100

CAACCAACGC TACCACTGAA GGAGAAGCGA CGGCGATGTA TTTATCACGG CTCATTAAAC 5160

CAGCTGGAAT CACCGTTACT CGTTTGGCAC ACGGCTTGTC TGTCGGCAGC GATATTGAAT 5220

ATGCaGACgA AATAACCTTG CTAAAAgCAG TCGAAGGACG TCGAGAAATT TAAAAGAAAC 5280

ATCGTTGGAA AAAGAGGTTG AAATCGAAGT AGGTCGTTTC TGTGAGAAAA TGTCGCAGTT 5340

CAAGCTTTwC CGTACTTCTT TTTCAGTTTC TTTTTTTCAG CGGATGCATC ACTTTAAATA 5400

GGACATAAGG ATTTCAATTA GACTGATAGA GCAAGATGAA GTATAATAGA TACAAGTAAA 5460

AATTGATTGG AGTTTTCGCA CGTGCAGGgT ATTTTTATCA CATTTGAAGG TCCAGATGGT 5520

GCTGGtAAAA CCAGTGTTTT GAAGGAAGTT TCTGAAnATT AGCTAAAGAG TCAAAACGAA 5580

AAATTGTCAC TACAAGAGAA CCAGGCGGCA TACCGATTGC CGAAAAGATT CGAACAGTTA 5640

TTTTAGATCC AAGAnATGAC AGAATGGATG AGCGAACAGA AGCTTTACTT TACGCAGCTG 5700

cTCGTCGTCA ACATTTAGTG GAAAAAATAT TGCCAGcTTT GGAAGCGGGT CATTTAGTCC 5760

TATGTGATCG TTTTGTTGAC AGCTCATTAG CCTACCAAGG CGCTGGACGT nGAATTGGGG 5820

ATGGCCCCCA TGCCTCAATT AATGCGGTTG CGGCA 5855

(2) INFORMATION FOR SEQ ID NO: 106:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8519 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

GTTTCATAAC TGTACCAATT GGAAAAACTG ACATGGAACG GCGTAACTCT TTTTCTCTCA 60

TCTCCATCCT CCTCTCTTAG TTATTTCTTA GAATAACAAT AGAATGACAA ATTGTCAAGA 120

ACTTATGTTA GATTTTCTAA CATAAAACAA AAAACCTGAT AAATATGGAC TTTTCCATCT 180

TTTCACGGCT GATTTTCTCC TACTAATCAA AAAAATAGAA ATGATTTGTT TTTACAAATC 240

ATTTCTAATC AACTTACCAT TTTTTATTCT TTTTCGGATT CAGCCAACCA TTTAGACTTT 300

TTTTTCGCAA AACCTTTCAC GATATACCGA TTTTCATTTT CAGCATATTC TTCTGAAATC 360

ATTAGTGTCA TCCGGCGAAG TTCACTTAAT AGTTGTCCGT CAGTTGGGCT AATTTCTAAT 420

TGATAAGGCT CGAGTAATTC CATCATTTGG GCACGGATAG CTTGAACCAG TCGTTCTTTA 480

TCTTCAGAAA TTTTAGTTGA GATCAATACA TTGGGAAATA AGGTTGGGAC AAATTCTTTA 540

CTGTCCACTT GATCTCTTTT ATTATAAACC GTCAAGCACG GAATATTTTC CAACGCTAAT 600

TCTTTCATCA ATGTTTGAAC TGTGCGTTCA TGTTGTAAAC GATCTGGCGC ACTTGCATCC 660

ACCACATGGA GCAGTAAATC CATCGTCCGA CTTTCTTCCA ATGTCGACTG AAAGGCTTCA 720

ATCAGCTGTG TTGGTAAATC TTGAATGAAG CCAACCGTAT CGGTCAGCGT AACTTCCATG 780

CCTTGAGGCA ATTGCCATTT TTTGGTTAAA GGATCTAACG TTGCAAACAA TTGATCCTCT 840

GAATAAGTTC CTGCTGTTGT CAACATGTTG AGAATAGTTG ATTTGCCAGC ATTGGTATAA 900

CCAATCAAGC CAATTTGAAA TAGTTCTGAA TTTTGTCTTT TCTGGCGGCT TCGTTCACGA 960
TGAGCAGTGA CTTCTTTTAA CTCACGACGG ATGCCTAAGA TTTTATTCCG AATGTGGCGG 1020
CGGTCCGACT CTAATTTCGT TTCACCAGGC CCGCGTGTCC CAATACCTCC TCCAAGTCGG 1080
GACAACTGTT TTCCTTGGCC AGCCAAGCGA GGCAATAAAT AATCCAATTG TGCTAATTCA 1140
ACTTGTAACT TTCCTTCTTT GGAGCGCGCA CGCAAAGCAA AGATATCCAA AATTAATTGT 1200
ACACGATCAA TTACTCGTAC AGCTAACGCT TCAGATAACA GTTGAtTTTG ACGCGGTGTG 1260
AGCTCATGAT TAAAAATCAC AAGATCCGCT TCATAGGCAT CTGTTAAtTG AATTAATTCT 1320
TCCAATTTCC CCTTACCGAT TACGGTTTGT CGATCGACTT GTGGTCGTTT TTGCGTTAAA 1380
ACAAACACCA CTTCTCCATT AGCGGTTTCC GTTAAATTCG CCAGTTCTTC CATTGATGCA 1440
GAAAATGTCT GATAATTTTn TTCCGTTTCA ACTCCTACTA GAATCACTTT TTCATGAGTG 1500
GTCATCGCCT ACTCACTCTC CTTCTGTTGT AGCCATTCCG CTACTTCTTT TTCCAACTTG 1560
GGcAAATCTG TTGGcTGCTG AACCAAATTC CACCAATGCG CAGCCATTCG ATTTCTAAAC 1620
CAAGTCAATT GACGTTTGGC ATATCGCCGC GATTGCTGTT TGACTGTTTC AACCGCCATC 1680
TCTAATGATT GTTCACCAGA AAAATACGGG AAAAATTCTT TATAGCCAAT ACCTTGAGCT 1740
GCTTGTGCAT GTGGTTGCTG AAACATCTGT TTTGCTTCTT CCAACAGCCC TTCTGTCATC 1800
ATTTGATCCA CCCGCTGATT AATACGCTCA TACAATAAAG CGCGATCTGT TTCCAAACCT 1860
AACAAATAAT AATCATAGAG ACGAGCTGGT TTTTCCTTCG GGGTCAAAAT ACTATAACCG 1920
GTTTTATCAA AGACCTCTAA CGCACGAATC ACTTTTTTGC GATTATTAAA ATGGATGCTG 1980
TTAGCCGCTA GAGGATCCTT TTGTTGTAAT AACAGCCATA AGGCTTGATT CCCTTTTTCT 2040
TCCGCAAATA AATTGTACGT TTCTCTAATT TCAGGCGAGT CATCTATCTC ACGACTTCCT 2100
AGTTGAAAAT CATAAAGGAG TGATTGAATA TACAAACCTG TTCCTCCAAC GATAATTGGT 2160
AGTTTCCCTT TTTCAGTAAT TTCTTTGATT TTTTGACGAC CCTCTTTCTG AAAATCAGCG 2220
GCAGAATATG TTTCGCTGAC TTCACGGCAA TCAATTAAAT AGTGAGGAAT TCCTTCTTTT 2280
TCTGTTTCTG TGACTTTAGC GGTGCCAATG TCTAACGAGC GATAAACTTG CATGGAGTCA 2340
CCGCTAATAA TTTCACCATT AAATTTTTTT GCTAAAGCAA TACTTAATGC TGTTTTCCCC 2400
ACCGCAGTGG GTCCTACAAT AACTAATACT TTCTCCAATT ATGTTTCTCC TTCTTCTTGC 2460
CAATTTTGAC GAGCAAATTT TGCTTCCTCA GGAAAATCAG TATGAACACC CGCCAATTTT 2520
AGCTGAAAAC AATTTTCCAT TTCTTCCATC CGATTAACAG TCCACGGCCG CAGCGGTTTG 2580
CCAATTGTGA TAAGTTCTTT TTCATGAGCC AACACCCAAG ACATTTTTGG ATGTAAACTA 2640
TCGACAAGCG CGAATGCCGG TCCCGCTTGA CTTAAATCCT CTGCAGATTC ATAAATAAAG 2700
GCAATTTCTG TCTTAGGATC AGCTTTTTTT AATTTAACTA AACTTGGAAA ATAAAAACTG 2760
GAATATAAAT AACGAAACGG CCAGTTTTTT TGCTTCATTT GTTGGACGAT TTTTTTCTCA 2820
ATTCCTTCAT AATGGATAAT ATCGGTTTTA AGTTCAATAT TCAAAAAACC ATTGAATTGT 2880

TCTTCTACAA GACAATTTAA AACATCTTCC AATGTTGGAA CACATTGAAC AGAAGGATTT 2940

TTCTGAAACC AGCTTCCAGC ATCTAGTTGT TTTAATTCGT TTAATGTTAA CTGTTGGATT 3000

TCACCATGAC CATCTGTCGT ACGATCCACT GTTTCATCAT GCATAACGAT TAAATAACCA 3060

TCTTTAGATA GGTGAACGTC TAATTCAATT CCTTCAGCAC CAACACGAAC AGCCTCTCTG 3120

AAAGCAATAC ACGTATTTTC AGGATGGGTG CCTTTACTTC CTCGATGTGC AATAATATCT 3180

GTCATGCTCT CACTCCAATC TTGTTTATTG TACCATTGGA GTGAGGTGGG TTCAAATAAT 3240

GTAGAAAAAA GAAGCCTCTC ACAAATGAGG GGCTTCTTTT TAAGATATAA CAACTAACGC 3300

ATTCCTGGAG CATTTTTTTC TGTAATATCA AAATTTACCG AATCAATAAC TTTGTTAGTA 3360

TTACTATCAA TAATCACAAA GTTCAAACCT TTTTTATTTC TAGACATTTC TGTATCATTT 3420

ATTTTTACAG AAGCAAATGT TTGATTTTCA TAATTAGTAT AACTTGAGAA TACATCATAA 3480

TTTAAAGCTT CAATTTTACC AATTAGTCTC GATAATTCTG CATTTGTTTC AAATGTTTTA 3540

TTATCTTTTG AAATCACTGC ACTTAAAGTC CACGCATTGT TACCtTCGAA TTTTACGTCT 3600

ACCCCTAAGT CTTTTAAAAG CGACTCTTCC GCTTGTGAAA GATAGACTTG ATTACCATAA 3660

CCACTTAATA AAATAGTATA TTGCGTATCA ATTAATTTTT GTAAATAATC ATTAAGATTA 3720

AAAATATTCA TCAACTCTGG GTTACCATTT CTATATATAT TGTCCCAATC TTTTTCTGAA 3780

ACACTTCGCA CAGACTCCAA GCCGTAATAT TTTGCAACAG AACGATTAAC TTGTGCATCT 3840

GAATTAGTTT TTACATGCGA TAAATGATTA GAATAAGCTG AATATTTTGT TGTATTACTA 3900

TATGATATAT CTGGAAACAC AGGATTGATA TTTCCTTTTT TCTCTTGCTC AACTAAATAA 3960

CTATATTGTT TTGTCAAATC TAATTTTGAC AAGAATACAT CACTTACACC AACCATTACG 4020

TTCATGAAGA ATGTACATAC CAAAATCACA TACAAAACAC TGTAAATTGG AGAAAACCTG 4080

ATATTTACTT TATAATTTGG CAGTGACATA GAAAAAGCGA TAATCAAGAA TAATACGGCT 4140

CCATAATATG AACGTCCGTA ATCTAATGCT GCTGGAGAGA TGGCTAAAAC TAGCATAGTA 4200

GCCATTCCAG AAATAAAATA TACAACACTT AGGTATACTC TTTTTTTATC TTGACTAATA 4260

TAAGCATATA AAACAAGCAA CAAGATCAGT ATCAATAAAA ATAATGATAA GTGTTCTTTC 4320

ATTTGCTCAA ATATTGTAAT AATTCCTGTC ATTAATTTTC TTGGTAAAGA CCAAGTACTC 4380

CTAGCAAAAT AAGTAGCGCG AATCTTATTT CCTGGTGCTA AAGCCATTAA CGCTAAACCT 4440

ATAGTATTAC CAGCAACACC CGTAAACATC CATAAAGAAA GTTTTCGTTT TGTATAAAAC 4500

TGCCAGCCAA GATACCCTAA AACGATAAGC AAAGCTCCTC CaGAGGTATT TTCATTACAC 4560

CAACCAGCCA AAATTCCTAA AGCAAACATC AAAAAAATCA TTAATTTTGT TTTCGTAAAG 4620

ACGTGTTCTT CAATGATTTC TCTATGATAC ACATAAAGGA ACGAAAGAAT AATTAAACTT 4680

CCCCAAAGAT AATTAGCTGA GCCTGTCTCC CACAAAAACA CTTCTCCAAA GGCTGGGGTA 4740

AACAACCAAA ATAGCACTTG AATAATAATA AATTTAAATG AATTATACTT TTTATCTTGC 4800

TTTTGTAAAG TAAGTCTATA AACTAAATAT GTTACTATTG TGTATGCCAA AGCATTAAAA 4860

ACATTAAAAA TCCCTTTTGG TAACAATAAA AATATTCTTG CAATTACATG GACCACTGAA 4920

CGTCCTGTCC ATGTCATGTA TTGATTGTAT TCATCTATAA GTATCGTCAT CCAACTTTTC 4980

GTTTTATAAA GGTATTCATA GTCATCTGCA AGTAATGGTG TTAAATTATT TAAAGTAAGC 5040

ATCACCAAAA ATGTAATAAT AATTGAGCAT ATAAATATCT TCTTTTTATT CAATTTAACT 5100

TtCATCTCTG CTTACCTTCC GCTTTTAGTT CATACTTAAC ATATTTATCC AAACCAGATT 5160

GAATACTATA TTTCGGTTCA TAACCAATGG CACGTAATTT rGATATATCC GCTAAGGAAT 5220

CTTkGATGTC CCCAGCTCGC GCTTCTTTGT ATTCAACTGG CAAAGTGACT TTCATAATAT 5280

CATTTAAAGA AGAAATTAAC TCATTTAAAT CGGTTGCTTT CCCAGTCCCG ACATTATACA 5340

CTTCCCCAAG TGATTGGTCA GAATGAGCGA CAAGATTCAA TGCTTGAACC ACGTCCTCAA 5400

CAAAGACAAA ATCACGAGAT TGTTTTCCAT CGCCAAAGAT ATTGAATGTC ACTTCTTGAT 5460

TGGCTAATAA GCGTTTGTAA GAGTCCATCA CAATCGAAAT AACGCCAGAA TAAGGCGAAT 5520

TAGGGTTTTG GTTTGGCCCA TAAACGTTAA AGAAGCGAAC AGCGCTTGTT GGAACATCAT 5580

ATAAATGACA ATAGTTTAAC ACATATTTTT CAGAAGCAAA TTTATCCACA GCATAAGGCG 5640

TTAACGGACG AATTACTGAT TCTTCCTGTT TTGGTAAAGT TGGCTCATCG CCGTAAACCG 5700

CTGCAGAAGA AGCAAACACC AAACGTTTTA AATCTTTCTG GTATTTCCGA ATTAATTCCA 5760

GTAATTGTAA CACACTCTCA AAATTCACTT GGTGTGTTTC CACTGGACGT GCTACTGAGT 5820

CTGCTACACT AGCAATCGCG GCTAAGTGAA AAATGTAATC GAACTGATAC TCTTGTAATA 5880

CTTTTTCCAT TAATTGTTGA TCAGTCACGC TACCTTCAAT GAACGTGATG TTTTCAGATG 5940

CGTTTAAATT TTCAGTTTGG CCCATTGATA AATCATCAAT GACCACAACT TGATTGTCTT 6000

TACTATaATa ATTTGCTAAG GTTGAGCCGA TAAAGCCGGC GCCGCCTGTG ATTAAAAAGT 6060

TCCTCAAGTT TCCCACTCCA ATTAAAATAA TACATTAATT TTACCATTTT TAGGTGTCAA 6120

AGACTACAGT TTGTTTTTTC AATAAAAAAG GAGAGAAAAT AACTTTCTCA ACCTTTTACT 6180

AAATATTTTA AATTCTTTGC GTTATTTACA AAGCTCTTAT GAATGAATAA GAAATAGATA 6240

ATATACACAA TTGCATAAGA AATGAAACTA TAAATACTAC CGCCAAACCA AAGATTACTA 6300

AAGATAAACA ATAATGTAAC CAACGTTTCC AATGCAGTGC TATTTCCTAT TTTAATATTC 6360

ATATCTTTGC ATAGAAAAAT TTCCGCTAAA TTACAACGAA ATGCTAAACT TACTAATATC 6420

AATCCAACAG TTAAATTCAA GTTCCCAACA ACGAAAATCA CAAACAACGA TAAGATTAAC 6480

GACAAGGCAA GTGTCmATaC GtTAACAAAT AAAATTGTTT TTTCTTTTCT CAACGTTTTT 6540

AAATACGTGT TAATTAGTAA AGACATCCGG CCTTCATAAA TAACTATAGG AAATAGAATT 6600

CCCATAAATT TCAAACTCTC GCTATATTCG GGTAACCACA TACCTAAAAC AAATTTGACG 6660

GGTACATATA ACAGCAGAAT TGCATACGTC AACGGAACAA AAACACCTCT TAATGTGACA 6720

AATAAAGACG GcAATCGCTC TTTATTTGTC CGTCTTAACA ACGGAAACAT TACAATTCCA 6780

ACAGCATTAA TAAAAGTTAA AAACATATTT GATAAaCTTA AAGTAAAgGm CAaTTTACCA 6840

AAGGTTTCGA TTGTCCAACG CTGTTGAACA AAAAAGCGAA TGGTTCCCAT TATTnGCnTA 6900

CTTGCTATAC TACTGAGCAT TAnTTnACTA CCTATGTTAA TATTATCAaC TATTTCAGGA 6960

GTAACTTCCT TCAAACTCAT CATGTTCACA CGCaACATGT CTTTAACTCT AAACATTCCC 7020

CAAACCGTAA TAATCAGTTT AGAAAGTATA TCCATAATGA TTAGCCAATA AAAATCTCGT 7080

CCTCCTAATG CAAAGTAAAT TGCAATGAAC AGGACATATA AATAGCGATC ATTTCTGGAT 7140

AATTGAGCAT ATTCCTTAAT ACGATTTGTT GACTGAAAAA CATACAGAAT AAACGTTTTT 7200

GCAATTGTAA TTACTGAAAC AATGGCCGTC AATATCAAAA TAATCCCTTG ATATTCTTTC 7260

GGCATGAAAA AGTAAGCCCA AAGAATGACA AAAGCAGAAA TACAACATTC AAAAATCGCT 7320

AAATACCAGA ATTGAGAACC GAGATTTCTT TTATCTAAAT TTTCGTATTC TTCTCCCCCA 7380

ATTTTTAAAT AGATACCATC AATCCAGCCT AAGTGGAGAA ATCCAACATA TGAAGAATAA 7440

AAAACATAAA GTTGCCAATA ACTGTATTCT GTAACGCCTA ACAACTTAGG GACAAATAAA 7500

TTTAAGAGAA CCGAAATACC CAACGTTGCA AAATTTGCAG CCACAGTATA ATAAAGGTTT 7560

TTAGCCACCA TTTTTGCTTT TGAATTCATA GCATTATGCT CACTTTCTAA TTTTATATTT 7620

CATGATATGG TAGATATCTG TGCTAAAATA AAACTTGTGA TGATACTTAA AAAAGGCACA 7680

TAGAATCGGG ATTATTTTTT TTCTTTCTAA TAAAGAAACT CGCATATTCA ACTCTTTAAT 7740

CAAATTTGAA AGAACTTTGT CTTTATATTC AAATTCTTCA ACATGTGGTT GATTATTTAA 7800

ATAATTCTTT GCAGCCTCTG TACGCTGTAA TCTCCGGCGT AAAGTATTCA ATCTTTCCTC 7860

AAAGTGATTT TGTTTCTTAG CTAAATCAAT TTCTTTTTTA AAGGAACTTT GTCCATGTTT 7920

ACGCCATCTA ATTAATTCTT CCCTAACAAT ATAGAGCCCA TCACTTAATA AAGAACTTCC 7980

CCACATCGCC ATATCATGTA CAGGATTTTC CATTTCAGAT GCGTATAGAT TAACATTGGC 8040

AATGAAGCTC TTTCTAATTG CATATACACA TCCCGGGCGA CGTAAAAATA AATTGTTCTC 8100

ATTAAATATA ATCTGTTCTG CTTTTTGATT AGGCAATTTA ACTGAATCAA TTTTTTTTAA 8160

TTCCTCTGCT AATCCCCCAG GTTCTACTAA TTCAGTATAA TCACTTACTA AAACTTTAAC 8220

ATCAGTGTTT GCAAAAGCTT CACACATCAC TTCAATTTTA GATGCATCCC AAATATCGTC 8280

TTGATCACAC GTAAAAACAA TTTCTTTAGT AGCAGCATTC AATAGATTAA AGAAATTTTT 8340

GCGCCAACCT AAATTTTCCG kGTTGATTtG CAAaTGCcAC GTTtCACCTA AwTTAwTTTC 8400

AGAAaTAAAG TtCTCAATtA TTTtAACTGT rGAaTCyGTA GrGCAGTCAT CCCAAATTAA 8460

TACTTCATCA ATTTTTTTCG ACTGATTTTT TAAGGAATCC AACTGTTCTA AGATATTTT 8519

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1508 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

TTTTGGGGGG ATAATCAAAT TCCCCTGnGT TCCGCCACAA GCCACACAAT GATGCCcATT 60

CTGGTCGTAC ACCAAACCGT TCTAACAATT GAATTTCAAA AATATTTGTA ATGATTTCCg 120

GATCTTTTTG CTCaTTTAAA GCGGATAATG CCATGTGGGT AAATTGAAAT AAGTTCGGAT 180

CATAGTGCTG ATCTTCAATT GCTGCATCGA CTAAGTTTAA AATATATGTT CCATATGCGT 240

TAATAAATAT GTCTTGTTGA ATGTTTCGAA AAGGTTGAAC TTCTTTGCTG CTATTGAGAA 300

AAGATAAGCC TTCTTCTCTA AAGTTTCCGA TGTAGACGGC TTCTGTAAAT GGCAGAATCG 360

CTGGCAAAAG AGGATTATTT TTTCGATGCG CACCTTTCAC AAAAAACATC AATTTTCCAT 420

AGGATTCTGT AAAAATTTTG ACGAGTTTAT CCTTCTCTTT AAAATCCTTC GTAAATAAAA 480

TAATGCCTTT GGTCTCCCCT AATGCCATTC AATTCACTTC CATTTCTTAT GTAAATTAGT 540

GTAGTGTTTG GAAAGTTCGT TTATACACTC GCTTCCCAAA CACTACTTTC TTAATATTCT 600

TCTTTGCGAT AACCAAAATC TTGTAAATAG ATTTTTTTGT CACGCCAGTC TTTTTGGACT 660

TTTACCCACA GCTCCAAATA CACTTTGTCA TCTAATAAAT ATTCAATATC TTGACGCGCC 720

TTGGTACCAA TTTGTTTCAG CATTTTTCCG CCTTTGCCAA TAATGATACC TTTTTGACTG 780

TCCCGTTCCA CAATAATGGT TGCTTGGACA TGGACTTTAT CGTTTTCATC CCGTTTCATT 840

GAGTCCACAA CAACAGCGAC AGAATGAGGa ATCTCATCTC TTGTTAAAAG TAACACCTTT 900

TCACGAATTA GTTCAGAAAC GATAAAATAT tCTGGATGAT CCGTAATTTG GTCaTCaGGG 960

AAATATTGTG GTCCTTCTGG CATTTGCtCA ACGAGAACAT CCATCAAATG CTsCAAAATT 1020

ATTTCCTTCA GTGGCAGAAA TGGGAATAAT TTCTTTAAAG TCCATTTGTT TCGAATAGTC 1080

TTCAATAATA CTCAATAAAT CATCTGGATG GACTTTGTCA ATTTTATTAA TAATCAAGTA 1140

GACTGGACTA TTATTATTTT TCAATCGTTC AATAATAAAG TCGTCACCTT TCCCGCGTTT 1200

TTGATCAGCA CTTACCATAA ATAAAGTAGC ATCTACTTCA CGCATCGCAT TGTAGGCACT 1260

TTCCACCATG AAATCGCCTA AGCGGTGTTT AGGTTTGTGA ATCCCTGGAG TATCGATAAA 1320

AACAATTTGT GCTTCAGGAA TTGTATAAAT TCCTTGAATC TTATTACGAG TTGTTTGTGC 1380

TTTGTCACTC AtAATTGCAA TTTTTTGGCC GACAATTCGA TTTAACAATG TTGATTTnCC 1440

AACAnTTGGT CTGCCGACAA TGGGCTACGA AACCAGATTT AAGGTTCAGT TGTCATGGAA 1500

TnCGCCGC 1508

(2) INFORMATION FOR SEQ ID NO: 108:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5918 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

ATGAAGTATT CAAAAGGACG GTACGTCCTT GTAGTCTTAG TAATGGTCTT GATTGCTTGG 60

```
TTAATTTTTA TTTTGTCAGG TTTAGCGAAC GGTTTAGCTC AAGGAAATCG TTTAGCCGTA     120
GATCAATGGC AAGCGAATCA AGTGGTTTTA TCAAAAGAGG CCAATAGTAA TTTAAATGTA     180
TCAGTGTTAG ATGAAAACGT GAAAGAAACG ATTTCAGGAG GCAAAATTGC ACCGATTGGT     240
CAACAATCTT TAGCCATCCG CCCAGCAGAT GATAAAAAGG CTGAATTAAC GAATGTTAGC     300
TTATTTGGGA TTGAAAAAGA AAGTTTTTTA ATGCCAAAAG TGATTGAAGG AAACGCATTT     360
ACTGATAAAA ATCAAGTGAT TGCTTCaGAA ACATTGAAAA ATCAAGGATT TAAAATTGGC     420
GACAAATTAA CTGCAGGGAA ATATGATGAG CAATTAGAAA TTGTCGGCTT TATTTCTAAA     480
AGTAGCTATA ACATTGTGCC AGTTATTTAC ACTTCTCTGG ATACTTGGCG GTCAATTAAA     540
TATGGTGACA ATCCAGCAAT GGCCAAAATG GTTAATGGTT TCATTGTTCG CAGCAAGGAC     600
AACACGGAAG TTAAAACGAC TAATAAAGAC AGCCAAGTTC TTTCAATTTC AGATTTTATT     660
GAAAAACTGC CAGGATACAG CGCTCAAAAC TTGACCTTGG ATGGCATGAT TTATTTCTTG     720
ATTGTGATTG CAGCGTTCAT TATCGGTATT TTTATCTTTG TGATGACTTT ACAGAAAACA     780
GCAATGTTCG GCGTCTTAAA AGTTCAAGGG GTGCCAACTA GTTTCTTAGC AAAGGCAGTC     840
ATGCTACAAA CCGCTTTATT GGCTGTTCTA GGAGTGGCGA TTGGTCTTGC TTTAACGGGA     900
ATTACCGTGC TCTTTTTACC AGAAGCGATG CCTTATGCAA CTAATGGACC ACGCATGATT     960
TTGTTTAGTG TACTATTGAT TTTATCTGCA TTAATTGGCG GGGCATTTTC AATTCGAACG    1020
ATTGCTAAAA TTGATCCGTT AATCGCGATT GGAGGTTAGA GACATGGCAA ATGTTTTAGA    1080
AATGAAAAAT ATTTATAAAA AATATGGTGA AAAACATACA GAAGTGATCG CGTTAAAAGA    1140
ATTATCTTTT GCGGTTCAGC CAGGTGAATT TGTCGCAGTA ATTGGTCCTT CTGGTTCTGG    1200
TAAAAGTACC TTTTTAACGA TTGCAGCTGG TTTACAGGCA CCGACAAGTG GCGAAGTCAT    1260
TGTTGGCGGG CAATCACTAA ATAAGTTAAC GAAGAAGCAA CGTTTGGCGC AACGTTTTCA    1320
AAAAATCGGC TTTATTTTAC AAAGCTCTAA TTTGGTGCCA TTTTTAACAG TGGAAGATCA    1380
ATTTCACTTA ATTGAAAAAG TTGATAAGTC ACGTAAAAAT AGTGAATTAA AAGAGCAATT    1440
GTTAGAGACG TTGGGTTTAA AAGAATTACG AAATAGTTAT CCTCGTGATC TCTCTGGTGG    1500
GGaAAGACAA CGGGTAGCCA TTGCGTGTGC ATTGTATCaT GAGCCAGACG TAATTTTGGC    1560
AGATGAACCA ACGGCTAGTT TAGATACAGA AAAGGCGTTT GATGTTGTCA AATTACTAGC    1620
GAAAGAGGCC AAAGAGAAAG ATAAAGGGAT TATCATGGTG ACACATGATG AGCGTTTGTT    1680
GAAATACTGT GATCGAGTGG TTCGTATTCG TGATGGTGAA TTGACAGAGT AACAAAACAA    1740
ATAAAAGaGA ACATGATCGG ATTGCTGGAG CAATCCGATC ATGTTCTCTT ATTTCTTTTC    1800
AATACAGAAA AGAACAGGTG GCTGATTTTT TTGATTGATA AACTCATAAC GTAAGACACT    1860
GTAAACATCT TGTGGTAAGG CTTGCGTGTA ATTCGTGACC AAATCAAGTT CCTCTGAACC    1920
ACCTTCGTGG CCATAATAAA CGACTAAAAT AATGCGACTT CCTTTTACCA AATGAGGCAG    1980
TAACCCATCT AACGCCTGTT TAGTCGTAGT GGGTTTTGTA ATAATTTCCT TATTACTTTT    2040
```

AGGTAAATAG CCTAAGTTAA AAATAGCGGC AGTGATTTCT GTTTCTTCAT CTAGGACAGC 2100

ATGAAGATGT TCATGTCCTA ATGGAAATAA CGTAGTTTGT GGGAGTAAAT TAAGTTCTGT 2160

TAACTTTTGT TCTGTATTGA TCAACGCTTC TTTTTGGATA TCAAAGGCGA AAACTTCACC 2220

AGAAGGGCCA ACCAGCTCTG CTAAGAATGC TGTATCATGA CCATTTCCCA TTGTTGCATC 2280

AATGACGAAG tCGCCAGGTT CCACCACTTC TTTTAAAAGT TGGTGACTAA AATGTAGGGC 2340

TGTTTGTAGC ATTAATTGAA AACTTCCTTT CCTGAAATGA CAGTATATTT TCCTTGATAA 2400

CTGTTGCGGC GTTTCATTTC CGCATCAATC GCATTGAGAA CTTCCCATTT TTTCAAGCTC 2460

CACATCGGTC CGATAATTGT TTCAAAGGGA GCATCGCCTG TCAAGCGATG AATAACAATT 2520

TCGGGTGGAA TCATTTCTAA CTGATCACAA ATCACTGAAA CATAGGCTTC TTTGCTCATC 2580

AATTGTAAGC GTCCTTCATT ATAATCCCGC ATCATTTTGG TATTTGTCAT TAAATGAAGT 2640

AAGTGCAGTT TAATCCCTTG GATATCTGAA TCCTGAATAG TTCGCCGCAC ATTTTCGCGC 2700

ATCATGGCAG GTGTTTCCCC AGGTAGCCCG TTAATTAAAT GTGTACAGAC CCGAATCCCG 2760

TGTTTCCGTA ATTTAGCCAC ACCATCTAGA TATGTTTGAT AATCATGAGC ACGATTAATG 2820

GCAGCGCTAG TTTCCTCAAA CGTGGTTTGC AAACCCAACT CTACCCATAA ATAAAAGCGT 2880

TCATTTAATT CGGCTAAATA GTTCACTACC TCATCAGGCA AACAATCAGG ACGGGTACCA 2940

ATGGATAAAC CAACGACACC TTTTTCATTG ACTACTTGTT CAAACCGATG ACGAATAACG 3000

TCAACGGGCG CATGTGTATT GGTGAAATTT TGGAAGTAAA CAATGTATTG ATCGACAGTT 3060

GGCCATTTTT GATGCATCAA TTGGATTTCT TTTTGAAATT GAAGCGGCAG GGGATCGCTA 3120

GGTGCAACAA TCATGTCGCC TGAGCCAGAA ACACTACAGA ACGTACAACC ACCTTTGGCA 3180

ACCGTCCCAT CACGATTGGG GCAATCAAAG CCACCATCAA TTGGCACTTT AAAAATCTTT 3240

TCACCAAACT GTTGACGAAG TGCATAGTTC CAAGTGTGAT AGCGTTTATT TGGATCATCT 3300

GAATAGGGAA AAATTTGCAT AATCAAGACT TCTTTCTAGC GTTGTTTGCT TTGAAATCGC 3360

CAAACAAAGC GTTTTTCTAA ATAAATAGGG TAACTTAATA ACAACCAAGC TAAACCAAGA 3420

CAAAAACCAC CTAAGATATC GCTAGGGAAA TGGACACCTA AATAGATTCG ACTAACCCCA 3480

ATCATGAAAA TCCCAATACC TAGTAAAATT TGCACACAAA GGCGGACCGT TTTATTTTTA 3540

ATGAAAATTG GCAGAATCAA TAAAATAGTG CCGTAGAACA AGGTACTTCC TACAGCATGT 3600

CCGCTAGGAA AACTATACGT ATGTTCAGTT ACTAAATGCT CCAAAGTAGG ACGTTGGCGC 3660

ATAAAGAATA ATTTTAATAA AGAATTACCT ACGCCAGCAA TTAAGGCAGT ATTAATAAAG 3720

AGCCATAATG CTTCAGCATA ATATTTCCAA ACAATAAACA GAGCAATGAC AACGATTGCT 3780

AAGATGATAA TCGTCAATGA ATTGGCAAAT TTGGTGTACC AAATAAAAAA GTTATTTAGC 3840

TGTGGAAAAG GACGACGAAC AAAAGAAGTA ATCGTTTGAT CAAAGCCTTT GAGCCAGGTT 3900

GGATAAAAAC GAACAACATA GCCTAAAAAC ATAAATACTA AAAGAAAACA ACTACCTGCA 3960

AATTGATAAT AAAGTTTGTT TTTCATAAAA AATCCTTTCA AAATTTAATC GCAACAATCC 4020

GTATTATACC ATCAAAAGGG TAAAAAATTA AAAGATAATT TCATTGCCTG ATTTGGAGGG 4080

GATACAGATA TTAAAAAAGt TCTGGCACCA CAATCTCTCA TTGTGGTGCC AGmACAGATG 4140

TTTTAACTGG CACATGTTTA TTGATTATCA TCATCAGGAG TATCAACGGT AATATCTCCG 4200

AATTCGCTAG AAAGGTTAAA CTCTGTATCC GCACGACCAT TCTTATAAGA TGTTCGCTCT 4260

TTGTTTAAAA GTTGAACATC GCCTAATTTT GATTCCCCTT TGATTGAAAT GTTACGCAAA 4320

ATTTCTCCAG AATTTACATA AATATCTCCA TGGACAAGAG CGACAGTTAA TTTTTTAGGC 4380

ACAGCAGGCA TGTCTAAAGA AAAATCACCA TTTTCACTTG AAAAGTTAGC TTCTCCTTTG 4440

AAATCATAGA CGCTAATATC GCCATCTGTC GTTGTTGCTT GGATGTTTCC ACGAGCATTT 4500

TGCACTCGAA TATCACCAGA AGTAGTCGTA TTTTTGATTT CATCGAAATA GGCTGAATCC 4560

AACGTTATAT CGCCGTCAGT GGTTTCAACT GTTGCTGTTT TAGCGGAAGT ATCCGCTGAA 4620

AGGTAGATAC CATCTGAAGA GTTAAGCGTC ATTTTTTTGG cTTTAACACC CGATAGACTA 4680

ACATAAGAAT TTGGtAGATT AGTTGTTAAT TCGTCTACGT TAAAAGCATT CAATGAAACT 4740

GGTGAATGGG TTGAGCCATC AATCACTAAC TTTTTCACAT TATTAGGAAT CTGGACACTC 4800

ACGTTCGGTG TGAAGTCACT AAAAATACCA ATATTAAAGA ATGACAATTC GGATTGGTTT 4860

TGTGTACGAG TGGTATTGAT GGTTGCTGTT ATTTTATTTT TTTCTTCTTT GACATCTAAG 4920

GTGCCTTCAG CTTTTGAATA AAAGGCATTA TTCGGTGATC GTAACCCCAT TTTGTTACTA 4980

TCAGTTTTAT AAATGGTATA AGGTCCTTTG CCTTTTAAGG ATAAGTGAAG TTCTTCGCCT 5040

GCTTGTTTAT TTTTTAGTGT ATAACTTTCT TTTAGCGGTG TGACTGATTT TTCCGCTTGC 5100

TTGTAATAAT AGACGGCGCC AGATCCACCA ATTAGCATCG CTAAGACAGC AATGGATAAG 5160

AGAATCCCTG TTGTTTTTTT CaTTAGATAT CACCTCGTAA AACAGCAATG TTCCATTGGA 5220

AATAACGCTT AAAGACATGT GCTGAAAATT TTAAAATTGG TAGAAACAGT AACCAACCAA 5280

TGATTGATGC ACCAAATAAA AAGAGACTCA TAGATAATTG GAAAAAGGAA GCGGCATTTG 5340

CTTGGGTAAT AACACTATAA ATaCCATAAA TTGGTGAAAA TAAAGTTGCA GCTCCGGCAA 5400

TGCmACAACC GACGTACATT AAGGCAATCC CTAAAATAAC CCAAAACATG AAGrAAAAAT 5460

TTAAACAGAA AATACCAAGC ACTTGGAAAA ATCGGGTTAG TGGCCGATGC TGTGGTCGTT 5520

GATAGTATGG AGAATCATTA TCATAGTAAA ATTCAGATTC TTCTTCATAG GGATGATCAT 5580

AAGCAGGTGT AGCCGCTGGA ATTTCTTGCC AGCCATTATT TTCAATTTTT CGTTCCGTCA 5640

ATTCAATGCC TTGTTCTTTC AAAATTTCTT CTGCGATAAC CCGTGGTTTC CCCAATGATT 5700

TTGCCACTTG TTCTTCTGTT TCGCCTTCCG CAACACGAGC ATCAAATAAC GCCTTGTATT 5760

TCGCAAGAAT AATAGCTTGA TCTTTTGGAT TTAATGCTTT TAAATAAATT TTCAGCTCAA 5820

TAATAAAATG TTCTTTATTC ATTCCGCTCC CTCTTTCTAC ACGATTCATT TCTCATCTTA 5880

AACTATAATT TCTTTTTGAG CGAATAACAA GAGGAATT 5918

(2) INFORMATION FOR SEQ ID NO: 109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2350 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

CATACCCGCT CAATCATGAG GAGACTGGCT AAAAGTAATC CTAGCGTTTG CATCTTTTTA 60

TTTTTATTAA AGAAGTGCCC TGCTAACCAA CCTTGTCCGC CATGAATCAA TAACGAAAAA 120

ACGAACCATT GTGGATAGCC TGACAATAAA TCAATTAAGC CACCGCTTAA AGCTCCGACT 180

GTCAAGCCAC CGATCGGACC AAAAAGTAAA CTAGCAATGT AAATCCCTGC ATCCGCCAAT 240

GTAACAAAGC CATTTGTTGC TGGGACGGGG ATAATAAAGA CCATTGACAC AGCCACTGTT 300

AAAGCGGTAA TTACTGCATA TAGACTAATT TTTCTTGTCA TTATCCGCCA CACCTCCTGC 360

TAGTAATCCT TCATGCCACA CACTTCCAAA AGCTGCATTT AGCTGGACCC CAGAAACAAT 420

GGCTTCATGA ACGTATGTTT TTGCTAAATC AACTGCTTCA ATTAACGGAT ATCCTTTAAC 480

TAAATTCGCA CCAATTGCAG AGGCAAAGGA ACAACCAGCG CCATCAATGG TGTGGGCCAT 540

TACTTTTTCT TTGACCAATT GAATATGTTG TTGTCCATCA TAAAAAAAGT CACGTGCTTC 600

TTGACCAAGC CAACGATTAC CACCTTTAAT CACCACCGTT GCCGCTCCTA AAGAGTTTAA 660

CTCTTCCGCC GCTTGTTGGA CTTCTTCGAC GGTTGTTAAT TTTCTGCCAA TGAGTAATTC 720

TGCTTCTTTT AAATTAGGCG TTATAATCGT TGCGAATGGA AAAAGTTTTT CAATCAATAA 780

CTGCATGTAT CTTTTTTCAT AAACTGAAGT CGTTTCTTTA AAAGCCAAGA CAGGATCCAA 840

CACCACCGGC AGACCTGCGT GCTTTTTCAA AAAGGTTACA ATTTGTAACA TACTTTCTTC 900

ACTGTGAATC AAGCCAATCT TGATGCCGTC TAAACTCCCG GcAGGTATGG AAGCGAGTTG 960

CTCTTTccAC CAAAGAAgTC TCaAGGTCAT GAATCTGAAA ATGATTCTTG GGGGCTGTAG 1020

CAAGGCAAGT TAAAGTTGAT AAGCCAAAGA CTTGATGATT TTCAAACGTT TTTAAATCTG 1080

TTTGAATCCC GCCACCTGAA AAAGGATCGG AACCGCCAAT TGTTAAAATT ATTTTTGTCA 1140

TTTGTTTCGC TCCTTTATAC ACTCAGTTAT CATACAAAGT TTTCTAAAAA GAAAAAAACAG 1200

CCAATTGGCA GATTTTTAGC TCAGCCACTT TTTAAGTCTT TCGTTATCTT TTTCCTTTCC 1260

AAAAAAAGAG TGACCTCATC CGTCAATGAA GTCACTCTTT CTTTGGATTT AACACTTATT 1320

CGTCAATGGC TGTTACATCA ATGTTTAATT CAAACAATTG TAATGGAGAA ACTAAACTTG 1380

GTGCGTTGTT CATTGGGTCA ATTGCTTTGC CGTTTTTAGG GAAGGCAATC ACTTCACGAA 1440

TATTTTCTTC CCCAGCCAAT AGCATTGCTA AGCGGTCTAG CCCTAAAGCA ATCCCTCCAT 1500

GTGGCGGGAA GCCGTAATCT AATGCGTCTA GCAAGAAACC AAATTGGTCT TGTGCTTCTT 1560

CTTTTGTAAA ACCTAACGTT TCAAACATTT TTTCTTGTAA TTCTCGTGTA TGGATCCGTA 1620

ATGAGCCGCC GCCTAATTCG TAGCCGTTTA AGACAACATC GTAAGCTTCA GCATACACAG 1680

AGGCTGGGTC TGTTGCTAAG CGGTCCACAT CCTCTGCTTT TGGTTGTGTA AATGGATGAT 1740

GAGCAGAAAC ATAGCGACCA GCTTCTTCAT CATATTCAAA CAACGGCCAA TCCACAATCC 1800

ACAAGAAATT GAATTTCGAT TCATCAATTA AACCTAATTC TTTACCTAAA CGTGTACGTA 1860

CCGCACCTAA AGCAGCGGCT ACGATTTCAG GTTTATCGGC ACCAAACATT AAAATATCAC 1920

CGACTTCTGC ATTGGTAGCA GCAATTAATT CATCACTAAC TTCTGTTAAG AATTTAGCAA 1980

TCGGTCCTTT TAAGCCATCT TCTTCAACTT TTAACCAAGC TAAGCCTTTT GCGCCAAATT 2040

GACTCACATA CTTACCTAAG TTATCCATAT CTTTACGAGA ATACTTATCT GCGGCGCyTT 2100

TTGCATTCAA AGCCTTCACA TGTCCACCAT TTTCCAACGC TGCTTGGaAA ACTTTGAAGT 2160

CTACGTCTtT CACTACCTCT GCAACATCAA TTAATTCCAT AGCAAAACGA GTGTCGGGTT 2220

TATCGyTACC ATAGCGTGCC ATTGCTTCGT CATAGCTCAT TCGTGGAAAC GGAACAGArA 2280

TCTCAATTCC TTTTGTCTCT TTCATCACTT CAGCCAACAT ATTTTCTGTA TATGTTTGAA 2340

TTTCTTCTGG 2350

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21591 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

TGAATTGGCC TCGATGACAG GTCTTCTACG AGGATACGTG CAACAATTAG ACGTCCCTGA 60

ACAGCGTGCA CTAACAGAAG AACTAGCCAA AATTTGCGAG TTAATCTACC ATGTCAATCC 120

AACGACCCGA ACAAAATTAA CAGTCACAGA GGACGAAATA GCCTGGCTTT TGGAACGTGT 180

CAATGCTATG AATGAACTGA CATACGAAGA AAATCGTCCT TTTGTTTTAC CGATGGGCAC 240

AATCTGTTCC TCTTATGCAC ACATTTTAAG AGCCAAAGCC AAAGATATTG TTCGTTTACT 300

TTATCGCATG GACTATGGTG GTAAAAAAAT TGATCCGCAG CTGTATGATG TAGTCAACTT 360

ACTCTCAGGT TACTTTTTCA TGTTAGCCTT GTACCTAAAT CAGCTCGAAA ATGGTGAAGA 420

AGTTCCCTTT GTTTCTCGCA ACTATTCTAT TTAAAAAGTA CACCGACCGC TAACGACGAA 480

TTGCTAGCGG TCGGTGTACT TTTATACAGG TAAGTACGCT TCTCCTTTTG CAAATAGTTT 540

AGCAGTACCA CCAATAATGA CTCGATTTTC TTTCACTTCA CAAGTTAAAA AGCCGCCTCT 600

TGGGGACACT TGATAGGCCG AAAGGGTCGT TTGATTGAGC CGTTTTCCCC AATAAGGAAT 660

TAAATTGGCG TGCGCAGAAC CACAAACTGG GTCTTCGTTA ATTCTTAGTT TCGGGAAAAA 720

AGTTCGTGAA ACAAAATCCA CCGAGTCTCC TGATGCCGTT ACAATGACTC CTACTCCTAA 780

ATCCAATGCT TTTAATGCAG AAAAATCTGG TGTGATTTTT GCCACTGTCT CTTCGTCTTT 840

TAAGACAAAA AATAAATCGC GACCCAAATA AGCTTCATAA ATTTTCGTGC CTAAAGCGGC 900

TTCATACTCT GGCAAAATTG GAATCCTTTC CGGTAGGATA TAAGGAAAAT CTAGATAATA 960

GTATTCCTCT TTTTTGGTCA CTGCGAGCGG TCCACTTTGA CTAGTAAAAT GGAGTGTTTC 1020

TTCCGCAACG GAATAATAGT TAAACAACAC AAAAGCCGTC GCTAATGTAG CATGTCCACA 1080

GAGATCGATT TCTCGTTCCG GCGTAAACCA GCGCAGGGCG TAACTTTGTC CCTCCTTGAC 1140

CGTAAAAGCA GTCTCTGACA AATTATTTTC AATCGCAATG TTTTGCATCA CCGCTTCTGG 1200

CAACCATTTT TCCAAAACAT AAACAGCGGC GGGATTGCCT TTAAAAACTT CCTCCGCGAA 1260

AGCATCGACA ATATAATAGG GATAACTCAT TTTAATTCTC TCCTTGTATT TTTTGATAAA 1320

TGTATTCCGC GCAACACAAA TCCTCAACTG CTAAGCCAAC CGCATCAAAA ATCGTGCGGC 1380

CTTTCTGCTC CGCCTTTAGT TTAACTTTGC CAGTGACTAA TTCGCCGAGA GAACCTACAA 1440

TTTTTTCAGC TGAAAAGGCA CCTTCTGAAA TCGGAATTAA AAGATCCCCG CTTTCTTTTA 1500

ATGCAGCCGC ATACTCATCT ACAAACACTT GACTGTTTTT AACAAGGTCG CTTGTTAATT 1560

CTCTGGAAGT CGGCGTGAAT GTACCAATTG CATTGATATG GGCATTCGGC AAAACGTCTT 1620

TTTCTTCTAA AAATGCGTCC TGACTGGAAG TTAGTGTACA GATAATTTCC GCTTCCTGGA 1680

CGGCTTCTTG GACAGAAGAA CAACCAATAA AGGTACATTC GGGGTAATTG TGCTGTTGCT 1740

TAGCGATTAA TTTTCGTGCA CGTTCTGCTG TGCGGTCATA TATAAAAACC TTTTCAATCG 1800

CCCGAACGCT TTGTAGCGCT GCGAGATGAG ATGAAGCTTG TTGACCTGCA CCAATCAAAG 1860

CGATACTTTT TGCCTTTGGA GCTGCTAAAT AATCGGTTGC CAAAGCAGAA ACAGCTGCCG 1920

TGCGAAsCCA CGTAATCGCA TTGGCGTCCA TCATCGCGAC TGGTCTTCCA GTAGAACTAT 1980

CAAACAATAA AATTTGACCA AGATGCGAGG CCAAATTTTT TTGTGGATTT TCTGGAAAAG 2040

CTGTAATGAT TTTTGCCCCA AAAAAGCGAT TTTTTCCTAA GTAAGCAGGC ATTAATGCAA 2100

AAAGATTTTG ATTTTTCCCA TCTGGTAAAA CCTCTACATG GCGCTCACTT TGCGAAATTT 2160

TTCCTTTTTG CAACTCTGCA AAGCAACGTT TCATCACTTG AATTGCTTCG GAAAACGACA 2220

ATTGTTGAAT AACCCTCTCA TAAGAAATAA TGTCCATCTG AATTCCTCCC TTTGCGCTAT 2280

CATAAACCAG ACAAACAATA CACTCAATGC TATAATTGTA TGGAATACAA TTTTATAGAA 2340

AGGAAGTTCC CCTCATGCCG ATTAATTCCT ATGAGAACTA TGTGTTGTCA TGGCGCCCTG 2400

TTAAAGAACG TTTAACAAGA CCTTATTACC AATCATTGGT TCAACAATTA GAAGCGGATA 2460

TCTTATCTGG TGCGCTTCAA AAAAATGTGA AATTGCCTTC GCAAAGAGAA TTAGCTGATT 2520

ACTTGGATCT GAATTTCACG ACGATTGGTC AAGCCTATAA ACACGCTATG GAAAAAGGCT 2580

TACTTTATAC CAACATCGGT TGCGGCACTT TCGTTTCTCC CAACGCGTTT CACTCAATTA 2640

CAATCTCTAC CAACCAAGTA GCGGACCATC TAATTGATTT AGGGCTGGTT AGTTCCTTTG 2700

ATATGTGTAA TCAGCGGATT TTACCTTTTA TTGAATCCGT TAGTAAAAAT GCGGCCCTTA 2760

ACAGTTTGCT AAATTATCGC GATCCTTTAG GTACGCACTT TCAACGAGCA ACCGCTGCCG 2820

AATGGCTTCA GACACAAGGC GTTCGGACCA ATGCCGAAGA AGTTGCCATT GTATCTGGTG 2880

TCCAGAATGG ACTGGCCGTG ACGTTAGCCG CCGCTTTTTC TCCAGGTCAG CGGATTGCCG 2940

TAGATCGATA CACGTATTCA AATTTTATTG AACTCGCCCA GCTTTATCAT TTAGAAATCG 3000

TCCCGATTGG CTATGATTCC GAGGGGATGG ACCCAGAGCA TTTACTGCAA GAATGCAAAA 3060

AGAAAAAAAT ACATGGTATT TTCTTAATGC CCGCATGTAA TAACCCGATT GGTTTCCAAA 3120

TGTCTAGCGC CCGCCGAATG ACTTTAGCTG AAATCATTCA GCAAGAACAC TTATGGGTAA 3180

TTGAAGATGA CATTCATTCT TTTTTAACAA CTTATGCACA GCAAGCAGTT TTACCAACGT 3240

TTCAAGAACT CTTACCTCAA CAAACGATTT ACTTAGCGGG GATGACAAAA TTTGTCTGTA 3300

CAGGcTTACG CATTGCGTAC CTTGTTTTCC CTCCGCTTCT TCGCCAAGAA ATAGAGCGTG 3360

CGATTTTTAA CATTAACGTT AAAACTTCTG GCTTTGATGC TGAAATTGTC ACCCAAGTCT 3420

TACGCTCACC AGTTGCAAAA GAATTAGTGA TAGAAAAATT GGCTTTAACA AAACAAGCCA 3480

ATGACTTATT TGATACCATT TTTGGGTTAG CTAGACCTAG CAATCCCTTG CCGTACTACC 3540

GAACTATTCC CCTTTCTACC GAAAAAACGG CGCCGCAAAT TGAACAAGAA TTTCTCCAAA 3600

ATGGGGTACG TCTTTTTCAC TCGAGCCGCT TTACTGTGCA AAATCAGCCA GATGCTTTTT 3660

TACGAATTTC CTTGGCCTCC AATCAGCTTG AGGTATTGGC GAAAGGATTA ACCATTATCC 3720

AAGAACTCTT ACCGACATTA AACGAAAAGA AAGGACATTC CCTATGAATT CATTGCCAAT 3780

ATTTGAAAAA GCAGTGGGCA CGATTACCCT GCTTTCCGAA GAAAAACAAA ATGCCGAATA 3840

TGAAGGCTTT TTATGTCAAG TTAATGAATC GCAACAACTT ATCCGCAGTC GCTTAGCCAA 3900

AAAAACGCCG AAAAAAGAAG GCTATTTTGT AGCCTTCTGG GAAAAAAATC AACAAAATCA 3960

GAACGAAGCC TTTGATGCAA CGGAAGCCCC TGAAATGTTA GCAATTGTCA TTGCTGATCA 4020

AGAAAAGCAA GGTCTCTTTC TTCTGCCAAA AGAATGTTTA ATCCAACAAA AAATTCTCAA 4080

AACGCATCAA CAAAAAGGCA AAATGGCGGC CCGTTTTTAT CCTTCTTGGT GCCAAAACCT 4140

TAATCAAACC GCCAAAAAAA CACAAAAATG GCAACTAACT TACTTTACTG ATTTATCTAA 4200

ATACTAAAAA CGTCAGCTGG CTTCAGAAGC CAGCTGACGT TTGTTTCTTC TACTGTTTAA 4260

CCAATTGGTG GCGTACCTTC TGCGTTCCCA GCAATCGCTT CGATTTGAAT CTGAGCACCT 4320

TTTGGTAAAT CCGTCACACC AATCACTCTT CTTGCAGGCG TGCCTTCAGG GAAGTACGTT 4380

TGGTAGACGT CATCTACTGC CGCCAGCTCT TCAATCTCTT TCACAAAAAT GTTCACTTTC 4440

ACCAGATCAG CTAAGCTGTG ATCCACACTT TCAATAATCG CTTTAATATT TTCTAAACAT 4500

TGCGTCGTTT GTTCCTTGAT TCCGCCAGCG ACTAACGCAT TTGTTTTAGG GTCCAATGGT 4560

AGCTGAGCTG ATAGATGGTT GTAATGAGAA AAAGCAACAG TTTGTGTCGA AAATGGACAC 4620

TTCGGTGCGT GCTCTGTATT ATTTGCTTCA ATAATTAATC CATGTCGATC TTCAATCGCT 4680

TGTGGTGGCG TACCATCCCC ATGAGAAACC ACCGCTTCTA CTTGCACCGC CGCGCCCATT 4740

GGTAAATCAG CAACAGCAAT CACTGTTCGT GCTGGCATAT AATTGACCGT GCGCGCAATT 4800

CCTGAATCTG GGAAAAAAGC GGCATGTACT TGATTGACAG CTTCCAGCTT AGATAAATCT 4860

TTTAAATAAA TGTTGATCTT CACAATATCA TCAAAAGGCA CATCAATACT GGTTAAAATG 4920

GCTTTTATGT TTTTCAAACA TTGCTTTGTT TGCGTTTTAA CACAGCCAGC CACAACTCTG 4980

CCCGTTTGAG GATCTATAGG CAATTGTGCA GACAGATTGT TATAGTGAGA AAACGCGACC 5040

GTTTGACTGG CTAGAGCATC AATTGGTGCC TGAACAGTCT GATTTGTTAA TTTAATTAAA 5100

TCCCCAGCTT GAGGTGCATT AGGGATAGTT CCTTCCCCAT TTGAAACAAG TGCCTCCATT 5160

TGAACAGAAG CgCcTAACGG CAAATCTGCG ACGGCGACCA CTGTTTTAGC TGGATAATAA 5220

TGGGTAAAAA AGGCCGCATA GACCTCCTCA ACTGCGGCAG CATCCCGCAT ATCTTTGACA 5280

AAAATGGTCA TTCTGACAAT GTCACTCATC ACATGGTTGA TACTTTCAAC AACCGCTTGG 5340

ATATTTTTCA AACATTGTGT CGCTTGGGCC CGAATGCCAC CTTCTACCAA TTGTCCAGAG 5400

GTTGGCTCAA TAGGTAATTG TGCCGACAAA TGATTATAGT GGGAAAATGC GACTGTTTGG 5460

GCTGCGAGAT GATGCTTTGG TGCACGCTCT GTATTTCTTG CTAAGACGAT ATTTTCATTA 5520

CTCATAACTC TACTCCCTTT TTTAGCATTC CAACAAAAAT GATTCCGCTT ACATTTTTAA 5580

GCATTAAACT TCCAACAAGA TCCTTTTTAT TTTCCTGCTT TTCTATTAAT GGTGATAGCT 5640

TTCCACCACC ACTTCTTTCA GTCTACAATA AGAACAGGAA CATTTCAATA AGGCATTCCT 5700

AAATTGATTG TTGGCCCGCA AACAAATCTT CTTCAGTTAA AGAAGACTGG CACGAATCAA 5760

GCTTGATTCG TGCCAGTCTT TATAGaTAAC TCGTTAGCTT TAACTAGTTT TATttAaCCA 5820

CCATATACTC ACGATAATTT TCCGAGAACG TATCCATTAG CGTATAAATT CGTTCAAAAT 5880

CATAGTCACG ACGAGGGTCA TCATACCAAA CACATTGAAT ACTTCCAATA ATTGGAAGAT 5940

TCGGGCTACC TGCTGGCACA TCTGTAAATG CTTTTTTCGA CATATTTGCT AAAGCATCAT 6000

CATATTGATA AATGCCAGAA TCAACATTGC CAGCGACCCA ATACCACGCA TCATTGGTAT 6060

TAAAAATTTT GTGTCCTTTT TGAACAAAGT ACTCAGGTTT GGCTACGTCA TATCCGCTCC 6120

AACCGGCTGT CCAATAAGAA ATAATAATCT CTGGGTCAAA TGTACCGAAA GAATCGTCGC 6180

TGTTGTAATA AATGCCATCA TTGAAGCTCA TTGGCTGCAT GCCCGCATCT TTAATTATTT 6240

TAGCTAAATC ATTCGCATAA GCGACAAAAT CTTTGTAGCG CCCAGAAGAT TGCAGTTTCG 6300

CCCAACCGCC TGTATCGACA TCATTtGCAT ACTCATCGCC GCCAAAATTG AAAATTTCAC 6360

TATGCGCGGA AAAATAAGCC ACGTACTTAC TGATAATCGC CTTTGTGAAA TTCACTGCCT 6420

TTTGGTTCCC TAAATCTACT GTTCGTTTTG AGCCATCAAA AGCGGGATTT TTAATCGCTA 6480

GTTTTTCCAT GGCCACTAAC AAGGCATCCA TATGTCCTGG ACTGTTAATC ACGGGAATGA 6540

TGTTGATGTT GCGCGCTTTC GCAAAAGCCA ACAACCGATC CATTTCTTTT TGTGTTAACG 6600

CGTTGCCATT AGGATCATTG TAATATGCGT TGTTGCCTCG TTGGATTGCT TTTGAAACCC 6660

GGTTGTGGTT GTATTTTTTA CCATTCACAT TGACCGACAT ATCATCCAAG ATAAACCGTA 6720

AGCCATCATT TCCTAAAATT AATTGGACAT CTGTGTACCC ATTTTGACTA GCTTTGGCCA 6780

CTAATTCTTC CAGTTGCTCC ACGGAAAAAT ATTTTCTTCC CGCATCAATG GAAAAGACAC 6840

TTTTTAAAGG TTGTGTTTGT TTCTCTTTTA AATTATCCAA AGCTGTTTGG ACCTCCGTTT 6900

GCCAAGTAGT CACTTGTGCG GCTGTATACG TTTTTCCCGC AAGTTCTGTG GCGACTGTCT 6960

CTTTTGTTGT TTGTAGTGCT TGAATCGTTT CTGGTGTATA GCTTTCTGCC TTTTCAATGG 7020

CTGTAACTGC TTGATCCAAC GCACGATATA AAGTGGTTAA GTCAATGGCT AACAGGCCTT 7080

TTTCTAATTC AGCATCATAG GTGTTTGAAA TCCCTTGGCC ATAATCGCCA CCTAAAATAG 7140

CTTTATTCAC ATCATAAAGC GTTGTTGCTT GTGTCACTGC TTCTGCGGCT TTATTTTGTG 7200

CTTCAAGGTC TGTGTTGGCG TAgCTAATTC GTTTTAAATA GTGTTGCGCA GCcGCTTTCA 7260

TCTCTGCAAG AGAAACGCCT TTACTCTCTG CAATGGCCGT TTTTGTCCAA AGTAAATTAG 7320

AAGGTTTTAT CTGATTTAAA TCGTCTTCAT TATAGGTGCG ACCATCGCGA TCTAAGGCAT 7380

AGAGGAACAT GCCCCCTAAA TTATTTTCAT AAGAATAACG AGCTACTTTA TACATGTTAC 7440

TTTCCATATA CGGCTCTTTT GTGTCATACC AGCGGTTGCG ATCTTGCTCT TCTGGGAAAG 7500

TTAATCCTGG AACAAATCGT TCCCCATTGA GAACCGGACT TAAATTATTT AATGCTCGTT 7560

GCGTGCGTTG GTCATCGCTG CCATACTGTT GATACGCGAG AAAATCGAAA CAGTCACTGA 7620

CATCTTGTAA AGGTGGTAAA TATTCTGCAT TGGTATCATA CAAAAATGGA CGATCCGTTC 7680

CCGACTTCGG TCCAATGTAT TTTGATAATG CACGAATGAC ACCATTAGAT AGAACAATAT 7740

CTTTTTCACT TGGACGAGTT TCCATGTCAA TATCTAACCC ATCAATTCCT AAATCATCGA 7800

CAAATTTGGT TAACAACTCT TTCGCATAAG CATCAAATTC TGCTTCTGTA GGCGTTGTTC 7860

CTGCATAAGG AACTTTTAAT AGCTCGCTGT AATCAATCCC ACGAACTAAA CGAACACCTC 7920

GTGCATGCAG TGctGGCGCA TACTCATTTT TTAACGTATC ATAAAATGGC TGTGCAAGTG 7980

CTTCTTGTCC TTTTGGCACA TAACTAAAGA CATTCACAAT GTCAATACCA TAAGGAATAT 8040

CGTGCATCGT TAGCCAATTT TCATCTGGCA ATGTTGTATT AACTCCTTGC ATCGTTTTGT 8100

CACGCCAAGC CCGATAATAA ACCATAAAAT GTTTCTCTTG GGTCGGCTCA TTGGCTTGGC 8160

TGATCATTGG TTGATGAAAC AGACTAAATA AGATAAACGC CAAGCCAACC AATAACGAAA 8220

GACCCAGCCA ACTGACAATT GTTTCTCTGT TTTCTCCTTT TCTTGTCGAT AAATTGTTCC 8280

CCACTTTGAA CACCATTCCT TTCTGCACTC CGTTCATTTT TGTGTATCCA CCTAAAAAGA 8340

AAGCGCATAC ACAACCATCA ATAGAATAAC AAAAAATAAC GAATTTGCAA AATAGAAACT 8400

GGAAAAATGA GTATTTTCTA TAAAAATCTT TTCCCGCCTA AAATCACTTT TGTTAGACGG 8460

CGTTTAAATG GTAAATTTTG AGACTTGCAA GAAAAACAGG CCTATTTTTC TTGCAAGTCc 8520

CCTTTACTTA GTCGTGATCA ACCAATAAAT AATCACTGTT TGAATTTTTT GGGATCTACT 8580

CGTTTGAAAT TTATCTCATC ATCAATAACT GCTCTATGCT TTAAATcTAA ATGATATCGC 8640

ATATCATGAT TTATATTTGA ATTTTTTTTC AGATTATCAT TATATGTTTC TAAAAATAAT 8700

TTTAAAACTT TTTCAGGTTG CCTTATACCT AATACTTTTC TTGATATTTC TAAACCATTT 8760

ACCTTTTTTA TAAACCATTT TTGATTGTAT TTCATTATAG ATGACAAATT AAAATAAGCT 8820

TGGAATACAG CTCCCGCATC TTTAATTTGA GTATTTTTTT TCAATTCATT GTATTCATTC 8880

AATAAATTCT TTAGCTCTTC ATAATTGTAA AAACTGAACA TTCCATTCTT AACTTCTGTC 8940

ATCATTTTAT TCTGCCTAGT CCAAGATCCC ATGCATACAG AAACTAACTC AACTAAGGGA 9000

ATTTTGTTAT TCTCTATATG TAAATCTTTT AACTTTTTGT ATTCTTCATT TCCTAAAATT 9060

ACGTATTTAT TAATATAGTC TATTACCCGC CACTTACATG CTGTATTATT TATTTTAATA 9120

ATGTCATTCA TTTCTTTGTT TACGGAAACA TAGTATGGCA CGGGTAAGCC ATATTTTCTT 9180

GCAATACTAT ATCTATGTTG CCCATCGATG ATCTGCATTG TGGAGTTTAC TATAATGGGT 9240

CGAATAATTC CTTTTTGTAG AATAGAATTT TCCAGTTTCT TGTTCAGTGT AACTTCACGA 9300

TTACTTCTTA AAAAGCTGAA CATTTCATAT TCTTCCGTTA TATAAATATC ACTCATTTCT 9360

CTATCCATTT TTTATCTCCT TTATTAAAAA CATACTTTGA ATATCTCTTA TTACTGGCTT 9420

CCCATCAATT TCGATTTCAT ATTTAATTGC TTTGTTTGAA TTTCTATGCA ATTTATAATT 9480

ATATGCTTTC AAAAACTCTT CCAAAACGAC TCCTTGGTGA AAAATTCCTT CTATAACATT 9540

TTTAGTTTTT ATTTTAAATC CATTAATTAA CCTATAATAA TCGAAATCTT TTACAATAAA 9600

TAATTCTACA TAAGCAAAAA ATGTTTGCTG TGAACTTCTC AAGTTCAACT CTTCACAAAA 9660

GTTTCCGTAA GAATCTAATA ATCTTATAAA TTCCTTGTAG TTATAAAATG AAAATTGACC 9720

TGTTTTTACT AATTCTTGTG AATTACCATC ATAGTGTTTA TTTCCCATAG CCAAAGATAT 9780

TAAAGAAGAG GGTGAGATAT TTTTATACTT CTTTACAATA AACTGTAGTT TTTTATATTC 9840

GTTCATACCA TCCATCGAAT ATTTCGTATG AAATCTCCAA TTTCCAGTCT TAATGTCGAA 9900

TTAAGCTCAA TTACATCATC TAAGTTCAGA CCTTTCGAAA TAATATAAGG AATGGACTTA 9960

TTCAATTTTC TGGCTATCAA AAAACGAGTT TGTCCATCAA TAATTTCAAA TTTCTCATTT 10020

ACATCTATAG GAACTAAAAT ACCTGATTTT TCTATAGATT CTATTAAACT ACGCTTTTCT 10080

TTAATTGCTC TGTTTCCCTT TAAAAATTTA AATTtTGAAT AATCCTTAGT TTCATACAAT 10140

TTACCAATCA ACTTTCCCAT ATTCTTCACT CCTTATAAAA TATATTTTTA ATTTATATAT 10200

TTTATAATAT CGTCTAAAAA AAAGAAATGC AATTCTTTTT TAATCAATAT AACACCAACC 10260

ACACATTAAA ATAACATTTT GGTTAACCAT CCATGTTTAT TAATTCGGCG ATTTTCACAA 10320

CAATTTTCCC TTGCGCTCGT CCTGTCAATG AATAATCTAC CGCTTCTTGA ATTTGGCTAA 10380

AAGGAACGAC TCGGTCAATT ATCGGCTGTA ATTTTCCCTG TTCAATAAAC TCAGTGAGTA 10440

AAGCTAATTG CTCGCCATCT GGACGCATAA ACAAAAAATG ATAACTAACA TCTGTGGCTT 10500

GCTCCAATCG ATGAATCTTG CGGGTAGCTA TTTTAAAGGC CCATTGTTTC CAAAGCGGCA 10560

AGCCATACTC TTTAGCAAAA CGTTCGTTGG GAATGCCTGA CAATGTAACA ACTTTTCCCT 10620

GAGGTTTAAC CACTGAGAAA GCTTTTTCTA AGATTGTCCC CCCCATTGTA TCAAACACAT 10680

AATCATAGTC GGATAAAACT TCTTCAAAAT TTTGTGTCCG ATAGTCAATC ACTTCATCTG 10740

CTCCCAACGC TTGAACCCAT TCTTTATTTT TACTACTCGT TGTGGTGGCA ACGTAAGCGC 10800

CTGCTAGTTT TGCTAATTGA ATCGCAATGG TTCCAATCCC TCCTGAACCT GCTTGAATCA 10860

GGACTTTCTG GCCTGGCTGT ACATTCATAA TATCATGTAG CGCTTGATAA CTTGTCAAAC 10920

CGACTAACGG AATGGCCGCA GCTTCTTCAA ACGTTAAATT TTTCGGCTTC ATAGCTACAG 10980

CCGCTTGATC CACTGCAATA TATTCAGCAA AAGTTCCCAC GCGATTTTTG GGTACTCGCC 11040

CGTAAACTGC ATCGCCCAGC CTAAAATTTT GAACATTTTT TCCAACAGAG ACAACAATCC 11100

CAGCAAAATC ACTGCCTAAA ATCAGTGGCA TTTGATAATC TAAAAGCATC TTTACTTTGC 11160

CATCTTTTGT TTTTAAATCA ATCGGGTTGA TACTGGCAGC GATAATTTTG ACCAAAACAT 11220

CATTGTCATG AATGGTGGGT AAnGGCACCT CCTCGATTGC TAATTCtTTT TGTCCGTATT 11280

TATGAATTAA AGCGGCTTTC ATCATTGAAT CTTTCCTCTC TTTCTATCCC TTTTATCGGT 11340

TTAAAAGTTT AAACCTTTAA TCAATTAGTA TCTTAACGAA GAATGGGATC CACGTCAAGT 11400

TTCAGCCCTC AGCGTATTTT ATTGACAAAA AACACGAAAA ATGGTTTAAT CGTTTAAAAA 11460

TACGAACCAA AGAGGCGAAA GAATATGAAA GAATTAGAAG AAATTCAAAC AGAATTTACT 11520

GCGCTTAGTG ATTTTTTAAT CGCATTGGGT GATGAAAAAC GCCAAGCAAT TATTATTGCA 11580

CTGTTAGAAG ACCAAGCCTG TGTTGGCCGT CGTGTCACTG AGTTAACAGA GGCGACACAG 11640

CTCTCACGCC CCGCTGTTTC TCACCACTTA AAAATTTTGA AACAAGCCAA ATTAATTGAT 11700

TGTCGCAGTG AAGGAACTAA AAATTATTAC TCACTTTCTC ACGATACCAC CAAAATTGAG 11760

CAACTCCAAC AACTATTGAC CCACATTACG TCGATTATGC CAGGGAGGAA GCAAGCATGA 11820

AAAAAATTCT CGTGGTCCTC ACAAACGTCT CCCGTTATCA CGGCACCGAA GAACCAACGG 11880

GCCTCTGGCT AGGTGAAGCC ACGGAGTTCG TGGAAGAAGT AACCAAAGCA GGATTTTCAG 11940

TTGACTATGT TAGTCCCCAA GGAGGCTATG TCCCTTTGGA TCCCCGGAGT ATGAAATATG 12000

TAGATTCATC CATTATGGCA GTTTATGAAA GTGCTGATTT TCAAGAACGA GCGCTAGCAC 12060

ATTCCCTCTC TCTTGAAGAA ATTGAGCCAG ATGAATACGC AGCAATTTAC TTTACAGGAG 12120

GCCATGGCGT TATGTGGGAC TTTCCTGATA ATCCTCAGTT GCAAGCAATC ACTCGTAGCC 12180

TTTACGAACA GGGCAAATAC CTAACTTCTG TTTGTCACGG AATTGCTGGC CTTTTAAATG 12240

TCAAAGATGC GTCAGGAAAT TATTTGATTG CAGATAAAAC CATTACTGGC TTCACTACTG 12300

CAGAAGAAAT TTTAGCCATG AAAAAAGCAG TTGTTCCTTT CTTAAACGAA ACAATTGCCA 12360

AAGAACATGG TGCTACTTTT GCGAAAAAGC GTCCCTATAA AGAATTTGCG TATCAAGATG 12420

ACCGTATAAT CACTGGGCAA AATCCTTTTT CTGTCCGCGC TGTCGCAAGA TTGTTAATTC 12480

AACAACTATC AAAATGAAAC AACAGCGACT ACTTTTAGGT AGTCGCTGTT GTTTCATTGC 12540

TTCTTCTTTT TTCCGAAAGC CGAAGCCCTA ACATAAAAAA TTTTTGATTA GCTAAAGAGT 12600

AACATCGCGC CAATTAAGAT AATTAGGTTA ATTAAGCCTA AGACAATCAT TGGTTTGGCC 12660

ATGAATTTTG TCCAAGTGGC ATATTGAACT TTGGAAATTT CCAAACCACC CATTACGACA 12720

CCAGATGTTG GCGTGATCAA ATTAATCAAG CCACAACCTG AAGCAAAAAT CATAACCATC 12780

ACTTCTGGAC TTAAACCAAT GTTTTTAGcT AACGCACCCA TTACAGGAAT AGAAACGTAC 12840

GCTAAACCAG AAGTTGAAGG AATTAGGAAA GATAAACCTA AGTATAAAAC AAAGGCACCG 12900

ATGACAAATA AGACTGGGGA GAGTCCTTGT AATAAGCCTG CCGCTTTATC AAGGATAAAT 12960

GAATCTAAAT ACGTTTTAGA CATCAAAACA GATGCCCCGC GAGCAACTAC GATAATTAAA 13020

ACAACGGATA ACATATCTTT GGCACCGTCG ATAAATGCTT CAATGGTTTC TTTTTCAGTA 13080

AAGCGACCAA TCAAGGCAGC AATGATCCCT AAAAGGAAGA ACCACATCGC TAAGTCGCCA 13140

AAGTACCAGT CACCGAAACT GTTGCCTGTT AAAACAGACG TCCAGCCTTT GAAAATCGTC 13200

ACTCCAAAGT CGTTCCATGG AATTAAGGAA ACAATCATAA TGACAAAACA GAAAGCGAAT 13260

ACCATTAAGA TTTTTTTATG ACGATTCGTA AATGGCATTT CTTTCGACGC TGCTTGTCCA 13320

TACGTTTTTT CCATATCCTC TTGTTCTTGT AAAGATAAAA TTGTTGAACC TTTGTCAGCC 13380

TTCACTTTTT TCGCATAACG CATCACAATG AAAATAGAAT AGCTAGTAGA GGCCGCCCAC 13440

AAAATAGCAC CGACAATTAA AATAATTCCT GTATTCGGCT GAATGCCAAT ACCTCGCAAC 13500

GCATCCATCG CTACCCCTGT GGAGAAAGGG TTAACCGTTG AACCGATAAC ACCTGATCCT 13560

GCACCCAATA GCACAGTACC CACTGCGACA AACGTATCAA ACCCAGCGGC GACCATGGTA 13620

GCACTCAATA AACCATAAAA AGGAATTGTT TCTTCTGCCA TTCCATAAGT CGACCCACCA 13680

ATAGAAAATA AAATCATTAA AATCGGAATG ATTTTCAATT CGTTCCCTTT TAACTTTTTA 13740

ACGACTGATT GAATCCCTGC TTCAAGAGCC CCAGTTTGCG TCACAATATT TAAAAAGCCA 13800

CCGAGTACTA AAATAAAAAT ATTAATTTCA ATGGCATCTT TAAACCCATT GAATGGAGCC 13860

ATGATAATAT CCGATAATTT GGCCCCTACC ACGTGGTCAA CCATTTTGCC ATCAATTTCT 13920

CTTGGGGTAA ATGGCTGGCC ATTTAGCAAG ACCGTTACGA TACCTAAAAC AATTAAAATA 13980

ATAAATAAAA TACTAAAAGA CGAAAGCTGT TTTTTGTTTT TCTTCTTTTT TTCTGCCATA 14040

TCCTTTTATA CAGGAACCTA CTTTCCTGTA TCTCCCTTCT CTATGATTAA TTACTCAATC 14100

GATCAATAAA GAATGGCGTA TCTAGAAAAG TTAAGAATTT ATGACTGTAG CTAATACGGT 14160

TCATTTGGAC TAATTTTTTC ATAACACTGC TTGCTGTTTC CCGCGTAGTT CCACTCATTC 14220

GTGCAATGTC ATTAATGGTA ATTGGACAAG TAATTGCAAC GCGTTCTCTG TTGTCAATTC 14280

GTTCTCCTAA ATCATTTAAT AACATCGCCA GTGATGTCAC AATCCGTTCA CATGCATCAT 14340

TGGTTGTTCC TTTTTGGATT TTGATCATAC TTTCTTTTAA AAGGACCGAT TGTTCTTGAA 14400

TCCAATGAAT CAATTGCTGG GGATTATTTT GGAGAAAACG CTCAAACAGT TTAACTGGCA 14460

CAAACACAAC CTCAATATCC GTATAGGCAA TCGCGGAGTC GTAATGAACC TCATCTTGAA 14520

ACAGTCCGAT TCTCGGAAAA AGAGCCTGCT TTTTCATAAA ATGCAGATAG GAAAAATTAC 14580

CACTGCGATC ATTTTTTTCA ATTTTAATGA CACCTGAAAC CAATAAAATT AAATGATTTC 14640

TCGGATCACC TGGATAAAAA AGGACTTGGC CTTTTTTATA GGAGCGAACG AAACTCTCGT 14700

TGCGCAATTC TGTTAGTTCT GATTCAGAAA AGTATTGAAA ATCTTGATTA TTTTTAGTGA 14760

TTATTTTCTC TATCATCATC TTAATTCTCC TAAATACTAT ACTTTATTTT GTCACAACCG 14820

```
TTCCTACGAT TTCATCGCCG CTCATACTTC CTAAGTTTTC CAAAGAAGTG ATAATTGCTT    14880
GTTTATTGGG TTGGCTTTCT ACAAACTGAA TGGCGGCTTC AATTTTCGGT AACATACTAC    14940
CTGGTGCAAA ATGACCCGCT TGTTTGTATT CTTCCAACTC TGCTACAGTG ACATTGGTTA    15000
ATTGTTTTTC ATCTGGTTTG CCATAATTAA TACACACATA ATCCACGCCA GTTAAAATAA    15060
CTAGTGCATC CGCATCCACT AATTCTGCCA GTTTTTCAGA AGCAAAATCT TTATCGATTA    15120
CTGCTTCGAC ACCTTTTAAT TCTTGTCCTA CGACAGGGAT GCCGCCACCA CCACATGAAA    15180
TGGTAATTAT ATCATTTTTT ATTAAGGTAT TAATAGTCTC AGCCTCGTGG ATGTCAATTG    15240
GCTTAGGACT TGGAACGACT TTGCGCCAGC CACGTCCTGC ATCTTCTTTA AAAATAGCAC    15300
CTGCTTGCAT TGCTTCTTTG GCTTCAGCTT CTGTTAAAAA TGGACCGATC GGTTTTGTTG    15360
GATTTTTGAA TGCCTCATCT GCTGGATCTA CGACCACCTG TGTTAAAACA GTAGCCACTT    15420
GTTTCTTGAT TCCTGCTTTG TTTAATTCTT GATTAAGCGC ATTGGACAAC CAGTAGCCGA    15480
TACTTCCTTG TGTCATAGCA ACACACGTAT CTAGCGGCAT CGCTGGATTT TTTTCAGAAT    15540
CAGCTGCTTG TTGTTGCAAT AATAGATTTC CAACTTGTGG ACCATTGCCA TGTGAAACAA    15600
TCAACCGATG CCCTTGTTTA ATTAAATGAA CTAAATAGGC AGATGTTTGA ACTAATGCTT    15660
GCTGTTGTGC ATGTGCGCTA GCATCATTTG AAAGGATGGC ATTTCCGCCT AGTGCAACGA    15720
CCATTTTTTT CCCCATTACA AATCCTTCTT TCCTATCATA ATAAAGAATA GGCCGCCCTT    15780
CAAAAGAAAC TCTTAAAGGG CGAAACAACT ATTCTTTTTT CAATGTTTTT TCAGTTGCTT    15840
CCTACACACG AGGAATGAAT AAGTTGCCTA AAGTAGCTGC CATGATCGCT TTAATAGAAT    15900
GCATTCTGTT TTCTGCTTCT TCAAATTGAC GACCATATTT ACTGCGGAAG ACTTCATCCG    15960
TAATTTCCAT TTCAGTAATG CCAAATTTTT CAGCAACCAT TTCACCGTAT TGTGTTTTCG    16020
TATCATGAAA GGCTGGCAAA CAGTGTAATA CGATTAAATC GCCATCCATG TTTTCGGTTT    16080
TTTCTAACAT AGCCATATTA ATTTGATAAG GTTTCAACAG ATTGACACGT TCTTCAAATT    16140
TATCTTCTTC ACCCATGGAC ACCCAAACGT CTGTATATAA CACGTTGGCT CCTTTAACGC    16200
CTTTAGCAAC GTCGTCTGTA ATCATTAATT CTGCACCAGA TTCTTTCGCA AATTCTTTTG    16260
CATAATTAAC GACTTCATCA GAAGGGAACA ATTCTTTTGG TGCACAAATG CGAACATTTA    16320
CGCCTAAAAT AGCACCTGTC ACTAACAAGC TGTTTGCCAT GTTGTTACGG CCATCGCCGA    16380
CATATACTAA AGTGATTCCT TCTAAACGAC CAAAGTTTTC TTTCACTGTC ATAAAGTCCG    16440
CAATCATTTG TGTTGGATGC CATTGATCCG TTAAACCATT CCAAACAGGA ACACCAGAAT    16500
ATTTCGCTAA ATCTTCCACT ACTTCTTGAC TAAACCCACG GAACTCAATC CCGTCAAACA    16560
TACTACCTAA AACAATCGCT GTATCTTCAA CAGATTCTTT TTTACCTAAT TGAATATCGT    16620
TGGCGCCTAA ATACTCAGGA TGTGCCCCTA AATCAATCGC TGCTGTAGTA AATGCAGAGC    16680
GTGTCCGTGT CGAAGTTTTT TCAAATAAAA GAGCGATATT TTTTCCTTCT AAATAGTGAT    16740
GTGGAATCCC TTTTTTCTTC AATTCTTTTA AGTGAATACT GAAGTCTACC AAATACTCTA    16800
```

ATTCTGCTCT TGAAAAATCT TTTTCTGCTA ATAAACTACG TCCTTGAAAT ACTGAATTCA 16860

TGATGATTCC TCCTATTTTT GGGTGTTTTT TTAAAGATCT TCACGGTAAA GTGGCATACT 16920

CATACAACGA GGGCCACCAC GGCCACGTAC TAGTTCACTT CCTGGAATGT AATTCAATTT 16980

CACGCCTGCT TCTTCTAGCG CTTTATTCGT AATCGTATTC CGATCGTAAA CAACAACTTC 17040

CCCAGGGGCA ATTGCTAATG TATTTGAACC gTCGTTCCAT TGTTCACGGG CTGCTGCTGT 17100

TAAATTTCCA GCGCCGCAAC GGATTAGTTG AACATTGTCT AAATGCAAGT ATTTGCATAA 17160

AATGTTATCT AATGTATCTT TTTCTTTTGT AATTTGGATG TCTCCATCTG CTTTTTCAGT 17220

GATCGAGTAA ACAACCAAAC CACCTTCGAT TTCTGGATGA ATCGTAAATT TATCATAGTC 17280

AATCATGGTA AAAACGGTAT CTAAATGCAT GAATTTACGG TGTTCACCGA TATCAAAAGC 17340

CAAGATATTT TTAAACCCTA ATTTTTGTTC AAAAATATTT CTCGCAATTT TTTCAATTGA 17400

CGCGGCATCC GTTCTTTGAG AAATCCCAAT GGCCACTACT TCTTTTGAAA GAATTAATTC 17460

ATCGCCACCT TCAATTCTGG TTGATTCTGA ACGATCATAG ACTCTAGGAA CCTCTTTTCC 17520

AGCAAAACGA GGATGATAAT CAAAAATGTA TTGCCCAAAA ATGGTTTCCC GTTGTCGAGT 17580

TACTGAATAC ATATGATTTA AAGAAATCCC GTGCCCCATT GTCGCAAAGT TATCACGCGT 17640

GAAGTATAAG TTAGGCATTG GATCAATAAT GAATGGATAA TCCGCTTCAA CCATATCTGT 17700

TAAAAATTCT TGCTCATATT TAGGTAATTC TTGTTTTTGA ATGCCTGCAA TCGCTTTTTG 17760

AATCAGTTCT TCGTTGTCGT CAATTTCTAA CATTAACTCT CTAACTTTTT CTTTTGCTGA 17820

TTCGCTGCGA ATATTGGCTT CTTCTAAGAA TTGGTCAATA AATTGTCGGC GGACCTCTTC 17880

ATTAATCAAC GCTTCAGCAG CTAAGTCCTC TAAATAGACC ACTTCGATAT CTTTTGATCG 17940

TAACAACTCT GCAAATGCAT CATGTTCTGC TTGTGCTTTT TCTAAAAACG GAATATCATC 18000

AAACAACAGT CTCTCCAGAT AATCTGGCAT TAAATTTTCT AATTCCTTAC CTGGACGATG 18060

TAACATCACT GTTTTCAATT TCCCGATTTC AGAAAATACA TTAATTGGAT GACTCATGTG 18120

AAATAACCTC CTCAACTTAA TTACAAACAC AGAATAACAG ACCTGATTAA CCTAGGACGT 18180

GAGCCATTTG CCAAACAAAA TACCAAATAT ACAACGGTAT TCTTTCTTAT GCATATTTCT 18240

AGCCTAAAAG AATGCGCTTT CATAGATATT CTAAAAAAAT AATGCATAAT TAAATAAAAA 18300

AAAATGCATT TATTCAAAAA AAATTATTTC TCCACACACT TCATAAAGTT ATAATCTTAT 18360

AGAGGkAAAA GGAAGTGAGC CAGAAGTCAA AACGACTTCT GGCTCACTTC CTTTTATGCG 18420

TTACTTTTTG ACAAAGGTGT TGGCTGAATA AATTGTTGAA AATACTGCTC TAGTTCTTGT 18480

GCTTTTTCTT CGGATGATGA AACAATAATA ACAGTGTCGA AGcCAGCGAC AGAACATTTC 18540

TTTTCACACA AAGACACTTC ATCAAGTACA GCTGAAAAcA ATGTTGCGTT GTCTGGTAAT 18600

GTATGAACCA CTAGTAAAAA GGCGACACGT TCAACTTTGG TCACTACCTC ACGAATTGCT 18660

TGTATCAACT GTTCTTTTTT TCGTTCTTCT AAAGAAGTGA ATTGTTCAAA AATCGTTAAC 18720

CGCATCTCGC CAGACTCATC TAATGCTTTG ACGACTTTTA ATTCCCATAA ATCCCGTGAA 18780

ATCGTTGCTT GTGCGACATT TATCTCGCGC TTTTTTAATT CTCTTAACAA CTCATTTTGA 18840

CTATGAATTA CGGTTGTTTG AATGATTTCT TTAATCAAAC TTTGGCGTGT TTGCTTATTC 18900

ACAAATAACA CCCTTCCTCC AGACATTTAC AAAGCAATAC TAGCAGAATC TCTCTAATTA 18960

GAATAGAAAA AGGGTCGTTT AGCTAAAAGG TTCTCATTGA TTTTGTCGGA TATTTCGCAT 19020

TTAAAAAAGT GAGACTTGCA CAATTTAGTA AAACGTTGTA CAGTGATTTT aGGGAATTAC 19080

TGGCACCGTT TTAAATAACA GCAATGATGT TTTTTTGCAT GATTAACTGA ACTGGAGGAA 19140

AGTAATGAAC CGATTAAAAC GCCAAACGAT GATTAAGCGA TTGGTCACAG AAAATGAAAT 19200

CCATACACAA AGTGAATTAC TAGCATTATT GGAAAAAGAA AATATTCACA CCACACAAGC 19260

AACTATTTCT CGTGACATTC GTGAGCTGAA TTTAAACAAA GTAGACAATG GTCAAGGAGA 19320

AGCCTATTAT CGTATCTTGA ATAATTCTGC CCTTCGTCCA AAATATAAAA CAGATGAAGA 19380

ACGTCTTGTT GACATTATTA TTGAAACTGG TGTTTCTTTA ACACAAATAG AGTTCACAAA 19440

TCTTTTAACC GTCTTACCTG GCAACGGGCA AGCCATTGGG GTTTTAATTG ATCGCATACG 19500

TATCCACAAC CCTCATATCT TGGGCTGCAT CGCTGGGGAT GATACCATCT TAATTCTTTC 19560

AAAAAATAAA GAAGATGCCC TTGAAGTTAA CAACTATTTC CAGCAGTATT TGTACCACCC 19620

ATAATTCATC ACTTAGTAAA ATCGACTATT TTTTGACATC ATtGCTATTT TTATTAAAGA 19680

TAAATAAAAA AAGGCGTCTC AGTGaAAGCT GAGGCGTTTT CTCATGAGTT TTTCTTTTGA 19740

ACACCGTCTT TTTTTTGTAT AATAATCAAA AAGCAAAGAA AgGGkTTCTT ATGACTGAAC 19800

AATTCATACT CACTTCAAAC GATCAACAAA CTCAGCTCAA TGTACGTCAT TGGCCCTGTC 19860

CTTCACCCAA AGCCGTAGTT CAATTGATTC ACGGCATGGC AGAGCATATC CAACGTTATG 19920

ATGAATTTGC TCGTTTTTTA AATCAATTAG GATTTACCGT AATTGGTCAT GATCATCTAG 19980

GTCATGGCGA ATCTGTAsaA CCAAACGCAC CGATCTATGG TTTTTTTGGA GAACAAGGGC 20040

CTGAAAATGT CGTTACAGAT ATTCATCAAG TGAAACAATG GGCTGTGAAT CGATATCCAC 20100

AATTGCCCTA TTTTATGATG GGTCACAGCA TGGGGTCTTT CGCCTTGCGC AACTATCTGC 20160

AAGACTATCC TGTGACTGTG CAAGGAGTCA TTTTCATGGG CACTGGAACA AGTCCGTTAC 20220

CTTTAACCGC AGCATTGCCT TTTATTAAAA AAATGGCCGA GAAACAGCCG AAAAAACCAG 20280

CTCCGTTCAT TGATAAGCTG GCATTTGGTT CTTTTAGCAA AAAATTTCCT GAAGCAAGTT 20340

CCTTTAATTG GCTTTCTAAA AATCAAGCCA ATGTGGCTGA CTATGAAAAT GACCCATTAA 20400

TGGGCTTCGT ATTTACCAAC AATGGTTTTG CTACCCTTTT TTCGCTTGTT AAACGAGCCA 20460

ATCAAAGGAA TTGGTACCAA GCGATCCCAA AAGAATTACC TATTTTAATC ATTAGTGGCG 20520

CTGAAGACCC AGTTGGTGAT TTCAGCAAGG GGCCAGCGAA AATTCAAAAG CAATTAAAGC 20580

ATGCAGGTTT TCAGCACGTG ACGTTACGAC TATTTCCCAC ATTGCGTCAT GAAATTTTAT 20640

TAGAAACGGA AAAAGCTACA GTTTTTCAAG AAATTGGGCA TTGGTTAACG GATTTAACGA 20700

ATTAAAAAAG AAGGTTGATG TTGTGACATC AACCTTCTTT TTTAATCTAT TTTCGTTAAA 20760

CCGCCCATAT AAGGACGTAG AACTTCTGGA ATCGTTACGG TGCCATCTTC ATTTTGATAA 20820

TTTTCCAAAA TGGCAGCCAC TGTTCGACCA ACAGCAAGTC CTGACCCATT AAGAGTATGG 20880

GCATATTGAA CATGACCCGT TTTATCACGA TAACGAATCA ACGCCCGACG TGCTTGGAAG 20940

TCTTCACAGT TTGAACAAGA ACTAATTTCT CGATACGTTT TCTGTGCCGG AATCCAAACT 21000

TCTAAATCAT AGGTTTTCGC TGCCGAGAAG CCCATGTCAC CTGTCGAAAG AGTAATCACT 21060

CGATAAGGGA GGCCCAATTT TTCTAAAATG TCTCCAGCAT TGTTGGTCAT TTTTTCTAAT 21120

TCTTCATAGG AGTGTTCTGC ATCACTAAAT TTAACCATTT CTACTTTATG AAACTGGTGT 21180

AAACGAATTA ATCCTCGGGT ATCACGACCT GCACTACCCG CTTCCGAGCG AAAAGAAGGA 21240

CTTAAAGCAG TAAAATAAAT CGGTAACTCT TGCGCATCTA AAATTTCATT ATTATAGTAG 21300

TTAGTTAACG GGACTTCCGC TGTGGGGATC AGCGTCAAAT CACTTTCTGC AACTTGGAAA 21360

ACATCTTCTT TGAACTTAGG AAACTGACCG GTGCCAAACA TTGCTTTGCT ATTGACCAAA 21420

TAAGGAGGAA TCACTTCGGT ATAACCATGT TCATTAACAT GTTGATCTAA CATAAAATTA 21480

TAGACCGCTC GTTCTAAACG AGCCCCTAAG CCTTTGTAAT AAAGAAAACG GCTACCAGAT 21540

ACTTTCGCGC CTCTTTCAAA ATCTAAAATT CCTAAGGCGG GGATCCTCTA G 21591

(2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6590 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

CAATTGCCCC GATTGGAATA AATGAAGGGC CTTGCGCAAC TGGTAATCTC ACACAAAAAT 60

GAGATTGAAT AATCGTGGCA AGTCCTGCTG CTATAAATGT TGCTTGGATT AAAGAAGACG 120

TTTGTGCTGT GCTAATCCCA ATAATTGAAG CAATGATCAC AGGTACGACG TACACGTCCA 180

TTGCTAAGAC ATGCTGAATC CCTAACATAG TGGCCTGCCC AAAAGATACT GAATCTTCTG 240

GCCCCACTGT TAAATGAAGA TTATTTTGTT TCTTTTCACT CATTTTTTAT ACTCCTTTAG 300

TTGCTGACAT TAAATTTTTA TCATGTACTT TTTCACCTTG TACCCAGACT TCTCGAATAT 360

TTTCTGGTCG TGCTAAATAT AAGATTTTTT GGAAAATATC AAGCAATGAT TCCCCTGGAT 420

TTTTAGGAAT TCGTGCTTGT GACAAGCTTG TATTAATAAT TTGAACATCC CAAGCATACC 480

CTTTTTCTAA TTTTCCAATT GGTAAACTTA GCGCTTGACC ACCACCAGCA GTAGCTAAAT 540

AAAACGCTTC ATTTACGGTA ATCCGTGCTG TTTTAACACC TCGTACTTCT GGTTTTTTTG 600

TAGCATCGAC GCCATCTTCT AACATTCTAG AAGAAATAAC CGCTTGCTTT AAATTGTCAA 660

AAAGACTTGG TGAAAAACCA CCTGAAAGAT CACTACCTAA ACCAACTTCT ACTTGATGTT 720

GATGTACTAA TTTTGCAATC GGTGTTACTG CATTACCAAA ATAAGCATTA GAGATTGGGC 780

AGTGCGCCAC AGCCGTCCCT GTTTCATGGA ATAAATTCAT ATCTGCTTCT TCTAAAAAGC 840

```
CTGCATGTGC CATCACTGcT TTTTCTGtCA ATAAACCAAA ATCATTCAAT GCtTGAGCAT   900
CATTTTTTTC AAAACGTTCT TGaACAAATT GATGTTCCCA ATCACTCTCA CTACAATGAC   960
TTTGAACAkG GACTTGATAT TTCGCTGCTA ATTCTCCTAA TCCTTTAAsT GCTTCCTCTs  1020
TACAACTTGG AaTGAAAcGC GGaGTTACAa CGGGrTACAC GCCTTGTTTT GTTTGCTGCG  1080
CTAATGTTTG AATTTCTTGA ATAAACCGTT CCGTTTCTTC TAATGCTTGT TGCGTTGTTT  1140
GATCACGATA GTATTCTGGA TTCTCATTAA GATCATCCAT TACAACTTTA CCAATTAAGC  1200
CACGCTGCCC TTTTTCCGCA CAAATCTCTG CCAAACGACG GCTAGCTTCA TAATGAATGG  1260
TTGCAAAATA AAGGGCTGTT GTTGTCCCAT TAGCTAACAA TTGGGTTACC AATTCATCAT  1320
AGACTTGTTG GGCAAATGCT AAATCTGCAA ACTTGGACTC CAAAGGAAAC GTGTACGTAT  1380
TCAGCCATTC ATACAATGGA ATATCCAAGG CGGTCCCTGA TTGTGCCCAC TGAGGTGCAT  1440
GGACATGTAA ATCAACAAAG CCGGGTAAAA GATAAGAATC TTGGGGCAAT TCTTGAAAAC  1500
GTTGTGCCTC TTTTTGTTCT GCTAGCACAG TTTCATACGC TTCAGCATCT GGTTCAATTA  1560
TTTCCTGAAT TACTCCGGCC TCATCTATAC AAAAAAGTGT CTCTTCATAA ACTTTAATCT  1620
CTTGCAAATT TTTACTAGAA AAACCTGTTC CTTTTATTAC AAAATGAAAC ATTTAATACT  1680
CCTCCTTACT ATCACATGAA GATTATACAT AAACTTTCGC TTAATAACTA TAAAAATGTG  1740
ATTTTTTTAA AAAAAATCCG TGTTAAAACA AGTAAAAACC GAACATTCTA AGaATGrTTT  1800
ATTTTTTCTT TGGATAAACA AGAAAACTTT CTATTCTTAA GATTCTTCGG TATACTTGAA  1860
GAGAATCGTT AGAACTTTAG GAGGTCCTGA GATGATACAG ATCGaACAAA TTAGAACAGG  1920
TGTTATTCAA GAAAACTGTT ATCTTGTTTA TAATGAAGAA GCGTTATTAA TTATTGATCC  1980
AGGTGCAGAA GGAGCAAAAC TAATTCAAGA AATTACCCGT TTAGGGAAGA AACCAGCCGC  2040
TATTTTATTA ACACATACTC ACTATGATCA TATTGGTGCG GTAGAAGAAT TGCGCCATCA  2100
TTATCAAATC CCTGTTTATG TTAGTCCATT AGAACAAGCG TGGCTTTCCA ATCCTATTTT  2160
GAATCTTTCA GGATTAGGCA GACATGATGA TATAGCAAAT ATTATTGTTC AACCCGCCGA  2220
ACATGAATTT AAACTAACCG ATTATGAAAT TGGCGGCATG AAATTTTCCG TAGTCCCAAC  2280
ACCTGGTCAT TCAATTGGTA GCGTCAgcTT TGTCTTTGAT GATTTTGTTG TTGTTGGGGA  2340
TGCATTGTTT AAAGGTAGTA TTGGTCGTAC CGATTTACAC ACTGGTGATA TGCAACAATT  2400
ATTACACAGC ATTACGACCT ATCTATTCAC TTTACCAGAA GAATTTCCTG TTTATCCAGG  2460
CCATGGTGAT GCCACGACCA TTCAACATGA AAAAGCCACA AACCCATTTT TTAATTAGAG  2520
ATTTATTCAC TATTTAAAGA AACTTAGGAG GAGTAGCAAT GACTGAAGCA ACGACACTTT  2580
ATATTATTCG TCACGGAAAA ACGATGTTTA ATACGATTGG CCGTACCCAA GGCTGGTCAG  2640
ATACCCCCTT AACCAAACAA GGAGAAGAAG GCATTTATCA TTTAGGCTTA GGTCTTAGGG  2700
ACATAGACTT TAAGGAAGCC TATAGTAGTG ATAGTGGTCG TGCCATGCAA ACCGCTCAGA  2760
TTATTTTACA AGAACATCAA AACCACCAAA AGATTCCTTA CCTAACCGAT AAACGGATTC  2820
```

GTGAGTGGTG TTTTGGTTCT TTAGATGGTG GCTATGATGG TGAACTTTGG GGTGTGGTGC 2880

CACGCATTTT AGCTTTTAAA AGTTATGAGG ACATGATGAC CACGAAGATT ACCTATCGTG 2940

AGCTAGCAAA TGCTATTATT GAAGCCGATA CGGCTGACTG GGCGGAGCnT TATGAAGTTA 3000

TCCGTGATCG TGTTTGGTCT GGTTTCGAAG ATATCGCTCA TCATCGAGAA AAAAATGGTG 3060

GCGGCAAAGT CATGGTGGTT TCCCATGGAC TAACAATTtC CTTnTTACTT tCCTTAATgA 3120

TGCAAtTTGC CTATGCAGAT GGcGCTGGAG AATGGTAGCG TGACGACCTT AACTTACGAA 3180

AAAGGAACAT TTACCATnCA AGGAATTAAT GATATTTCTT ATATTGAAAA GGGAAAAAAG 3240

ATCGCTGAAA AGCgCCGTTT GTTGTAAAAA AAGAGAAGGT TATCTGTCAA AATGACAGAT 3300

AACCTTCTCT TTTTATTTAT TCTTTTAAAC GTTCTACGTT TTTTCTTGAA ATTTGAATTT 3360

TTTCTAAGGT TTCCTTATAG TTTTTTGCTG CATAAGCAAA AAATAAAACA TCGACTACAT 3420

ACAATTGTGC CACCAAGGAA ACAGTAGCCG CACTTCTCAA TGGCACGTCT TCCCCACTAG 3480

AGGTTTGGAG GACGATATCA CTTTTTTCAC CAATCACTGA TTTTTCATCT TGTGTGATAC 3540

TAACAATAGG AATCCCTTGc TGctTCGCTA AATCAGCTAA AGTACTTGCT TCTTTATTCG 3600

TTCCAGAATT AGAAATAACA ATAAAAACAC TAGGCTTTTC TGTTGAGCCT AACATGGAAG 3660

CAAATAAATG ATGATCCAAT GTTGTAAATA CCGTTCGGCC TAATCGCGTA AATTTTTGAT 3720

AAATATCTTG AGCAACCAAC GAAGAGGCGC CCAACCCATA GACAAAAATT ACTTCTGTTT 3780

TTTCCAACAA ATCAACGGCT TCATTCACTG CGACATCCTC TAATGTCTGc aCgctTCGCT 3840

CAACCACATG AACAAAGCGC GCCTGTAACT TTTGTTTAAT GTCAGCTGTT GATTCCCCTT 3900

TAGTAACTTC AGTATACATT TGCTGCTTCA TTTGGCCAAG ATTTGCAGAT AATAAAAGTT 3960

TCAACTCTGT GAAGCCATTC ACATCCaTAG AATGACAAAA GCGAATAATT GCCGCTGGGC 4020

TTGATCCTGC TTGTTTCGCT AGCTCTTGCG CATTCATTGA AATAACTGTT TCTGTATTTT 4080

TTAAAATATA TTCTGCGATT TTTCGTTCTG ATTTTGGTAA TTGTTGTAGA CGATCTTGGA 4140

TAGATAAGAC AATATTTTGT TGCATAAACA ACCCCTCCTG ATACACCATT TCCTACGCAA 4200

TTAAATTTTA ACAAAGAAAA AGGGAGAAGC CAAGTTCTCC TTTTTCTATT TATTATTTAG 4260

TTGTTTGCAT ATCTTTAGGA ATCCCAAAGA ATAACGTTGC TACAAAACCA CCTGCGTAAG 4320

CAGCCAATAA ACCTAATACA TATGCAAGCC ATTGGTTGTT GGCAATTAAw GGAATTAAAG 4380

CCACCCCACT TGGTCCAATA GCAATCGCAC CAACATTACC GAAgGCCCCA ATCACCGCAC 4440

CACCAATACC ACCGCCGATA CAAGCAGTAA TGAATGGACG TCCTAATGGT AAAGTAACTC 4500

CGTAGATTAA TGGTTCACCA ATACCTAAAA TCCCAACTGG CAAGGCACCT TTAATCATTT 4560

CAACTAATTT TTTATCTGAT TTACAACGAA TCCATAATGC TAAAGCTGCA CCGACTTGTC 4620

CCGCACCAGC CATGGCCAGA ATTGGTAATA ACAAAGTCAT CCCTGTTTGA TTAATCATTT 4680

CAATGTGGAT TGGCGTCAAA ATTTGATGTA AACCAAACAT TACCATTGGC AAsAAGGTTA 4740

AACCAAGAGT AAAGCCAGCA ACCATGCCAC CTTTTTCTAA GACAACATTG ATAATACCAA 4800

CTAAGCCATT AGAAATAGCT CCGGCTACAG GCATGATTAA GAAAATAGTT GCTAAGCCAA 4860

TCACTAAAAG TGAAATGGTT GGCGTTACGA TAATATCAAT TGATTCTGGG ACAATTTTAT 4920

GCAATTGTTT TTCTAACAAT GATAATAACC ATACGGCAAA AATAACACCG ATAATTCCGC 4980

CTTGTCCTGC AGATAATGTT CCCCCAGTGA AAATATTACT GATTGGAGCA TCCGGATTCA 5040

TTCCTGTTAA CATTGTGACT GCCCCAATAA CACCACCCAA CGCTGGCGTT GCACCAAAAA 5100

CGCTGGCACT GTTAATCCCA GTGTAAAGAG CTAAATAAGC AAAAATCCCA TTTTTAATGA 5160

TATTTAAGAC ATCAATATAT TGTTGCCATG AAGCTGAAAT ATCACCAGCA ACAACTAAGT 5220

TAGACATAAC AGCTGCAATC CCGCCAATAA TACCAGCCCC AACAAAGGCA GGAATCATCG 5280

GAACAAAGAT ACTAGAAATG GCTTTTAACA CTTTTTTAAA TGGTGAAGTG TTATTTTGTT 5340

TCGCTTTTTG TTGTGCTTTC ATTTGAGCGG CTTTTTCTTC TACTAAATCT TTTCCAGATT 5400

TGCCTGCTGA AGCATCAGTT GTGGTGCCAT GATTAAATGG TTCTCCCAAT TTGACACCTG 5460

CTTGATCTAC CATTTCTTGT GCAACCTTGT TAACAGTCCC TGGACCAACA ACTACTTGTA 5520

ACGTATCATC TTCTACGACG CCCATAACGC CATCAATTTT TTTCAATCCT TCAATATCCA 5580

CTTTGTCATA ATCGCGGATA TCCATACGAA CACGAGTCAT ACAATGAATG ACTTTGTCCA 5640

CGTTTTCTTG ACCACCAACT TCTTTATAAA TCCCATCAGC AATGCGTTTA ATTCTTTCTT 5700

CTGCCATGTT AAAACACCCT CCTAATTCCT TAATTTAATG TATTCTTGAT AAAtTGaTTC 5760

CCGTCAATTA ATTTTTGTGT TGCTTCTTCT TTAGTACTAT CAGTTAAAAT CATTACAATT 5820

GCTAACTTCA CATTTTCTTC TGCTTCTACA AATTTAGTT CGGCAACTTC ATAACTACAA 5880

CCCGTTGCTT CCATAATAAT CCGTTTGGAA CGCTCCACTA ATTTTTCATT GGTTGGTTTT 5940

ACATCAACCA TTAAGTTATT ATAGACTTTC CCAATACCAA TCATGGAAAT AGTTGATAAC 6000

ATATTTAAAA TTAATTTTTG AGCTGTCCCA GATTTCAAGC GAGTCGAACC TGTTAAAAAT 6060

TCAGGTCCTG CATCTACTTC AATCGGCATT TGCGCATATT TACTAATTTC GGCATTTTTA 6120

TTACATGCCA CAGTTGCGGT AGCTGCTCCC ACCGTTGTAG CATATTCTAA ACCGCCAATT 6180

ACATAAGGCG TCCGACCACT GGCTGCAATC CCAACAACAA TGTCATTTGC TGATAATTTT 6240

AGATCAACTA AATCTTGACG ACCTAGTTCT TTTGAATCTT CGGCACCCTC TACTGCGACA 6300

GTCATTGCTT TTTGACCACC AGCGATGAGG CCTTGAACCA TCTCTGGCTC GACGCCAAAA 6360

GTTGGTACGC ATTCGGCTGC ATCTAAGACA CCTAACCGAC CACTCGTACC AGCGCCCATG 6420

TAGATCAAGC GGCCGCCTTG AwTGAAGCyT TTGaTGAkGG CTTCTACCAC TGGTTCAATG 6480

GCTGCTAATT CTTGACCCAC TGCCaTTGCA ACTTTTTGGT CTTCTTGATT CATTTTTTGT 6540

AAAGCTTCTT TGACACTCAT CTCATCTAAG CCCATTGTaT TTTCATTTCT 6590

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3660 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

GACATCTGTC AAAACGACCG ATGTCCTACT CTCTAACGAT TAAGATAGTC CGCTAAATGC 60

GTTTGTTTTA GTCTTTTATA AAAGTTAAAT ACTGAGAATT TATTGTGTAA TTCGTGTACC 120

ACTTTCTCCT TTGATTGCTT CCGAAGCTAA TTCTAAAAGC CCGATAATCG CCTGGCGGTT 180

TGGTTTACTT TCTGCAAAAC TGATGGCCGC TTCCACCTTA GGTAACATAC TACCTGGCGC 240

AAATTGATTT TCTTCTATAT ATTGTTGCAT TGCCGTAACC GTAACATCGG TCAAAGCTTG 300

TTGATTTGCT TGACCAAAGT TAACAAAAAC ATGTTCCACT GCAGTTAATA CCACAAAAAC 360

ATCTGCTTCA ATGAGTTCAG CCATTTTTGC TGCCGCAAAG TCTTTATCAA TGACTGCATC 420

TACACCATGA TAGACTGGaC CATCATAGGc AaTAGGrACC CCGCCACCAC CACAAGCAAT 480

TACGATTTGT CCTGgCATCA ACAAGTGTTC GCAGCCTAAC TTTTTCGTAA ATATCAATGG 540

TTTTGGnTGA AGGAACTACC CGTCGcCAGC CTCGaCCCGc ATCTTCyTGa TAGACGTTCC 600

CAGTTTCCGc CATCmACTGt TTGGckGTTT CyTCGtCATA ATAACCACCA ATTGGTTTTG 660

TTGGTTTTGT AAAcGCTGGG tCTTTTGGaT CCcACCACGA CTTGTGtAAT TAACGTAACA 720

ACTGGTTGTT CCCCTAAATT TCTGGCTACT AGTTCTTCAT CAATTGCTTG TTGTAAATGA 780

TAGCCAATAT AGCCTTGGCt CATTGCTGTG CATTCaGGAA ACGGCACGAC ATCTTCTGGA 840

GATTGACTCG TTTCCTCAAA AGCTAAGCGA ATTTTGCCAA CTTGTGGCCC GTTGCCATGA 900

GCAATCACCA CTTGATGGCC TTGTTCAATT AAATCAACAA TGGATTTTGC AGCGGTTTGT 960

GCTTTTGCCA ATTGTTCTTT CGAAGAATTT CCTAACGCAT TTCCACCTAA TGCAATCACT 1020

ATTTTTGACA TCTGTTTCCC TCCTAGCGCT TTGGTTCTTG TTGTGTGATA CAGTGAATAT 1080

TGCCGCCCCC ATAAACAACT TCTACCGTAT TAACGCCGAC AATTTTTTTA TCTGGGAACA 1140

TCGTTTGAAC TTGCTCTAAA GCCAAATGAT CATTTTCATC ACCATACTGT GGCACAATCA 1200

CGCCATCATT TGTAATTAAA AAGTTCATAT AGGATGCGAT ACAAATATCG CCGTCTTCTC 1260

GTGGCATGGT TCCTTCCACG AAATCAATTT TGAAGGAGCC TTTAATGGTT ACATTTTTGA 1320

CAGGACAGCA TAATTTGTGG ACTTTTAATT GGCGTCCTTT AGCATCAGTC ATTTTAAGCA 1380

AACGTTGATA CGCATCTTGT GCCGCTTCAT AAAATGGGCT ATTTTGATCT TCAGTGTAAA 1440

TACAAGCCAC TTCACCAGGT GCAATAAAAC AArCTACATC ATCCACATGA CCATTGGTTT 1500

CTTCTGGGTC AATTCCATCA CCTAGCCAAr GAaCTTTTTC AACATTTAAA TAGTCGCATA 1560

ATTTTtGTTC GATTGCTTCT TTTGAAAGTT GTGGATTTCG ACCTTCACTT AATAAGCACA 1620

TTTCTGTAGT CAAGACTGTT CCTTGTCCAT CTACGTGGAA TGAACCGCCT TCTAAAACAA 1680

AATCATCTGT ACGATAAGAA TCTACATGCT CAATTTCACA AATTTTTTGT GCAACCAAAT 1740

CGTCTTGGTC CCAAGGGAAA TATAAACCAT CTACTAATCC CCCCCAAGCA TTGAATGTCC 1800

AATCCACACC ACGGATTTCT CCATGGTCGT TAATGACAAA TGATGGTCCA CAATCACGAA 1860

CCCAGGCATC ATTGTTTGAC ATTTCGTAAA CCGTAATTTC TGGCGGCAAT TGGCGACGGC 1920

AATTTTGAAA TTGTTGTTGA GAAACCACCA CATTCATCGG TGTAAATTGA CTAATTGCTT 1980

TGGCCACATT AGTAAAAGCT TCTTGGACAG GTTTACCACC ATCACGCCAA TTATCCGGTC 2040

TTTCTGGCCA AATCATCCAA ACTTTCTCTT GTGGTTCAAA TTCTCCTGGC ATTCTAAAGC 2100

CATCTTGtTT AGGGGTACTT CCTACAATTC GTTTAGCCAT TTTCTTTCGT TCCCTTCTTA 2160

ATTTTTTTCA CGTAAATAAT TACTTCACCA AACAGAATAA AAATGATTGC CCCAATTGTG 2220

ATTGGTAATT TTTCATTTAA TGCTGCGGGA CTAAAATCTA AAGGcACTGC TGTGAAAATC 2280

AGTGAAATAA TAATCATAAT CATTGGCAAA ACCACTAAAA CTTTTAAAAA TCCTGGTTTT 2340

CCGTTAACTT TAAACGGACG TGGTGTATCG GGATCAATTT TTCTAAGCTT ATAAAAAGCT 2400

GGGAAAACTG GGATATACGA TAACAAGAAT AGGACGACAT TCAAGGAGAA AAAGGCCCAG 2460

AATAAATCTT GATTAGGTAA AATTGGTGCT AATACAATGA CCACCGATGC CACAATGCCA 2520

TTCATCAAAG CAGCCCCAAT TGGCATGTCA TTTTTCTTAC TACGTTTTTC AAAAAATTTA 2580

GGCATGTCAC CATTTTCAGC AGCATAACAA GCTGTATTAT TGACTCCTAA TGACCAAGAA 2640

ATCATGTTTC CGAACAACGT AAGAAGGAAT AGGAAAGCCA TCAACATAAT GAACCAACCT 2700

TCTGTCGAGC CTGTCAATAG TTTAAAGCTA TCCATCATAC CACTGCCGGT ACTAATTTGA 2760

TCGGTGGGAA TCGCTACCCC AATTCCAAAG GCTGAAAAAA TATAAATGGC CGCAATGACT 2820

AACCCCGCTG CTACAATAGC TTGTGGAATT TGCTTCTTAG GATTTTCCAT ATCATCTGCA 2880

AACGTACAGA TGACTTCAAA CCCTAACAAA TTAAAAATGA TTACAGAAAT AAAAGATAAA 2940

CTCCTTAAGT CAAACGTTGG CAGCATTGAT TTCAAGGTAA ATTCATTGGC AACACCTTTC 3000

GTCAAGGCTA CATACAATCC TAGCCCACCA ATTAAGACAG CTAAAACCAT CTTGATTACG 3060

GCAGCTCCAT TTAGAATCCA GATACTGTCG CTGACTGGAT AAAAACTAAT CCAAACCACA 3120

ATCCAAGTAA AAATTAATTC AACAATAATC ATCATCGGTG TTGAAAACTT AATACCTGTA 3180

ATTGTCGTTA ATAGTTCTGG CGTCATAACG GCTAATGAAG CTAGCCACAA TGGATAATTA 3240

ATCCAATAAT ACCAAGAAAC TCTGGCACCC CAACGATGAC CATATGCTTT GGTCACCCAA 3300

TCATAGATGC CCCCTTCACC AATATAAGTT GTTCCTAATT CTGATGAAAT TAACCCATAT 3360

GGCAATAGGA ATGCAATTAA TAAGAAAATC CACCAAAAAA ACTGAGAATT ACCAATTGCA 3420

GCAACAGGCG CCGCTGCTTC AGCAACGAAA ACCACACAGA TCACTGATAA AATGGCACTC 3480

GTTAAACTAA ATTTCTTTTT CCCCTCCATG AAGACTCTTC CTCTCTTTAT AAACTGTCCA 3540

CTGATTCATT TTTTCAATCA GTGGACAGTT GGCCGTTAAA TGCTTTGAGT GTCTAAGAAA 3600

ACTTCTAGTT CTTTTTTCGC TTCTGcTTGC TTGATTAAAT CATAAGGATT TTTgCTTCAT 3660

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 827 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

AATGACAACA TTTACAAAAA AGTTTGAGGC AAAAATTAAA ACGGCGGACA wTAAaCTTaA 60

AGTAAAAaTT TTGTTTTTAA AAaTTGCGAA CTTAATAAGT GGGCTTTTTC TTTTTCGTTC 120

AACCATTATA AATAATCCTA ATGCGATTAC TGCAATAATG AACAaTAAaT AGGATTGTAA 180

CGAGCCaAAC CCGCTCTCTT GTCCGAGGAA AATCCCGCCG AAGAACGTCA TAATCGCAAT 240

GGCAATACAA GCAAATCCTG AAAAATCGAT TTTTTCTTTT GTTTTTGTAA TATCTTTTGG 300

TAGGAACTTT TCGCCAATTA AAATTGTAAC TAGCCCCACG GGAACATTAA TCCAAAAAAT 360

ATACGACCAA GAAAAGTTTG ATAAGATCAA TCCACCAATC CCAGGACCAG CAATCGCGCC 420

TAAAGAGACA AAGGCACCTA CTGCACCAAG CGCCCGCCCT CTTTCATTTA AAGGGAATAC 480

TTCCGTAATA ATTCCTGAAT TGGTCGCCAT GGTCATGCTA GCACCAATAC TTTGAACGAC 540

TCGCGCAAAT AATAGAAAGC TCAGCGATTG ATTAAAGCCA CAAAGTAACG AACCAATCGT 600

AAAAATAACC GTGCCAATTC GATAAACTTT TATTTTTCCA AAACTGTCGC CAATTTTTCC 660

AAATAATAAT AAACAAGCAC AGACAACCAT TAAATAAATA GATACGACCC ATTCCGCTTG 720

ATTCATGGGT ACGGACATTT CTTtTGAAAT AGTCGGTAAa GCAATATTTA CAATGCTGGA 780

ATCTAATGTA GACATGAAGG TAAACATAGC CACAGAAACT AAGATCC 827

(2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13574 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

AAGGCATGAA GTAATnGGGA TGGATTTATT AATCGrATGA AGyCAtATCa GGAAGAAGAA 60

GCAGTTTGCT TTACAGATCG TTTGAAAGAG CAGGTGGAGA AAAATGGCTA GAAAAAGTCG 120

TATAAAAACA CCGATTGcGG AACAAGCACC TACTATCAGA AAGAATAATT TTGAAGAGGT 180

TTGCTTAGGC TATACATTAG CAGAAGGGCA AGAAGAAGCG ATTCGCTGTT TACAATGTAA 240

AAATGCACCT TGTATTGCTA AATGCCCCGT GATGATTGAT ATTCCTGGAT TTATTTTAGC 300

TATTCGTGAA GGAAACATGG AACAAGCAGC CGATATTTTA AGTAAATATA CGAATTTACC 360

TGCTATTTGT GGACGCGTTT GTCCGCAAGA GAAACAATGT GAACAAGTTT GTAAATTAGG 420

GAAAGCCAAA AACTTTGAAC CAGTGGCAAT TGGCAAATTA GAACGTTTAG TTGCCGATTG 480

GGCATTAGAA AATCACAAGT TTCCTAAGAA AACAGCGGAA CCAACAAAAG GGAAAATCGC 540

AGTGGTCGGT TCTGGTCCTT CTGGGTTAAC AGTTGCGGGG GATTTGTCAA AGCTAGGCTA 600

CGAAGTAATT ATTTTTGAAG CGCTTCATGA AGCTGGTGGC GTATTAACCT ATGGCATACC 660

TGAATTTCGA TTACCGAAAA AAATTGTTAA ACAAGAAATT GCGAGTATTG AAGCCTTAGG 720

TGTAACAATA GAAACCAATG TGGTAGTAGG TAAAACGATT ACGATGGAAG AAATAATGTC 780

AGAGTTTGAT GCCTGTTATC TATCGGTTGG TGCAGGGGCA CCAAATTTTA TGGGAATCCC 840

AGGAACTAGT TTAAATGGGG TTTATTCTTC TAGTGAATTT TTAACACGAA TCAATTTAAT 900

GCATAGCTAT GAGTTTCCAG AATATGATAC ACCGATTAAA CGTGCTAAAA ATGTTGTTGT 960

GATTGGCGGC GGGAATGTTG CAATGGATGC CGCTCGTTCA GCCAAACGAT TAGGCGCTGA 1020

AAATGTGAAC ATTGTGTATC GTCGTTCGTT AGAAGAATTG CCAGCAAGAA TTGAAGAATA 1080

TCACCATTCC CTTGAAGAAG GAATTAACTA TTATTGGTTA ACGAATCCAA TTGCTTATTT 1140

AGATGACCAA CAAGGAAATC TAGCCGGTGT TGAATGTGTC AAAATGGTTT TAGGTGAACC 1200

AGATGCGTCT GGAAGACGAC GGCCAGAACC AATTCCAAAC AGCACGTTTA CCATTCCTGC 1260

TGATGCAGTA ATCGAGGCAA TTGGCCAAGG GGCAAATCGA GTCTTGTTAT CAACATATCC 1320

AGAAATAGAA CTAAACCAAT GGGGTTACAT TCAAGCTGAT CCTAAAACAG GAGCAACCTC 1380

AATACCTGGC GTTTTTGCTG GGGGCGATAT AGTCACAGGT GCTGCCACCG TTATTTTAGC 1440

GATGGGCGCT GGGAAAATCG CAGCTAATGC AATTGATCAG TATGTCAAAA CACAAAAAAC 1500

aACAGTCACA AGCAATGTTT AGGAAAGATA AAAGGTGAAA AAATGAAAAA AACAATGGAT 1560

GGAAACACCG CAGCGGCTTA TATTTCATAT GCGTTTACAG AAGTTGCCGC TATTTATCCT 1620

ATCACACCAA GTTCAACAAT GGCTGAGTTG GTAGATGAAT GGGCGGAAAG TGGTTTAAAA 1680

AATATTTATG GTCAAAAAGT TCAAGTTATT GAAATGCAAT CAGAAGCTGG GGCAGCGGGA 1740

GTAGTTCATG GTTCGTTAAA AACAGGTGCG TTAACCACTA CTTATACTGC TTCACAAGGT 1800

TTACTGTTGA TGATTCCCAA TATGTATAAA ATCGCAGGAG AATTGCTGCC TTCTGTGTTC 1860

CACGTGGCGG CTCGGGCAGT AACGACAAGT GCGTTGAGTA TTTTTGGTGA TCATGGCGAT 1920

GTgATGGCTA CCCGCCAAAC AGGTTTTTGT ATGTTAGCAG AATCGAGTGT TCAAGAAGTG 1980

ATGGATTTGT CGGCAGTTGC TCATTTAGCT AGCTTAGAAG GTAGTTTACC TTTTGTAAAC 2040

TTTTTTGATG GTTTTAGAAC GAGCCATGAA CTACAAAAAA TCGAAGTGAT TGATTATGAT 2100

GACCTGAAAG AAATGGTGAA TCAGGAAGCT GTAACTCGTT TTAGAATGGG AGGCATGAAT 2160

CCTAATCATC CCACTGTTTC AGGGACCGCT CAAAATCCTG ACATTCATTT TCAACAAAGA 2220

GAAACAGTCA ATAAAAATTA CGAAGAAATG CCAAAAATTG TTCAAAAATA CATGAAAAAA 2280

ATTAACAATC TTCGTGGAAC AACGTATGAT TTAACCGATT ATTACGGGGT TGAAGATGCA 2340

ACAGAAGTCA TTATTTCAAT GGGGtCGgct TCGCCAGTAA TCAAACAAAC AGTCGACTAT 2400

CTCAATCAAC AAGGACGCAA AGTTGGTTTT ATCAATATTC ATTTATATCG TCCATTTCCT 2460

ACCGAAAACC TTCTTGAAAA ATTACCAGCT ACAGTAGAAA AGGTTGCGGT TTTAGATCGA 2520

ACAAAAGAAT CGGGAGCCGA AGGGGAACCG TTATTGCTAG ATGTTCAAAG CGCTCTCTAT 2580

CAACATGAAA ATCGACCAAT TGTGATTGGT GGACGTTATG GGTTAGGCTC TAAAGACGTA 2640

GCGCCTAATC AAATTAAAGC GATCTATGAT CATTTGTTAT TGCCATTTAA AGAATTAAAA 2700

CAACGTTTTA CGGTAGGGAT CGTTGATGAT GTAACCTATC AATCATTGCC AGAAGAACCT 2760

ACTTTAGATT TAACGCCAGA TACTACCTTC CAAGCAAAAT TTTGGGGCTT TGGTTCTGAC 2820

GGAACTGTGG GAGCAAACAA GCAAGCGATT AAAATTATCG GAGATCATAC CGACCTTTAT 2880

GCACAAGCTT ACTTTAGTTA TGATTCTAAA AAATCAGGCG GTTTAACCGT TTCGCATCTG 2940

CGTTTTGGTA AAGAGCCAAT TACTTCCACG TATTTAATTG AACAAGCAGA TTTCATCGCT 3000

TGTCATAATG CTTCCTATTT GCATAAATAC GACCTGTTAA AAGGTTTAAA AGATGGCGGA 3060

ACGTTTTTAC TGAATACTAT TTGCGATCAA GAAAAGATTC ATCGTTTATT ACCGGCTAAG 3120

TTAAAAAAAT ATATTGGGGA GCACCAAATC AAATTTTATA TTATCAATGC GGTTGAGTTA 3180

GCTCGCAAAG TAGGCTTAGG CCGTCGGATC AATACAGTGA TGTCTACTGC CTTTTTTGAA 3240

GTAACAGATA TTTTAACGAA AGAGCAATAT TTACCGTTAC TTAAAGAAGA AGTGAAAAAG 3300

ACTTATGGAA AAAAATCAAT GGAAATTGTT GAAAAAAATT ACCAAGCGAT TGATGCTACC 3360

TTTGAATCGC TGCAAGAAAT TGTAGTTCCA GAAGATTGGG CAACCATTGT TTTAGAGCCA 3420

GAAAATCAAG AGTCGGAaGC TCCAGCTTTT GTGACAAATA TTTTAGAACC AATCAATCGT 3480

CaAGAAGGGG ATCAGCTGAC CGTTTCAGAT TTGATTGAAA ACGGAATGAA GGGCGGTGCC 3540

ATTCCTATGG GCACTGCGGC TTATGAAAAA CGTGGCATTG CGTTAGAAGT GCCTGAGTGG 3600

CAAATGGATA AGTGTACGAT GTGTAATGAA TGTGCCTTTG TTTGTCCACA TGCGGCTATT 3660

CGACCGTTTT TAGCGGATGA AGAAGAACTG AAAGAAGCAC CTGCAGGTTT TGCCATGCGC 3720

GAAATGCGTG GCACAGATGG GTTGATGTAT CGTATCCAAG TTTCTTTAGA AGACTGTACA 3780

GGTTGTGGTC TTTGTGTTCA GGCATGTCCC GTGAAAGACA AGGCGATTTT AATGAAACCT 3840

TATGAAACTC AGAAAGAGCA AGCGATGAAT TGGGCCTTTG CGATGACTCT TAAACAAAAA 3900

GCCAACCCAG TCAAAAAACT TTCTGTTAAA GGTTCCCAAT TTAACCAGCC ATTGATGGAA 3960

TTTTCAGGTG CTTGTGAAGG CTGTGGCGAA ACACCTTATA TTAAGTTGTT GACGCAATTA 4020

TTTGGTGATC GAATGATGAT TGCTAACGCA ACAGGCTGTT CATCAATTTG GGGTGGTTCG 4080

TCACCGGTGA CGCCTTATAC AACGAACGAA TGTGGCCAAG GACCAGCTTG GAGTAATTCG 4140

CTGTTTGAAG ATAATGCGGA ATATGGTTAT GGTATGTATA TTGCCAATCG AACGAAACGT 4200

CAACATTTAG CTAGTCTGGT TGAAGAAAGT CTTGCTAAAA ACGTCGGATC AGATTCACTT 4260

CAAGCCTTGT TAAATGATTG GTTGGAACAC ATGGCTGAAG GGGAAGGAAC ACAACAGCGA 4320

GCCACCAAAC TAGCGGCAGC ATTAAGTGAA GAAGCCGATG AAGATCCACT TTTAACAAAA 4380

ATTTACGAAC AAAAAGATTT ACTAGTGAAA ACGAGTCAAT GGATTGTCGG CGGCGATGGT 4440

TGGGCCTATG ATATCGGGTT TAGCGGCATT GATCATGTTC TTGCCAGTGG GGAAGATGTT 4500

AATATTTTTG TTATGGATAA CGAAGTGTAT GCCAATACAG GCGGACAAAC ATCTAAAGCT 4560

ACACCAGCTT CAGCCATTGC CAAATTCTCA GCGGGTGGGA AACATACCTC TAAAAAAGAT 4620

TTAGGGATGA TGGCAATCAC TTATGGGAAT GTCTATGTTG CACAAGTGGC TTTAGGCGCA 4680

AACTCAATGC AAACGATTAA AGCAATTGAC GAAGCGGAGC GGTACCCAGG CCCATCAATT 4740

ATCATCGGTT ATACCCCATG TATTAATCAC GGGGTCAAAG GCGGAATGGT TCGGACGTTA 4800

GATCAAGCCA AAGAAGCCGT TGAATCTGGC TATTGGCCAC TTTATCGCTA TAATCCAGAG 4860

TTAGTTAAGA AAGGCAAAGA TCCGATGGTC ATTGACTACA AGAAAACAGA TTTTGACAAA 4920

ATGCGTGATT TCCTTGAAAA ACAAACACGT TACTCTGCGT TGCATACAAT TAAACAGAAT 4980

CAAGAAGTAG TCGAACATTT ACTTGATAAA ACTAAAGAAG ATGCAATTGA ACGCTCTGAA 5040

AGTTATAACA AACTAGTCAC ACACATGAAA AAATAAAAAA ACAAAGCCTC TTACTTGAGA 5100

GGCTTTGTTT TTTTATCCAA TAATAATTTT TTCTGAAGGG TATTCGTAGC CTAAGTCACG 5160

TGTTTCTTTT GGTACAAAAA TCATTAAGGT ATAAAGCATA CCAATACGTC CAATAAACAT 5220

TAATAAGGCA ATCATTAATT TACCTACTGT TGTTAAATCG TCTGTAATAC CCAAAGAAAG 5280

ACCGGTTGTC CCAAAGGCAG AAGCGACTTC CACAATGATT GAAATCAAGG GCAGATTTTC 5340

AGTGGCTGAC AAAAAGACGA CAGCGAAAAA GCACATAATC AAAGATAACA TGAAGACGAC 5400

AATCGATTTT TTGACATCAT CGTCATCTAT TTTACGGCTA AAGACGCTAA TATCATCTTC 5460

ACTTTTTAAA AAGGCTAACA AATACAAGCC AATAATCGCA ACAGTCGTTG TTCGAACACC 5520

ACCACCAACA GAGCTAGGAC TACAACCAAT AAACATTAAC ACAGAGAAAA TAATCAACGT 5580

TGTAATTTGG AAGTCCCCTA AATCATTAAT TTGGAGGCCA GCGTTTCGTG TAGTCATCGA 5640

ATAAAACATC GAACTTACCC AACGGCCACT TTCATTCATT GTCAAAAATA GGTGATCTTT 5700

TTCTAATAAG TAAATCAAAA TAGTCCCACC AATAAACAGA ACAATAAACG CTAGCAAAGC 5760

AATTTTACTA AACAATGAGA AGCGGAAAGG CAAGCCTCGT TTCTTCTTAC GAAAGTTCTT 5820

TTTCTTAAAA TGTAACCACT CATGAATTTC CAGTAAAACG GGAAATCCAA TCCCGCCGAT 5880

AAAAATCAAG AACATAATTG CCAGTAAAAA GAGATAATCA TGAGCAAAAG GAATAATTGA 5940

ATCACCTGAA ATATCAAAGC CTGAATTAGT CACTGCAGAA ATAGCTTGAT AAAATCCATA 6000

AAATATCGCA TCTCGCCAGT TCGTATAATA ACCACTTATA TAAAAATAAA GCGAAAAGAG 6060

CGAGCCGAAG ATTACTTGGA ACCAAATTAA AATCAAAAAA GTGGTGCGGA TCAGTCGAAC 6120

AATTCCGCTC AGTCGCGGCT GATTCATGTC AGTCATAATT AAcTGACGTT GCTTTAATGT 6180

GATTCGCCGT TTTGATAAAA TAAAAAAGAC CGTTGAAATC ATCATAATCC CTAAGCCACC 6240

AACTTGGAAA AGAACCTCCA ATAAGACGAT CCCACGATCA TTAAAAACCG AATTAATATC 6300

GACTGTGGTG AGTCCaGTAA CACTGACTGT ACTGATTGCT AAGAAAAATA AATCAATAAA 6360

AGGAACATGC GAGTCAGGTT CTCTGAAAAT AGGGAGGCAA AACAATAAAT AAGAAATAAC 6420

CGTCATTAAT ATGTAATAAG CGACGATGAT TTGAATAGAT GAAAAATTTT TGCGAATAAA 6480

TCGATCGCTt cTctTTCGCT GTGCCAGCCA CAGCAAACGT GTATGGTTCA CTAGATaACy 6540

TCCCaATTCT AATTAAACTA mGTATACCAA AACTTTTCGC CTCAAACGAA TAATTCTGCC 6600

AGAAAATCCT TGAAATGACA AATAAAAATA AGAAATTCGT TAGAAATAAT TGGAAAACAT 6660

CCACAAACAA GCGAGAACAT GATATTTTAT GAAAGAATAG AAAAATAGGT TGTTCCCCAA 6720

ACAACTTATT AATAATTGGT CAGTTTGTCT AACGATACAA CTTTAAAAAG TAGTGGTAGC 6780

TTTTTTATTT CGGTTGATAT GTGAATGTGT TCACTTTATG TAACAAACTG TTATTTTTAT 6840

TCAAATGAAA AATAAGAAAT TGTGTTGGTA GAAATCAATA GTTTATAAGG TGGGAAAACA 6900

ATGAAAAAAG TGTTAATGGG AGTTCTTTCC CTAGGGTTAT TATTAGGTGC AGCAACAGGT 6960

TGTACAAGCG ATCAAGAAAA AGCCGCTGGT AAAACAAAAG CGTCTTCTGA AAAAACAGAA 7020

GCAACATCTG GTGCCTCTGC TAATGGTTAT ACAGATCCAA GTGAACTAAA AGACAGTTAT 7080

GATGTTGTTA TTGTTGGTTC AGGTGGTGCG GGGATGACCG CTGCATTACA AGCCAAAGAA 7140

GCTGGCATGA ATCCAGTTAT TTTAGAAAAA ATGCCTGTTG CAGGTGGGAA TACAATTAAA 7200

TCGTCTTCAG GAATGAATGC TTCACAAACA AAATTCCAAG AAAAAGAAGG CATTAAAGAT 7260

AGCAACGATA AATTCTTTGA AGAAACACTA AAAGGCGGTA AAGGAACCAA CGATCAAGAA 7320

TTACTACGTT ATTTTGTTGA TCATTCAGCT GAAGCAATTG ATTGGTTAGA TACAAAAGGr 7380

ATTACATTGA GCAATTTAAC GATTACAGGT GGAATGAGCG AAAAACGCAC ACACCGACCT 7440

GCTGACGGAT CAGCAATTGG GGGCTATCTA GTTGACGGTT TAGTACGTAA CGTTCGCGAA 7500

GAAAAAATTC CATTGTTTGT TGATGCTGAC GTCACAGACT TAGTTGAAGa AAACGGTCaA 7560

ATCGATGGCG TAAAAGTGAA AATGAAAGAT GATAAAGAAA AAACAGTTAA AGCAAAAGCA 7620

GTCGTTGTCA CAACAGGTGG TTTCGGTGCC AcGAAAAATT AATTACACAA TACAAACCAG 7680

AATTGAAAAA CTATGTTACA ACGAACCAAG AAGGCACAAC AGGTGACGGT ATCCAAATGA 7740

TTCAAAAAGT TGGcGGTGCG TTagTGGATA TGAAAGAAAT TCAAATTCAC CCAACTGTTC 7800

AACAAAGCGA TGCTTTCTTA ATTGGTGAAG CAGTTCGTGG AGAAGGCGCA ATCTTGGCTT 7860

CTCAAAAAGG AGAACGTTTT GTTAATGAGT TAGATACACG CGACAAAGTT TCTGCAGCAA 7920

TCAATGCCTT ACCAGAAAAA TCAGCTTATT TAGTCTTTGA CCAAGGCGTT CGCGACCGTG 7980

CAAAAGCCAT TGATTTCTAT GATCAAAAAG GTTTCGTGGA AAAAGGTGAA ACGATTGAAG 8040

AATTAGCTGA AAAAATTGGC ATGCCTGCCG ATACATTGAA AGCAACGATT GATACTTGGA 8100

ACCAAGATGT TAATGCCAAA GATGACAAAC AATTTGGCCG GACAACTGGA ATGGAAGCTG 8160

ACTTAAGCAC AGCCCCTTAC TACGCAATTA AAATTGCACC AGGGATTCAC CACACAATGG 8220

GTGGCGTGAA AATCAACACG AAAACAGAAG TTTTACGTGA AGATGGCACA CCAATTAAAG 8280

GTCTATATGC TGCTGGTGAA TTAACAGGTG GCTTACATGG TCaAAACCGT ATCGGAGGCA 8340

ATGCAATCGC CGATATTATC ATTTATGGTC GTCAAGcAGG TACGCAATCA GCAGAATTTG 8400

CCTCTGCTCA AAAATnACTA GCATTAAAAA GTCTGAAATT GTAGTCAAAT AACTACAATT 8460

TCGGGCTTTT TTATTGCATA CTAGCAGGAT CATTAGTTGG AGTATAAATA GCCCTACTAA 8520

AATGACTTTG GCGTAGATTC TGTTTTTTTC AGAAAAATGC ACGACCATTT TAGCTAAAAA 8580

TACTGCTTTT TTGCTGAAAA TATAGTAAAA TACGTAAGGA ACTAACGGAA AGGTGTAGTG 8640

AACATGGCTC AATTATTTTT CAAGTATGGC GCAATGAACA GTGGTAAAAC CATTGAAATC 8700

TTAAAAGTAG CACACAATTA CGAAGAACAA GACAAACCAG TTGTTTTGAT GACCAGTGGA 8760

CTAGATACAC GTGATGGTGT TGGGAAAGTA TCCAGTCGCA TCGGATTACG CAGAGATGCT 8820

ATTCCAATTT TTGAAGAAAC CAATGTCTTT GACTTGATTA ATGATTTGTC TTACAAACCT 8880

TTTTGCGTAT TAGTCGATGA ATGTCAATTT TTAAACAAGC ATCATGTCAT TGAATTTGCA 8940

CGAATTGTTG ACGAATTAGA TATTCCAGTA AtGGCATTTG GGcTAAAGAA TGATTTTAGA 9000

AATGAATTAT TTGAAGGCTC AAAGTATTTA CTTTTATATG CAGACAAATT AGAaGAaTTA 9060

AAAACGATTT GCTGGTTCTG TCATAAAAAA GCAACTATGA ACTTGCACTA TATTGATGGC 9120

AACCCCGTTT ATGAAGGGGA TCAAGTCCAA ATTGGTGGGA ATGAAGCCTA CTATCCTGTT 9180

TGTCGCAAAC ATTATTTTCA TCCAAAAACA GTGAATGAAG AACAGTAAGG AGAGACACTA 9240

TGTACGATCA ATTwCAATCT ATTGAAGATC GCTATGAGGA ATTGGGCGAA TTATTAAGTG 9300

ACCCAGCAGT CATTAGTGAT ACAAAACGTT TCATGGAATT ATCGAAGGAA GAAGCTAATA 9360

CACGTGAGAC GGTTGAGGTT TATCGCCGAT ATAAACAAGT TGTTGAAGGG ATTAGCGATG 9420

CCGAAGAACT TTTAAGTGAA AACTTAGATG CGGAAATGGC CGAAATGGCG AAAGAAGAAC 9480

TTTCTGATTT GAAAAAAGAA AAAGAAGTGT TAGAAGACCG CATTAAAATT TTATTATTAC 9540

CAAAAGATCC AAATGATGAT AAAAACATCA TCATGGAAAT CCGTGGTGCA GCCGGTGGCG 9600

ATGAAGCCGC ACTTTTTGCT GGCGACTTAT TTAATATGTA CCAAAAATAT GCAGAAGCAC 9660

AAGGGTGGAA AGCGGAAGTT TTAGAAGCGA ACGTCACTGG AATCGGTGGA TATAAAGAAG 9720

TAATTATGAT GATTTCAGGA GATAACGTCT TTTCAAAATT AAAATATGAA AGTGGCGCTC 9780

ACCGCGTGCA ACGAGTTCCT TCAACAGAAT CACmAGGACG GATTCATACA TCGACAGCAA 9840

CCGTTGTAGT TATGCCGGAA GCGGAAGAAG TCGAAATCGA ATTAGCTGAT AAAGATATCC 9900

GTGTCGATAT TTATCATGCA AGTGGCGCTG GTGGACAGCA CGTCAATAAA ACCGCTTCGG 9960

CTGTTCGATT AACCCATTTA CCAACAGGAA TCGTAGTCGC AATGCAAGAT GAACGTTCAC 10020

AGCTAAAAAA CCGTGAAAAA GCAATGAAAG TTTTGCGTGC TCGTGTGTAT GACCAAATTC 10080

AACAAGAAGC ACAAAGTGAA TATGATGCTA ACCGTAAATC GGCGGTGGGG ACCGGAGATC 10140

GTTCAGAACG AATCAGAACG TATAACTTCC CGCAAAATCG AGTAACGGAT CATCGTATCG 10200

GTTTAACCAT CCAAAAACTA GACCAGATTC TTGCTGGCAA ATTAGATGAA ATCATTGATG 10260

CATTAGTCTT ATATGATCAA ACATCGAAAT TAGAAGAGAT GCAAAATGGC TAAACGTTAC 10320

TTTGAAGTCC TTAACTGGGC TTCTTCTTTT TTAGAGGCGC AAGGCAAAGA AGGCTACGCT 10380

ATTCACTATG TTTTTTTAGA ACGTAAAGGT TGGGATAAAA CGCAGTGgCT CCTGCATATG 10440

CAAGAAGAAA TGCCCCAAGA AGAAGAGGAA CAATTAAAAA CGGATTTAGC CCAATTATTG 10500

ACAGATTACC CAGCACAATA TTTACTAGGA CAAGCTGAAT TTTATGGACA TTCCTTTATT 10560

GTTAATGAAC ATACTTTGAT TCCTAGGCCA GAAACAGAAG AATTAGTCGA GCGTTGCTTA 10620

AAAGCCAATC CAGATACTCC TTTGACGGTC GTGGATGTCG GAACAGGGAC AGGTGCCATT 10680

GCGGTTAGCT TGAAACTGGC ACGGCCCAAC TGGCGAGTTA TTGCCATTGA TCTTTCAGAG 10740

GAAGCGTTGA CTGTCGCTAA ACAAAACGCG CAAGCATTAG GTGCTGGCAT CGAATTTTAT 10800

CATGGGAATG GCTTACAGCC AGTTGCTTCT GAAAAAATTG ATCTGCTTAT TTCTAATCCG 10860

CCCTATATTA GTGAACAAGA ATGGTATTTA ATGGATGCAA GTGTCCGAAC GTATGAACCT 10920

AAAACCGCTT TATTCGCTGA AAATAACGGT TTGGCTTTAT ACCAACAATT AATTCATGAA 10980

AGTCAAACGA TGCTTAAAGC AGACGGTAAA ATTTATTTTG AAATAGGGTT TCAACAAGGC 11040

GCGGCCCTCC AAGAATTATT GAGCGCCGCC TATCCACAAA AAACAATTAA AATTGAAAAA 11100

GATTTATCTG GCAACGATCG CCTCGCGATA GCCGAGTAAC TAAGGTTCTT CGTCGCCACT 11160

TTTCTAGAAG GAGCGATTTG CAATGGAAAC AAAAAAATTA ACCAAACAAG AAATAGCCCA 11220

AGCAGCAGCA GCCATTCAAG CAGGCGAGCT AGTGGCTTTT CCAACAGAAA CTGTTTATGG 11280

CTTAGGTGCC GATGCCTTAA ATGAAGCGGC AGTAAAAAAA GTTTACCAAG TAAAAGGCCG 11340

CCCCAGTGAT AATCCATTGA TTGTTCATGT GAACAACGTT GAGATGGTCA AAAAATACGT 11400

GGCGACCTTA CCAGAAGTCG CTACAAAATT GGTGTCTGCC TTTTGGCCGG GACCATTAAC 11460

TTTAATTTTT AATGTGCCTA GCGATACATT TTCTAAAACA GTAACAGGTG GTTTAGAGAC 11520

AGTCGCTTTT CGTATGCCAG ATAACCAAGC GACCTTAGCT TTGATTGAAC AAGCCAATGT 11580

TCCGTTGGTC GGACCAAGTG CGAACACTTC TGGCAAGCCT AGTCCAACAT CTGCAGAGCA 11640

TGTTTATCAT GATTTGCAGG GGAAAATTGC TGCCATATTA GATGATGGCC CAACGCAGAT 11700

TGGGGTGGAA TCAACGGTCT TGGATTTAAC GGCGGAAGAT GGTGTGCCAG TTATTTTACG 11760

ACCAGGAGCA ATTACCAAAG AGCAGTTGGA AAGCATTATT CCCCAAGTGA AAATCGATAC 11820

ACATTTGATT AGTGAAACTG CAGCACCCAA AGCGCCAGGG ATGAAATACA AACACTATTC 11880

TCCTGATGCG GAAGTTTGGA TTATAAGCGG TCAAACAGCA GAATGGCAAG CGGCGATTCA 11940

GCAAGCAAAA GAACAACAGG AAAAAATCGG CTTATTTTTA AGTGATGAAC AGGCAGCACA 12000

ACTAGATACA GAAAATGCGT TTGTTTATTC TTATGGTGCA GAAACGGTTG AAAATGCCAC 12060

AAAAGAACTA TTTGCAGGGT TAAGAGCGTT AGATGAGCAA GGAGCTACCA CGATTTTTGC 12120

GCAAgGATTT GCAGAAACTG GTTTAGGCAC TGCTTACATG AATCGCTTGA AAAAaTCAGC 12180

CAATCAAAAA TTTTTTGAAA AATAAGAGGT ATTTGTAACG AAGCTGACAC AGAAGTATGA 12240

TACAATCAGA CAAAAGAAAA CAAAATGGAG TGTTGTTTGA ATGGmTTACA AAACGTATGA 12300

CCCAGATTTA TGGAATGCAA TTGCAAGAGA AGAAGAGCGC CAAGAAAATA ACTTGGAACT 12360

AATCGCATCT GAGAATGTCG TGTCAAAAGC AGTTATGGCT GCCCAAGGAA GTATTTTAAC 12420

GAATAAATAC GCAGAAGGTT ACCCTGGCAA ACGGTACTAT GGTGGTTGTG AATTTATCGA 12480

TATCGTAGAA AATTTAGCTA TCGATCGTGC CAAAGAATTA TTTGGTGCAA AATTCGCGAA 12540

TGTACAAGCC CATTCAGGTT CTCAAGCCAA TACAGCGGCA TACCTTTCAT TGGTTGAACC 12600

AGGTGATACC ATTTTGGGGA TGGATTTATC AGCTGGTGGT CACTTAACAC ATGGTTCGCC 12660

CGTTAACTTT AGTGGAAAAA CCTATAATTT TGTCAGTTAT GGAGTGGATC CTTCAACAGA 12720

AGTAATCGAT TACGATGTCG TGCGAATTTT AGCAAGAGAA CATCGTCCAA AACTAATCGT 12780

TGCAGGCGCA AGTGCGTATT CACGAACGAT TGACTTCAAA CGTTTCCGTG AAATCGCTGA 12840

TGAAGTAGAT GCCAAGTTAA TGGTTGATAT GGCGCATATT GCAGGCTTGG TAGCTTCaGG 12900

GTTGCaCCCA AATCCaGTTC CgTATGCTGA TATCgTAACA AGTACgACCC ATAAAAcATT 12960

GCGTGGTCCc TCGTGGCGGT TTAATTTTGA CAAATAGCGA AGAATTGGCG AAAAAAGTAA 13020

ATAGTAGTAT TTTCCCAGGC ATTCAAGGTG GCCCATTGGA ACATGTGATT GCCGGAAAAG 13080

CCGCAGCTTT TAAAGAAGCA TTGGATCCAA GTTTTGCTGA ATATAGTCAG CAAGTGATTG 13140

CAAATGCGCA AGCGATGACG AAAGTTTTCA ATCAAGCGCC AGAAGCTCGT TTAATTAGTG 13200

GGGCAACAGA TAACCACTTA TTACTGATTG AAGTGACAGG TTTTGGTTTA AATGGAAAAG 13260

AAGCAGAAGC TATTTTAGAT AGTGTAAATA TTACTGTCAA TAAAAATTCA ATTCCGTTTG 13320

AACAATTAAG TCCATTTAAA ACAAGTGGTA TCCGTATTGG AACGCCTGCA ATTACTTCAC 13380

GTGGTTTTAA AGAAGAAGAT GCGGTAGAAG TTGCTAAATT AATCGTTCAA GTCTTAAAAG 13440

ATCCAGAAAA TACAGCAGTT CATGATGAAG TCAAAGCAGC TGTAGCAGCG CTAACTAAAA 13500

AATATCCATT ATATAACTAA CATCAGGGGA TAGCTGAAGC GCCGATTCTC TGGTACAAAT 13560

AAGAGTCGGA GCTT 13574

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4528 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

ATTTACAGCG TTACTGAnAC AAGCAAAAGA CGCTTTAAGC AAATAAGCTT AAACAATCTT 60

TCACACCCTA CTCAAGGAGT AGGGTGTTTT TTTAATAACT CCGTTGGTAA CTTTCATCTA 120

TAAATGAAAA AACTTGTGAG AGCGGAATTT CTTCAGAAAG ACTTAACCAA TGCTCTTTAT 180

TCATATGATA GGCAGGAAAA TAGCCAGGTT GGTTCTTTAA TAAGGCAATC AATTCAGGCG 240

CAAGTTTAAC GGTTAAAATA GTGAGCTGCT TATTCTCGGT AGCGCCGAGC AGTTGTGCAT 300

CCAGTGGCAT AATTAAAGCA AACCATTTTG CATTTCTAGG ATGACGAAAG ACGAGATAGT 360

CTGGGTGTTT TTTAAAAGGA AAATCGGGCG TTATCTGATA ACGTTCTTGG ATATAGGCGC 420

TAATAGTTTC TTGAAAAGTC ATAGGTTTCT TCCTTTCGTT AAGGCTAGGC ATATCTATTT 480

ACATCATAGC AAAAATTACA TAAAATTACT TCGTTCAAAG GAATGGGAGG AGGGAGCAGA 540

```
TGACATTAAT TTCAACAATT TTAGTAACCT TGGTGGCTGT AGAATTTTTC TACATTATGT    600
ATTTAGAAAC GATTGTACCA ACATCTGAGA CAACAAGCCG TGTTTTTAAA ATGGATAGGC    660
AAGAGTTACA AAGArAATCA GTCACTACAT TGTTTAAAAa TCAaGGAATC TATAATGGAC    720
TAATCGGTGC AGGACTTATT TATAGTGTCT ATTTTGCGCA AgCAACGATG GAGATGACAA    780
AATTTTTATT GATTTATATT ATTTTAGTGG CGGCGTACGG TAGCTTAACT AGTGATAAAA    840
AAATCATTTT AACACAAGGC GGTTTAGCTA TCGTGGCTTT GATTAGTTTA CTTGTTTAGC    900
CATCAGAGTT AATAGAGAGA GGCAGTCAAA TCGACAGAAA AAACACGGTT GATTTGACTA    960
TTTTTTATAG AATTGAGAGA GACAAAATGT GGTATACTTT GTTTATTAGA TAGGGAGGAA   1020
TTAAAAATGC TGAACCAAAA AGAAGTAAAA AAAACCGATT ATACTCAATT AATTTCTCTA   1080
TGGCGTTCTT CAGTAGAAGC TACCCATTTG TTTTTGAGTC AGGCAGATAT TGATAAAATC   1140
GAAGTGGTTT TACCTGATTA TTTTCAACAA GTTCAATTAA GTATGTGGTT AAATGAAGAA   1200
CAGAAaTGTG TGGGCTTTAG TGGAACAAAT CAACAAaCGT TGGAAATGTT ATTTATTGAC   1260
CCAGTCTATT TTCGTAAAGG CTATGGTGGa GAAATTATCC mAAAGTTAAT AGAGCAAGAA   1320
TCAATTATCT TTGTTGATGC TAATAAACAA AATGAAGGGG CAGTGAAGTT CTATCAATCT   1380
CAAGGGTTTC AAGTAATTGG TGAATCAAAG GAAGATCCTC AAGGAAATCC GTTTCCTATT   1440
TTACATATGA AACGGATATA GAAAAGAAGG GAGAAGCCAT GCGAGTAATC GTTACAGAAT   1500
ACCAACCAGC CTGGGTGGAA CAGTTTGAAG AAGAAGCCCA AGCGTTGAAA CAGATTCTAA   1560
AGGAAAATTG TCTTAAAGTT GAACATATCG GTAGCACATC TGTGCcTAAt TTAGCAGCCA   1620
AACCAATTAT TGATTTTTTG GTCATTGTTG AAGAAATTGA AAAAGTAGAC CTGTTACAAT   1680
GGGAATTTGA AAGAATCGGT TATGAATATA TGGGGGAATT TGGGCTATCA GGACGTCGTT   1740
ATTTACGAAA AGGCCCGATA AAAAGAACAC ATCATGTGCA TATTTATCAG TTTGACAATA   1800
CGCAAGAAAT TTTGCGCCAT CTTGCTTTTC GAAATTATTT ACGAGAAAAT CCGGCAATTG   1860
CTACAACCTA TGGTACATTA AAAAAGCAAT TGGCCCAAGC ACACCCCGAT AGCATTGATA   1920
AATATATGGA TGGCAAAGAT GCGTTCATTA AGAAAATAGA AAAAGAAGCG TTGAAGAAAT   1980
ATTGGGAAAA ATAAAAAAAG TTGCTAAGGC TAGCAACTTT TTTTTGGTTC TTTTTTGAAA   2040
AAATGATAAA GAAGACCACT TAAAAGAACA CCTAGAATAA TTCCAATCGC TAAATTATGC   2100
GTATAGACGA TCACACCCAC TGTCACGAGT AAAATAACAA TTTCAGTAAT TTCAAAAGTT   2160
CGGAAAAGCT GCAAACTTTC CCAATCAAAG GTGTCGACTG CCACTGTCAT CATGATCCCA   2220
ATCAATGCAG CTGTAGGAAT GGCTACCATC ACGGATTTAA AAACAAAAAT GAATAGAAGG   2280
AGGGTAATAC CTGAAATCAA AGTAGAAAGC CGTTTTCGTC CTCCAGATTT AACATTAATG   2340
ACCGCTTGGC CAATCATCGC ACAGCCAGCT TGGCCACCAA AAAAGCCGGT AATTATATTA   2400
GCGAATCCTT GTGCTTTAAC TTCTCGTTGG CTATCACTTT GGCTCGCTGT CATTTTGTCC   2460
ACAATGGGGA TCGTCAATAA AGATTCAATT AACCCGACCA TTGCTAAGGC AACAGAGCTA   2520
```

GGTAAAATAA TCCACAACGT TTCTAAGTTG ACGGGTACAT TAGGCCAACG GATCGCTGGA 2580

TGAAACTGAC TCATTTCACC AAGATCACCT ACCGTTTGCA AATGACCTTT TAAAAAATAG 2640

GACAGGATAG TCATGAATAA AATAACAATT AAAGCAGGAG GCACAATTGT AATAAACTTA 2700

GGTAATAGAT AAATAAGTAA GATACTACAG ATAATCATAC CGTAGGACCA CAGTGTTTGA 2760

TGGGGCAATT GTTGTACTTG CGCCATGAAA ATTAAAATGG tAACGCATTG ACGAATCCGT 2820

ACATCACTGT CTTGGGAATA AAGCGCATCA ATTTGTGGAT ACCTAAATAA CCTAAAATAA 2880

ATTGTAGAAA GCCTGTTAAA ATAGTCGCAG CAAGCATGTA TGACAAACCA TGTGCTTGAA 2940

TTAAGCCTGC aAGGACTAAG GCCATTGAGC CAGCAGACGC CGAAACCATT GCCGgACGGC 3000

CGCCCGTAAA AGTAATCACA AGTAAAGTAA cAGctGAAGC AAACACTGCA GAAAGCGGGT 3060

GAACACCAGC AATAATCGCA AACCCGATTA CTTCTGGTAA AATGGCGACA GAAGAAACGA 3120

GACCGGCAAA GAGATCATTT CGTTCATTGC CTAACCATTC GCTTTTTTTA ATACTAAATA 3180

ACATCGAAAA ACTCCTTTAA GTTAATAAAA CCTTATTCAT AATAGCATGA ATAAGCTGAA 3240

ATGAGCGTGG GACAAAAATC ACTTTGGATT TTTGCTCCaC GCTCaAAAAC TGATAAACGG 3300

CGGGAACAAA AGCAACTCCT TCGGAAATAA GCCGGAATTC TCCAAAAATT ACAGAACAAT 3360

TTTCGGAAAT TCCTTCTTAT TTCTCGGAGT TAAACGCTTC TGTCCCGACC TCATTAGAAC 3420

TGTAGCAGTG CCAACAACAA AATGGCTAGG ACGACTTGAA CAGTTCCTAC TGATAAACCA 3480

TAGATTAAGA AACGTTTGCC TTCTTGGAAA AACTTTTTAA AATCGAGTCG TAACCCGATT 3540

GCTGCCAAGG CAGTAATTTC AAACCAAGAA CTAAAGAAAT GAGCAGTCTC ACTGATCACG 3600

ACGGGGAAAT GAATCAAACT ATTAAAGACA CAGGCAATGA AAAAGCCAAC TACATACCAA 3660

GGTAGGGCGC TGCTTTTTTT GGTGACTTCT ACCAACTCAG CCTCTGATTC TGCCGTCTTA 3720

CTTTGCTTGA AACGTCCAAA TAAATAGACA ACTGCTACGA GTAGGACAAT CCGCATAATT 3780

TTAAAGAGCA TCGCAAGTTG AAcTGCGTTC TCATTGACCA TATTGGCACT GGCGACAACC 3840

TGACCAACTG ATTGTAAAGT CCCTCCGATT AGTGCGCTAC GAGCCAACAA ATTGGTACCA 3900

TAGAGAATCC CACTTAAAAT GGGTAAGGTC AACATTAAAA CGGTTCCTAA TAAGTTTACT 3960

AAAGTAATGA TTTGGCCTTT TTCTTCCTCA TCTGCTTGAA TTGCTGGAGC AATCGAAGCA 4020

ATAGCAGAAG ACCCACAAAC GGCATTACCA CCAGCCATCA ATAAAGACAT ATTATCAGAA 4080

AAAGCCAGTT TTTTACCGAT CAAATAGGCA AAAATAATGG TCAGCGACAT TTGAATGAGA 4140

ATAAATGCGA CACCTTGTAA GCCAATTTGG GCAATCGTTT GAAAGGTCAC GGTGGTTCCC 4200

AATAAAACAA CAGAAAATTC TAGTAATTTA CTTTCGGCTA CTTTGGTTCC GCGATTTAAA 4260

TTGGCACCGC GAACAAAGGT ATTTCCCAAA AAGATTCCTA GTAAAATAGC TATCGTAGCG 4320

GCGCCTAAGC TAGGTAACCA GATGGCAAGA AACTTGCTAA GGCAGGCGAC AAGGAAAGCT 4380

GTTAGCAAGC CGGGAAGAAT TTGGAATAAA GATTGGATAT AATCATTTTC TGAATTTTTC 4440

ATTGAAAAAA TCTCTCTTTC TTGTTCATTA CTTCCAGTAT AACGAAATTT TTTACATTTG 4500

AGAnATAAAT AGTTAAnATA ATATTTAT 4528

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5147 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

GAAGTCAGTG AAGCGAAATT AAAAGAAGCA GCCAATCTTA TGCCAGAAGA GCCGTTAATT 60

CAATTTGCTT TGGGAGAACT ATACTTTACA AATGGTCAAT TTGTCGAGGC TATTACTCGT 120

TATCAAAGCA TTGTAGAAAG TGGGACAGCG CAAATCTCTG CTATTTCCTT GAATGAACGC 180

TTAGGAAGTA GTTACAGCAT GTTAGGTGAT TTTGAAGAAG CGGTTCCTTA TTTGGAAGCC 240

GCAGTCAAAG AAGAACAAAC AGACGATCGT TTATTCCAGT TAGCTTTCAC TTATTTACAG 300

TTACATGAGA ATCAAAAAGC CATTGCGTTG TTCCAGCAAT TAAAAACACT AAACCCACAT 360

TATCAATCAC TTTATCTATA TTTAGCGGAA GCGCTCCAAG AAGAAGAGCA ATTAGAGGAA 420

GCAAAAGCGG TGATCGAAGA AGGAATTGCG GAAAATCCCT TCCAAGTAGA GTTATACCAA 480

TTTGCCTCTG AAAATGCCTA TCGCTTACAC GATATTGCTG GTGCTGAAAA CTGGTTGTTA 540

AAAGCTTTGG AATTAGGTGA GAAAACAGAT GAAAGTCGTT TGACCTTAAG TAATCTCTAC 600

ATCACGGAAG AGCGTTTTGA GGAAGTAATT ACTATCTTAA ATGATTTAGA GGAAACGGAT 660

CATCCTTATG CTGAATGGAA CTTAGCAAAA GCGTACAACG AATTAGAAGA CTTTGCCGTC 720

GCTAAAGTTC ATTATGAACA AGCAGCGCAA GAATTGTCTC ACGAACCAGA ATTTCTAAAA 780

GAATACGCAT TGTTCTTACG AGAAGAAGGC GAACTTGAAA AAGCAAAACA CTTGTTATAC 840

CACTATTTGG AACATGAACC TGGTGATGTC GACATGCAAT CTGTGCTAGC AGATATGGAA 900

GAAAGATAGG AAACGACGAT GTTTGTTAGT TTATCAGAAA AGAAAAAATT TTTAACTTGG 960

TTAGTGAATA CCGCGCCTTT TGGAAGAAGA GAAGTACTGT GGATTTTAAA TTACTTATTA 1020

ACGCACGATG CGATTTTAAA TAATGTTCAT TTTGTTGAGA ACGTAGAAAA AACCGATCGG 1080

GGGATTCGTG TGGTAGCGGA TGGTTTAGGT AAGGAGCCGT TATTGCTTTT TATCCAAGCA 1140

CAAGAATTTA CAGATCCGGA ACAAATTTTT CACGAGATTC GGATGAATTG GCGGAAAGCT 1200

CTTTATTTAG AATGTGTATT TCCAGAGGCT TGGCAAACGA GCCAGTATTT ATCTGTTTTA 1260

GAAGATAATC CGTTTGCGCC ATGGAATGAA CAAGTGGATC AGGAAGTTGC CAGAGCGATT 1320

GATCAATATT TTAAGCAAGA AGAACAAACG CAGCGAATGG cATTACTTAA AGCTCAAATT 1380

GATGATGCTT TAGAAACAGG CAACAAAGAA GCGTTTTTAG AATTATCCGA TGAATTAAAT 1440

CGTTTGAAAC AACAATAAAT AAAACCCAGC ATTTATTGAA AATGCTGGGT TTTATTTATT 1500

GTTTTAATTT AAAGCGCCGT TGTTTGCTGG GTTTTGCATA AGTAAAGCCT TCGCCAATTG 1560

CTTCATGTAC GTTAATGATG GACACAAAGG CTTTTTCATC TATTTCATTA GCAATCCGTT 1620

TGACTTCCAT TATTTCTCGA GGGCTGACTA CCGTGTAGAG AACCCGTTTT TCCACTTGTG 1680

AATAAGCACC TTCACCATTT AAATATGTTA CCCCACGTTC TAAAGAAGCC ATTAAACTTT 1740

GACCGATTTC TTCGGAATGA TTGGAAACAA TTAAAATTCC TTTCGCAGCA TAGGCTCCAT 1800

CGAGTACAGA ATCAACAACA CGACTGAAAA CAAAGGACAC AATTAAAGTA TACATCATTC 1860

GTTTGACATC AATATAGGTC AACGAGAGGA GTAAGACTGC AACATCAAAA AGTAACAACG 1920

AGCGGCCCAT GCTAATCCCA AAGTATTTTT CGAAAATGCG GGCAATGACG TCGGTCCCAC 1980

CAGTGGTTCC GCCTACGCGG TAAACAAGGC CACTACCGAA TCCTGCGGCC AAACCAGCTA 2040

AGAGAGCAGC AATCAATAAA TCATGATCTA AATTAATTTC AACAGGAAAT TTTTGCCAAA 2100

ACCACAAAAA TACCGATAAC GACACGGTAC CAATTAGCGT ATAATAGAGG GAACGTTTGC 2160

CTAAAATTTT CCCACCAATT AATAAGAGCG GAATATTGAT TAATAATGTC GTATATGCTG 2220

GGTTAAAATG AAATAAGGCT CTTAAAATCA AAGTAATCCC CGTGACACCA CcTTCGGCCA 2280

GGTCATTGGC AATATTGAAA GTAACCAAAC CAAAGGCATA GATACAGCTT CCTAATACAA 2340

TAAATAACAC ATCTTTAAAA TAAAATCTTT GTTCTTCCAC TTAATCACGT CCTTTTACAG 2400

TATAAACTGT ATCATTCTCT ATATAAAACA ACAAGTAGAA CCTTTCCAGA TAGCCAAGTT 2460

TTTGATAAGA TGTAGGAGGT AGAGTAAGAA AGGTTGGGGT AAAATGTCAG AATCAGAACA 2520

ACGTAGTTTG CAATCGATGC AAACAGAAGT TGATGATTAT ATTCAACAGT TTAAAACAGG 2580

CTATTTTTCA CCATTAGGAC AAATGGCTCG TTTGACGGAA GAGGTCGGAG AATTAGCCCG 2640

TGAAGTCAAT CATTATTATG GTGAAAAACA GAAAAAAGCA GACGAAAAAC CAAAGACAGT 2700

ATCAGAGGAA CTTGGAGATG TCTTTTTTGT ATTGATTAGC ATGGCAAATT CTTTAGAGAT 2760

TGATTTAACC GAAGTTTTTG AAGAAAATAT GGCCAAATTT AACAAGCGGG ATCGCTATCG 2820

CTTTGAAAGA AAGGATGGGA AAACACATGA TTAAAATAAT TGTCGCTGGG TTTAAAGGAC 2880

GGATGGGCAG TACTGCTACA CAAATGGTCT TAGAAACCGC AGACTTTGAA TTAGTAGGTG 2940

TGTATGATCC TCATGAAGCA CAAGAAACTG TTTCTTTCAA TGATGAAACA GCAATTCCCG 3000

TTTTTCAGCG CTTAGAAGAA GTCCTAGCTG TCAAACCAGA TGTTTGGATT GATTTTACGG 3060

TTCCTGAAGC CGCCTATCCA AATACGCGTT TTGCGTTGGA ACACGGTATG GCGCCTGTTG 3120

TTGGGACGAC AGGCTTTACG GAGGAACAAA TCAACGAATT GACAAACCTA TCTCGAGAAA 3180

AAGCCATCGG CGGCTTGATT GCGCCAAATT TTGCAATTGG GGcAGTTTTA ATGATGCAAT 3240

TTGCCCAAAA AGCGGCGCAG TATTTTCCAG ATGTGGAAAT CATTGAATTA CATCATGATA 3300

ACAAATTGGA TGCACCGAGC GGAACGGCTA TAAAAACTGC AGAAATGATT CAAGAAGTTC 3360

GGCCAGCAAa AAAACAAGGG AAcCGCAAGA GGTAGAATCA ATACCAGGGG CACGCGGGGC 3420

TGATTTCGAG GGTCTGCGAA TCCATAGTGT CCGCTTACCT GGTTTAGTGG CACACCAACA 3480

AGTCCAATTT GGCAGTGTAG GTGAAGGGCT CACTATTCGG CATGATTCTT ATGATCGGCG 3540

TTCATTCATG ACAGGAGTAG CGTTAGCTTG tCGGCAAGTG GTACAGAGAA CAGAACTACT 3600

TTACGGATTG GAACAGATGC TATGAAATTG ACAACTATTC CAAACGAATT TAAAGAAGCC 3660

GCTCCCGTTA TTCGTGAAAT CAATGCACAG GGCTTTGAAG CGTATTTTGT CGGGGGCAGC 3720

GTCCGTGATG CTTTATTAAA TAAACCAATC CATGATGTCG ATATTGCGAC AAGTGCCTAT 3780

CCAGAAGAAA TTAAGCAAAT TTTTAAACGA ACCGTTGATG TGGGCATCGA ACATGGCACA 3840

GTCCTTGTAT TGATGGAAGA CCAACAGTAT GAAGTAACAA CCTTTCGTAC CGAGTCAACT 3900

TATCAAGACT TTCGCCGACC AGATGAAGTC ACCTTTGTTC GCTCCTTGAA AGAAGATTTA 3960

AAACGGCGAG ATTTTACGAT TAATGCCTTG GCGCTCGACA GCACAGGTGA AATCATTGAT 4020

TTATTTGATG GTATCGAAGA TTTAACTAAT CAAACCATTC GTGCAGTCGG TAATCCCCaT 4080

GAGCGGTTTC ATGAAGATGC TCTGCGAATG ATGCGGGGcT TACGCTTTGC TAGCCAATTA 4140

GATTTTAAAA TTGAAGAAAA AACCTTAGCC gCTATCGCTG AATTTCATCC TTTATTGGAA 4200

AAAATTTCAG TAGAACGCAT TACAATTGAA TTTGTAAAAA TGCTTTTAGG GGTTAATCGA 4260

CAAGGTGGCT TAGCTCCGTT TATTGAAACA GAATGCTATC AATACTGTCC AAAATTACGA 4320

GAACAAGGGG CTGGACTTTT TCGTTTAATG GACTTGCCTG CTCGTCAAAT TGAAACAGAA 4380

GCAGAAGCGT GGACGTTACT TATCCAATCG TTGAATTTAC CAGAAGCAGA AATTCGCTCC 4440

TTCTTAAAGG CTTGGAAACT TTCAAATCAA TTAATCCAAA ACGTTTCACA ATTAGTACGC 4500

GGTTTGCGTT TTCGCTTGAG TAACGACTGG CAGCCAATGA TGCTTTATGA ATTAGGCGAA 4560

GAATCAGCAG TACTTGTTGA AAGGCTTTTA TATTATTACC AACAAGAGAG TCAAGTACAA 4620

GTAACAAAAG AGTTAGTTAA GGCATTACCC ATCCATCAAA GACATGAATT GGCCATAACA 4680

GGAAAAGACC TTTTAGCCGT CTTAGAAGAA ACACCAGGCA AATGGTTAGG GGAACTGATT 4740

GCTGAAATCG AGCAACACGT TGTTGAAGGT AGCTTAGAGA ATAAACAAGA AGTTCTGCTT 4800

TCGTTTGCCA AAAAACAACG TTCAAAAGGA GAGAAAGCAT AATGCCAACA GAAATGTTTA 4860

CAAAAGAATT CTTGGTGAAA GAATCACAAA CGGCTArAAT GCTCGGTTCT GGTGACTTAG 4920

AAGTTTTAGG AACCCCTGCT TTAGCTGCCA TGATTGAACA AaCAGCCAAA GAaGCTGTGA 4980

AGGACCAwTT ATCAGTTGGG GAAaCAaCTG TCGGAaCGGT GTTAGAATtG CGCCACTTAT 5040

TTCCATCCGC GGTAGGGGCG ACAATCGTTG TTACGATGAC CAGCATTGAA CAGACTGCAC 5100

ACAAAATACG CTATGAATTT GTTGCTTATG AAGGAGAACG ACAAATT 5147

(2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12022 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

AAGAATTTAG ACAACTATTA GCCGCAAAAG GCATTGAATT GACAGACCAA CAGATGACAC 60

AATTTGACCA ATATTTCCAT TTACTAGTAG AATGGAATGA AAAAATGAAT TTAACAGCCA 120

TTACAGAAGA AAAAGAAGTT TATTTGAAAC ATTTCTATGA TTCTGTTTCG TTGGCaTTCT 180

TTGAAGACTT TGCCAGTGAC AAAGCCATTT GTGATGTTGG CGCAGGCGCA GGATTTCCAA 240

GTATTCCTTT GAAAATTGTT TTTCCTTCAT TAAATGTCAC AATTGTTGAC TCTTTAAATA 300

AACGAATTAC TTTCCTAACA GAGTTGGTGA ATCAACTAGG TTTGGCTAAT GTTTCTTTGT 360

ATCATGATCG TGCAGAAACA TTTGGTCAAA AAGCAGAATT TCGTGGCGCG TTTGATTACG 420

TTACAGCGCG AGCAGTTGCA CGATTAAATG TTTTAAGTGA ACTGTGTTTA CCCTTAGTTA 480

AAAAAGAAGG CTATTTTTTA GCCTTAAAAG CATCGAAAAG TGAAGAAGAA ATTAATGAAG 540

CAAAACCTGC AATCGCTACA TTAGGCGGTC AATTTCAAAA AGAAGTTGGT TTTACTTTGC 600

CGATTACGGC AGACGAACGC CACATTGTCG TGATTCAAAA GAAAAAAGAA ACCCCTAAAA 660

AATATCCCAG AAAACCTGGA TTACCAAATA AACAACCGAT TAAATAAGAT GGAGGCACAC 720

ACATGACACG AATTATTTCT GTAGCGAACC AAAAAGGCGG CGTTGGTAAA ACAACGACCA 780

CTGTAAATTT AGGCGCATGT TTAGCAAACT TAGGCAAAAA AGTCTTACTG ATCGACATTG 840

ATGCACAAGG AAATGCTACC AGTGGCATGG GTGTTCCTAA ACCAGATGTC GCACACGATG 900

TGTATGATGT CCTTGTAAAT GAAGAACCAA TCACTAGCGT GGTTCAACAT ACCAGCCGTG 960

AAAATTTAGA TATTGTCCCT GCAACTATTC AATTAGCAGG CGCTGAAATT GAATTAACTT 1020

CAATGATGGC TCGAGAATCG CGCTTGAAAT TAGCAATTGA TGAAGTAAGA GACATGTATG 1080

ATTTCGTTTT AATTGATTGC CCGCCTTCTC TAGGGCATTT AACAATCAAT GCTTTTACAG 1140

CAAGTGATTC GATTCTAATT CCTGTTCAAT GTGAATATTA CGCATTAGAA GGCTTGAGTC 1200

AATTGCTAAA TACGATTCGC TTGGTTCAAA AACATTTTAA CCCTGAATTG AAAATTGAAG 1260

GGGTCTTGTT AACCATGTAC GATGCTCGGA CCAATTTAGG TGCCGAAGTC GTTGAAGAAG 1320

TCCGTAAATA TTTCCGTGAA AAAGTCTATG ATACGATTAT TCCAAGAAAT GTCCGTTTAT 1380

CCGAAGCACC AAGTCATGGT TTACCAATCA TTGACTATGA TATTCGCTCA AAAGGTGCCG 1440

AAGTGTATCA AGCACTAGCA AAGGAAGTGT TGGAGAATGA ATAAAGGCAA AGGTTTAGGc 1500

AGAGGCATCG ATGCCTTGTT TCAAGATATC GCAAaGCTAG AAGATGTCGA CGTTAAGAaC 1560

GAACAAGTGA CAGAAATTCT ATTAAACGAA CTTCGTCCCA ATCCTTACCA ACCGCGGAAA 1620

ACATTTGATG AAACCTCTTT ACAAGAACTT GCAAACTCGA TTCTACATTC CGGTGTGTTT 1680

CAACCGATTA TTGTTCGTAA ATCGGCAGTT AAAGGCTATG AAATTATTGC AGGTGAGCGT 1740

CGGTTCCGGG CTTCTAAATT AGCCGGCAAA GAAAAAATTC CCGCAATTAT TCGCGAATTT 1800

GATGAAGAAT CCATGATGCA AGTTGCTGTT TTGGAAAACT TACAAAGGGA AGATTTAAAT 1860

CCGCTAGAAG AAGCAGAAGC CTATGAAATG CTAATGAAAA ATTTAAAATT AACACAAGCG 1920

GAAGTCGCAG AACGTTTGGG CAAAAGTCGC CCGTATATTG CCAATTATTT ACGGTTGCTA 1980

ACTTTACCAG ATGCGGTCAA AGCAATGGTT CAAAAACAAA GCATGTCAAT GGGTCAAGCA 2040

CGGACGTTAT TAGGCTTGAA GAACAAAGAA CAGCTATTGC CATTAGCTAA TCGTTGTATT 2100

AAAGACAATT TAACGGTTCG ACAATTAGAA CAACTTGTCG CTGAATTAAA CGAAACACAA 2160

GGCAAAAAGG GCAAAAAAGC GAAAAAAGCC ATCAAGGAAA AACCCATTTA TATTCGAGAA 2220

AGTGAAGACC GCTTAATGGA TAAATTTGGG ACAACCGTAG CGATTCAAGA AAAAGAAGGC 2280

AAAGGAAAAA TTGAAATTGA GTACTTGTCC TCTTCGGATT TGGCTAGAAT TTTAGATATT 2340

TTAGATATTC ATTTTGACGA GGAATAAAAA GCAAAGTGAG AAAGTAAGAC AGAAAGTCAC 2400

AAATCCTTTT GCTAGTTTCG GTAAAAAATG GTATTGTAAT CATGGTATTT TTATATAAGT 2460

AAGAGAAGAA AGAAGGCTTG GAACATGTAT GATTTAGGCG ATATTGTCGA AATGAAAAAG 2520

CCGCATGCTT GCCAAGCAAA TCGTTGGCAA ATCATTCGCA TGGGCGCTGA TATCAAAATC 2580

AAGTGCACGA ATTGCGGGCA TATTGTCATG ATGCCGCGTC GTGATTTTAC TAAAAAATTA 2640

AAAAAAGTAA TCGAGAAAAA AGCAGACTAA TGCTTTGATG CCTCTAATAG AAAAGAGAAT 2700

AGACTGAGAG AAAGAGTGGA AAACAAATGG CACTAACAGC TGGAATTGTC GGCTTACCAA 2760

ACGTTGGTAA ATCGACCCTA TTTAACGCAA TTACAAAAGC AGGAGCAGAA GCTGCAAATT 2820

ATCCCTTTGC AACGATTGAT CCGAACGTCG GCATGGTTGA AGTACCAGAT GCCCGTTTAC 2880

AACGCTTGAC AGAACTAGTA AAACCGAAAA AAACAGTTCC CACAACGTTT GAATTTACAG 2940

ACATCGCCGG AATTGTTAAA GGTGCCAGTA AAGGAGAAGG GCTAGGAAAT CAGTTTTTAA 3000

GTCACATTCG TCAAGTAGAT GCAATCTGTC ACGTGGTTCG TTGTTTTGAT GACGAAAATA 3060

TCATGCGTGA ACAAAATCGG GATGCGGATT TCGTTGATCC TTTAGCAGAC ATTGATACAA 3120

TTAATTTAGA GCTAATTTTA GCAGATTTAG ATTCAATCAA TAAGCGCTAT ACTCGCGTGG 3180

CTAAAATGGC GAAAGCAAAA GATAAAGAAG CGGTGGAAGA ATTAGCTGTT TTAGACAAAA 3240

TTAAACCAGT TTTAGAAGAA GGTTTATCTG CTCGTACGAT TGAATTTACG CCAGAAGAAG 3300

AGAAAATTGT TAAATCTCTT TTCTTATTAA CCACAAAGCC GGTCCTTTAT GTAGCAAATG 3360

TTTCAGAAGA TGAAGTGGCG GATCCAGACA ACAACGAGTA TGTACAACAA GTTCGTAATT 3420

TTGCAACCAG TGAAAATGCG GAAGTAATCG TAGTTAGTGC GCGAGCAGAA GAAGAAATTG 3480

CTGAATTGGA TTCTGAAGAA GACAAGGCTG AATTTTTAGA AGCAATGGGC ATTGAACAAT 3540

CTGGCTTAGA TCAATTGATT CGTGCGGCTT ACGATTTGTT AGGCTTAGCT ACCTACTTTA 3600

CAGCTGGTGA ACAAGAAGTT CGTGCATGGA CGTTCCGTAA GGGAATCAAG GCACCTCAAG 3660

CAGCTGGGAT TATCCACAGT GACTTTGAAC GAGGCTTCAT TCGTGCGGAA ACAGTTTCTT 3720

ATGAAGACCT AGATAAATAT GGCAATATGC AAGCTGCCAA AGAAGCGGGC CGTGTTCGTT 3780

TAGAAGGAAA AGACTACGTC GTTCAAGATG GCGATGTTAT GTTGTTCCGT TTTAACGTTT 3840

AAAAAAATGT GCCCAACAAG CAATTTGTTT TGGGATAAGC CCACATATTT AAGGCCATTC 3900

AGCCGTTTAG AAAGAGGTTC CCTCTACATG AGGGTGAGGA GGATACAACT ATGGAAGCAG 3960

AAGCATTACG AGCAATCGTG GCTGAGAATC GTCAATTAGA ACAGAATTTG ACTAAAAGAA 4020

```
ATGAACAATA CATTTTTGAC TTAAAAAAAT CATTAAAAGC GGCTAATTTA TCTGAAGAAG   4080
AGCTAGCATT GGCTCTACAC GGTATTTTGC CAGAATTAGT AGCCGGTCAA AAAACAGGTA   4140
AAACAGCCCG TCAGTTATTT GGCACTGTTT CAGAACGGAC GGAAGCAATT TTAAATAAAC   4200
CAGCAGAAGT CAAAGAACCA GCGGGCTGGA TGATTTGGCT CGACAATACC TTACTATTAC   4260
TTGGTTTATT AACAATCATG TTAGCGGCGA TGTCTCTATT TTCAAAAGGA ACCGCACAAC   4320
CTCTCGGATT AACCACATAT ATCTTAGGTG CGATGGCTGG GGGTTATGTT TTCTATCTAA   4380
TGCACAAATA TGTTTACCGT TTTGATCGTC AAGGTGGCGA CAAAAGCAAA CGTCCTGGTT   4440
GGTTAAAAAC GACCGCCATT CTATTTGGTG GGATGTTCTT GTGGATTGCA GTTTTCGCTG   4500
GTTCAGCGAT GCTACCACCA GTGATCAATC CCATCTTAGA TCCGATGATT GCTTTAGTAA   4560
TCGGTGCGTT AgcCTTTGTA GCACGCTATT TCTTCAAGAA AAAATACAAT ATTCAAGGTA   4620
GTTTCATGAC ACGTCAATAA ACGGCGCGCG CTCAATTGTG AATGCGCTCG GcTGACGCAG   4680
AAACTGCAAC AAGCAAAGTC GCTCTCTCAA TAGAGCGTTT TAACAGTTAA AGCCTGGGGG   4740
ATTAGTCGAA AAAACTAGTG TCCCAGGCTT CTGTTTTCAG AAGTTTGCTA AAGAAAGGAG   4800
CCTCTTATGG GACAACAAGT GAAGTCTTGG TGCCGTCAAC ACGACCGGCT ACTCATTTTT   4860
ATTAGTTTTC TCTTATTAAG TGGACTTTCT CTCTATTTTG TCTATTTTCA AGAGAATTTT   4920
TTACGAGCAA GTGATTTCCG CTTTCATCAA AATCGGGTGG AAGGTCTCGC ATTAGCAATT   4980
AAAAACAACG ACTGGTTTCC TAAAATCAAT TATTTTTTCT TGGGAGGCTA TGGCTATGCC   5040
TCTAGTTTGT TTTATCCAGA TGCGTATCTT TATTTACCTG CGCTTTTGCG TGTTCTGGGG   5100
ATTTCCTTTG TCGCTAGTAT GGCGATTTTT GTTTTTGCGG TAAACTTGGC AACATTTAGT   5160
CTTACTTATT ATGCAGGCCG ATTAATGGCG CTCTCTAAAA AAAGAAGCTA TCTCTTTGCC   5220
ATTTTGTATG GCTTATCTAT TTACCGCATG CAAGACTTGT TTAATCGCCA AGCGTTAGGA   5280
GAATTTTTGG CGCTAAGTTT TTTTCCGTTG GTCTTAGCGA GCCTGTTTTT ATTACGAAAG   5340
GGAACCACGA AATGTTGGCC GTTGTTAACG CTGGCGATGA CTGGGATTGG GTTGGCTCAC   5400
TTTATTTCTA TTGAAATGGT TTCTATTTGG ATTGGGTTAT ACATTCTGTT TTATTGGCAG   5460
CAGTTTTTTA AAAAGGAGGT CCTGTTGGCA TTAGCGAAGG CTGCAGGTTT GACGCTGCTT   5520
TGGCTGGCCT TTTATCTTTT ACCAGTTGCT GAACAAATGA AAAATCAAGT GTTTAAAGTT   5580
ACGTCCAATC CACTCACTTA TATTTCCGAA AGAAGTTATC CTATTGTTTC TCTTTTCATT   5640
AATAGTTTAA AGAGCAGTGT TTTTCACGCG AAAACAGCGA ATCTAGGAAC CCTTTTGTTT   5700
GTGGGGCTTG TTGTAGCAGT GGTTTCGCTT GCCAGCAAAA AAATTCAGAA CAAACGGTTT   5760
ATTGGTTTGA CGCTGGTACT CTTATTGATG GTTACCACGC TTTTTCCTTG GCACTGGCTG   5820
AACCACACGC CTTTAAATAC AATTCAGTTT CCTTGGCGGT TTTTAGGGAT TCTTTCTGTA   5880
ATGCTCGCTT TTTTCATTGC ACAAGACGAA TGGGGCGTTT TCCGTAAATC TTGGACCGTG   5940
GCTTTGTTGG TTTTCCTTGC CATCAGCAAT TTGGGCATTT ATCAATATCA AAGTATTCAG   6000
```

TCACAACAAG GCCGCCTGTT AACAAAAGCT GAGTATGAGC AGCCAGCGCC TTTTTATATC 6060

GGTGCGGGCC ATGAATATTT GCCAGATGAA ATCAATTATC AGGAATTATT GAAACAGAAA 6120

AAACGACCAC TTGATTATTC GGCAGAACAA GTAACCATTA CAAATGTTAG AaTGCCTTAT 6180

GGGAAAATTT CATTTGACTA CCAAGTGGTG AtCAATCAGC GAAGGTGACC GTTCCTTTTA 6240

TTTATTACTT AGGCTATCAA GCCACGATTC AAATGAAAAA CCAAACGGGC GCTAAAAAAA 6300

TGAGCCTAAC CAATCAAGGA GGCTTAACCG CTCTTTCTTT ATCTGGTACA GGACATGTCG 6360

ACATCCGTTA CCAAAGAACA AAAGTGCAAA AAATAGGGAC AATGATTACC TTACTTTCTG 6420

TTGGTGGTTT TGGATTTAGC CGATTTTTAC AACAAAAAAA GAAACACAAG ATAAAAGAAC 6480

AACGATGAAA GTAGGACCTT AATTAGGAAG AGTTGAATAA AGGACAAACC GATTTTGGCG 6540

TTTgTCCTTT ATTCCTCATT GATTATTTAA ACAAATCTAG TTGACTTCCT TAGTATTATC 6600

TGTTATTTTC ATATATAAAA TTAAGTCGAG GAGTGGACAA AAAAATGTCA AAyTGGGAAA 6660

CAAAATTTGC GAAAAAAGGT CTTACATTTG ATGATGTGCT ATTAATCCCT GCAGAAAGCC 6720

ATGTCTTACC AAATGATGTA GACATGAGTG TACAATTAGC TAAAAATATC AAGTTAAATA 6780

TTCCATTGAT GAGTGCCAGT ATGGATACTG TTACAGATAG TAATATGGCA ATTGCGATGG 6840

CTCGTCAAGG TGGACTAGGT GTCGTACATA AAAATATGAC TGTCGCCCAA CAAGCAGATG 6900

AAGTACGTAA AGTCAAACGT TCTGAAAGCG GCGTTATCAT TGATCCATTC TTTCTAACAC 6960

CAACAAACTT AGTAGCGGAT GCAGAAGAAT TAATGAGTCG TTACCGTATT AGCGGTGTGC 7020

CAATTGTAGA AACAATGGAA AACCGTAAAT TAGTCGGAAT TATTACAAAC CGTGATATGC 7080

GTTTTGTAAC GGACTACCAA ATCAAAATCG AAGAAGTCAT GACAAAAGAT CATTTAGTCA 7140

CTGCGCCAGT TGGCACATCT TTAAAAGATG CAGAAAAAAT CTTACAAAAA CATAAAATCG 7200

AAAAATTACC AATCGTTGAT GAAGCAGGTC GCTTAAGTGG CTTAATTACA ATTAAAGACA 7260

TTGAAAAAGT GATTGAATTC CCAAATGCTG CCAAAGATGA ACACGGTCGT TTACTTGTGG 7320

CGGCAGCAGT CGGTGTCACT AGTGATACGT TCGAACGTGC GGAAGCATTA TTAGAAGCAG 7380

GTGCGGATGC GATTGTCATC GATACTGCAC ATGGCCATAG TGCCGGCGTT ATTCGCAAAA 7440

TTAAAGAAAT TCGTGAAACG TTCCCTGAAG CAACATTAAT TGCTGGTAAT GTTGCCACTG 7500

CTGAAGCAAC CAAAGCATTG TATGATGTCG GCGTTGACGT TGTCAAAGTT GGTATTGGAC 7560

CTGGTTCAAT TTGTACAACA CGTGTCGTTG CTGGTGTTGG TGTTCCTCAA TTAACAGCGA 7620

TTTATGATGC TGCTTCTGTA GCTAGAGAAT ATGGTAAAGC AATTATCGCT GATGGCGGTA 7680

TTAAATATTC TGGTGATATC GTGAAAGCCT TAGCTGCTGG TGGACATGCT GTAATGCTAG 7740

GAAGTATGTT AGCAGGTACA GATGAATCTC CAGGCGAATT TGAAATTTAT CAAGGCCGTC 7800

GTTTCAAAAC TTACCGTGGC ATGGGCTCAT TAGGTGCAAT GGAAAAAGGA TCAAGTGATC 7860

GTTACTTCCA AGGAAGTGTC AATGAAGCAA ACAAATTAGT GCCAGAAGGA ATCGAAGGAC 7920

GTGTTGCTTA CAAAGGAAGC GTTTCTGATA TTGTCTTCCA ATTAATCGGT GGCTTGAAAT 7980

CAGGGATGGG TTACGTCGGT GCAGCAGACC TGAAAGCGTT GCGTGAAGAA GCACAATTTG 8040

TTCAAATGAG TGGTAACGGA TTGAAAGAAT CACATCCACA TGACGTTCAA ATCACAAAAG 8100

AAGCACCAAA CTATTCTGTT CAATAAAATA AAAGCTGATT TTTTTGAAAT TTAAAAGAGG 8160

AAGACGTCAT TTCGACGTCT TCCTCTTTTT CTGTAGAATG AAGGATTTGC TTCTGTTTCC 8220

GCCGTTTATC AATTTTTAnG CCTGTAACAA AAaTCCAAAG TGATTTTTGT TACAGGCtCC 8280

TTTTCCTATA GAATAAAAGC CGTTATCTGA TATCAGACAA CGACCTTCAT CTTCCCTTTA 8340

TTTAATAACT GTTAGGTTAC CCATGTAAGG AACTAGGACT TCTGGTACAG TTACAGAGCC 8400

ATCTTCGTTT TGGTAGTTTT CTAAAATAGC AGCGACTGTC CGTCCTACGG CTAAACCAGA 8460

ACCATTGAGT GTATGCGCAT ATTGGACTTT ATCATTTTCA TCACGGTAAC GAATCATCGC 8520

ACGGCGCGCT TGGAAATCTT CACAGTTTGA GCATGAACTG ATTTCACGGT ACGTCTCTTG 8580

AGCAGGAATC CAAACTTCCA AGTCATAAGT TTTCGCAGCT GAGAAGCCCA TGTCACCTGT 8640

TGAAAGAGCC ATGACACGGT AAGGTAATCC CAGTTTTTGC AGAATTTCTT CGGCATTATT 8700

CGTCATTTTT TCTAATTCCT CATAAGAATG CTCTGCATCA CTAAATTTAA CCATTTCGAC 8760

TTTATTAAAT TGGTGTAAAC GAATTAATCC TCGTGTGTCA CGACCAGCGC TACCAGCTTC 8820

AGAACGGAAA GAAGGGCTCA aGGCCGTAAa GTAAATCGGT AAaTCCTTGC CATCTAAAaT 8880

TTCATTGTTG TAGTAATTGG TTAAAGGGAC TTCCGCAGTT GGAATTAACG TTAAATCCGT 8940

ATCTTGTAAT TGGAAGACAT CTTCTTTAAA TTTAGGAAAT TGGCCAGTCC CGAACATGGC 9000

GGTGTCATTA ACGATATAAG GCGTAATCAT TTCTGTATAG CCATGTTCAT AAACATGTAA 9060

ATCTAACATG AAGTTGTATA ACGCACGTTC TAAGCGTGCA CCTAAGCCTT TGTAATAAAC 9120

AAAACGGCTA CCAGCTACTT TGGCACCACG TTCAAAATCT AGAATGCCTA AGTTTTCAGC 9180

GACTTCCCAG TGTGGTTTTG GTTCAAAAGC AAAAGTTCTT GGTTCACTCC AACGACGTAC 9240

TTCTACATTT TCTTCTTCAC CAGCACCAAC AGGCACAGAA TCATCTGGCA AATTCGGTAA 9300

AGTAGTTGTG ATCCCACGTA GCTCTTCGTC AATAGCGTTA ATTTCTGCAT CTAATGCTTT 9360

AATATTGCCG CCGACTTCTT TCATTTCCGC GATTTTGGCC GTTGCATCTT CTTTGTTTCG 9420

TTTTAATTGA GCAATTTCAG CAGACACATC ATTCCGGTAT TTTTTCATTT CTTCGACTTT 9480

GACTAATAAA TCACGGCGAC TTTCATCTAA ACGCAGAAAT TCCACTAAGA TTTCTTCTTT 9540

CACGCCACGG GTTTGTAGCT TTGCTTTTAC TTCGTCAAAA TTTTGACGCA TCATTTTTAC 9600

ATCTAACATA TTTTTTCCTC CTTAAAATTA GTACATAATC AATAATAACT GTTGTAGAAA 9660

ACAACAAAAA AGCCATCTCA CCCCAAAAAT AAATATTCTT GGGACGAAAT GACTGGAAAT 9720

CTCGCGGTAC CACCCAAGTT CAGTAATGTT ACTGCCCTTG ATTTGCAGAT AACGGCTGCA 9780

GACCGTTCCA ACTTATTTCG CTGGACACCT GGAAGAGACG GACTTCTTTA TAGATGAATG 9840

TTTGCTCGCA CCCCCGCAAA CTCTCTGGAA TTCAGTGCCA TAAATACTTT TTCTCTTAGT 9900

TTTCTCTTTT CTATAGAATA CCGAAAAAGT TCAGGGATGT CAATTAACTC TCTGATTTCT 9960

TGTTGGTACG CTGTAAGGGT AGATGAATAA TAAAGGAAGT CCACTTTTGA TCAGACTTCG 10020

CATAAATATA TCCACCATGT AAAGCGACAA TACTTTGTGC AATAGCCAGA CCTAAACCAG 10080

TACCCCCCGT TTCTTGTGAT CGGGATTCTT CGACACGATA GAAGCGGTCA AACAACTGAT 10140

CTAAAGAATG TTTAGGGATA GCTGGTCCAT CATTGCGAAC AGCAATAATT GCTTCGGTGC 10200

CGACTTTGTC TACTTCCATT ACAATATGAT GGCCACCTTT GCCATATTTC AAAGCGTTTG 10260

AAAGAAGGTT GTTAAAGACT CGGACTAGTT TTTCCGTATC GCCTTCCATC ATTAAAGAAG 10320

CTGGATTGGC TTTTACTTGG ATTTGCATAT TAATCTTTTT CGCTTCCAAT TCAAAATCGG 10380

CAGCAAGCTG TTCAATTAAT TGTGCCATAT CAAAAGTCGT CGTATGAATC GGTACACTTG 10440

GTTGACGGAC TTTTGTATAT TCAAAAAGAT CATCAACTAA TAATTTCATT TGTTTCGCTT 10500

TGACGTAAGC GGTATGCGTA TATTTCAGTA AATCTTCTTG GCTATGAAAT TGACGATCTT 10560

CAATCAATCC CAGATAACCA ATAATAGAAG TTAACGGTGT CCGAATATCA TGACTGACAT 10620

TCGTAATTAA TTCATCTTTT GATTTTTCAA TTCGCCGTTC ATCTTCAATG GCTGCGACAG 10680

TACTATCCAC CAGACCGTTA ATACTCGTCA CGACTTTTGC TAAATCACCA CTTAATTCAA 10740

ACGGAATGCG GTGATTGTAA TTACCATCAG CGATGTAATG TAACTCACTG ATAATATGAC 10800

GCAATTGCAT TTGATGGTAG CGACGAATTA AGCGCCAATA GAGAACAGCT GCATCTACTA 10860

AAAAGAAAAT GGGTAAGACA AATTTATTCC AGCTCCAAAA CAAATCAGTA TTTAATTCTT 10920

TAGCGACCGC ATCCTTAGTG GCCCAGATGA CACCAGCTAA GGGCGAGTTT TCACTAAGTA 10980

GTTGTCCTAA AAGGAACATA ATTGAGACAT TTAATAGTAG TAATAAAACA ATTGTCACAA 11040

TTGCTTCTGC TAATAACTCG CTGATTTCTT TAGAAGTTAA GGTAATTCGT TTTTGTTTTT 11100

CAATCTTTTT AGGTTTAACG AGCATCGATC TTATATCCAA CTCCCCACAC TGTTTGAATG 11160

ACTTTTTCTC CACCAGTTGC TTCTTCAATC TTATCTCTTA AATGACTGAC GTGAACCATC 11220

ACGGTTTTAG CAGAAACAAT ACTTTCTTGT TGCCACACAC GCTCAAAAAT TTCATCCGCA 11280

CTAAACACAC GATTTGGATG ACTAGCTAAT AGATAGAGGA TACCAAACTC AAGAGCGGTT 11340

AATTGAATTT CTTTTCCTTC TATTGTTTTG ACTTCGTGAG AGTCCTTATT AATCATTAAT 11400

GGGCCGACTT CTAATTCATC GGGTTCTTCT TTGGTTAATT GCATTTGGCT TCTACGTAAA 11460

ATAGACTTTA CCCGAGCCAT TACTTCTAAA GGATTAAAAG GCTTGGTGAC ATAATCGTCT 11520

GCACCAGCAA CTAAGCCTTT GATTTTATCC ATATCGGTTG TTTTTGCTGT TAACATAATA 11580

ATTGGAATTT GCGATTCTTT TCGAAGTTCT TTCACTACTT CCATTCCATC CATGATTGGC 11640

ATCATAATGT CCAAGATTAA TAAATCGATA TCTGGTGTTG TGTGCAATTT GGATAACGCT 11700

TCTTTTCCAT CATAAGCTTT CACGACTTCG TATCCTTCAT TATGAATATA AATACTCAAT 11760

AGTTCTACAA TTTCTTTATC ATCATCTGCA ACTAAAATTT TCATAAGCGT TCATTTCCTC 11820

CATATATAAA AAATCTCTTT TCCATATTTT AACAAATAAT TGCTAAAAAG GACAAAATCC 11880

TTGATATATT ACGATaAATT TAGAAGATTT TGCTAAAAGT TAATAAACTA TAGCCAAAAA 11940

AGGATGAAAG GGGGTAAAAA ACGAACGAAA ATACCTATAC CAATCATAAT AATCGGTATT 12000

GAAAAAAAGT TTAAAAAAAG TG 12022

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 556 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

TAAAATAATC GCAATGACTG CATCTGTAAA TGCTTCTACA CGTGTTTTCG GCATCATTTC 60

CCCCTACTTT CATAAAAATT AAATAACTTA CCCAAAGTAA GTATAACGCA TTTCTGTAAA 120

AATTCAATTA TTTATGAAGT CTCTCGGTTG TTGAAGTTCC TAAAAACGAG TGCAAGAGCC 180

CCACTTAAAA GTAAGCCTCT TGCACTCGTT ATAGAACAAA TCAGTTATTG GAAGACGCCT 240

GCCTGATAAG GTGCTAACTG CATCCGTTTC TTCACCGTAA AATCTTGATT CCCTAATAAT 300

TGTTGCCACT TTCTGCCTAG TAATTTATTC GGCAAGTCAA ATTGGCACTT TTCTCCACTT 360

AAATTGACAA TCACTAACGC TGTTGCCTCA GTTGTTTTTC TTAAAAAAGC AAAAACAGAC 420

GGATGATTGG GcGCTAATAG TTCAAACTGT CCAtCTGTAA AGACCGGATT GTCTTtCCGC 480

AACGCAATTA ACTTTTtATA ATAATcAATA CGGATgCGCA TcCGTTtCTT GGGCAGCCAC 540

ATTTATnnCT TGAAAT 556

(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5998 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

TGTTAAAAAA ACCATTTTTA CTATTTTTTT CACTTTTAGG AGCTATTTTT ATTTTAGCTA 60

GTTGTGGCAT AGGAAAaGaT GCTGTCACAG ATACTAAGTA CAAAGTTAGT TTGCAGCAAG 120

CTGCTGAAAT CTATGAAAAA GAAGCTGGCA ACAGCAAACC ATTAGTAAAT GTCCAATTTG 180

ATACAGAACC AGCAAGTGAC TACAGCTATA TCTTTACTAA CGATACAGAA ACACTTTACG 240

TGAATCCTGA AACAGGAAAA GTCACCAAAA ATACTGAAGC AAATCAACTT GGCGAAAACG 300

AGACAGCCTT TTCAGCTGCT GAAGTCAAAG AATTAGGCGC TGTTAACGAC GTTTTAGCCA 360

AAGCAAAAAA AGAAGTTGGA GGACTTTCTC CACGTATTTT GACTTGGAAG TTAACCAAAA 420

ATAACAATAA ACTTGTTTAT ACAGTAGATG TTAAAACGAC TACGGCAGAT GAAAAAGTTA 480

CCTTAGATGC CAATAAATAA GCACAAGCGT ATAACTTATC ACGAAATAAC CAAATCTTAA 540

AATACTCTTT GGGACAGAAA CTACGCATTA GGAAACCTTA ACATTTTCCG AAGCCGATCG 600

TTCTGTCCCA ATTTCATTCT TTTAAGTAAT AATAATTGTT CGTTACCCCT GCTACCGTTA 660

TACTTAGTTT AAAAGAAAAG ATGGGAGTAT CAATTCATGG TTTATACAGT TGATTTTAAA 720

GAAGTAGCCA CTATTGGTTT AGAAAGTTCA CCAGTGGCAC CTGCCCTTGC TGGACTAAGA 780

GCGAATGAAG CTCGCTATTT TTGGAATAAA TACAAACATC GCTTTACGAC CGTCCCTGCC 840

TCTGAAGCTC CAGAAACATT GGCCTGGATT GAAAAAATTT TATTAGAACG AGAATTACAT 900

TTCCCTTACA AGCCACTTGA AATTTCTTCT TTCACTTTAG ATGGCATAAA AATGGCCTAT 960

GTATTTTATG AAAACGGCTT AGCAGTTAAC GTGATGTACA CATTAGCTGA GGAAGGCAAA 1020

CGCGCCGTTG GGTTTAAACT TTCAGATGGT ATGGCAATCC CTGCAGAATT TGAAGGAAAA 1080

TTCAAATTTG CTCGTCAAAA ATCAAAATTA GCTGGAACCA TTCGCGGCTC TTATTTTGTC 1140

ATTAAAGGAA CCTATTAAAA AAATGCTTTT GCCCTTATTA TTAAAGGCAA AAGCATTTTT 1200

TTATTTTAAT TCTTCAATAT AGTTAAATAC TTGTTGTCCA GCAATACCGC CTTCACCGAC 1260

CGCTGTCGTG ATTTGACGCA ATGTTTTTTC ACGGACATCA CCAATGGCAT AAACACCAGG 1320

AATTTTTGTT CTCATTTCTT GGTCTGTTTC AATCCAACCA GCCTCATTCG TAATACCAGC 1380

TTTTTTAAAT GGTTCTGTTA ATGGATCTAG CCCCACGTAA ATGAACACGC CGTTTGCTTC 1440

AATTTCAGAA ACTTCATCTG TTTTGACATT CCGTGCTTTC ACACCTGTCA CGACCATTTC 1500

ATTGCCAACA ATTTCTTCGA CAACTGTATC CCAAACAAAG GAAATCTTTT CATTTGCAAA 1560

GGCACGATCT TGAATGATTT TTTGCGCACG CAATTCATCA CGACGGTGAA CAATCACTAC 1620

TTCAGAAGCA AATTGCGTTA AATAGATTGC TTCTTCAACT GCAGAATCGC CGCCGCCAAC 1680

GACTACTAAG CGTTTATTTT TAAAGAAAGC ACCGTCACAA ACAGCACAAT AAGAAACACC 1740

ACGGCCAGCA AATTCTTCTT CGCCTTTTAC ACCCAATTTT CGGTGTTCAC AGCCTGTTGC 1800

AATAATTACC GCTTTAGCTT CATAAGATTT GTCATCACAA ATCACTTCTT TATAGGATCC 1860

GTGGTCTTCA ATGCCCATCA CAATACCATA CGCATTTTCC GTACCAAATT TTTCTACATT 1920

TTCATACATT TTATAGGCCA ATTCTGGACC CATAATAGAA TCAAAGCCTG GGTAATTCTC 1980

TACCTCAGCA GTGTTGTTCA TTTGCCCACC AGGCGCACCG CGTTCAATCA TTAATACAGA 2040

AAGATTGGAC CGTGATGCAT AAAGTGCTGC CGTCATCCCA GCAGGACCGG CACCAATGAT 2100

AATTACATCA TACATTTTAT TTCCCCCAAA ATTTTAAGTA AGTTTACTTT ACTCCTTAAA 2160

CAATTTCTTG TCTACTTTTC TGCCCTTTTT TGACATTCGG AGTAGGTGTG TCTATTACGT 2220

AAGTTATTGT ACCACGGTTT TGTGTTTTCT AGGAGGGGGA AAAGAAAAGT TTAAAAATAA 2280

CGATCACAAC GAATCATTGA TTTTTAGAAG CCGTTGTACT GCCGCATCAG AGAATTCTTC 2340

TTTATCATTT TTTATGCATT CTTGTTGATT TTGCCCTTGT GGAACAAAAC GAAATACTGT 2400

AAATGATGCA GATAACGTAT CAGGATTACT AATCCGATGA AGACTTTTAT GCATCCGAAT 2460

AAGTTCTCCT TCTGCACAAA TTCTTGAAAT AATTTGCTGA TTTTCTAGCT CTTCTATTTT 2520

CACTTTTCCC GACTGCACAA GTAAAACTTC TTCGATTTCG TGATGCTTAT GCCAACCTTG 2580

```
CACAGTTTTT CCTGGCAAGC TATTTATATG TATTTCAAAC TCTGGAAATA GAAAATAATT   2640
TACTTCTGTT CCATCTGGCT TTTTACGAAA GATGACCTCG TCTCTTGTTA AAATTTCTAT   2700
ATATTTCTGT AACATGGCAG GCTCCTTTTA ACTTTTTTTA TAGCCGAAAG CTATTCTTTC   2760
AAACTAAGTT TGAAGCTACT TTCTGAAAAT TCGAGTTCCT TTAAATATAC TTTTATTACC   2820
ATTTGAAAAT CACTTTTTTT AAAACAATTG ACTGTTCAGT ATCAATCTTc ACTAATCCAT   2880
CTGTTCACCA GAAAAAATAT AGTATTTCCG ATTATTCACG AACTTATTAT TCAACTTCAA   2940
AGGAAAATCA GTCAACATAC AGCCCACATA TCTAACATCA GAAACACTTC TACTTTTCTC   3000
AAGTACTGCT ATGGTATACA TTGACCTTAC TAGTAACATT ATAGCATCAA GTTTTAATCA   3060
TATGTAGATT TAAAAAAAAT GaTAGcACCC TAACAGaTGT TTTCGTTGCT CATTTAATTT   3120
ACATGATTAT TTTTTAGATG CTAATTAAAT GTTTGTCAGT AACCTGAATT TTTGATATTT   3180
TTATCCAGAT TACTTATGTC GGTTTATCAG TTGTGCTAAC AAAAAATGAT TGTTTTTAAA   3240
GATAATTTAA TTCATAATAT ATATAGAATG AAAAAAAGGA GGTTTTATTG TTGAAAGTTG   3300
CTATTTGTGA TGACAATCCA ACGTTaACCG AAAAAATAAA CACGATGCTA TTTAATTATA   3360
ACCCAAATAT ATTTGAAACC TATACCTACT ACAATCCTCT CAAGTTAGTC AATCAACTTG   3420
ACCAAGAAAA TTTTGACTTT TTTATTTTAG ATATAGAAAT GAGTGAAATG AGCGGAATAG   3480
ACTTGGCAAA AAAAATCCGA GAACGTGATA TTTTATCACC TATCGTTTTT TTGACGAGTT   3540
ACAAAGAATA CATGGAAGAA GTCTTCCAAG TACAAACTTT TGATTATTTA TTGAAACCGC   3600
tACGGAAGAT CGGTTGCACC AAGTTTTGGA AAAACTGAGA CAACAGCTAG AAAAAAAAAG   3660
AAATTACTTT GTATTTTCTA GTAACAAAGT TACTTATAAA ATACCAACCA AAGATATCAT   3720
TTATTTCGAA AAAGATAAAA GACAAGTATT GATTCATACA GTTGGAGAAa TCTACAAACC   3780
CtATATGTCA ACGAATCAGA TAAACGAACA ACTTGATACG AATTTTGTGC AAGTCCACTC   3840
TTCTTTCATA ATAAATTGCG CTTATATAAA AGAGTTAGGA AAAAATTTCC TATTAATGGA   3900
TTCTAAAGAA AAATGTATTG AAATACCAAT AAGTAGGCGC TTCAAAGCCG CCGCTCATAA   3960
AAGTATAGTT ATGTCCATGA GAGGAAAGAT ATAATGCTTG ATATAACATT ACGAAGCCTA   4020
AGTTTGTTAG TTTATTTATT TCTTATCTCT TATTTTTTTT CTAATTGGTC AACAGCTATT   4080
AAAAATCATA GAGAATTTTA TAATGTCATC TGGATTGTTA TAATCATATC AGCCATGGTT   4140
ACGATTATAT TATCCAATGC TTACATACCG AACATTTATT TACCGTATCT TAACTTATTA   4200
CTAAGTTTCA TTTTAACTTA CACTATTGCT ATGCTACACT CGGTCCCTTA TCTCGACTCG   4260
GCCGTATGGA CAATGACACT TATATCAATA AATTTAATCT GTGAAGTATT GTCGTTACAC   4320
TTTACGAAAA TAATTTTAAA TGCAGAATTA ACATCCTATG GTAGCCCTAC TTTTTTCATA   4380
ACTTCAATAA CAATTACAAC TTTGATTGGG GTAGCCTGGA TAGTCATACT TAAATTTTCA   4440
ATAGTTAAAG AAGCAAAGTC TGAATTATCG ACAAACTTAT CATTTATTGT GATTTTATTA   4500
CCTATTCCCA TACTATCAAT AATCATTTTA TTTGGGCTTT TAATAGGAAA CAACAATGAC   4560
```

AAAATTTCTG AGATTACTAT TACGATTAGT GTCATATTTT TAAACATCTG TGTAATTTTT 4620

CTTTATAAAA TCACTATAGA ATATCAAAAA AATCAATATA ATCTTACTTT GAGAAAAAAA 4680

AATATCGAAG TGGAGTATAG AATTTTAAAT GAAATAAAGA GAAACAGAAC TAACGTTTTA 4740

AAATTAAAAC ATGATTTGAA GAATCAATAC TTAACAATCC TCGGACTTAT CGAAAATGAA 4800

GAAGTTAATG AAGCTATCGA TTATATTAAG AGTAGTTTTG ATATTTTAGA ACCACCAACA 4860

AAAACTTATG CTGCCGATGG AGTATTAAAC TACCTATTGA ATGAAAAGCT TGCTGAAGCA 4920

AGAAAGAATC AAATTAATGT TGATCATCAG ATATTTGTCT CTAAAAATAT AAAAATTAAT 4980

AATGACGTTC TAACTATTGT AATTGGAAAT ATAATTGATA ACGCCATACA AGCATCGAAG 5040

AGAATCAAGC CGATTGACAG ATATGTCAAT ATAATTATCA AGCAAGTCAA TAATGATTTA 5100

TTTATTGAAG TATCtAATAA CTATAATTCt GaAGAnATTT TTaCaAGAAA ACaTAGaAAA 5160

AATaAAGGtT TAGGaATGaA AAACaTTGaT GATCTCTTAC AACAATTAGG TGGTATTCAA 5220

CGTCACTGGA CGAAAGAACG TAGATATTTT GTAACTATAG TTATATTTAA TGTATACAAA 5280

GGAAACAAAT GCGGTGATAA CTAATCAAGT AAATCACCGC ATTTTTTAGT ATACTTCCAT 5340

GTCGCATAAT TTATCAGAAA ATTTCAAAAA TAACCAAACA AGCATCAATC CCATAATCGT 5400

TGAAAATAGG AACGTTATAG TTGGAAAATT CATAAGACCC ATAGCAGAAA AATTCGAATA 5460

TATTAGCATT AGTACCACTG ACGTTATTAT TGAAGCAATT TCCGATTGTT TCCAAATTGA 5520

AATTATTTCT GACAATATCA AAATTACGAA TAGTAGACTA CAACTTGTAA CTATGGTAAT 5580

AAGACCAAGG TAAATTTCAA AAAATAAATT TGAACTATGA AACTTCAAAA TTAAGTTAAG 5640

CAAAACCTCA GTCACAAATG AAATAAGAAG CGCACTTCCA AAGCTAGATA AGATTATCAC 5700

GTAAAAAGAA CTTGTTTTTT TTACGAGAAG CTCTTTTCGA TCAATCGGAA AAAGAAACGT 5760

TCTATTCCGG TGATTGCCAA CATAAAATCG TACAAGCTCT TTCCTTAGTA ATACAGCTCC 5820

ATAAATTGAT ATGCCGCACA TAGAAACAGT GTTCGATAAT GCTATAATTC CTTGAAATGA 5880

CGAGATCGTA TCAGCACTTT CATCAAATAT TCCCACAAAT GAGAAAAmCA TAAAAAATAA 5940

CATTTGGATT ACCAGTGTAA TTAAGAAAAC TTTAAACCAA CATTnTAATG GnnCTAAT 5998

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1641 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

AAATCCGTAG CAACATTTAT GGAGACAATT ACAGAAAAAT GCGGTGAAGC ACATCAAGAT 60

TGGGCAAAAA ACTTTCAGGG CGGnCTTTGC CAATACTTTA CTAACAACAG TGAAACGTGC 120

AAGAAGATGG TACTACCTTT TTATTAACAG GCGATATTCC AGCAATGTGG TTGAGAGACT 180

CAACAGCACA AGTTCGTCCC TATTTAGTGA TTGCnAAAGA AGACGAAGAC CTTGCTCAAA 240

TGATTGCTGG TTTGGTTAAA CGCCaTTTCG tTATATTTGC ATTGATCCTT ATGCCAATGC 300

GTTTAATGAA ACAGATAATC ATGCAGGCCA TCAAACTGAT AAGACAGAAA TGAATGGCTG 360

GATTTGGGAA CGGAAATATG AAATTGATTC ATTATGTTAC CCGGTCCAAT TAGCCTATTT 420

ACTTTATAAA AATACTGGGA TGACCGAACA ATTTAATTCA GATTTTGTGG AAGGCGTCAA 480

AAAAATTCTC AACGTTTTTA CAACAGAACA AGATCACGCA CAGTCACCTT ATTTATTTGA 540

ACGAGATACG TGGCGTCAAG AAGATACGTT GGTTGAAGCA GGGAAAGGAA CACCAGTTGG 600

CAAAACAGGG ATGACATGGT CAGGCTTCCG TCCTAGTGaT GATGCCTGCC AATATGGTTA 660

CTTAGTTCCT TCTAATATGT TTGCAGTCGT AATTTTAGGC TATATTCAAG AAATTTTCTC 720

TGATGTTTTA GAAGATGCTA CGATTGGGGC GATTGCTAAA AAATTACAAG AAGAGATTGA 780

AGAAGGCATC CAAACATTTG GACGAACAAA AAATCAAAAT AACGAAACGA TTTACGCGTA 840

TGAAGTGGaT GGCTTAGGCA ATGCAAGCGT GATGGaTGAC AGTAATGTTC CCAATTTAGT 900

AGCAGCCCCT TATCTGGGTT ATTGTTCAAC AGAGGATGAA CAATATTTAA CAACGCGTCA 960

AACCTTATTA AGTAAAGAAA ATCCTTACTT CTATGAAGGC AAATACGCGA AAGGGATTGG 1020

TAGCTCGCAT ACACCAGAAA ATTACGTGTG GCCGATTGCT TTAGCAATGG AAGGAATGAC 1080

AACCAAAGAT AAAGCAGaAA AAGAACGAAT TTTAGATTTA TTAGTTGCCA CAGATGCGGG 1140

GACTCATTTG ATGCATGAAG GTTTTGACGT AGATAATCCT GAAAATTATA CAAGAGAATG 1200

GTTTTCATGG GCTAATATGA TGTTTTGTGA ATTAGTTATG GATTATTTTG ACATCCGAGT 1260

GGAAAAATAG GAGGAAAAGT AATGACAAAG AAAAAAGTCT ATATTGTTTC CCATAGTCAT 1320

TGGGATCGTG AATGGTATCT GCCTTATGAA GrACATCATA TGCGCTTAAT TGAATTAGTG 1380

GATAATGTAT TAGATTTAAT TGAAAATGAT CCAGAGTTTA ACAGTTTCCA TTTAGATGGA 1440

CAAACGATTA TCTTAGACGA CTATTTACAA GTTCGCCCAG AGAAAAAAGA AGCCGTGAAA 1500

AAAGCTGTCC AAGCAGGCnA ACTAAAAATC GGACCATTTA GTATTTTACA AGATGCTTTT 1560

nAATTAGCTC TGAATCCAAT GTCGAACATG CTGATTGGGC TTAGAAAGCn AAATnGGGGC 1620

ACAGTCAATT AGGTACTCCC G 1641

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8033 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

AGAGAAGGGA TCATATGTCA GCAAAACATA ATCATGAACA ACACCAGGAA AAACAGCCAA 60

CAATGTCCCA CGATAAAATG AATCACATGA ATCATGATAT GGCGCATGAA CAAATGGCTA 120

TGTCACACCA ACATGAAGGA ATGAAGGGAA TGGATCATTC AATGCATATG GGGAATTTTA 180

AACAAAAATT CTGGCTGTCA TTGATTTTAG CTATTCCGAT TATTGTTCTT TCACCAATGA 240

TGGGTTTTCA ATTACCTTTT CAATTTACTT TTCCAGGTTC TGATTGGTTG GTCTTAATTT 300

TAGCAACTGT GCTATTTTTT TATGGTGGGC AACCATTTTT AAGTGGCGCT AGAATGGAAT 360

TACAGCAAAA GAGTCCAGCA ATGATGACGT TAATTGCGAT GGGGATTTCG GTATCATATT 420

TTTACAGCTT ATATGCGTTT TACATGAATC ATTTTACGAA TCAGGCCCAC GTGATGGACT 480

TTTTCTGGGA ATTAGCCACT CTGATTGTTA TTATGCTTTT AGGCCATTGG ATTGAAATGA 540

ATGCTATTTC TAATGCTGGT GATGCCTTGA AAAAAATGGC TGAACTGTTG CCTGACACAG 600

TCAAGCGAAT GACTGAACAT GGGGAAGAAG AAATTCCTTT GCAAGATGTC CAAGAAGGTG 660

ATCGCTTGAT TGTTCGTTCG GGAGATAAAA TTCCGACAGA TGGTAAAATT TTGAAAGGAA 720

GCACAACAGT CGACGAATCG ATGGTTACAG GGGAATCAAA AACAGTAGAA AAAAACATTG 780

GTGACTCCGT GATTGGTGGA GCTGTCAATG GCAACGGAAT GATTGaAATT TCGGtAACTG 840

GCACGGGTGA AAATAGTTAC CTGTCAAAAG TTATGGAGAT GGTTAAACmA GcGCAATCGG 900

AAAAATCaAA ACTAGaATCC aTTTCAGACC gTGTAGCCaA GTGGTTATTC TATATTGCCt 960

TATTCGTAGG TGTGTTAGCA TTTATCGGTT GGTTAATGGC GACGAAAGAT TTGTCGCTAG 1020

CATTTGAACG AATGGTTACA GTTTTTATTA TTGCCTGTCC CCATGCGTTA GGCTTAGCAA 1080

TTCCTTTAGT GATTGCTAGA AGTACCTCCA TTGCTGCTAA AAATGGCTTA CTTTTAAAAA 1140

ATCGGAATGC ATTAGAACAA GCCAATAAGG TTGAATATGT GTTGCTGGAT AAAACTGGAA 1200

CTTTAACAGA AGGACAATTT ACAGTTACTG GCTTAGAATT AATGAGTAAG CAGTTTACTA 1260

GAGAAGAAGC ACTTAAATAT ATTGGGGCAT TGGAAAAAAA TGCAAATCAT CCGTTGGCAA 1320

TTGGGATCAT GAACTATTTA AtAAACAAGC CgTTCAACCT TATGAAGCCC ATAATCTGCA 1380

AGCTTTATCA GGCGTAGGAC TGGTTGCTAC TGTTCAAAAC CAAGAAGTCA AAATAGTCAA 1440

TGAAAAAGAA GTTGCACGGT TACAGTTAAC CTTCGATGAA ACAATTAAAA CAAACTATCA 1500

AGAGCAAGGA AATACCTTGA GCTATTTAAT CATCGCTGGT CAATTAGTCG CACTGCTTGC 1560

TTTAGGAGAT AAAGTCAAAC CAGAAGCTAA AACATTTATT GCAGAGTTAC AAGCGCAAGG 1620

AATTACGCCA GTCATGTTAA CAGGGGATAA TCAAACAGCC GCCAGTGCAG TTGCTAATTA 1680

TTTAGGAATG AAGGAATACT ACGCAGAGCT CTTACCAGAA GATAAAGAAA AAATTGTGCA 1740

ACAATATCTT ACTGAGGGAC ATCAGGTCAT GATGGTAGGA GACGGTATTA ATGATGCACC 1800

AAGTTTAGCC CGTGCTTCAA TCGGCATTGC GATAGGCGCT GGTACAGATG TAGCGATTGA 1860

TTCGGCAGAT GTTGTACTGA CGGATAGTGA TCCTAAAGAT ATTTTAAGGT TCTTAGATTT 1920

AGCCAAACAA ACACGACGTA AAATGATCCA AAATCTTTGG TGGGGAGCTG GGTACAATAT 1980

TGTCGCCATT CCTTTAGCGG CTGGGGTTCT TGCACCAATT GGGATTGTTT TAAATCCAGC 2040

TGTAGGTGCT GTCTTGATGT CATTAAGTAC AATTATTGTA GCAGCAAATG CTATGACACT 2100

GCaTATTTcA AAAAAATAAT TCGATaTAAT TGTaGAGAGC GGAAAGTAAT CAATACTTTT 2160

TGTTCTCTAT TTTTTTCTTC TTAAAAAAAT TTTTGCCTTA TTTGACGTAT TCTTATwACa 2220

AGGGAGTGaA AGGGGACGAT ATATGGaTTT CaAGGaATTA TTGgTCTTTk GAmCATAAAA 2280

ATATGTTGGA ATTGTTGGGA TTTACTGCCA ACAAGAGCAG ACATTTTCTG TTTTGGATGG 2340

CTGCCAGAAA ACTAAAGGTG AGCGACAAAA CCTCCATTGT TACATTCAAC GTTTTGAAAA 2400

AAAGCAGGAA ACGTTTAAGA TAGAAGGGCT TTGTAAGATT TATACAACGA CCAAAGGAAT 2460

TGTAAAATAT CGTGAATTAT GCGTTTGTGG CTTGGCTCGA TTTAGACAAA AATTTTGTTA 2520

TTTTATTCCC GAATTTTATA TTTTACGTTG TCTAATTGAA GAGCGAATCA ATTATGCACA 2580

ATTGACCCAG GTATTGGAAT TACAGGAATC AACTCTTCGG AAGAAATTTT CTAATATTCG 2640

ACGTTGGCTA GTTAATTTTG ATATTATTGT ACGTCAAAAA CATTACGATT TAACTGGTAA 2700

TGAATGGCAA ATTCGACAGT TAATCCTTTG TTTTTATTTG TTTTTCCAAG AAAGCTGTCT 2760

GGAAGAAAAC AGAGAAATGA CTCGGAAAAT CATTACCTTT TTTGAGTTGG ATTTAAATGT 2820

AGCACAACAG AATCATCTTA GTTGGCTAAT TTATATTTGG GAGAAACGAT ATAGAGGCGG 2880

GCATGAAATT TCAGTGCCAA ATGCTAATTT ATTTCAACAA ACAAGTGCTT TTTTTTATTT 2940

ATTTCGTGTG GAAGTGTTGA GCACTTCTTT TATGTCTCTT AAAGAACAAA AAGCGTTATT 3000

TGTGATTCTT GAAGCACATT TTGGTGGTTG TTTTGGCAAG AGAGCGCGAA AGTATTTTAT 3060

TCATGAGCAA ATGAAAATAG AAAGTTTATG CTTAAAAACA GCTATTTTTA TTATGAAAGA 3120

AATTAGGAGA AATTTTACAC AACATCATTT TAATTATCAA GAAATTCATT TATGTCGATT 3180

TTTAAGTACG CATATGAATA GTTTATTAGA CGGACAAGCA TGGTTGCCAG CTCATAAACA 3240

GGAACAAACA CTCGCTGCTC GTTATCAACA AACCTGGCAC AGATTACAAA AACTTATTAG 3300

GCTTTTAAAA CGCCTATATC CAGTCTTTAC TTCAGTAAAA GAACGGGAAT TGACGAGTTG 3360

CTATTTTTAT CACATCCTAG ATTTATTTAA CCCGATTTTA TATGAAAAAA AATATATTAT 3420

TTGCCTATTG ACGGACTTTC CTCCAGAAAA GGAACAAGCG TTGGGGCAAT CTATAAAAAG 3480

CTATTTCAGT GAAAAAAAGA ATATTACAAT TATTCATGGA AAACCAACGT ATCAACTTCA 3540

CCAAGTACAC CTCTTAATTG TCAATCATTT GTTTCAGATG AACGTAGCGC TTTCTTCTAA 3600

GACAGTGGTT TATTTGCCAG AGGAATTATC TCCTGCTTTT TTTGAAAAAG TCGAAGCGAA 3660

TCTATCGTAA GAAGTCCAAT TTTGTCTAAA TAGTGTTAAA ATGAAGAAGT AAGAGGAATG 3720

GATGAGGAGG ATATTAAAAT GAAACAACTG cAATTTGGTA CAAGTGACGA AACAGTTTCT 3780

TCGGTTATCT TAGGATGTAT GCGCTTAAAT GGTGCTGAAA ATCCTCAACA GGTAATTGAA 3840

ACTGCTTATG ATCATGGCAT TACATTTTTT GATCACGCCG ACATTTATGG TGGCGGCGAC 3900

TGCGAAACCA TTTTTGGTAA GGCCTTAAAA GAAAGTACCA TTCGTCGTGA AGATATTTTT 3960

ATACAAACGA AATGTGGGAT TCGTCAGGGC TTTTTTGACT TTTCAAAAGC ACATATTTTA 4020

GAAGCGGTCG AAGGGAGTTT ACAGCGTTTA GGTGTAGATT CAGTCGATGC ATTATTGTTG 4080

CATCGTCCAG ATACTTTAGT GGAACCAGAA GAAGTAGCTG AAGCTTTTCA TTTATTAGAA 4140

AAACAAGGGA AAGTTCGCTA TTTTGGTGTA AGTAACCAAA CGCCAGGACA AATTGAACTT 4200

TTAAAAACAG CTGTTAAACA ACCGTTATTA GCCAATCAAT TACAATTTGG CATCAAGCAT 4260

ACTGGTATGG TAGATCAAGG CTTACAAACA AATATGGAaA TATCAGGAAG CATTGATTAT 4320

GATCATGGTA TTTTAGATTA TTCACGGTTG AAACAAATGA CTATTCAAGC ATGGTCACCG 4380

TACCAATATG GTTATTTTGA AGGTGTCTTT ATTGGAAATG AAAAATTCCC TGAGTTAAAT 4440

CAAAAATTAA GCGAATTGGC TGAAAAATAT CAAACAACGC CAACTGGTTT AGCCAGCAGC 4500

TGGATTCTGC GCCACCCTGC TAATATGCAA GTCATCGCAG GAAGTATGAA TCTTGGTAGA 4560

ATTGAAGAAA TTGCTAAAGC AGCAGACATT GTGATTAGTC GCGAAGATTG GTACGATATT 4620

TATCGTGCAG CGGGAAATGT TTTGCCGTAA AAAGATTCTC ATACTTGTGA AGATCCAGCG 4680

TGTATTGCTG ATTGATTTTG GTATTTGATA TAGTTGTCAC TAAGGAACTA TGTCAAATAC 4740

CTTTTTCTTT CATTCTGCAA TCGCCGTCTA CTGCAAATTC TGTAGTTTTT TGTTAAAATA 4800

GAAAGGAAAC TAGTTTGCTA AAGATGAGGG GAAAAAAATG ACCAAAAAAA AATTATTACT 4860

CGTTGATGGA AACAGTGTAG CCTTTCGTGC ATTTTTTGCG CTACATAACT CATTAGAACG 4920

ATTCAAAAAT AAGAACGGCT TGCATACAAA TGCTATTTAT GCATTTAACA ATATGTTTGa 4980

AAATGTAATG cAAAAAGAAA TGCCTACTCA TGTTTTAGTC GCTTTTGATG nCAGGGAAAn 5040

CCACATTCAG AACaGrATTT TATTCaGAAT ATAAAGCCGG ACGTTCAAAA ACACCGGGGG 5100

AATTTAAAGA ACAAATGCCG TATATTCGTG AACTTTTAGA AGGTTTAGGC GTAAAATATT 5160

ATGAACTACC TAACTATGAA GCAGATGATA TTATTGGTAC ATTAGnCCAA CAAAGTTGAT 5220

AAAGATGAAT TTGATGTTGT CGTGTTATCA GGGGACCGAG ATTTGACGCA ATTAGCGACG 5280

GACACTGTCA AAGTAGATAT CACTGTTAAA GGAGTTAGCG ATATTGAATC ATATACGCCT 5340

GAGCATATTG CTGAAAAATA TGATGGGTTA ACGCCTAAGC AAATTATTGA TATGAAAGGG 5400

TTGGCTGGCG ACACGTCGGA TAACATTCCA GGTGTAACGA AAATCGGTGA GAAAACCGCG 5460

ATTAAGCTGT TAAAACAATA TGGCTCAGTT GAAGGGGTTT ATGAGAATAT CGATGAAATG 5520

AAAAAAAGCA AAATGAAAGA AAACTTGATT AATGATAAAG AGCAAGCCTT TCTTTCTAAA 5580

CGTTTAGCGA CAATTGAAGT CAATGCACCA GTCGAAGTCA ATGTGGaAGA CTTGGcATAT 5640

GAAGGGAAAA ACTTAGaAAA ATTGGTTCCT TTTTATAAAG AAATGGACTT TAAACAATTT 5700

TTAGCTAAGT TGGACATTAC AGAAGAACCT GTTGAAATGG AAGATATTTT ATTTGAAGTA 5760

GTCGAAGATC AGTTAACCAA TGAAATGTTT ACAGACGATA TGGCTCTTTA TGTTGAGATG 5820

ATGGAAGATA ACTATCATAC GTCACCGATT GTCGGCCTTG CTTGGGGCAA CAACAAGAAG 5880

ATTTATACCA CGAATAATTT AGCTGTTTTT GAAAGTCAGC CATTTATTGA CTGGTTGATG 5940

GATGAGACAC GTAAAAAAAA TGTCTACGAT GCaAAACGGA CGTaCGTTGC ACTAAATCGC 6000

TATGTAGGAA AAATGACAGG GATTGCCtTT GATGTTTTAT TGGCTGCTTA TTTACTTGAT 6060

ACCAATGACA ATAACGCTGA TATTGAAGGT GTAGCACAAC ATTACGGATA TGATGCGATT 6120

CAATCTGATG AAGCCATTTA TGGAAAAGGG GCCAAAAAAG GCTTGCCAGA GGACGAAGAA 6180

GTTTTCTTTG GTCATTTAGC ACGTAAGATT AAAGCAATTC AATTTTTAAC AAGCAAATTA 6240

GATAGCGAAT TAACAGAAAA AAATCAAGCT GATTTATTCT ATAAAATGGA ATTACCACTT 6300

TCTCGTATTT TAGGCGATAT GGAAATAACT GGTATCCGTG TTGATGCCAC ACGCTTAAAA 6360

gAAATGCAAG TAGAATTTTC AGAACGCTTG AAAGAAATCG AAGAAAAAAT CTATGCAGAA 6420

GCTGGCGAAG AATTTAATTT GAATTCACCT AAACAATTAG GCGTTATTTT ATTTGAAAAA 6480

ATGGGCTTAC CTGTGATCAA AAAGACGAAA ACTGGCTATT CAACAGCGGT AGATGTTTTA 6540

GAACAACTAA AAGAACAAGC ACCAATTGTG GCCGATATTT TAACTTATCG TCAAATTGCC 6600

AAAATTCAAT CGACCTATGT TGAAGGACTT TTAAAAGTAA TTCAGCCAGA TAATAAAATT 6660

CATACTCGGT ATGTACAAAC ACTGACaCAA ACAGGACGAT TAAGCTCAGT AGATCCCAAT 6720

TTACAAAACA TTCCTATCCG TTTAGAAGAA GGCCGAAAAA TTAGAGAAGC GTTCGTACCA 6780

AGAGAGGATA ATTGGTTGAT TTTCTCTTCT GATTACTCTC AAATTGAGTT ACGTGTGCTC 6840

GCACATATTT CTAATGATGA GCATTTAAAA GAAGCATTTG TGGAAGGACA AGATATCCAT 6900

GCCAGCACTG CAATGCGAGT TTTCGGTGTT GAAAAAGCAG AAGATGTCAC ACCTAACATG 6960

CGTCGTCAAG CTAAAGCGGT GAATTTTGGA ATTGTCTATG GTATCAGTGA TTATGGCTTG 7020

TCTCAAAACT TAGGCATTAC TAGAAAGCAA GCGCAACAAT ATATTGACAC CTATTTTGAA 7080

AAATATCCAG GCGTGAAACA GTACATGGAA GAGATTGTCC GCGATGCCAA AGACAAAGGC 7140

TATGTGGAAA CATTGTATCA TCGCCGTCGT TATTTAAGTG ATATTAATTC AAGAAACTTT 7200

AATCTACGTT CGTTCGCTGA ACGTACAGCG ATTAATACAC CGATTCAGGG AAGTGCCGCA 7260

GATATTTTAA AAATTGCGAT GATTGAAATG GCGCGTCGTT TAAAAGAAGA AAAATTACAA 7320

GCGACAATGT TATTACAAGT GCACGATGAA TTAGTTTTTG AAGTACCAGA ATCAGAATTA 7380

GAACAATTAA ATCAACTCGT CAAAGAAGTG ATGGAGCACG CAGTTTCCTT ACATGTGCCA 7440

CTTATCACCG ATAGCAGTTG GGGAAAAACA TGGTATGAAG CAAAATAATA GCTTACCCAT 7500

TAAGAACGAT GAATAAAATA AGGATAAGTA GGATCTGGAC GTTGATATCT AAGCGTTCAG 7560

ACCTCTTTAT ATAAAGGAGC AAAAATAATG CCAGAATTAC CAGAAGTAGA AACAGTGCGA 7620

AAAGGACTTG AAAAACTTGT TGTAGGAAAA ACCATTCAAG AAGTCATTGT TTTTTGGCCA 7680

AGAATTATTG AATCACCAGA AGTTGATGTA TTTCAAGGTC AATTAGCTGG TCAAACGATT 7740

GAAGGGATTG AGCGAAGAGG AAAATTTCTG ATTTTTAAAT TGTCTGATAA TGATATGATT 7800

TCTCACTTAC GGATGGAAGG AAAATATGAG TTTCATCAGG CTGATGATGA AATTGCCAAA 7860

CATACCCATG TAATGTTTAC TTTTACAGAT GGCACGCAAT TACGTTATTT AGATGTACGG 7920

AAATTTGGGC GTATGACATT GGTTCCTAAG AATCAAGGTC ATCAATATAA AGGGATTTTA 7980

GCTTTGGGCC CAGAGCCAAC ACCAGACGTT TTCCAATTAG CTACTTTTCA ACA 8033

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9813 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

GTTTGGATAA TCGGCGTTAC TTTAGCAATC AGTGCTTTTC TTTTAGCTAG CTGTGGGCAA 60

ACCACACAAT CGCCAAAAGA GAAGAAAGAG CTCACGGTAA TGACTACTTT TTATCCGATG 120

TACGACTTCA CAAAACAAGT CGTTGGTGAT GAGGGAGAGG TTGAGCTGTT GATTCCAGCT 180

GGAACGGaAC CACACGATTA TGAACCAAGC GCAAAAGACT TAGCGAAAAT TACTGATGCG 240

GATGTCTTTG TTTATaATAG TAAAGAGCTA GAAACTTGGG TTCCGAATGT TATTGAAAAC 300

TTGGACACGA AAAAAGTTTC AATCGTGGAA GCCAGTCAAT CCATTCAATT GATGCAAGGA 360

ACAGAAGAGG AAGAATCTGG CGaAGAAGGT CATGAAGGCC ATAACCATTC CCACGAACTG 420

GATCCGCATG TCTGGTTAGA TCCCGTGTTA GCGCAAAAAG AAGTAACAGC GATTCGGGAT 480

GCGCTTATCA AAAAATATCC TGAGAAAAAA GCCGTATTTG AAAAAAATAC CGTTGCTTAT 540

TTGGAAAAAT TGACTGCGTT AGATAAGGAA TACCAAGCTG CGTTTGCAGG CGCTAAAAAT 600

CGGACCTTTG TGACGCAACA TGCTGCATTT GGCTATTTGG CGAAACAATA TGGCTTGACC 660

CAAGAACCGA TTGCTGGGAT TTCTCCAGAC CAAGAACCTT CGCCAAGTCG TTTAGCCGAA 720

TTAAAAAAAT ATATTAAAAC AAACAATGTT TCTGTGATCT ATTTTGAAGC ATCGGCTTCT 780

ACAAAAGTAG CCAAAACATT GGCAGACGAG ACAGGCGTCG AATTAGCCGT ATTGAATCCT 840

TTAGAGAGTT TAACTCAAAA AGAACAAGAA GCAGGCGAGA ACTACGTGTC TGTCATGAAA 900

GAAAACCTCG CTGCTTTACA AAAAAGCATT CATTAATCTG GCAACAGAAA GGAGCCAACT 960

TATGCATTAT ATTGAAGTCG AAAATTTAAC CTTCTATTAT GATGATGAAC CTGTTTTAGA 1020

AGATGTTTCT TATTATGTTG ATCCAGGGGA ATTCGTTATT TTAACAGGAG AAAATGGCGC 1080

TGCCAAGTCA ACGCTGATCA AAAGTACCTT AGGGTTATTA AAACCAACAA GTGGCAAAAT 1140

TACGGTCGCT AAAAAAAATA GTGCTGGTGA AAAAATCAGT ATTGGCTACA TTCCACAACA 1200

AGTCGCTTCT TTCAATGCGG GCTTTCCTAG CACCGTCATT GAATTAGTAC GTTCAGGTCG 1260

CTTTCCCAGA AATCGTTGGT TCAAACCGTT GACGAAAAAA GATCATCTTC ATGTGGAAAA 1320

AGCTTTGAAA TCAGTGGACA TGTGGGAAAT GCGCCACAAA CGAATTGGCG AGCTTTCTGG 1380

AGGGCAAAAG CAACGAATTA GTTTAGCACG AGTGTTTGCG ACCGATCCAG ATTTATTTAT 1440

TTTAGATGAG CCAACAACAG GTATGGATGA ACAATCGCGA AACGAATTTT ATCAATTGTT 1500

GCAACACAGT GCGCATGAAC ATGGAAAAGC TATTTTGATG ATTACGCATG ATCATGAAGA 1560

TATCAAAACC TATGTGGATC GTCAAATTCG TCTTGTCCGC AAAGAAGATT CGAAATGGCG 1620

TTGTTTCCAT ATGAGTGAAG AATCGTATAC GTAAATTTTC CAATAAAACG TTATCTGGCT 1680

CGATCATAAC GAAAGCCTAG GAACGGCTTA AAGAAATGTG GGAAAAATTT AAAAAATAAG 1740

CAGGAATTAA AAAAATAATT CGCTTAAACG TCGCTACAAT GGTAGTGTTC ATTTTAGCGG 1800

TAAGGAGTGA ATCGTATGGC AGAAATGCTT TCTTATGCAT TTATGCAAAA GGCCTTTTTA 1860

GCAGCACTGT TTATCTCAGT GATTGCCCCA ATGCTCGGCG TCTTTCTAGT TATTCGCCGA 1920

CAATCTTTAA TGGCAGATAC CCTTTCACAT GTGTCATTAG CCGGTGTGGC ACTAGGCTTC 1980

TTTTTTAATT GGAATCCTAA TTTAATGACC TTAATTGTCG TGATTGTGGC TGCAATCATT 2040

CTAGAATATT TACGAATGAT TTATAGCACC TATTCAGAAA TTTCGATTGC TATTTTAATG 2100

TCAGGCGGTT TGGCTTTGGC GTTAGTTTTG ATGAATTTAA CAGGAGGCAA TTCAGCTGCT 2160

AGTATTCAAT CGTATTTATT TGGTTCCATC GTCACGATTA CGTGGGATCA AGTGGTTATG 2220

TTGGCAATTT TATTCGTAGT TTTAGTTCTA TTGTTTATGT TATTTAAACG TCCAATGTAT 2280

GTTTTAACAT TTGATGAAGA TACTGCTCAT GTTGATGGGC TACCTATTCA TTGGATGTCG 2340

ATGCTTTTTA ATGTAATTAC TGGTGTGGCG ATTGCTGTGA TGATTCCGAT CGCGGGAGCC 2400

TTGTTAATTT CAGCAATTAT GGTCTTACCA GCTGCAATAG GTATGCGAAT TGGTAAAGGC 2460

TTTAACACGG TGATTATTAT CAGTGTGTTT ATGGGCTTGA TTGGCATGCT AACAGGGTTG 2520

ACTAGCTCGT ATTATTTGGA AACACCACCG AGTGCAAGTA TTACCCTAAT TTTTATTGGT 2580

TTATTCTTAT TAGTCAATAT TTATCGCCGA GTGGTTGTCA TGGTCCAACG AAAACAAAAA 2640

ATGCAAAGAA ACTAAGGAGA AAAACGAAAG TTTTTCTCCt TTATTTCTTT TTTTTAAAAG 2700

AAATGGCGAA CTATTTTCAG TTATTTGAAA AAAATCCTCG AATTTTTCGG TTAAACCCCT 2760

TATAATAAGG AACGATAGAA TGAAAAGGAG TGACGAATTT GAAAATTCGG AGAAGTGAAC 2820

GATTAATCGA TATGACCCAA TATCTTTTAG ATCATCCCCA TACATTGGTT TCTTTGACAT 2880

ATTTTGCTGA CCGATACCAA TCTGCCAAGT CTTCCATTAG TGAAGATTTA GCAATTGTTA 2940

AAAAAACATT TAAAGAACGT GGAACAGGTA TTTTAGAAAC GATTCCAGGC GCTGCAGgcG 3000

GTgTCCgCTT TATCCCTGAA ATTCCTTATG AGGAAGCGGA CAGTTGATTA TGGATTTATG 3060

TGATCGTTTG TCAGAACAAG ATCGCTTATT ACCTGGGGGC TATGTTTATC TTTCAGATTT 3120

ACTTGGGGAA CCGAATTTAT TAAGACAAGT GGGCCGGATT ATTGCTTCTA AATATTTAGG 3180

CAAACAAATT GATGCGGTAA TGACCGTTGC GACTAAAGGA GTGCCAATTG CGCAAgCTGT 3240

GTCTTATTAC TTAAATGTTC CGTTTGTGAT TGTTCGTCGC GATTCTAAGA TTACTGAAGG 3300

TTCAACGGTG AGCGTGAACT ATGTATCAGG GTCTTCAGAA CGTATCGAAA AAATGGAACT 3360

TTCAAAACGC AGCTTAAAAC GCGGCTCAAA AGTTTTAGTC GTGGATGACT TCATGAAAGG 3420

CGGCGGCACC GTCAATGGTA TGAAGAGTAT GATTGAAGAA TTTGAAGCAG AACTTGTGGG 3480

CATTACCGTT TTTGCTGAGT CAAAATTCAA TGGCCGTCGT GCAATCGATG ACTATACTTC 3540

GCTATTGTAC GTGGAAGATG TGGATACACA AACAAAAACG ATTAGCGTAG TCCCAGGAAA 3600

CTATTTCACA GAGAAATAAT TTGAAAAAGA ACCTGAAAAT GGTCCATTCC ATTTTCAGGT 3660

TCTTTTTTTC TGTTTTATTC CTTTAAAAAA AGATGAGAAA ATAAATTGGT CTACCTTTTA 3720

AATGCCTTAT TTTGTTGAAA AAGACAAAAG AGTAAGATAA ACTAGAAAAG GTATAATGAG 3780

GACGATTTAG AAGAAAATCA AACAATTTAG TTAATTTGAG GTGGCAAACG TGGAAAATAG 3840

ATACGCAATT ATTTTAGCAG CTGGTAAAGG AACACGCATG AAATCTAAAC TTTATAAAGT 3900

ATTGCATCCA GTTGCTGGTA AACCAATGGT CGAACATATT TTAGATCAAG TAGAACAAAC 3960

AGAACCAACA GAAATCGTGA CAATCGTTGG ACATGGGGCG GAAATGATTA AAAGCCATTT 4020

AGGCGAACGT AGTCAATATG CCTTACAAGC TGAACAATTG GGAACTGGGC ATGCAGTCAT 4080

GCAAGCACAA GAGTTATTAG GTGGTAAACA AGGAACAACA TTAGTTATTA CAGGGGATAC 4140

GCCGTTATTA ACTGCGGAAA CCTTGAAAAA TTTATTCGAT TACCATCAAG GTAAGAATGC 4200

AAGTGCGACT ATTTTAACAG CGCATGCGGa AGACCCAACA GGCTATGGTC GAATTATTCG 4260

TGATCATGTG GGCATTGTTG AACGAATCGT GGAACAAAAA GATGCCAGTG AAGAAGAAGC 4320

ACGTGTTCAA GAAATTAATA CAGGAACCTT CTGTTTTGAT AATGAATCAT TGTTTGAAGC 4380

GTTAgcGnAA ACAGATACAA ACAATACACA AGGGGAATAC TATTTAACAG ATATCATTGA 4440

AATTTTGAAA AAAGAAGGCA AAGCTGTCGC TGCTTACCAA ATGGCAGACT TTGACGAAGC 4500

AATGGGTGTA AATGATCGTG TTGCTTTATC CACTGCTAAT AAAATTATGC ATCGTCGTTT 4560

AAATGAAATG CATATGCGAA ATGGTGTTAC ATTTATTGAT CCAGACACAA CGTATATTGA 4620

TGAAGGCGTG GTAATTGGTT CAGACACAGT CATTGAAGCG GGAGTCACTA TCAAAGGAAA 4680

AACAGTGATT GGCGAAGATT GCTTGATTGG CGCACATTCA GAAATCGTTG ATAGTCACAT 4740

CGGCAATCAA GTGGTTGTTA AACAGTCTGT GATTGAAGAA AGTGTGGTTC ACGAGGGGGC 4800

CGATGTGGGT CCGTATGCAC ATTTACGTCC TAAAGCAGAT GTGGGGGCAA ACGTACACAT 4860

TGGTAACTTC GTGGAAGTAA AAAATGCAAC AATCGATGAA GGCACAAAAG TGGGCCATTT 4920

AACATACGTT GGTGATGCAA CATTAGGCAA AGATATTAAT GTCGGTTGCG GCGTTGTTTT 4980

TGTTAATTAT GATGGCAAAA ATAAACACCA AACAATCGTG GGTGATCACG CTTTTATTGG 5040

CTCTGCAACG AACATTGTTG CGCCAGTCAC GATTGGTGAT CATGCGGTGA CTGCTGCTGG 5100

TTCAACCATC ACAGAAGATG TCCCTTCAGA AGATTTGGCG ATTGCCCGGG CACGTCAAGT 5160

GAATAAAGAA GGCTATGCTA AAAAGTTACC TTATATGAAG GATTAATCTC AAATAAAAAA 5220

GAACGCTCTT TTGCTTAAGC AAAAGAGCGT TCTTTTTCCT CCTTAAAGAC TAGACCAATT 5280

GCTGCTAGTA TTTGTTTAGA GACGTTCAAG GACTAGACCA ATGCTGTTTA TTCGGTATTT 5340

TTAGTTATTT TTTTCAAAAA AATAGTTTTT GTTCCGCTTT TCCAATATAA AAAGACTTTC 5400

TTTTTATGTA AAGTTCGTTT TTTGGTCAAA ATTACTATTT TTTTCACAAA CAAATCCGCT 5460

ATATTGTCCA AAAATAAAAA CAATTTACAG GAAAAACATC TTATTTTAAC GGAATTTACT 5520

GTTTTTTATT TAACGGATAT ATTATGATAG GAAAAGAAGA GCAGCGTGAT GGGGGAAGGC 5580

TCTTTTTGTT GTTTGTTACT TTGACAATTG TTTGTTGCAA ATGGGTTGAA ACACAAGGAG 5640

GAAATTTGTG AAAAAGTTGA GCTTTAAAAA AGTGAAGTGG GGCATGCATT TTTTAATGGC 5700

TGTTGCGTTG ATAGCGCCAA GTGTTACTAG TACGGCATAT GCAGTAGAAA CAACGAGTCA 5760

ACAAAGTTCA GAAGCAGTAA CAAGTACCAC CGATTCAAGT AGAAAACAAG AACCAGTCAT 5820

TACACAGGAA ACAACAGACA TCAAACAAGA AGCACCAAAT CAGGCTACGA GTGACAGTGT 5880

CAAGCAGTCA CAAGAAACCA CAGCACCAAC AGAGACGACG AATTTAGAAA CGTCAATCGC 5940

TGAAAAAGAA GAAACGAGCA CGCCGCAAAA AATAACAATT TTAGGTACGT CAGATGTTCA 6000

TGGTCAATTA TGGAATTGGT CTTATGAAGA TGATAAAGAA CTACCAGTTG GTTTGTCCCA 6060

AGTAAGTACA GTCGTTAACC AAGTCCGGGC ACAAAACCCA GCAGGCACCG TTTTAATTGA 6120

TAATGGCGAC AATATTCAAG GCACTATTTT AACAGATGAC TTGTATAATA AAGCGCCTTT 6180

AGTGAATGAA AAGACCCATC CAATGATCAC CGCCATGAAT GTGATGAAGT ATGATGCAAT 6240

GGTTTTGGGA AATCATGAGT TTAATTTTGG TTTACCGTTA ATCAAAAAAA TTCAACAAGA 6300

AGCCACTTTT CCAATCTTGT CTGCGAATAC CTACAATAAG GAAGATGGTC TTCGTTTTGT 6360

TGAAGGGACT ACCACGAAGG AACTTGATTT TAATCAAGAT GGGCAGCCAG ATTTAAAAGT 6420

TGGGATTATC GGCTTAACAA TTCCGCACAT TCCTTTGTGG GATGGCCCTC GTGTTACTTC 6480

GCTTAATTTT TTACCTTTGA AAGAAGAAGC AGAAAAAGCA GTTACTGAGT TGAAAGCTAA 6540

CGATCAGGCT GACATTATTG TTGCCTCGAT TCATGCGGGA CAACAAAATA GTGATCCGGC 6600

TGCCAGTGCC GACCAAGTAA TTGAAAATGT CGCGGGGATT GATGCGTATA TTCTGGGTCA 6660

TGACCACCTT TCTTTTACCA AGCAAGGAGC AGCGCCGAAT GGAAAAACTG TACCGGTAGG 6720

GGGACCGAAA GATACGGGGA CAGAAGTTGT CAAAATTGAT CTTTCAGTTG CTAAAAATGC 6780

CGATAAGTGG GAAGTGCAAG AAGGTACAGC AACGATTGTA CCAACAACGA ATGTTCCAGC 6840

AGATGAAGCA GTTAAGGCAG CGACAAAAGA ATACCaTGAA AAAACGCGAG CGTTTATTCA 6900

GGAGGAGATC GGCACAGCAA CAGCTGATTT TTTACCAAAA CAAGAAATTA AAGGAATTCC 6960

CGAAGCACAA TTACAACCAA CAGCGATGAT TTCTTTAATT AATAACGTTC AAAAAGAAGT 7020

AACGGGCGCA CAATTAAGTG CGGCAGCGCT GTTTAAATAC GACAGTAAAT TACCTGCGGG 7080

GAAGATTTCC TATGCCACGA TTTTTGATAT CTACAAATAC CCGAATACCT TAGTGAGTGT 7140

TCCCATTAAC GGTGAAAACT TACTGAAGTA TTTAGAAAAA CAAGGGGCGT ACTATAACCA 7200

AACACAGCCA GATGATTTGA CCATTAGTTT TAATCCAAAC ATTCGTGTAT ATAACTATGA 7260

CATGATTTCT GGAGTGGACT ACAAGATTGA CATTTCAAAA CCAGTGGGTG AACGAATTGT 7320

AGATGCGAAA ATTGACGGCC AACCGCTGGA TCCTGCCAAA GAATATACGA TTGCTATGAA 7380

TAATTATCGT TACGGCGGTT TAGCTAGCCA AGGGATTCAA GTAGGGGAAC CTATTAAAAA 7440

TTCTGATCCA GAAACCTTAC GAGGAATGAT TGTTGATTAT ATTAAGAAAA AAGGAACTCT 7500

TGATCCAGAA CAAGAAATCG AACGAAATTG GTCAATTATT GGGACAAATT TTGATGAAAA 7560

ATGGCGTGCC AAAGCAATCG AATTAGTGAA TGACGGCACT CTTCAAATTC CGACTTCTCC 7620

```
TGATGGACGT ACACCAAACG CCGCCGCTAT TACGAAACAA GATGTCCGTA ATGCGGGCTT    7680
TGATTTAGAT AATGCATATA CCATTATGCA CACAAATGAC GTTCATGGCC GACTAGAAGC    7740
AGGGAAAGgC GAATTAGGTA TGGCGCGTCT AAAAACCTTT AAAGACCAAG AAAACCCAAC    7800
CTTGATGGTG GATGCAGGGG ATGTTTTCCA AGGATTACCA ATCTCCAATT TCTCCAAAGG    7860
CGCGGATATG GCCAAAGCAA TGAATGAAGT TGGTTATGAT GCCATGGCGG TGGGAAATCA    7920
CGAGTTTGAT TTTGGTTTAG AGATTGCACT AGGTTATAAA GACCAACTGA ATTTTCCGAT    7980
TTTATCTAGT AATACGTATT ACAAAGATGG CAGTGGACGG GTTTTTGATC CGTATACAAT    8040
CGTAGAAAAA TCCGGGAAAA AGTTTGCCAT TGTAGGTGTG ACGACCCCAG AAACAGCAAC    8100
GAAAACACAC CCGAAAAACG TAGAGAAGGT GACATTTAAA GACCCGATTC CAGAAGTAGA    8160
AGCAGTGATT AAGGAAATTA AAGAGAAGTA CGCGGATATk CAAGCTTTCG TGGTTACTGG    8220
GCATTTAGGC GTAGATGAAA CGACGCCGCA TATCTGGCGT GGTGATACGC TAGCAGAAAC    8280
CCTTAGTCAA ACATATCCTG AGTTAGATAT CACTGTGATT GATGGACATT CGCATACAGC    8340
CGTCGAAAGT GGCAAACGTT ATGGCAAAGT GATCTATGCT CAAACAGGTA ATTATTTAAA    8400
TAATGTTGGG ATCGTCACAG CACCAGAGAG TGAACCAACT AAGAAAACAA CAAAATTGAt    8460
TTCagcAGCA GAGCTGCTAG AATTGCCAGA AAACCCGGCA GTTAAAGCCA TCGTTGATGA    8520
AGCACGTACG AATTTTAACG CTGAAAATGA AAAAGTAATT GTCGATTATA TTCCATTCAC    8580
ATTGGATGGA CAACGAGAAA ATGTGCGCAC ACGAGAGACC AACTTAGGGA ATTTGATTGG    8640
TGATGCGATT ATGTCATATG GCCAAGACGC GTTTAGCCAA CCTGCTGATT TTGCAGTAAC    8700
TAATGGTGGC GGCATTCGCG CTGATATTAA ACAAGGGCCA ATTAAAGTTG GGATGTCAT     8760
TGCTGTGTTA CCTTTTGGCA ATAGCATTGC GCAAATTCAA GTAACCGGCG CCCAAGTTAA    8820
AGAAATGTTT GAAATGTCTG TTCGTTCGAT TCCACAAAAA GATGAGAATG GCACAATTTT    8880
ACTAGATGAT GCTGGCCAAC CAAAACTTGG CGCAAATGGT GGTTTCCTAC ATGTTTCAAG    8940
CTCCATTCGT ATCCACTATG ATTCCACAAA ACCAGGTACT CGCTTGGCTA GTGACGAAGG    9000
CAATGAAACA GGACAAACGA TTGTCGGTAG TCGCGTATTA GGAATAGAAA TTAAAAATCG    9060
GCAAACACAA AAGTTTGAAC CATTGGATGA GAAGAAACAA TACCGGATGG CTACCAATGA    9120
TTTCTTAGCT GCTGGTGGTG ATGGTTACGA TATGCTAGGT GGTGAACGAG AAGAAGGGAT    9180
TTCACTAGAT TCTGTCTTAA TTGAATACTT GAAAAGTGCA ACCAGCTTGC GGTTGTATCG    9240
TGCAGCAACG ACGATTGATT TAGCACAATA TAAAGAACCA TTCCCAGGCG AACGAATTGT    9300
TTCTATTTCG GAAGAAGCTT ACAAAGAGTT AATCGGTGGA GGAGAGACGC CAAAACCAGA    9360
TCCAAAACCA GACCCGAAAC CAACACCAGA AACACCAGTA GCAACCAATA AACAAAACCA    9420
AGCGGGAGCA AGACAGAGCA ATCCATCCGT AACAGAGAAG AAAAAGTATG GCGGCTTTTT    9480
ACCTAAAACG GGTACAGAAA CAGAAACGCT TGCATTATAT GGTTTACTGT TCGTTGGACT    9540
TTCTTCTTCT GGCTGGTATA TTTATAAACG ACGTAACAAA GCTAGTTAGT CATTTTTTAA    9600
```

GAACCGTCcT TCGGGATTGG TTCTTTTTTT ATTTTAATGC TTGCCATTTG TTTTTAATTT 9660

TAGTATTATA AGAATATtCT AAAAATTAAA AAATGAAGAG TTTTTTCTGT GAAGGGAGAA 9720

GGGCTGAAAG TAAACGTATA AAGGACTTAT CGGGTTAAGT TTACAACGCT GTTAGTTTAG 9780

CTGCCTGTGG TGGTGATGGG GGACCTGGCG CCG 9813

(2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32768 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

CATTGGGAAA AGTGGAAAGC CTTAAACCTC GCAATCGCCC AAGGTTTACC AGCTGAACTC 60

GGCATCACTA AGCCgAAAAT TGAAAGCTGG AcCAACGGCT GGAATTTACG CAgTCATTTC 120

TGGGCTGCGT ATCGCAGTGA GCAACGTCAA GCGGAAAATG CTTGTTTGGC GTTGTTATTA 180

AATAAAAAGC AATTTCAAGT GTATTTGATG TTTCAACATT ACAAAAGTGA GGAGCGGACA 240

GGCTCAGTCG TCGCTTACAA TCAATTGCTT AATCGTTTAG AAGCTTGGAG TCAAACCATC 300

GACTGCGAAG GCTATTATAT TTGGCCGCAG gAGGAGCATG AACTAGTGGA CCATTTGCCT 360

TTAAAAGATT ATTTAGCAGA TGAAAAACAA CGGAAAATTC TAGAAGAGCG GATAGGCAAT 420

CGCTCATTTC AGATCGGTAA GTTGTGGTTT GCGACGGAAA AGCTTGAAAC GATTGAAGAA 480

AAAAGTTGTC AGGCGCTTCA AGAATTAGCG CCGTTGTATA ACGCATTATC TGAAAAATGA 540

ACGAAACAAC GTCTCTGACC AATTACCTGG TCAGAGACGT TGTTTGTGTT TTTATTGATT 600

TAAATCTGTT AGTAGTAGCT CGTTAGAAAC AGCAAAAGTA GTAACTTCTG TTAAAGCTGG 660

ATCCAAAGTG TATTTTGTAA ATGAGGTAGT ACTTTCTGGT TGGACATACA TGGTTAAATT 720

AGTTACATCA TATCCTGACC AATAATGTGT GAAATCAGAC AGAGAATTAC TTTGACCTAA 780

ATGGCCAATT CCCCGAGGAA TCATCACTGT TTGGAAGGCG CGATAGAGTT GCATAATCCC 840

TTCAGAGCTA TCAAATTCTT GGAGATTGTC AGAATAGTAT AACGCACGAA TGAAGCGTGA 900

CGGCGAAGTG AAGTCTCCTG GGATTCCTAA TAGTCCGTAA GAAGTCCCAC CTTCAATTGG 960

GTCAAGTGTT GCGCCTGATT TTAAAGTGTA AGGTTCCTTA ATAGCATAGT TGCGCATCCC 1020

gATATAGTTT CGTGCATTGG TTAAATGATA ATTGTATTCT GGACTGTTTG TGACTACGCC 1080

AATCGAATCA TAAATGACGA ACGCTCCGTT ATCTGTTGGT TGCAAAACGA TACTATTATT 1140

ATTTTTATCA TTAATGGTCC AGTGTAAGGT AAAATGAACT TCCCCCAATG AATCAAGTTG 1200

GAATTTTTGA TCTAACAGGC CGATTTTTTC AAAAGCTACT TTTACTTCGT CAACGCTTTC 1260

AAAATTACTC AAAATATAAG CAAGTACTTC TTCTGCAATA TGAGGTGTTA AACCACGATC 1320

TGCTAAACTT TCAGCAGGTG CTGTTGAAGA TTCCATTAAA TATTGAATAT CACCAGATAA 1380

GCCGGCTTCA TTGATACCAT CTGTGATAGC AAGGCTGACG GTTTTACCAT CGTTCGCTGG 1440

TTGATCTGTG CCTGACATTG CTAGTCCCAT AAAGGCATAT TTTGCCGTCC AAGGATTTAG 1500

TTGGCTGTTT AGAACTGTAC CTTTTGGAAA TTGAGCAATT AGAGTAGGGA TTTTTGGTGC 1560

AATTGGATCT TCGTTGCCAA AAAAGTCAAA CGTAAAATCC ATCGTACGTC CATAAAAAAT 1620

ATCATTTGTT TTGGAAATAA TCTTGATGCC TGTACACATA ATAGTTTCCC CAAATTAATG 1680

TAAGTTTGTG AAGCCTGTTT ACTAGTTAAG CATATCATAC TCTAGGAAAC AGCGGTGGCT 1740

AAATTCCTAA AATATTAGAC AATTTTGTCT AAATGAGGTG AAAAAAATTT TTTAGCGATT 1800

TCCTTCACAA AACACCTTAT TATCTTTCCG AAAATATGAT ATCTTTAGAA TGGAAATTTC 1860

ATGAAAAGAG GATGCGAATG TTTAAAAAGT ACACACAAAT GATAAGGGAT GAATTTAAAG 1920

GGTATAATCA AAAAAAAATG AAAAAAGATT TCTTgGCaGt TTGACTkTCG CAGCCGTCGC 1980

GTTGCCATTA GCGTTAGCGT TTGGGGTCTC TAGCGGTGCT ACTGCAGCAG CTGGCATGAT 2040

CACGGCGATT ATCGCTGGAA TTGTGATTGG GAGCCTTTCG GGTGGGTTTT ATCAAATTTC 2100

AGGACCAACA GGTGCGATGG CAGCTATTTT AATGTCAATT GCGGCAGTGC ATGGCATGCA 2160

AGGGATTTTG TTGGCCACCT TTTTGGCGGG GATTTTCTTG CTTTTGGCAG GGATTTTGCG 2220

CTTAGGGAGC TTGACTTCTT TTATTCCGTC ACCGGTCATT ACCGGGTTTA CTTCAGGAAT 2280

TGCTATTATT ATTGCGTTGG GGCAGCTTGA TAATTTCTTT GGGGTTCgTT CAGAAGGACA 2340

CAATTTAATT GAGAAAATCA GCAGTTATCG TTCATTGGGC TTTGATTTAA GTATACCAAC 2400

GACGGTTATG TCGATTTTAG TGGTGCTTGG ATTAGTGTTC TTTCCGAAAA AGTGGAATGC 2460

GATGATACCA GCTTCATTGG TGATGATTAT TTTAGCTACT TTAGCAACAG TGGTTTTTCA 2520

ACTGCCAGTG GCCACTGTAG GGAAAATTCC GACAAGTATT TTTAGTGATA CACGTTTGGA 2580

ACTTGGTTCA CTTCAATTAA GTGCGTTTCA GGATATTTTG GTGCCTTCGA TTAGCATTGC 2640

TATGTTAGGT ATGATTGAAA GTTTGCTTTG TGGCGCTTCG GCTGGTCGCA TGACAGGGAA 2700

ACCATTAGAC AGTAATCAGG AACTAGTGGC ACAAGGAATT GGCAACTTAA TTTTGCCATT 2760

TTTCGGTGGC GTGCCTGCGA CGGCAGCGAT TGCCCGGACG AGCGTAGTGA TTAAATCTGG 2820

TGCGCAAACA CGTTTAGCGG GTATTTTTCA TGCAGTCTTT TTATTGTTAT CGATGTTGCT 2880

TTTATCGCCA GTCATGTCTA AAATTCCCAT GCCTGCGTTA CTGGTGTTTT aGTGGTGACA 2940

GCGTGGCGAA TGAATGAATG GTCGGTGATT AAAGAATTCT TCCAGAAACG TTTCCACACA 3000

GCTTTATGTT TATtCTTCTT AACGATGGTA GCGACCGTTA TTTTTGACTT GAGTTTAGCG 3060

ATTGTtTTAG GGGTCTTAGC GGGTTGCCTG TTTTyCATTG CGAAAAGTGC CGTGATTACG 3120

ATTAATGTCG AGGAAGTTGA TTGGTCAAGA ATGGGGCTAG TACCGACCGA TTGTACGGAA 3180

AATTGGGTGG TAGTTTATTT GAGTGGCCCG TTATTTTTCA TGTCTTCTGA ACGATTGAAA 3240

CAAACGTTAC TGCAATTGGA AACAAAAAGT GGTGTAGTAT TTTCAATGCG AGGAGTTCCT 3300

AGCATTGATA GTACGGCGTT AGATATTTTA GAAGAATTTT GTCTCGCAGC AGAAGGTCGT 3360

GGACAACAAG TTCAATTTGC CGCTCTGCAA CCAGAAGTTG AAAAAATGAT TCGAACCATT 3420

CAAGGAAATC AAGAGAAAGA ATACCATTTC TCAGTCGCAG AATTTTTACA AACGGTGCAT 3480

GTTGCAGAAC AATAATTCAT TTTCAAATGA AATGTTCTAT CTTCTATCGC GGGCTTAATG 3540

GGTGAAGAAA TAAGAAAAAT AAGTGGGGAC TAAAATGAGT TTCTAGAGGA GTTCCTAAAA 3600

AATAATAAAG AGAAAACTTC TCAGTATGAA GTTTTCTCTT TATTATTTTT TTGGTTATTT 3660

ATTGAATTAC TTCTAGCATA ACGTCTTTTT GATCATCCGT AACATTGCCA GCGTAGTTAG 3720

GAGTGATTAA AGATATCGAA CCATTTTTAT TGTTGAATAA GTATAAAGTG CCCGATCTAT 3780

TTTGTTCATC GTTTGAAGAA ATATTAGTTC CTAAAATTAA TTCTGTATTT ACTTTAACGG 3840

TGCGAATACT GTTATCGGCC GCTGAAAATA TTCTAATCTC TTTTGTAGGA ATTGTATTTA 3900

TAGATAGCTC ATATTTATTA ATGGCTTCTT TCACTTGACC ATTTTCAGCC CCAAACAATT 3960

TAGTTCGGAA GGTAGCAGTT GGTGTTGAAT TTAAGTTCTC TAACGTAATT GAAGTTGGCA 4020

CATTCATTCC TTTAAAATTA AATACGAGTG GATTGTTTAA GCTACTAAGA TCGACAGCAT 4080

AAGGTGTTGC TTCTAAAGAA TTTGCTTCTT CGACATTTGT AGATACTTCA TTTTTCGACT 4140

TTGACGAAGT TTGAGTGGTG CTGGCACTTT CTGAAGTTTC TTTTAGTTTA TTATTAGAGC 4200

TTGTTTCTAT GCTAGAACTG TAATGATTAG TACTGCTTTT TATTGATGTT GTATGATTGG 4260

TAGAATTTTC TTGTTTTTTA GGTTCATTCG CTTTTTGAGA ACTACAACCT GTTAAAAAAA 4320

TAACTAAAAT AACACAAACT AATTTTTTCA TACTATTCCT CCTATAGACG CTATAACTAA 4380

TTTTATCTTA CTTGTTAATA TAACAAAAGA AATAATTATT TTATTTTCAC GAATGCTCAA 4440

ATTGTTTAAA TCTGTAGTTT TTATACTTAT TTTTGTTGCA TTTGCAACAA AAATAAGTTA 4500

AAGTAATGAC AATCATAAAA TATTAGCGAA AGGAAGTGTC CTAGTGACAG ATATTTTAAG 4560

AGAAATTGGC ATGATTGCGC GGGCCTTAGA TTCAATCAGT AATATTGAAT TTAAGGAATT 4620

ATCTTTAACA CGAGGACAAT ATCTTTATTT GGTCAGGGTC TGTGAAAATC CTGGGATTAT 4680

CCAAGAAAAG ATCGCCGAAC TGATTAAAGT TGATCGAACC ACAGCGGCGC GGGCGATTAA 4740

GCGTTTGGAA GAACAAGGCT TTATTTATCG TCAAGAAGAT GCAAGTAACA AAAAAATTAA 4800

GCGGATTTAT GCCACAGAAA AAGGGAAAAA CGTTTATCCA ATCATTGTGC GCGAAAATCA 4860

ACACTCGAAT CAAGTTGCTT TGCAAGGCTT GTCAGAAGTG GAAATTTCAC AATTGGCAGA 4920

TTACTTAGTC CGGATGCGTA AAAATGTCAG CGAAGATTGG GAATTTGTCA AAAAAGGGAA 4980

CACTCGAAAT TATTAAGGAA AGGACGACTG ATTGATGGGA GAAATAAAAC AATTAACAAT 5040

AGAAGATGCA CCAGCGTTAC AAGAAATCAG CATTGAAACT TATACAGATA CTTTCGGACC 5100

ATATAATACC CCAGAAAATA TGCAAGCCTA TTTAGTAGAA GCCTACAATT TAGAAAAATT 5160

ACAAAAGGAA TTGGCTAATC CCCACTCTGA CTTTTATTTT ATCTTTGTCG ATAATCAATT 5220

AGCTGGCTAC ATGAAATTAA ATCGTGAGAC GGCACAAACG GAACCGATGG GGCCAGAAAA 5280

GCTCGAGGTT GAGCGATTGT ATATTTTGCC TGCGTTTAAA GGaAAGAAGC TGGGGACACA 5340

ATTATTAGAA CTGGCGGAAG AGAAAGCTCG TGAATTTGGT AAAAAAGCAC TGTGGCTAGG 5400

TGTCTGGGAA CACAATGATG CGGCGCGGGC ATTTTATAAA ACGATTGGCT ACCGTTATTA 5460

TAGTCAACAT TCGTTTTTTA TGGGAGACGA TGAACAGACG GATTTTATTT TGTTAAAGGA 5520

ATTGTAGCcA AGAAAATTTA GAAACAAAGA TTTCTGATAA AATCAGAAAT CTTTGTTTAT 5580

TTTTTTGACT AAAATTAAGC ATTTTTTTAT TTTCAAAATG TTATTTTTAA AAACGTAAAA 5640

ATTGAAAAAG AAGATTTTGA ATCTGTGAAA AGTTGTCGAT CTTGTGAGAA AACTCTTTTG 5700

TCAGTGTCTT TTTTCAGGAT GGGTCTCTTC CTGAAAaTCC ACTTCAGTTT ATCTGTTTTG 5760

TGTGTGTTTG ATTAAAAAAT GTTGTGTTGC ATTTCATTTA ACTGTATAAT TTaTTTATGT 5820

CATTTTTAAT AATAAGCGTA TTTTAGCATT TGamAAGAGA ATTTGTACAT AATATGTGAA 5880

TAAAAAGATT TGTTTCTTAC TTTCTAGAGT GAGAGGACGA TATAAAAAAG GTTGCTATTT 5940

TTTCGTTCAA AGTGAACAAA TAATCGTTTG TTATTTAGTT GGGGTAACTA ATGATTATTG 6000

TTATAAAAAG CGCTTACAAT TCGTGTGATG AGAGTAGTAC GCATGTTTGT GCAAAAGTGA 6060

CAAAAAAGAG AAGGCAGGGT GTTTTAATAG AAAAAATTTT TTAATAAAAA GGAGAGTTTG 6120

TCGTGGAGAA AAATGTGTCT CAAGTAACAG CTCAAGAAGA AACAGCGCTG AAAGCACGTG 6180

GCTACAATGA GGATTTGTTG CCAAGTAGCC CGAAACAACG GACGATGGGC GCGCGCAATT 6240

TTTTTACATT ATGGATGGGT TCGATTCATA ATATCCCGAA TTATGCAGCA GTGGGCGGCT 6300

TTATTTTCTT AGGTTTATCG CCGCTGCAAG TTATGCTGGC AGTTGTGTTA AGCTCATTCA 6360

TTGTAGCAAC TTTCATGAAT TTGAATGGCG TGGCTGGCTC TAAATATGGC ATTCCTTTTG 6420

CGATGCACTT ACAGTCTACT TACGGCAGTC TCGGGGCAAA ACTGCCAGGG TTCTTAAGAG 6480

GTTGTGTGGC CGCGATTGCT TGGTTTGGTC TGCAAACATT TACAGGATCG CTGGCTTTAT 6540

TAATTATTCT CGGGAAATTT TGGCCGAACT TTTTAGAAAT TGGTGGATCA TTCCAGTTTT 6600

TTGGTCTCAG GCTACCTGAA TTAATGGCGT TTACTCTTTT CTGGCTCTTA AATGTGGCAA 6660

TAGGTTTCGG TGGTAGCAAA ATTTTGAACC GGTTTACCGC CATTTTAAGT CCGTTGATTT 6720

ATGTCGTTAT TATTGGATTA ACGATTTGGG CCATTCGAGC TGGGGGCGGT TTAACACCAA 6780

TTTTGTCCTA TCAAGTTTCT GGTGCGATTC GTTCAGTCAA TCCACTGGTG GCGTATCTGA 6840

TTATTTTTAA CTCGGTTGTG GCCGTTTGGT CAGCGCCTGG TGCGAGTGTG GCGGATTTTA 6900

CTAAGAATGC GCGCTCGACA CGTGCTCAGG TTGTGGGGCA AACAACAGGT TTAGTTGTTG 6960

GTTATGGGAT TTTCGCTTTT TCCAGTGTGG TTATTCTGTT AGGTGGTTCG CTGTATTTTG 7020

GGATTCAAGA ATGGAATATT TTAAACATTA TTGATCGTTT AGATAATGTG GCAGTGGTGG 7080

TCTTAGCCAT GTCGGTTTTC TTATTAACGA CTATTTCAAC GAATGCAACG GGGAACATCA 7140

TTCCGGCAGG TTATCAATTG GCGGCGCTAT TTCCTAAAAA AATGACGTAT AAAAAAGGCG 7200

TCATGATTGC CAGTGTTATC AGCTTTCTCA TTATGCCTTG GAAACTGATG GAAAACGCCG 7260

ATAGTATTTT TATTTTTCTG AACGCAATTG GGGCTGTTTT AGGACCAGTC GCTGGTGTGA 7320

TGATTGCGAA TTATTATTTT GTTCAAAAGC AACAAATTGA TTTAAATGCT TTATACGTAG 7380

ATAAACACAA AAAAGAGGAA GCCAATCCTT TTTACGGATT AAACAAGCCG GCGTATGTTG 7440

CTACTATTCT GGCGTTGGTT CTTTCCTTAA GTGGTCAATT TATTCCACAG GTTAAAATCA 7500

TTGCTGATAT TTCGTGGTTT GTGGGCTTTG CCACAGGTTT TGTGCTGTAC TTAGTGTTAA 7560

AAAAATGGAC TTGGGATTCA AAAAAAGTGA AAGAAACAGC TTATCAGGAG GGAAAATAAA 7620

TGCAAGCAGA GTTAGTGATT AAAAACGGGT TAGTTATTTT AGAAACTGGA GAAGTCATTA 7680

CGGATGTCGC CGTTCAAGGA GGAAAAATCG TAGCGATTGG GCAAGACTTG TCAGGAGAAC 7740

GTGTGATTGA TGCGACGGGT TTAGTCGTTA GTCCTGGGAT GGTGGATGCT CATGTTCATA 7800

TTACCGATCC TGGTGGTGGT TACCGCGATG AATGGGAAGG GTATGTCACA GGAACGGCCG 7860

CTTGTGCCAA AGGGGGCGTG ACGACCTTTA TGGAAATGCC TTTAAATCAG ATTCCAGCAA 7920

CTGTGGATAA GACCTCTCTT GAGATCAAAT ACAAGGCAGG AGAAAATAAA CTAAAAGTAG 7980

ATGTAGGGTC ATTTGGGGGC GTGGTCCCGA CCAACTTGGC TGATGGGATT CAGGAACTTG 8040

ATGAAGGAGG CGTTTCTGGC TATAAATGCT TTTGGGAAC GTGCGGCGAT CGTTCAATTG 8100

AAGGAGATTT TCAAAATGTC GACGATTATT CATTATATGA AGGAATGAAA CAAGTCGCCA 8160

AAACTGGCAA AGTGCTAGCA ATCCATGCAG AAAATGCGCC AATTACGGAT AAATTAGGCG 8220

CGGTAGCTTA TCAAAATGGC GAAACGACTT TAGCGGCATA TGTAGCTACA CGTCCTGTGT 8280

TTACCGAAGT TGAAGCGATT CAAAAAGCCA TTTTATTTGC CAAAGAAACA GGTTGTCGGA 8340

TTCATATTTG CCATGTGGCT TGTCAAGAAG GCGTGGAAGA AGTTTTGAAA GCCCAAGCAG 8400

AAGGGGTAGA TGTGACTTGT GAAACATGTA CGCATTATTT ATATTTTACC ACGGATGAAC 8460

TTGATGCCAT TGGTCCAGTT GTCAAATGTT CACCGCCGAT TCGTGACGCG GACCAACAAG 8520

CCGCCTTATG GAACCACGTG CAAACAGGCG GCATCGCGTT TGTAACCTCT GACCATTCTC 8580

CTTGTACGCC AGACTTAAAA GATACAACCA ATGCTTTTGA AGCTTGGGGC GGCATTTCTG 8640

GTGTTCAAAA TAACGTCGAT GTGTTATTCG ATGAAGCCGT CCAAAAACGT GGTTTATCCT 8700

TGAAACAATT CGCGGATATG ATTGCAGCAA ACCCAGCTGA CCGTTATCAC TTAGCACAAA 8760

AGGGCCGCAT CAGCATTGGT AAGGATGCCG ATTTTGTACT AATTAAACCT AATGCGCCGT 8820

ATATCTTAAA AGCAGAAGAT TTAGAATACC GAAATAAAAT TAGTCCTTAC ATTGGTCGTG 8880

AAATTGGGGC CCAAGTGATT CAAACAATTT TAAGAGGCGA AACAATTTAC GCACAAGAAA 8940

CTGGGGTAAC CGAGGCGTTT AACGGCGTAT TTATCAAAAA TTAAACACCT ATTTTAGAAA 9000

TGAAAAAGGA CAGAAGCGTG ATTGAATTAG CGCCTCTGTC CACGAGTTAG AGGATGTTTG 9060

TTGGTTGTTT TGCACAAAAA AAGGCTTAAT TTTGTTTGAT AAGTACAATG AAGATGGCTA 9120

GGGAAACGTT TACAATCAGA GTGTAATCAG TCAGGAGGTT TTGTGATGGA CTTGGAAAAA 9180

GTTTTAAAAC AAAGAATCGA AGAGTTGTCA GCAATTGGGA GTGATCCTGC AGGAGGGATG 9240

ACACGCTTAC TTTATACAGA TTCGTGGTTG GCAGCACAAA AGTATGTGCA AAGCCAAACT 9300

GAGGCGTTTG GACTGGAAAC GCAATTTGAC GAAGTCGGGa ATCTGTTTTG TCGAGTGGCA 9360

```
GGAACGGAAT TTCCTCAGGA AACGATTTTA ACTGGGTCAC ATATCGACAC AGTAGTTAAT    9420
GGTGGGACCT TAGATGGTCA ATATGGTGTC ATTGCTTCCA TGACAGCTGT TCAATATTTA    9480
TTAGACAAAT ATGGGAAACC ACGGCGTTCT TTAGAAGTGC TTTCCATGGC TGAAGAAGAA    9540
GGGAGTCGTT TTCCAACGGT TTTCTGGGGA AGCAAAAATG TCGTTGGTGA AGCCAAACGT    9600
GAAGAAGTAG TCGATATCTC TGACAACGAA GGTAAAAAGT TTGTTGATGA GATGCATCGC    9660
CAAGGCTTTG ATTTTCGGCA AGAGCAAACT AGACGTGAGG ACATTGAAGC ATTTGTCGAA    9720
ATTCATATTG AACAAGGCAA TATTTTAGAG AACGAAAAGC TCCAAGTCGG TGTTGTTCAT    9780
AGTATTGTAG GGCAACGTCG CTACACGGTT AACTTAAAAG GGCAAGCCAA TCATGCAGGT    9840
ACTACGCCAA TGAGTTACCG TCATGATGCG GTGTACGGCT TTGCTAAAAT TTGTGCAGAA    9900
GCCATTGATC GTGCGAACGA AATTGGGGAT CCGTTAGTTT TAACATTCGG CAAAGTCATT    9960
CCGAAACCCA ATACAGTCAA CGTGGTGCCT GGAGAAGTGG AATTTACAAT TGACTGCCGT   10020
CATACCGACG CTGCTTTTTT ACGACATTTT ACAGGTGAAT TAGAAGAACG GATGAAGACA   10080
ATTGCTCAAG AATTAGGCTT GACGATTGCG ATTGATCGCT GGATGGATGA AGCCCCTGTA   10140
CCGATGAATC AAGCAATTGT AGAAGTTATT GAACAAAAAG CCAAAGAAGG CAAGTATCAG   10200
TATCGTATGA TGCATAGTGG TGCTGGGCAT GATTCGCAAA TTATTGCGCC TCATTACCCG   10260
ACAGCGATGA TATTTGTGCC AAGTATTGGT GGGATTAGTC ATAATCCTGC AGAAGCAACG   10320
GCATTTCCAG ATTTGGTTGA AGGTGTCAAG CTTTTAGCAG ATACCTTATA TGAATTAGCG   10380
TATAAATAAG ATAGAAAAGG AAGAATTGTG ATGGGTTATA AAAATAATCG AGTTGGTTAC   10440
CAAAAAGAAT TGTTAACATC AAGAGCTGTG ATAAAAAAAG ACAATTACGC CATTATTCCC   10500
CATGATGGTT TGGTGCAAAA TGCAGTACCT GGTTTTGAAA ATGTCGATAT TTCCATTTTG   10560
GGATCACCTA AATTAGGTGC GACATTTGTG GATTACATTG CTACGTTTCA TAAAAATGGT   10620
CAACAAACAA CTGGTTTTGG TGGCGACGGG ATCCAAACGT TGGTTTATGT CATTGACGGT   10680
CGTTTGCGCG TTTCAGATGG TCAAGAAACT CATGAACTTG AAGCCGGGGG ctACGCCTAC   10740
TTTACGCCTG AAATGAAAAT GTATTTAGCG AACGCACAGG AAGCGGACAC AGAAGTTTTC   10800
TTATATAAAA AACGCTATCA GCCGTTGGCA GGCCATCAAC CTTATAAAGT GGTTGGGTCA   10860
ATTCATGACC AACAACCCGA AGAATATGAA GGAATGACAG ACGTGTTACT TTGGAGTTTA   10920
TTGCCgAAAG AGTTTGACTT TGATATGAAT ATGCATATCT TATCTTTTGA ACCAGGTGCT   10980
AGTCATGCCT ACATTGAAAC GCATGTCCAA GAGCATGGCG CGTATTTAAT TAGTGGCCAA   11040
GGGATGTACA ATTTGGACAA CGAATGGTAT CCAGTGGAAA AAGGGGACTA TATTTTCATG   11100
AGTGCCTATG TGCCACAAGC AGCCTATGCG GTTGGTCGTG AAGAACCGCT AATGTACGTT   11160
TATTCCAAAG ATGCCAATCG TGAACCAGAA TTATAAAGAA AAGCCGAATG TAAAGTAATG   11220
ACCTCGGAAC AATAGGATTT GAACACGCAG tGgTGAACGG TAGAGTCCTT CTTCTTGGTA   11280
GACGTTACTT TTATTCGGCT ATTTTCCTGT TAATTATTTT TCTGAAAGTT CTAACGCTAC   11340
```

CCGTAAGAGT AAACTCGTCT CAGGCTCTTG GATAGAGGTG CCTAATAATT TTTCACATTG 11400

ATTCATGCGA TATTTAATGG TGTTTCTGTG AAGGTAAAGA GCGTTGGCCG TTTTAGTAAT 11460

TTCGCAATTG AAATCTAAAA AAACTTTCAA GGTTTTTCGT AATTCTTGTA ACGTCGGTTC 11520

CGTCGGATAG GCCAACTCTT TCAATTGTTG CTGGCAAAAG TATTCAACAT CTTCAGTGCC 11580

AATCTTTTCA AAAAGTCCAG CTAGGCCTTT GGGATGGAAC AATTGCACAG TCGCTGGTTT 11640

TTGAGCGTGT AAAGAAGCCT CTAATGTCGA GGAAGCTTCA ATATAAGAAT TAGGCAAATC 11700

TTCTAGGTTT TCATAAGCAT TGCCTAAAGC AAAGCGAATG GTAATAGGCA AAGCCTGTTG 11760

CAAACGTTCA GCGAGGTTTT GCAAAATCAT CAAATGGTCA TTTTTTTTAG ATTGAAAAAT 11820

CAAGATACTT TTATTTTGAT TTTTTAATTT AAACAAGGCG ACATCTGGCA GGATTTCCGG 11880

CAATTGTTCT TTGAGCCATT GGAAAATTAG TTGCCCTTCC GCTTGCTGAT AACGAATATG 11940

GGTTTCATTT TCTGGTTTAG TCACGCCGTA GACAATCGCT AATTGATAAT AATCACTGTT 12000

GATTAAACGA TAATTTTTCC AAAGTTCCAA CCAATTTTGA TGTTTACTCA GCGCTTCTTG 12060

ATGTGTATCA AGCAAGCGGT CCAAAAAGTC TGTTTTGAGA TGCTCAAAAG ATTCAGCAAT 12120

TTTCTGGTTT TTAAATAACA TAAAAGTCAA AACTAAAATG GCCTGATCAA TGGCAAATTC 12180

AGAAATAGGA TAAGGAATCT GTTCAGGAGA TAAGACCAAT AGATAATAGG GAAAATAATC 12240

GTTCACACGA ATTGGAAAGC CAGCTACTTG AATGTTTTCT TGTCGCTCAT CCTGTAAACG 12300

AAAGATTTTC TTTTCTTGTG CCAATTGTTG AAAGTGGTCT TTTGATAATT GTTCGATGAA 12360

AAACTCCGCC GATTTTTGAT TGCCATAAAA ATAATTCGAG TGGGCAATCA CCTGTTGCCA 12420

AGGACTTAAT AAAATAATTG GCGCATTTAA AATTTTCGCA AATTCAGCGA TGAAGCGCGT 12480

AGCATCAACA TCTTGCATTA ATAATGTTGA AAAACGCTTT TGTATGTCAA ATGCTACAGA 12540

CATCTGTTCG GTTTTGCTAT CACGAAGATA GCCTACAATT TCATGCAACA ACCCGCCTAG 12600

TGGCTGTGTG CTAGGAATTT TAATCAGGGG CAAATCAACG GCACTGGCAT AGGCAACAAT 12660

TTCTGGTGAA ATTTCGTCAA GGAACCGCCC AACTTTGATT CCTAAAGCCG TACACTCTGC 12720

TTGTTTTAGC GAGTCAATAA AAGGTTTCAA TTTTTCTTGA TCATCTTTAT AAATCATTGC 12780

CGTTGTTAAA ATCATGACAT TTTTAGGGAT AAAATCAGCC ACATCAGGTG TTTCCGTGAT 12840

TTCTACACTT TCCAAGGGCT GTGTTAAGTT TGAGTGAGAC GATAATAATT GTAAGTCAGA 12900

AAATCTTGGT AAAGCCAAGA CCTCAGCTAA TATTTTCATC ACAACAACCT CCTTTTTTTA 12960

TAGTATAGCG AAGAATAAAA AGGAGTACAA AAGATAGTTA TTGAATATGT CTAAAAAACA 13020

AAGTCAAAAT AGTGCAAATT GAACAATGAA TTTCTGGGTT TTTCCAAAGA TAATAAAAGA 13080

GAAGAATTAA AAATGAAGTG AGGGATTACT ATGTATCAAC AACCAAATTT TGTACGGCGA 13140

ACAATCATGA CGCCGGGACC AGTCGAAGCC CATCCATCGG TACTTCGTCA AATGGGTCAA 13200

CCAATTTTAG GGCAATTTGA TCCAGAATTT CTACAGATTA TGGACGAAGT TCGGGAAATG 13260

ATTAAAGTAC CTTTTGCCAC AAAGAATCAA CAAGCCTTTG CCATTGATGG CACGTCACGT 13320

TCAGGCTTGG AAGCGGGGTT ATTTGCGTTG ATTGAGCCAG GCGACAAAGT GCTTGTGCCT 13380

ACGTATGGCC GGTTTGCTTA TTTATTAGGT GAAATTTGCG AACGCGCCCG GGCCGAAGTC 13440

ATTTATTTAG AAAAGGACTG GCTCGCACCG TTTGAACAAA CAACGGTGAT TGAAGCGATT 13500

AAAGAACATC AACCAAAAAT TGTTGCGATG GTCCATGGCG AAACGGCGAA TGCACAAATG 13560

CAACCACTGG ATCAAATTGG CGCCTTTTGT CGGGAAAATG AAATTTTCTT TGTAGTAGAT 13620

ATGGTGGCGA CTTACGGCGG TGTTGAAACA AAGGTCGATG ATTGGAAGAT TGATATTGCG 13680

GTAGCAGGCA CTCAAAAATG TGTCAGCGTA CCTTCAGGCC TTTCTTTAAT TACCTACAAC 13740

CAACGTGTAG CCGATTATCT GGCGGGCCGT TACCAAAAAG AGTTAGGGCT TGGCGCGGAT 13800

GCGCGGAATG AACGGTTCAT TCAAAGTAAC TACTTGGATT TATCGCAATT AGAAAAATAT 13860

TGGGGACCAG AGCGCTTAAA TCACCATACA GAAGCGACTA CAATGATTTA TGGGCTACAT 13920

GAAGGGCTGC GCTTGTTACT TCAAGAAGGC ATGGAAAATG TCTATGCACG CCATCGGAAA 13980

AACAATCGTA TTTTAGTTGA AAGTCTGCAA AAAATGGGcT TAGAAATTTT CGGGAAGTTA 14040

GACACAAAAA CACCAACAGT CATTCCTGTG GTCATTCCAG AAGGAATTGA TGGTGAAAAA 14100

GTGCGGAGCT TATTACTGGA TCATTTTAAA GTTGAAATTG CCTCTTCcTT CGGTGATTTA 14160

AAAGGAAAAA TTTGGCGGGT CGGTAATATG GGTTACAGCA GTCGCGAAGA CAATGTGTTG 14220

CATTTCTTAA GTGCCTTTGA GACTGTATTG AAACACCAAG GCTATCAATT TGAAGGCGGT 14280

AGCGGTAGTA CTCATGCATT GGCTGAGTAT TTAAACTAAA AAAGAAGGTA GGACAAAGGT 14340

CACTACGTAA AAAGCGTAGT GCTTTTGTTC CTTGTGAAAC GTTAACCAAA AAAATGATTG 14400

GTGGAGGAAA AGAAAATGAC AGACAAGACA AAAAGTTCGT TGGCTTATTG GAAAAAAATC 14460

GTTATTTTAC TATGTATGGG CTGGGTAACA ATCTGGATTT ATCGTTCCGC GCTTTCTCCT 14520

GTTTATCCAC AAATTCGTGC GTCTCTAAAT GGTAACGCTT CCGATACGGC CCTAGGGTCC 14580

ATTTCTAGCT TTTATTTTCT CGGTTACGTA GTCATGCAGA TTCCTGCAGG CTTCTTGGTA 14640

GACAAAATTG GGAAAAAACG TGTATTAATT CCAGGCTTTA TCTTGTTTGC GCTGGCAGCA 14700

TTACTGATTG CCCAAGCCCA AACGATTGGA ACAATTTACT TAGGTAGCTT ATTGGCAGGC 14760

TTAGGATGTG GTTCTTTCTA TGGCTCGGCT TACTCGTTGA CGTCGCAAAA CATTCCTCAA 14820

GCCAAAAAAA GTTTTGCAAC AGCTATTGTC AACAGTGGTT CAGCCGTTGG TTCAGGTTTA 14880

GGGATGATTT TATCTAGTTA TTTAGTTGCC CAAAAAGGAT TACCTTGGCA AGTGATGATG 14940

TACATTTCCG CCTTTATGAT TATTTGTATG TTAGTGGCTT TTCAAGTGAT TATTCGTAAT 15000

CATAAAGAAG ACATGGCCTT GATTCAGCCA ACCGTTGTTG ATGAAAAAGA TAACGGAGCG 15060

GTAAAAGAAA AAGTGCCCTT TAAAAAACTG TTTGCGCCAC ATATGCTATT TGCGTATATT 15120

TTATATTTTG GCACTTGTTA TGCATATTAT ATGATTGTGA CTTGGTTACC TAATTTTTTA 15180

AGTACCGAAC GTGGTTTCCA AGGAGCAGCA ATAGGGTTAT CCTCTTCACT GGTCGCATTT 15240

GCCTCTATTC CAGGAGCGTT ATTTTTTAGT CGGTTGGCGG ATAAGTACAT GCATAAAAAA 15300

```
GTTCAATTCA TCGTTGTTTT AGAATTTTTA GCAACAGCCA TGCTCTTGTT AACTGTACAA    15360
GTAAGTAATG CGACGTTGTT ATTAATTGCT TTAATTATGT ATGGGTTCTT AGGAAAATTA    15420
GCTGTAGAAC CGATTATTAT TTCTTGGCTA GGTGAAAATG CACCAAAAAT TGGAATTGGC    15480
ACTACTTTGG GCGTGTTTAA CTTTTTCGGC ATGATGTCAT CTGTGATTGC CCCAGCTTTA    15540
ACTGGGAAAA TTTCTGACAC AACGGGTTCG AAAGTCATGG GCTTTTATAT CGCCGTCGTG    15600
ATTTTACTAA TTGGCACGAT TCTTTTTATG CTTGTCAATT TAAAAAAACG CAGGGCAGAT    15660
TCGTCTAATA AATAAAAAAA CAACGCAAAC TTATCTTTCG CTGAATAAGT TTGCGTtGTT    15720
TTAgcTTTTT GACAATCATA TTAATCGCTT GTGCTAATCG CTCTTCTTGT TCTTTAGGAT    15780
CTTTTTCAGG ATTTTCAAAG CAATGCTTTA AATTCTCCGC AACTATCATC CCCATTACGC    15840
GATCAATACT AGAACGGACC GCACTCAACT GAGTGATAAC ATCGATACAT TCTTTTTCTT    15900
CATCAATCAT TTTTTGAATC CCGCGAATTT GaCCTTCTGT GCGGCGTAAA CGATTCAGCA    15960
CTTTCTTTTT TTCTTCAGTC ATTTTGCCCC TCCTTAAGAG TAGTGATTGG TTTAAAATAT    16020
ACCTACACGA GGTATTTTTG TCAACTATCT TTTCTTTGAA TCTAAAAATG GCAGAAATCT    16080
CTTCAGTTGA CAAGtCTAaC TTTTtCATGT tAAATATmCC CCGTAAGgTA TATAAGAGAA    16140
GGAGAAAAaG rTGTTATTTC AATTGTTTAA AAAAACGCCA AGTATTTCrA CAAAGGAATT    16200
GCAGTCACGT TTATCTAAAG AGATAACGTT ATTAGATGTC CGGACACCTA GCGAATACCG    16260
CGCAGGGCAT ATTCCACAAG CGATTAATGT CCCACTTAAC AAAATTCCCG CCTACAACAA    16320
ATCCGCTAAC GAAGTCTATG TGATTTGCCA GTCCGGCATG CGCAgCAAAA ATGCGGCCAA    16380
AATTTTAGCG AGAAAAAACT ATCACGTGAT TAATGTTCGT GGCGGGATGT CCCAATGGTC    16440
TGGTCAAATT AAAGGAGGAA AATAAATGAA AATTGTAATT ATCGGAGGCG TGGCTGGTGG    16500
TATGTCAGCA GCGACACGTT TACGTCGTTT AATGGAAGAT GCTGAAATTG TTGTCATGGA    16560
AAAAGGCCCT TTTGTTTCAT TTGCAAACTG TGGTTTGCCT TACTACGTTT CTGGCGAAAT    16620
CGCAGAAAGA GAGCAACTGC TTGTTCAAAC ACCCGAAGCG TTAAAGGCAC GGTTTAATTT    16680
AGATGTTCGT CCTCACCATG AAGTCGTGGC TATTGATCCA ATAGAAAAAG TGATCACAGT    16740
GAAACATGAA ACAGAAATTT TAACAGAACA CTATGACAAA TTAATTTTAT CACCAGGGGC    16800
GAAACCTTTT GTGCCACCAA TTACAGGATT GGCAGAAGCC AAAAATGTTT TTTCATTAAG    16860
AAATGTTCCT GATTTAGATC AAATTATGAC AGCCTTGACA CCAGAAACAA AACGAGCCGT    16920
CGTGATTGGC GCAGGCTTCA TTGGCTTGGA AATGGCAGAA AACTTGCAAA AACGCGGATT    16980
AGAAGTCACT CTCGTGGAAA AAGCGCCTCA TGTTTTACCG CCATTAGACG AAGAAATGGC    17040
CGCTTTTGTC AAAGCTGAAT TAAGCAAAAA CAATGTTCAA GTAATTACGG GACAATCTGC    17100
GGTTGCTTTT GAAGAAGAAG GGCAAGTGAT TCGCTTAGAA GACGGTCAAA CATTAGCTTC    17160
TGATTTAACC ATTTTGTCGG TGGGGGTCCA ACCAGAAAAT ACCTTAGCAG TTGAAGCAGG    17220
TGTAGCAACT GGTTTACGTG GCGGTATTGT TGTTGATGAA CACTATCAAA CGAATCAACC    17280
```

CGATATTTAT GCGGTTGGGG ATGCTATTGT AGTGAAACAA CAAATCACTC AAGAAGATGC 17340

GCTGATTTCT TTAGCTTCTC CCGCCAATCG CCAAGGACGT CAAGTAGCGG ATGTGATTGC 17400

TGGGTTAGAG AGAAAAAATC AAGGAAGCAT TGGGACCGCC ATTGTGCGAG TCTTTGATTT 17460

AACCGCTGCT TCAACTGGTT TAAGCGAACG GGCTGCTAAA GCTGCTGGAC TAACAACAGC 17520

TGTTGTGCAT ATCAGTGGAA AAGACCATGC GGGGTATTAT CCTGGCGCAA CAGATCTTCA 17580

GTTAAAATTA GTTTTCCATC CTACGACAGG GGAAATTTAT GGCGCACAAG GAATTGGGGC 17640

AAAGGGCGTA GATAAGCGGA TTGATATTCT TGCGACCGCT ATTAAAGGAC AGTTAACTAT 17700

TTTTGATTTG CCTGAATTAG AGTTTACCTA TGCGCCGCCG TTTGGTTCAG CGAAAGATCC 17760

TGTGAACATG TTAGGCTATG CAGCGATGAA CCTTGTAGAA GGATTGAGTG AAAACGTTCA 17820

ATGGTATGAG TTATCCAACG AATTAGCTAA TGGAGCTGTT TTATTAGATG TCCGTAATCC 17880

CGCCGAACGA GCCAATGGTC AATTTAAAAA TGCTGTGTCT ATTCCTTTAA ATGAGTTAAG 17940

AGAACGTTTG GAGGAATTAG ACAAGTCAAC GGAGTACATT GTTAGTTGTC ACAGTGGTTT 18000

GCGTAGTTAT ATTGCAGAAC GGATGCTAAA ACAAGCAGGC ATCTCAGCCA AAAATTTAGA 18060

TGGTGCTTTT GCGCTATATC GAATGGTAAA ACCGGAGGAA CTAGAAAATG TATAACACGA 18120

TTGGCATGCC TGAATTTTAT CaGGAAGCAA AACGAAAAAA ATTACCGATT ATTGATGTAC 18180

GAGAAGAAAT GGAATACCAA CTGGGGCATG TCCCTCAGGC CATCAACTTA CCTTTAAGTA 18240

CTTTAACCGA AAATTATCAG CAATTAGATA AAAACCAACC CTATTATCTG ATTTGTCAAA 18300

TGGGGTCGCG CTCCGCACAA GCGTGTGCGT TTCTCAGTCA ACAAGGGTAT CAAGTAACCA 18360

ATGTTCTAGG TGGTACATCT GGCTGGTTGG GTCAATTAGA ACAATAAAAT AACAAAAAAC 18420

GTTTATTCTT TTACACAGGA TAAACGTTTT TTTGTTAGAT AAAGATGTGT CAAAACTGTA 18480

AGAGGAATGA TAGGAAGATG AAGGGACCAG TTTTTTTATT CATTATATAC TAAATAAAAT 18540

GACAAACTCG TCAACAGAAA TCAAGGAGGA ACAAATAAAT GAAAAAGTCA GGCCAGAAGA 18600

AGCGAAATAA AAAAGTGTGG TTCGGTATTG GTGCCGCAGT GGTCGTCGTC GGATTTATCG 18660

GAGCAAAAAC AGTTTTTTCT TCAAAAGAAG TTGAACCTGA ATACACAACG TATACCATTA 18720

CAGAAATGGC CTCATTAAAA TTAGATGGTC AAGTGAGTTT TCTAGATACT CGGGATATTT 18780

TCTTTGATCC TTCTTTAGGA AAAATTGCCG AAATAAATGT TGAAAATGGT AAAGAAGTCA 18840

AGAAAGATAG TCCATTGTTA ACGTACAATA ACTCAGACAT CCAAGCCACA GAAACAGAAC 18900

AAGCGAATGC TGTTAATCGG AACAATCTTC AAGTTCAACA AGCACAAGAG AACGTTAACT 18960

TAGCCACTCA AAAATACAAT GAAGCGTTAA ACAAAGTTGC TGCAGCCAAG CAAAAATTAA 19020

ACACCGCAAA AGAGGCTGAA GAGAAAGAAA CACTAAATGC GGAAATTCAG CAATTGAATG 19080

AAGCCGTTAG TGCTGCGAAT AGTGAAGTCG CCCAAGCTAA CCAAGCCCTT CAATTAGCCA 19140

ACAGTGATGC CGTCGGAGCA GCAAACACGC TTGAACAAAC ACGTGGCAAA GTCAACACAG 19200

TCGTTACAGC ACCAATTGAT GGCCAAGTCA CTGTTGATGC TAGTGCCATG AGTAGTACCG 19260

ACAAACCAGT CATTAAAATT GCAACACAAA AGAAAAATAT TCAAGGAAAA GTAACCGAGT 19320

ATGATTACGA TAAATTACAG ACAGGGGAAG AAGTAACAGT GACGACCGTT GGTTCAGGTA 19380

AATCAGCACC AGGCAAAATT GTTTCCATTG CGCAAACCCC AATAGCTAAA AATGAAGGCA 19440

ATCCAGTAGT GAGTTATCAA TTTACAGTAG AAGGTGATTT TCCTTGGGCA GAAGGCTTGT 19500

CTACTAGTAT CGCTGTGCCA CAAAAGCAAA TGATTATACC GACAGCAGCA ATCCAAAAAG 19560

AAGGCCAAAA AGAATTTGTC TATGTTTACA AAGCGGGCAA AGCGAAAAAA ACACCGATTG 19620

AAACAGAAAC AAATTTAGGT CGTAAAGTTG TCAAAAGTGG CTTGAATTGG AAAGACCAAG 19680

TGATTGCCAA TCCCAATAAA GAATTAAAAG ATAATCAAGA TGTTCAGGTA GCTGCCAATG 19740

ATTAAGCTAG CCCATATTCA TAAATATTAT TATTCCGAAG AAGAAACACT GCATGTGTTA 19800

GATGATATTA ACCTCCAAGT AGACGCTGGA GAATTTTTAG CGATTATGGG TCCTTCTGGT 19860

TCTGGGAAAT CGACTTTGAT TAATTTGCTA GGCTTTATTG ATAAAAAGTT TGAAGGAACG 19920

TATTTATTTG AAGACCGTGA AATTGGCGAC TTTTCTGATA AGGAACTATC CCGAATTAGA 19980

AATGAAGCAG TCGGCTTTGT CTTTCAGAAT TTTAGTTTAA TTGAAACACT AACAGTTGAA 20040

GAAAACATTG aATTACCTCT TTTATATAGT GGGTTAACCC CAAAAGAAGC CAAAGATCGT 20100

GTCCACGAAG TTCTGACAAA AGTCGGCTTG CCAGATAAAG GCAAGAAGCA TCCAAAACAA 20160

CTTTCGGGCG GACAACAACA ACGTGTTGCC ATTGCGCGTG CGATTGTCAA TCGGCCAAGT 20220

TTTATTATTG CAGATGAACC AACTGGGGCT TTAGATAGTA AAACATCGGA AGAAATTTTA 20280

ACGCTCTTTC AGCAATTAAA CAATGAAGGT GTCACCATTA TTTTAGTGAC ACATGATGAA 20340

GAAACAATTG AATACTGCAA TCGTTTGATT AAAGTTCGTG ATGGAAAGAT TTTAGAGGAG 20400

GTGCTGACAT GAAACGTTCA ATTATTTGGA AAACCGCCTT CCGTTCCATT TTGAAAAATA 20460

AGCGACGCAG CTTGTTAACG ATGCTAGGGA TTGTCATCGG GATCGCTTCG GTGATTACCA 20520

TCGTGGCTAT CGGTAACGGC TTTAAAGAAG ACATGGTCGA TAAATTATCT GCTGAAAAAC 20580

AAAAAGAAAA TGTGAAAAAG ATTTCTTTTT CGGCCTACAA CACTTCGGAT ATGTTTAGTG 20640

ACCAAGCGAT GTTTACCGAC AATGATTTAG GGGTCGTGCG CATGGTTCCT GGCGTAGAAA 20700

AAATTGATTT TGATCAAAGA GAAGTTGATG GCGCCCAAAA AGGTACACTG AATTTTCAAG 20760

CAGGACCGAA AAATCTTTCT GTAACCTATG AACTTGCCAA CAAAACCACC AAAGAATTGT 20820

TGGCGGGACG CCAACTGAAT GATGGAGACA ATGCAAAACT GGATAAAACA GTGGTGGTGG 20880

ATGAACTGGT TGCCAAAGAA CTCTACCAAG ACTCGACTGC AGCAGTTAAT CGAACTTTTC 20940

CATATAAAGA GCAACTTTTT ACCATTGTTG GCGTGACAAC CAATACCAGC GGTGCCATTG 21000

GTCCAGGTAA TGATGACTCA TTGCTTTATT TTCCCAAAAA GACCTATGAA CATTATTTCG 21060

GCAAGCTAAA AGATACATCT ACGTTGAAAC TAACAGTAGC ACCTGGCTAT CAACCAGATC 21120

AAGTATTGAA AGAAACAATA AAAACTCTCT CTCAACAAGG AACCATGAAA AACAGTGGGA 21180

CGTATCAAGA ATATAATGTT AAAGATACCA TCAAAGAAAT GGGCTCTTTA TTAAATAATT 21240

```
TAACCTTGTT TATTTCCGCA ATTGCGGCTA TTTCACTGGT AATTGCTGGG GTCGGCGTTA    21300
TGAATATGAT GTACATTTCG GTCTCAGAAC GGATTAAGGA AATTGGTATC CGACGTGCAT    21360
TGGGCGGGAC AGCCAGTGAT ATTAAAAAGC AATTCTTAAC AGAAGGGATT GCCTTGACCT    21420
TAATCGGCGG AATTACTGGT TATCTTTTAG GAATGATTAT TGCCTACTTA GCCTCAATGG    21480
CACTACCGTT TTCTGTTCGA CCAGATTTAA TGACCGTTTC ATTAGCCATT GGGATTTCTG    21540
TCTTTATCGG GGTTGTATTT AGCTATTTCC CTGCCTCGGC AGCTTCAAAA AAAGACTTGA    21600
TCGATATTAT GAAATAAATA TAAACAGCAA AGAAACAGCC ATTTTTGGCT GTTTTTTTTA    21660
GTGAAAGTAG AAAAAAATCC AAAGATTTCA AGAGAAAGTA GCTAACAAAC GTCAAAAGGT    21720
TGATATTAGC TACCAACCTT TTGACGTTTG TCAGTTAAAT AAAAAAGAAA AAAATTAAAA    21780
AAATAGCACG TTGAGGATTA AATAACAAAC ATTTTTTAAA TAAGACTATT GGAAATCCTT    21840
TATAAATCGG GGATTACAGG AATGACTGTT AATGACTGAT TTGCTTGTAA ATTTCTAAAT    21900
TTTAATTCGA AAAATTTCCA GTATACTAAA TATATCAAAA CAAAACACGA AAATTGAGAA    21960
AATTTTGAAC CAAAAAAGAA TTAAGGAGGA ATCAAGAATG AAAATAATGG TCATCACAAA    22020
AAATATTTTA TCAGAAACTG CTTTTCAAAA CCGCTTACAG CATTTAGATC ACCAAGTTTT    22080
TTGTACGACG TGTCTGATTG GTGAAATTAA GGAAAGTCAT TTGTTTTATC GAAAATTTGA    22140
TGCCATCATT CTCAGCGAAA CGATTCCTCG CAAAGAACAA TATTTATATC TAAAGGAATT    22200
AGCCAGAGTT AAAACGCCGT TGTTTCTTCG CTCGGAAGAC ATTCCTACCT CGGAAGAATT    22260
GAGCATTAGC TATCAATATG GCATTCTCAA TTGGTTACGA TTGGATGCGA CTTTGGATAC    22320
GTTAAGAGAA AATCTCCTTT TGGCAGAATT AAATAAACCC ATCCAACGAG AAGATGAAAA    22380
GTTAGATGTA GAGGGAATAA AGCTAACAGC CGCAGAACGT GCACTACTGG CTGTCTTGTA    22440
TAGTACGCCA AAAGAACACT TACCAATGGA GCGCCTGTGT AAATTAATAT ATGACACAGA    22500
AGAAACAGCC AATAATCGAA AAAAAATGAC CAAACAGCTG ACAAAATTAA ATAAAAAAAT    22560
TCAAGTAGGG GGCTTTAAAA AACAAGGTAT CCAAAATACA CGGTATAAAG GCTACAAACT    22620
GCATCCGCAg TTGCTTCCTT ACATTGGCGA ATTATTTACT TAAATTAGAG AAGCACTTGA    22680
CGAATGAAAA TCACCATCAC AGGATGACCT CTAAAAACCA TGCTATACTA ATACGGAATA    22740
GTATAGGAGG TGGGTCTGTA TGAAAGACGG ACTATATTGG GTACAACAAT TAAAAGAACA    22800
GAAAGTAAGT CGCATAGAAC TCTTAAAGGA TATTGAGAAA AAAGTCCAAA AGTTAAATCC    22860
AAAGATCAAT GGCTTTGTAA CCTTTGATAG CCAAGAAGCA GAAGCAACTT ATCAAAAAAA    22920
TAAACAGCAG GACACACTAT TTGCTGGTTT GCCATTTCCA TTGAAAATGT TAGGCCAAGA    22980
AAAAGCGGGG TGGTCAGCAA CGGCCGGCTC GCAACTTTTT AAAAATAGTA CTGCAACAAC    23040
GACCTCTAAC TTTGTCAAAC AAGCAGAAAC GATTGGCTTG ATGCCTTTAG GACAAACCAA    23100
CGCTCCAGAG TTTGGCTTTA AAAATATTAC GGACCCTGCG CTTTACGGGC CAGCCCGCAA    23160
TCCTTGGAAC CTGGAACATT CGCCAGGCGG TTCCAGTGGC GGCGCTGCCG CAGTTGTCGC    23220
```

```
TAGCGGGATT GTTCCTATAG CCGGCGCTAG CGATGGCGGT GGCTCCATCC GAATTCCCGC    23280
CTCTTTTAGT GGCTTAATTG GCTTAAAACC TAGCCGTGGC TCCATGCCTG TCGGCCCAGA    23340
AGGCTGGCGA GGCTGGCAAG GTGCCTCGAT TGACTTTGCG TTGACTATTT CTATGAGAGA    23400
CACCAAAGCG TTATTTTATG GCCTACGTGG TAGCCACAGT GGCGCCCCCT ACCAAGCGCC    23460
CTTGGCAGAA TGGCAAACAC ACCCTAAAAA ACAACGGCTA AAAATTGCTC TGTGTAcCGc    23520
TTCgcCTATC GGCTCACTGA TTTCGCCAGA AGCAACACAA GCGGCGAAAC AAGCGGCTGA    23580
TTTTCTGGCA GCCGCTGGCC ATGAAATCAT TGAGATTCCT TATCCAGTCG ATGGTGCAGC    23640
GCTGATTCGC TCCTATTATC AAATGAACGG CGCAGAAACT ACAGCGATGA TGAACAGCAT    23700
TCAACAGGGG CTCGGGCGCC CAATTCGTAA AGAAGAAATC GAATCTTTTT CATGGACCAT    23760
GCATCAATTT GGCCAAAAGA TTCCAGCAGC CACCTATGTT CATTCATTAC AGTTGTGGGA    23820
TCAAGCAGCA GTAACGATGG AAGAACTTTT TCAAAAATTT GATCTGTTTT TATCACCAAC    23880
AACGGCTTTT TCAGCACCAA AAATTAACGA AGATTTGCAA AGTGAGCACA TTCGTCAACG    23940
AATGGCGCAA GCAGCAGAAT TAACAGAAGC AGAATTAGCT GAATTAATCT ATGATTATTT    24000
TGATAAAAGT TTACAGCTAA CACCGTATAC TCAACTAGCT AACTTGACAG GACAACCAGC    24060
CATCAGCTTG CCCACACATG TTACGGCAAC GGGGTTACCC TTAGGCATTC AATTGCTGGC    24120
TGCTCGTGGA CGTGAAGATC TCTTATTCCA AGTAGGCGAA CAGTTTGAGC AAGAAGGCAA    24180
ATTTAAATTA CCAGAAAGTT ACCGCTAAAT AGGGTGTTAT TGTGACCAAA TTTCAGATAA    24240
AATAGACACA AGACAAAAAA CTGAGGTGGA AAAATGCTGT ACGTAGTGAG ACATGGTGAA    24300
ACAGATTATA ATGTGGCCCG TAGAATTTGT GGTCATGCGG AAGCGCAATT GACCGAAAAA    24360
GGCTATCAAC AAGCGGAATT GGTAGCTGAA AAAATTGCCA AACAAGGAAT TCAAATCGAC    24420
CGTTTACTGG CGTCGCCTTT AAAAAGGGCG CAAGAAACAG CCCGTAAAAT TGCAGAACGA    24480
AATCAATTGA CGATTGAAAC AGAACCTCGT TTAATTGAGA TGAATTTTGG CATTTATGAT    24540
GGTGAACCGA TTGAAACAAC GGCTTTCCAA GAAAATCGTT CTGAGATTTC CTTGCCGTTT    24600
CCAGAAGGGG AATCAGTCCT AGATGTCGCG GGTCGAATTT ATCCCCTGAT TGAGGAAGCC    24660
TTGGCTTCAG AGGAAACTTA CCTATTTGTT TGCCACAATG CGGTCATGCG AGTCATTGAT    24720
AACTACTTCA ATGGTAAAAA AATCCAAGAT TTCTTGGACT TTCATTGTGA AAATACACAA    24780
CTTGTACAAT ACGGAGAATA AAAATCAAGT CTTTAATCTT TTGGTTAAAG ACTTGATTTT    24840
TTTTATGGTA AATAAGAAAC CGCTTCCGTT TTAATGTATA ATAAATTAAC TTTTGGATCA    24900
AAAAAGAAAT GGGGAAAAGA CTATGAGAAT TGATAATATT CATCAAGCGT TAAAACAGTT    24960
TCACGTACAA AATCCAAGTA GCAAAGAAAC CTTGTTGCCC GCAACGGcTT CGCAATTAAG    25020
TCAGGCCTCT GGCTATACCC GAACCGCTGT TAGTGAAGAG CTTTCAGAAT TAGTGAGAAA    25080
AGACCAAGTC GTTAAAGTTA AATCAAGACC GGTGTTATTT TTTGATAAAG ACTACTTAGA    25140
GGAACAGTTT CCTTTTTTTA AAAGTAGCAC GTGCTATGAT TCCCTCGcTT CGCTGCAAGA    25200
```

```
GGACATTTAT CAACAGACAG AACAGATACA TAAGAAAAAG CAGCTACTGA AAATGAATCC    25260
ATTTGCGAAC GTAATTGGTT TTGATGGGAG CTTAGCAGAC GCCATCAAGA AGTTGAAAGC    25320
CGCTATTTTA TATCCGCCTA ATGGCTTGAA TATTTTACTT ACTGGAGAAT CTGGGGTGGG    25380
GAAAACCTTA ATTGCAGAAC AATTACACCA ATTTTATCAA GTTAAAATGA ACCAAGAAGT    25440
TCCGTTTATT TATTTTAATT GTGCAGAATA TTTCAACAAT CCAGAATTAT TAACTTCGCA    25500
TCTGTTTGGC TATAAAAAAG GCAGTTTCAC TGGTGCCGTC AATGACCAAA AAGGTTTAGT    25560
GGAGTTGGCA GATGGCGGTT TTTTATTTTT AGATGAAGTC CATCGATTAA CGAGCGAAGG    25620
CCAAGAAAAA TTGTTTACAG TTTTGGATAA AGGGTATTTT ACACGGATGG GCGArCgAAA    25680
AACAAAGAAA TGTATCGATA AAATTTATTT TTGCAACAAC CGAAGCAGTC AATCAACATT    25740
TTTTGAAAAC GTTTTTAAGA AGGATCCCTG TGACATTAGC CATTCCAGCT TTAGCCGAGC    25800
GCTCTTTGAA AGAAAAAATT CAATTGATTT TGTCCTTTTT CTTAGAAGAA AGTCAACGGA    25860
TTCAAAAGAA TATCCATATT CGTTATGAGG CGTTGAAACA ATTAACATTT ACTCATTTTG    25920
AAGCCAACGT AGGCGAATTG AAAGCGGCGA TTCAATTTAT GAGCGCACAA GCCTATTTAG    25980
AGGTGTTCGA TCAATTAACG GAAACTATTT GCATTGAAGC AAACGAGGGG ACAGTCAAAG    26040
AGCAACCTGA TGTAGAAAGT CTGCTGGCTT TTAATGAGTG GATACCTAAC GATGGCCTCA    26100
CGATTCAATA CAACCATTTT TATTTAAGCA GTCAAGAGTT ACTTAAATCA GAAAGTCTGA    26160
AAAATGATCG TTTTTATAAT TTTCTATTGA AAGAATATGC GGATTTGAAA AAATTAACTC    26220
TTTCGACAGA GGATGCGCTA GCAATCATGC AAAACAAAGT CGAGCAACTT TATGATATGG    26280
ATATTTTTGC TCAAGAACCG CTAGCCATAG AGGCCATTGA CGAAACATTG CAGCCAAAAA    26340
TTCAAGAAGC CATTGTCTTT TTTGAACAAT TAATCCAAGA AAAATTGACA GAGGATTTCC    26400
AGAATATTTT AATTGTCCAC CTTTATTCTG CTGTTTTTTA TTCTGCACAA ATGAATGAAA    26460
CATTTTATTC AGATACCAAA AATATCTATT CTGGCAAAGT CACCCAGTAT CAAGCTGCTA    26520
AGACCATGGT GGACTTTCTA CAAACATTGT TTGGGATTTC TTTTGCTGAT ACGGAAATCC    26580
TCTTTTTCGA TCTATTTATT CGTAAAGTCA CAGCTAAAAA AAGAGCGCAC TCAGATGAAG    26640
AAAACTGTGG CGTTATCGTC GTTGCACATG GTGCCGCGAC GGCCAGTAGC ATGGTTGAGT    26700
ACACCAACCT TCTGTTTTCA ACAACGGTCA TGCGTGCAGT TGATATGCCG ATTAATCAAT    26760
CAGTGGAAGA AACATTAGAA AAAGTCCGTA CAATTGTAAA GCAAAATCAG TATAAAAAAA    26820
TTATTTTAAT GGTGGACATC GGTTCCCTGG TTTACTTTGG TAATGCTATC AGTCAAAAGT    26880
TTCAAAGGGA AGTGCTGCTA ATTAGCAATA TCAATCTATT AACATTGCTA GAAGTGGCCC    26940
GAGAAGTCAT TTATGAATCG ACGGCGTTTG AGTATCTATT ACCAGTCTTA AAAGAAAAAA    27000
ATCATCGCGC CTTGATTTGC AGTAAAGGTC GCTTCAACGA GGCAAAAGTT CTCATTATTT    27060
CCTGTCTAAC TGGCATGGGG ACTGCTATTA AAATTGAACA GTTGTTACTA AAAACTTTTC    27120
CCAATGAATT ACTAACGAAT GTACGCATCA TTACTTTAGA GAAAAGAGAA GTGGAAGACA    27180
```

TTGATCGCTT GCATTCTTTT GTTCGGAAAG AAGAGAAAAT CGTTGGTATC ATTGGCACCG 27240

TAAAGACAGA AGTGCCAGAT ATTCCGTTTA TTTCTTTACA AGATTTACTA TCGAAGCGGG 27300

GCATTCAGCT AGTTTTCGAA TTATTAGGCC ATCATTTAGA AGAAAAACAA AACCAAGAGT 27360

TACTAAAAAA AGTTTCAGCT AAATATATTC AAGGATTATC AATCGAAGCT ATTACAGACT 27420

TACTAACGGT TTTGAATCCG ACAACCATTT CAATCGAATT AGCTAAGATT TATTCAGACA 27480

TTTGCTTGCA AACTAAAATA AAAAGTGACG AAAAGATTTT ACTTCGTTTT ATCATTCATT 27540

GTGCTTGTAT GTTAGAACGG CAATTGTTGA ATCCAGAATA TGACCCGACA GCTTATGaAG 27600

TGTCaTATCG AGAGTTGCCA GAAGCTGTAT CTGTCATAAA AATGGCGTTT CGACCGCTTG 27660

AAGTGTCCTA TAATCTGTTG ATTCCACCTT TGGAAGTGCG CTATATTTAT GAATTGCTCT 27720

TTGAAAAGGC TTAATTCAAA CGGAAATGAA AGTTGGCACA CTTATTGCTT TTAATAAAAT 27780

GATTAATTAG CAAAGGAGTT TGTTTATGGA AAATGAACCA ATTTTACTTT TAACACACAA 27840

TGGTTGGGGC TTGGAATTAG TCAAAAGCGT CCAGATGATT GTTGGAGAAG TCACAAATGT 27900

ACATGAAGTT GCATTACAAG CAGAGGATTC ACTGGGGGAT TATTTAGAAC GAGTCACAGA 27960

AAAAATTGAC CAATTAACGT GGCACAACCA GTTGCTAATT TTAACAGATG TTAAAGGCGG 28020

CACACCTAGT AACGTTGCAT TACGTTTATC AAAAGATTAT GACCTCCTGA TTGTGTCAGG 28080

ATTATGCACA TCGTTATTAT TAGAATCCGT CATGAAACAA AGTGGTCCTG GATTTCGTTT 28140

AGAAGATGCA GAGATGATTC AACAAGCAGC AGTAGACAGT TGTCAAATTT TAGAAATTCC 28200

AAAATAAGAA AGTACAGGTG AACAGAATGT CAAAAATCGT ATTAACAAGA GTCGATAGCC 28260

GTTTAATCCA TGGACAAGTA GTGACTAAAT GGTTACAACA ATCAGGTGCA AATGAAATTT 28320

TGGTAGTGAG TGATGAACTA GAGCAAGATG AATTTTTACA AAGTATTTAT TTAATGGCGG 28380

CGCCTCCAGG GGTTTCTGTA GTAATTAAAG GGATTGAAAG TGCCAAAGAA TATTGGGAAT 28440

CGCAAGAAAA GGAAGAGAAA AAGGTGTTAT TTTTAGTGCC AGATTTAACG ACCTTAAAAG 28500

CGATGGTGGA CCAAGAAATT ATTAAAGAGG AGATTCAAGT TGGTGGCTTA GGTGGCGGAC 28560

CAAATCGGAA AAATGTTTTA AAAAATATCA ATTTATCTGA AGAAGATGTT GCAATTTTGA 28620

CTGAACTTTT AACGAAGGGC ATCCATGTTT TCTTCCAAGC GATTCCCGAA GAAAATCCCC 28680

TACCAATTCA AAAATTAATT GAAAAATATC AATCACTATA ATCAGAGAAA AAGGAGAGAA 28740

TCATATGGGC GTTTTAGGTG TTTCAATTAT TATGGGACTT TATTATTGGT TTGCACGTTT 28800

GCGTTTTGGG TACACATTTT CAGGCATGCT GTCTCAGCCA TTAGGCGCAg cGTTGGTGGC 28860

TGGTTTAGCA ACAGGAAAAC TTAGTGAAGC AATGGTTGTG GGGGCGGGGA TTCAACTTGT 28920

TTATCTCGGA GTCACCTCCA CACCGGGCGG CAATGTGCCT AGTGATCCGG CTTTAGCTTC 28980

AGCAATCGCG ATTCCGATTG CTTTAGGTGT GGGAATGAAT GCCGAGGCGG CTGTTGCACT 29040

AGCTGTTCCT TTTGGTGTGT TAGGCGTCTT TATGGATCAA TTACGGCGCA CCATCAATGC 29100

GACGTGGGTG CATATGGCTG ATAAATATGC GGAAGAGTTA AATTATGGGG GGATTTACCG 29160

AGCTGCGTTT GTTTATCCAG CTTTAGCGGG ATTTTTAATT CGCTTTCCGT TGGTTTTTGC 29220

GGCGAACTTT TTTGGTCAAT CTGTGGTGAA TAAAGTCTTA GAAGTTATCC CAGCATGGTT 29280

GATGCACGGC TTTGAAGTAA TGGGTGGTAT TTTACCAGCG TTAGGCTTCG CCTTAACAAT 29340

TATGGTAATT GGCAAGAAAA ATTTGATTCC TTATTTTTTA ATTGGTTTTG TTGCAGTAGT 29400

ATATTTTGGC GCAGAAGTTA TGGCCGTCGC TATTTTAGGA ATCTGTATTG CCTTTTTAGT 29460

TAGAAACAAG GCATTGAGTG AGGAGGCAGC ATAATGGAAC AAATGACACA AGCAAATGAA 29520

CAACAATTGG AACAGAAAAA AATTACTAAA AAAGAATTAT CCAAAGCCTT TTGGATTTAT 29580

CAACTAGGCT GTGAATTATC AAATTCTTAT GAACGTCTCC AAAGTTTAGT TTTTTGTGCA 29640

TCGATGATTC CTGCAATTAA AAAACTTTAC GCAGACGATG AAGAACAACA AAGAGAAGCT 29700

CTGAAACGTC ATTTGAATTT TTTTAATACT GAAGGAACAG TAGGGGCTTC CATTCAAGGG 29760

ATTGCAATTG CTATGGAGGA AGAAAAATCC AATGGCGCAg cGATCTCTGA TACGGCGATC 29820

ACATCGATTA AAACTGGGCT GATGGGTCCA CTGGCTGGGA TTGGCGATTC GATTATTTGG 29880

GCAGCATTAA TGCCGTTAAT CATTTCCATT TTTATTCCAA TGGCTAAAGG AGGTAACGTG 29940

ATTGGAAGTA TCGGTCCGTT GGTTCTGTAC ACAGCCATTA CATTATACAT CAGTTGGACG 30000

TTAGTGAATA AGTCCTATAC ATTAGGGCGT AATTCAATCT TAAGTTTGTT AAAAGACGGC 30060

AAAATTAAGC AAGTCATTTA TTCAGCAAAT GTTTTAGGAA TGATGATGAT GGGCGCTTTA 30120

AGTGCGAGTT ATGTTAAAAT TGCTAGCCCG ATGACGTTTA AAGTTACTGG CGGTGCCACG 30180

ATTGTTTTAC AAGATATCCT AGATCAAATT ATGAAAGGGT TACTTCCTTT AGCAGCAGTC 30240

ATGGCTATTT ATTTCTTCAT GGTAAAAAAA GGCCCACGTT ACGGTATCAT TATCGGCACG 30300

ATTGTCTTGG TTAGTTTGGT CACTTCATTC TTTGGTTTGT TATAAAGGAG AAAATAATGA 30360

ATAGTTACGA AAAATTCCAC TTAAAAGAAG TCATTAATGC TTCTGGCAAA ATGACCATTC 30420

TAGGTGTCTC CAAAGTTTCG GAGGCAGTCC TAGCAGCTCA ACGATTCGGC GGCGAACATT 30480

TCTTTGAAAT GAGTGAACTT AGTGTGCAAA CAGGGGCCTT TTTAGCCAAC TTGTTGAAGG 30540

TAGAAGATGC CCAAATCGTT TCCTCGGCTT CGGCGGGGAT TGCCCAATCG GTGGCTGCCT 30600

TGATTGGAAA AGGCAGCTTG TACCACGCGT ATCATCCCTA TACAGAAAAA ATCGAACAGC 30660

GGGAGATTGT TTTACCTAAA GGGCACAACG TGGACTACGG CACGCCTGTA GAAGTCATGG 30720

TAGCGCAAGG TGGCGGCCAA GTGGTGGAAG CTGgcTACGC CAACATGTGC AGTCCAGAAC 30780

ACGTGGAAAT GATGATTTCT GAAAAAACAG CGGCCATTCT ATACATCAAA AGCCACCATA 30840

CCGTGCAAAA AAGTATGCTG ACGGTGGCCG AAGCGGCCAA AGTGGCGCAA CGCCACAAGG 30900

TGCCGCTAAT TGTTGACGCG GCAGCAGAAG AAGATCTTTT CAAATACACT GAAGCAGGCG 30960

CTGATTTGGT GATTTACAGC GGCGCCAAAG CAATTGAAGG GCCAAGTGCC GGTTTAGTGG 31020

TCGGGAAAAA AGAATACATT GACTGGGTGC GCCTACAAGG CAAAGGCATT GGTCGAGCAA 31080

TGAAAATTGG CAAAGACAAT ATCTTGGGCT TTACGCAAGC AGTGGAAGAA TACTTGGCAC 31140

ATGGCAGTGA ATCAGGGGCT TCGATGCAAG AGCGCTTAAA ACCGTTTGTT GAAGCCATCA 31200

ACAACTTATC AGATTTAACC GCCAAAATCG TCCAAGATGG CGCTGGCCGA GACATTTATC 31260

GCGCCAGTGT CAAAGTGGAC GGCAGAAAAA CCGCTAAGGA AGTCATCCAA GCATTAAAAG 31320

CAGAAAGTCC AGCCATTTAT ACGCGAGAAT ATCAAGCCAA CAATGGCATC ATTGAATTTG 31380

ACATTCGTTC TGTCAATCAA GAAGAAATGA ACAAAATCGT GCAACGATTA CAAGAAATTA 31440

TGGACAAAAA GGAGAAATAA GTCATGTCAT TAAAACCTAA TTATTTAGAA GAACGCATTT 31500

GTTTAAACGT CTTAGCCAAT TCAGTAGAAA ATGCCCAAGC CTGTTATGAA GCAGCAGAAG 31560

GGCATGTTGT GTTAGGCGTC CTTTCAAAAA ACTATGAAAC CGATGAAGCA GCTATTGACG 31620

ACATGAAAAA ATACCAAGCA GCAACCAACA ACGCTTTATC TGTCGGCTTA GGGGCTGGCG 31680

ATCCTAATCA AAGTCAAATG GTGGCTCGCT TATCAGAAGT TCTACAACCG CAACACGTGA 31740

ACCAAGTCTT CACAGGTGTC GGcgCtTCGC GGGCCTTACT TCGTCAAGAC GAAACAGTGA 31800

TTAATGGCTT GGTTTCACCC ACAGGCAAAG TTGGCTATGT CAACATTGCG ACAGGTCCGT 31860

TAAGCTCTGG TGCGCCAGCG GCGGAAGTGC CTATCGAAAC CGCCATCAAA TTGTTGAAAG 31920

ACATGGGGGG CAGTTCCATT AAATATTTCC CGATGAAAGG CTTGGCTCAC AAAGAAGAAT 31980

ACCAAGCAGT AGCAGCAGCT TGTGCCAAGT ATGACTTTTA TTTAGAGCCA ACAGGAGGCA 32040

TTGATTTAGA AAACTTTGAA GAAATTGTTC AGATTGCGGT GGATGCCGGC GTTAAAAAAA 32100

TCATTCCTCA TGTCTACAGC TCAATTATCG ATCAAGAAAC AGGGGACACT AGAACAGAAG 32160

ACGTGAAAAC GTTACTAACC ATGATGAAAA ACACATTGAA TAAGTAGGTG GCAAAGATGA 32220

AAATTGCTGC GTTTGGTGAA GTCATGCTtC GTTTCACTCC GCCAGAGTAT CTAATGTTAG 32280

AACAAACAGA GCAGTTAAGA ATGAACTTTG TAGGGACTGG CGTAAATCTC TTAGCAAATT 32340

TGGCTCATTT TCAACTAGAA ACAGCATTGA TTACGAAATT ACCTGCCAAT CGTCTTGGTG 32400

AAGCTGGCAA AGCAGCGCTA AGGAAGTTAG GAATTTCAGA TCAATGGGTA GGAGAAAAGG 32460

GCGATCATAT TGGTAGTTTT TTTGCAGAAA TGGGTTATGG AATCCGCCCC ACACAAGTAA 32520

CTTATCAAAA TCGCCATCAG AGTGCTTTTG GCATCAGCGA GGCGAAGGAC TACGATTTTG 32580

AGGCCTTTTT AGCGGAAGTA GATATGGTGC ATATTTGTGG TATTTCGTTA AGTTTAACTG 32640

AAAAAACTCG TGATGCTGCT TTGATACTTG CTCAAAAGGC GCATGCCTAT CAAAAGAAAG 32700

TTTGTTTTGA CTTTAATTAT CGACCTAGTT TGAATACAGC AAATAGCGCA CTCTTTATGC 32760

GTCAACAG 32768

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6027 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

TTTATTTACA CCTGCATTGG AAGTCGAAGA TGCCGAAAAG AGTAGTTGGA CTTATCCCGT 60

TTATGGTTAT TTAGATAGCG AAAATCCTTG GGAAAAAATT GCGGCTTTAT TAAATAAACG 120

AACACAAGGC ACACCACGCT TTGCTTTAGA AAAACAAGCT CTTTCTTTAG CTCGTTTCGA 180

TCAATTGAAA ACCTATTTCC CAGCTAGCGA TTTTTCGCAT GATGTTACGC CTTTAATTGA 240

AAAACTTCAA TTAATTAAAA CAGAACCTGA AATTCAACGT TTACTAGAAG CTGGTAGTTG 300

GGCTGATGTC GCTTTTGAGA TTGGCTTCAA AGCAATTCAA GCAGGCGTTG CGGAACAAGA 360

AATTGTTGCA GAAATCGAGT ATCAATTAAA ACGTCAAGGT ATTCGCTCAA TGAGTTTTGA 420

TACGTTAGTC TTAACAGGAA AAAACGGAGC TAGTCCTCAC GGGGTACCTG GCGAAACAAA 480

AATTGAGCCC CATGACTTAG TGTTATTTGA TTTAGGTGTC GTTCATAATG GCTATTGTAG 540

TGATGCCACA CGAACAGTCA GCTACTTGGA ACCTTCTGAT TTCCAAAAAG AAATTTATGG 600

AATTGTCTTA GAAgCACAAT TAGCAGCAAC TGAAGCGGTG AAACCTGGTG TTACAGCAGG 660

cGAATTAGAT GATATTGCCC GAGGTGTAAT TACmAAAGCT GGTTATGGTG AATACTTTAA 720

CCATCGCTTA GGCCATGGCA TTGGAACCAC TGTCCATGAA TACCCATCGC TAGTTCATGG 780

CAATGATTTA GTCATTGAAG AAGGTATGTG TTTCTCTATT GAGCCCGGTA TTTACATTCC 840

TGGAAAAGTA GGCGTTCGTA TTGAAGATTG TTTACACGTA ACGAAAACAG GCTCCGAACC 900

ATTCACTAAG ACAACCAAAG AATTGCAAAT TATTCAATAG CATAAAAATA AAATGAGGCC 960

ATCAGTCAAC CGGACTGATG GCCTCATTTC TAAATTTTAA GGTAAGCCGC TGAACAAAAG 1020

CAACACGATC CGTTGCTAAC TTTCATAAGC TAAACGCTTG TTACCCTTTA AATTCTTTTT 1080

TTACGTCTTT CGCTGCATCT TTAACATCTT CTTTTGCTTC TGCAGCTTTT TCAGGAACAT 1140

CTTTGGTGGT TTCTkTTGCT TCATCGACAC CAGCAGAAAC AGTTTCTTTG GCCGCTGCCG 1200

CAGAATCTTT AACGTCAATA TAGATATCTT CTGCTGCATC TTCGGCAATT TCACCTAAAT 1260

CTTCTGCGTG ATCTTGAGCA TCGTCAGCTG CTTTTTTAAA TTGATCAGAT AAATCACCTG 1320

TTTGTTTTTT AAAGTTATCA AGCATGTCGC CAGAAACACC TTGTGCTTTA TCCAATGAAT 1380

CTTTTGTTTT TTCTTTGACA GAACCTGCCA AATCAGAGGC TTGATCTGAT AAAACGCCGG 1440

CTTTTTGTTT TGCGATTTCT GTTAATTCTG TTCCTTTTTG AACAGCATAA TCTGTGTAAT 1500

CTTGCGCTTT GTTTTTTAAA TCATCTGTTT GATTTGATAA ATCATCACGT AATTCTTTAC 1560

CTGATTTTGG TGCAAGTAAT AATGCGGCAA CGGCTGCTGC TGTTCCACCA ATTACTGCyC 1620

CTAAGAAAAA TCCGCCTTTT TTAGCCATAA AAATCTCCyC CAATTAGTTA TTkGTTTTTT 1680

TATTTTTTGA ACGCATTGCT CTAAAAGCAG AACCGCCTAC TTTACTCACT nAAACCAGcT 1740

TTCGCTGtGC CTTTyCCTAC GGTTCCTACC TTTGTAATCA AATGTTTACT TGAGGTGTTC 1800

AATTCAGAGA CACTTTCGCT TAAATCAGCA ACGGCCGTAA ACAACGGATC AATCGTTGCC 1860

ACTTTTTGAT TCACATCTGT CAATAGTTCA TTACTTTTAA CAAGCAAGCC TTCCACTTGG 1920

CGACTCAATA CATTCACATC GCTAGTCACG ACTTTGATTG TTGTATTTGT TTCATCAATT 1980

GTATCTTCAA TTTTTGAAAC AACTTGAGAG ACTTTATATA ACACAAGTAC TAAAAAGATG 2040

ACTAAAACAA CAAATGcTAC CGCTGCAATT AAGGCAGCAA TTTCCCCACC TGACATATCA 2100

TTACTCCCCT CTGAAATTTT ATTTATAATC tTAGGATAAC ACCAACTTAT ATTTTTCACA 2160

ACTCATAAAG CTAAGAATGA CAATTTATTT TATAAAATTC TCATGCCCTT TTCTTTAAGA 2220

AAAAAAGCGT AAAATTTCCT TCTAACGAAT TGAAATTTTA CGCtTATTTT TATTTATTTT 2280

TCTTTTGGTT GCTCTTTTTT CGCTAATTGA GCACCTAAAT TAATAATGTA TTCTCGTAAA 2340

TTATCTTTTA CTTCTGGATG AGTCAAACCA TATTCAATGC TTGTTTTCAT AAACCCAAAT 2400

TTATCGCCAA CGTCGTAACG TTCTCCTTTA AACTCACGAG CAAAAACGCG TTGTGTTTTA 2460

TTTAAAGTGT CAATCGCATC TGTTAACTGA ATTTCGTTAC CAGCGCCTGG TTTTTGATTT 2520

TCAAGTACAT CAAAAATTTC AGGGGTTAAT AAATAACGGC CGATAATCGC TAAATCACTT 2580

GGTGCTTCCT CGGGTTTTGG TTTTTCAACA AAATTCTCGA CATTGTATAG TCCTTTTTCG 2640

ATTTCTTTTT CTGGGTTGAT AATACCATAT TTTGAAGTAT CTTCATGGGG TACTTTCATT 2700

ACCGCAATTG TCGAAGCATG TGTTTGTTCG TAATCATCCA TTAATTGTTT CGTTAATGGC 2760

ACTTTATCTT CCATTAAATC GTCGCCTAAC ATCACGACAA AAGGTTCATT GCCAACAAAT 2820

GCTTTCGCTT GTAAAACAGC GTGTCCTAAG CCTTTCGGAT GGGATTGACG AATAAAATGT 2880

AAATTAACAT CCGTTGTTTC TTCGACTAAT TTCAATAAAT CGGTTTTATT TTTTTCTTTT 2940

AAGTTATTTT CTAATTCAAC ATTTGAATCA AAATGATCTT CAATCGGACG TTTGGCTTTG 3000

CCTGTCACAA TTAAAATATC TTCGATTCCT GAAGCTAGTG CTTCTTCTAC AATAAATTGA 3060

ATCGTTGGTT TATCAACAAT CGGTAACATT TCTTTTGCCA TCGCTTTGGT AGCCGGCAAG 3120

AAACGTGTTC CTAACCCAGC TGCTGGAATA ACTGCCTTTT TAACTTTCAT CTTTTTTCCT 3180

CCTAAAAATC AAGTGTAAAT TCGTTTTCGG TTTTGCCATC GCGCAACATG ATTTCTTTTG 3240

CTGCTTGTTG TACATCTTCG TCATTGTACA AGACATTGTA GATTGTTTCT GTGATTGGCA 3300

TTTCGACTTC TAGTTGTTGT GCCAATTCAT AGGCTGCTTT AGTTGTTGAA ACACCTTCCA 3360

CAATCATGCC CATGTTTTCT AAGACTTCAT CTAACTTATG GCCTTTTCCT AATAAATTAC 3420

CCGCACGCCA ATTTCGAGAA TGGACACTTG TACAAGTAAC AATTAAATCC CCAACACCAC 3480

TTAAACCAAT AAATGTTAAT GGGTTTGCTC CCATCGCAAC CCCTAAACGA CTAATTTCTG 3540

CTAAACCACG AGTCATAATT GCCGCTTTGG CATTATCACC AAAGCCTAAG CCATGAATCG 3600

CTCCGGCGCC CAAAGCAATA ATATTTTTTA AGGCCGCCCC TGTTTCTACA CCAATCACAT 3660

CATCATTCGT ATAAATTCTA AAGTAATCAT TCATAAACAA CTCTTGCACA TACACAGCGT 3720

CTGCTAAATT TTCGCTTGCT GCTGTAATGG TTGTAATGTC ATGAACAGCG ACTTCTTCTG 3780

CATGGCTTGG ACCAGATAAA ACCACAATCC CTTGCCGTTT TTCTGCTGGA ATCTCTTCCG 3840

CTATCACTTC TGAAATCCGC TTATGTGTAC CTTGTTCGAG CCCTTTACTT GCATGAATAA 3900

TAATTGGTTT TGTTTTTAAA TGTTGGGCCA CTTCTTGTGC TACTGAACGA ATGGCTTTAG 3960

TTGGTACCAC AAACAGTACC GCATCAACAT CTACTAAAGC AGTGGCTAAA TCAGTGGTTG 4020

CCTGAATTGA AGTAGGTAAG ATTAATTCTG GTAAATAATG TTTGTTGGTG TGTTTGTATT 4080

TATTTCATCG ATTTGCTCTG GTTTATTCCC CAAATACAGA CCTCGTGTCC ATTTcGGCTA 4140

ATACTgAGCT AAAGCTGTcC CCCAGGAACC TGGACCTAAA cAGCGACTTT TgTTTCATCG 4200

GTGAAACTCT CCTTTCGCTA AACAACGTTA CTTCTTTAGT TTACCATATT CTTCAATCAA 4260

GTTGGATTAA GAAGTTGTTA TTTTTTTCTT TTGTTTCTCC CGATTGTAAA ATTTCAAATT 4320

GGGATTTCTT CTGCGAACGA TCACAATAAC AATTGCTGCC ACGAACATCA CTAAAGACAA 4380

CAATTGTGAG ACACGAATGT TACTAAACGC ATACAAACTG TCCATCCGTA AGCCTTCGAT 4440

GAAGAAACGG CCGAAAGAAT ACCATATTAT ATAACCTAAA AAGACTTCGC CTTCTTTTAA 4500

AAAGTGTGGC TTTTTGCGTA ATAACACTAA AACGATAAAG CCTAAAACAT TCCAGACAGA 4560

TTCATATAAA AATGTTGGTT GATGATATGT TCCGTTGATA TTCATGTTAT CAATGATAAA 4620

CGTTGGTAAA TGGAGATTTT CTAAAAATTG TCTGGTGGTC GCTGGCCCAT AGGCTTCATG 4680

ATTCATAAAG TTGCCCCAGC GACCAATTGC TTGCGCTAAA ATCACACTTG GTGCGGCAAT 4740

ATCTAAAAAA GTCCATGTAG AAATGAAACG ATGACGCGTA AAGAAAAACA GCGCCaGTCC 4800

CCCACCAATA AGACCACCAT AAATCGCTAA GCCACCATTT CGAGTAAAGA AAATTTCAAT 4860

CGGATTGTCT ACATAATCTT GCCATTGAAA AGCGACGTAA TACAAGCGCG CGCCAACAAT 4920

TGCTGCAGGT AATCCCCAGA GCATAAAGTC AAAGACATCA TCTTCTTTTA AACCAACACG 4980

AACGGCTTCA CGACTACTTA GCCAAACAGC TAAAGCAATC CCTGAAACAA TAATAATTGC 5040

ATACCAATAG ACGGGAATAC CGAAAAGGCG AAATGCAATT GAATTTACTT GGGCTAACAT 5100

CATCACATCT TCTCCTTCTA TTCTCCTGAA TTTTCTTCAA TCAAACGGGT TAAGCGCTCT 5160

TCGAACGTTT TTGTTGCATC ATAGCCCATT GTTTTCGCGC GGAAGTTCAT TGCTGCTACT 5220

TCAATGATAA TGGCCACATT TCGGCCGGTC TTAACAGGAA TACGAATCTG AGGGACATCA 5280

ACGTTAGCTA TTTCAACCAT TGCATCATCA CTACCAAGAC GATCATATTT CTTGTCTTTT 5340

TCCCATGCCT CTAAATAAAC GACCAATTGA ACTTGCATAA AGCCACGGAC CGCACTAGCT 5400

CCAAACAGGT TCATGACGTC AATAATTCCG ATACCACGAA TTTCAATTAA ATGCTGCAAA 5460

ATTTTTGGTG GTTCTCCAAC CACGGTTAAT TCATCTTGCT GATAAACATC GACGCGATCG 5520

TCTGCGATTA GCCGATGTCC ACGTTTAATA AGCTCTAAAG CTGTTTCACT TTTACCAATA 5580

CCGCTATCTC CTTGAATCAA AACACCAAGT CCATAAACAT CAACTAAAAC TCCGTGGACA 5640

CTTGTCCGAA CAGCTAAACG GCCATCTAAA TAACTGGATA GTTCCCCTAG TAAACGGGAA 5700

GTTGAAATCG GTGAACGTAA TACAGAAACG CCATTTTCTT TTGCTGCTGT AATTAATTCT 5760

TCGGGGATTT CTAGTCCTCT GGATACTATA AAGGCTGGAG TATCTTTCGC ACATAAACGG 5820

CGCATCACTA ATAAACGCTC TTCAGGCATC ATTCGTTCCG CAAAGGTAAT TTCTTtGCTC 5880

CCAAATAATT GTAAACGGTC ATGGGAATAA TAATtArAAT ACCCaGTTAA TTCTAAACCT 5940

GGCCGGGAAA TTTCACCAGT TTTAATTGTT CTATTTAAGC TTtCTtCATC TCCGTATACT 6000

ACTTCTAAAG AnAGATTTTC TACCAGT 6027

(2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4915 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

CAACAGTCCC AGATGATAAG GTTCTGCTAC CATTACGAAA CTCTACATCA GTATATAGTT 60

TATCTGACAC TCGCATAATA ATTTGGTGCG GATTTTCAGT AAAGTAATTA TCTGAATTTG 120

ATACCAATTT TGTAACATAT GATTTGTTAG CAGTTATATA ACCTTTGGCA GkCyTTAATC 180

yAGGTAATCC GTTTTCGGTA TAATCAACAC TaAAACTTgA TTAATGTATC TTTAGAAACA 240

AACTCACtTT TTTACTGAAA TTTACATCTT GATAATAATA ATCATTTGTT TTCATAACTA 300

CATACTCTGG ATAATCATAT AGATAATCTT TGATGTTTGG CACAACTTTT TGTGCATAAC 360

TTTTATTAGC CGTTAAATAA CCATTTTGGG TAATCAAACG AGGAACTTGA TTTTCACGGT 420

ACTCGATTCC CACAACAGGA ACTAAACTTC CCTTGGAAAC AGCTGCTCCT TTATATAAGA 480

AGTCTGAATC CCCATAAAAA TTATCACTAT TTTTCATAAT AATTTGCCCT GGATTAACCG 540

TATGATAGTT AGCAATCCAA GATCCAGTAG CTACTACATA GTTCTTGTTA GCCGTGAAAT 600

AATACCCATT TGCCATCTTC AATCTAGGAA TTCCTGATTT TGTTTTCTCA ATAGCAGTTA 660

CTTTTAAAAC ACTGCCTGCA GATACTTTAT TTGCTTTTGT CTTGAATTCT ACATCTTTGT 720

AATAGGTATC TTCATTTCTA GTAATTACTT GTCCAGGATT AGATGTGTAG TAAGTATCTT 780

CTAATCCTTT ATATTGTTCA ACTATATTTT TATTTGCAGT CAAATATCCT TTCGCTGTTT 840

TCAATCGAAA TATACCGCTT TGAGTATAAG CTAAATCTTC AACTTCTACT ATTGTTCCCT 900

TCGAAACCTT AGACAACCTT TGCTTAAATT CTGGATCAGC ATAGAAATAA TCATCGCTTT 960

TCAAACGAAC TTTTTTAGGA TTTGTCGTAA AATACAAATC AATATTGCTT TGTGCCGCTA 1020

ATACATAATC TTTATTTGCT GTTAAATAAC CTTGGTCTGT GACCAATCGA GGAATTCCTG 1080

CTGAACTATA TTCAATCCCC TTAATGGTAA CTAGTGTATT TTTTTTAACT CGCCAACCAG 1140

GAGTTCTAAA CGCTACATCT TGGTAAAAAG TATCGTCATT TTTCATGATC ACTCGACCTG 1200

GaTTAGTTGT GTAATACTTA CTTAAGTcAG GCCCACTACT ATTACCCATA TTTAaaAAAT 1260

TACTATAGGT CwGGAAtATC AAACGGACGC CCTTcATAAT tAGCTAAACC AGGgAAATAC 1320

ATTTGCGAGA CCATTGCCAA GCAccTGgTC ATTATyCCmC TGCATGGAAC TATCTGGaGT 1380

ATATGGATAT tGCGCTACCC aTACTCTGTC TCTTCCgAAG GCGGAGGTGT CGATATAGCC 1440

ATTkGTTAAc CACCATTTGC CAACATACAA CGCATCATTT TTATACCCTA ATGCCTTTAA 1500

```
TTGCTGGTTA AACGCAACAG AATTAGcATG cGTTTCGCCC ATTGTTTGTA AGCGTTGGAT    1560
CTTCCGCATC ATTAAACATT ATGGTATTTT TATCTAAGCC AGAGTTAGCC GCCGCCTGAG    1620
CAAAATATCT TGCTTCATCT TGTGtGTCGC AGCAGAGACA TACATACTAT AGTGATATGC    1680
TGAAACTTTT AAGCCCGCAG CTTTGGCATT TCTAATTTGG CCCGCAGCAT ATGGATTAAC    1740
ATACCAAGTT CCTTCTGTCA ATTTAACCGA AACACCTGTT ACACCATAGC TTTTAATTTT    1800
TTGATATTCA GCTACTGAAA TGTCTCCATT CCAACTAGAA ACGTCAATAA AATTCTTGCC    1860
AGGCAAATTA GGATCATTCG CATAGACGTT CGCAACTGTT GCTCGCGCTA CCCTAGAAGA    1920
AAGGCCGTTT TTCATAGCAT ATGTTCCTTT TTTATGGTTC GTTCCTTCAC GTGCTGCTTC    1980
TTGATAATCC GCTTCATCTT GGTCAATGGT CTTTTTTTCT ACTTCAGAGC TTTGTGATGC    2040
TTCTGTTTTA CTATCTGTTG TGTTTGTTTC TGAAACAGAA GAATGTTCTT CCGTTGTACT    2100
ATTTGTTGAA GAATCAAGAG ACTCTTCTTT TGTGCTTGAA GTAGAACTAT CTCTTTCAGA    2160
GGTATTTTCT TCTGTAACAG AAGATGTATC CTCCGTTGAA TTAGTCGTTG TTTCTGACGT    2220
AGCTTGAGTG ACCAAGCCAG TcTCCCCCGT TGTTTCTGTT TTTACAAGCT CTTCTCCTGA    2280
TGCAATCACA CTAAAACTAT TTAATAATAC AGTACAAATT AACATACCTG AAATGATTTT    2340
TTCATGAAAC TCTCCTTATA TATTTAATAA TATGTATCTT TCTTAAACAC TTGCCATTCG    2400
CTCTCGAAAA ATGTGTATCC TATAACCTTC AACTAATTGT TTTCACATTT TAAAACTAAC    2460
ACCAATTATT ATCTCATACC TTTATTTCAA ATTCAATTAC GCTTTTTTAA CATTCCTATA    2520
TCAATTTTTT ACTTTTAAAT GGTAAAAAGA TGAAAACAGA TCTTTAATCG ATTGTTTTCA    2580
TCTTTTATTT ATTTTTTAAG ATATTCTGTG TACGATTTTT TAGTATCAAC AATTAGCGCT    2640
TCATGATGGG GTTGGCGATC TTCCTCTGGT GGCAATTGCA TATAATCCCC AAAAACCTCT    2700
GTTAAGTAGC GCTTGTAGCC CACTGGTATT GGCATTTTTG TATCTTCAAA TGGAACAAAA    2760
ATAGCAGATT TAAAATCTTC CAAATGATAA ATATTACCCA TGTAGCGTGG TCCGACACAT    2820
AATTCTGTAA CATAAGGAGT GGAGCCGAAG GAATACTTAG TCATTCTTTT TTCAGCAAAG    2880
CGCCAAATAT GATAGCGAAC CTTTTTCGAT GGAAAGATAT TCAATAATAC CCGACTCCCC    2940
GTTGCTAAAA TGCCACCATG TTTTTCTGGT ACCACTTGTG AACAGAATAA AGCATAAATT    3000
AATGCCCAAA TTTTTTGTTT CTTCCTTTGA AAATTTCCTT TTGGCGCACC ATCTAAAGGA    3060
AAAATATCGA TAGGAATACC ATGAGGAATC GCTAATCCAT CTTGATAGGT TTTAATAAAC    3120
GTTGTCTGTG CATCCCTAAT TGTCATAAAT GAATTATGAT CATTATAAGT TTTTGAAGCA    3180
ACTAACAATG GATAACATTC TGTATCTGCA TATTTAGGCC AAAGCACTTT CaATTTTTCa    3240
TAATCTGGTC TTGGCATAAA GAAATCTAAG TCATCATCCC aTGGAATAAA ACCTTGACTG    3300
CGGATTGCTC CGATACAACC GCCTCCACAA AAATAACAGG TTAGATTATG TTCTTCACAA    3360
AAAGCCACAA AATATTTTGA CATGCGTAAA ACGACTTGTT GAATTTCTAC TAATTGCGTC    3420
TGTGTTGTTT GTTCCACTTT TCCTTTTCCT CCCTATTCAT TCTTTGCATA TTTAAATGTT    3480
```

GTATAGAATA AAAAACGCAT TAAGTAGAAT GAACTTTTAA TAACTGAAAA CTGTTCATTT 3540

TGACGCAAAA TTTTCCAAAC CCAACTGATT GTATTCCATT TGTTAGAAGA TGTAGAGGTG 3600

GTTAGCTTAC GATAAATGGC TAATTTTTCG TTAATAAAAT AGACCGTATT TATTTCTTTC 3660

AAAATGTTTA ACCAAGTAGC ATAGTCTTCA TGTTTTAAAT CCGGCATTAA ATGGCGTTTA 3720

AGCAGATTAG ATTCGACTAC GACTGTTAAG CAGCCGATAG TATTTGTTTT TAAAAGATCT 3780

TGATAGGTTT GTGGTCCCTC ACTAAAATTA CGCTCATGAG TAATGTTGCC ATTTTCATTA 3840

CATACAAAAT AAGAACTACA TGTAAATAAA AGTTTTTTTT CTTCCATCCA CTTGACTTGA 3900

ACTTCTAATT TATTAGGAAG CCATAAATCA TCACTATCTA AAAAGGCAAT ATATTGTCCC 3960

TTTGAAGCTT CAATTCCTTT ATTTCTGGTC TTTGCAATGC CCATGTTCTG TTCATTTTTC 4020

AGTAAAAAGA TTCTTGAATC TTTTTCCATA AAGGCAAGyG CAATATCAAT AGAATCATCT 4080

TTACTTCCAT CATCAATGAG TAATAATTCC CAATTCTGAT AGGTTTGTTC CAGAACGCTT 4140

TCAATAGATT TTGATAAAAA TTTGCCAGCA TTGTACATGG GCATAATAAT TGAAACAAGT 4200

GCTGTTTCCA TkTGTTCACT CCTCGTTGCA TTTTATGTAA GAATGTAACA GTTTATATTA 4260

TCGCATAGCT CTTTTATTAT TTCAACGTTC CTTGTTTCCA AAAAAAGTTT GAGACATAAG 4320

CTTAAATAAC TTATGTCTCA AACTTTTAAC GGTAAATTCT AATAGCATTC TCTTTACAAT 4380

AATCCATCGT CAATTCTTCA CGTTCTTCAG ACTCATCCAA TCCTCTtGAT GATACTTTAT 4440

CAAAGAGAAT TTTGATGGTT tgAAGCAGAA TTTTCATATC CaATAGCAwG aATAATTCTT 4500

GATGTACAGT AAATCAAAGT TCAATTTACT ATTGTAGTCA GAAGCATACT TCCCATAAAC 4560

TTGGGCATAG CCAGTAATCC CCGCACGTAC ATTGTGACGT AAATAGTAAT TCGGGTTTTC 4620

CTGATTAAAT TGATCAACAA AGAATGGTCG TTCAGGACGT GGTCCGACAA TGGACATATC 4680

CCCGCGAATA ACATTGATTA GTTGTGGCAA CTCATCAATA CGTAAAGCGC GCAAGTATTT 4740

ACCAACTGTA GTGACGCGAC TGTCGTTACT TGTAGATAAG ACTGGGCCTG ATTTTGCTTC 4800

TGCTGTTGCC GACATCGTTC TAAATTTTAA AATCGAGAAC TCACGTTGAT CTAATGTAAT 4860

CCGAGTTTGT TTGTAAATCA CTGGTCCCGG TGAGGTTGCT TTGACTAAAA TGCGC 4915

(2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3149 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

ATTAATGCAA TAGCTGCCCC TACAAGCGCT ACTCCTAATA AAGGTATTTT CATATAAACA 60

GCCATTACAA AACCAATAAC AAGATAACTA AAATATTCTT TTGCAGGTAA ATATCGCAGT 120

AGTAACCCTA TACCGACAGC TGGTAGTAAA CCTCCAGCAA CAGTTAATCC ATTGGTTAGC 180

```
CATTCTGGAG TATACGATAG AATTGTATTA ATTAATGATG GACCGAAAAT TACTGCCAAT      240
GCTACTGGTA TGCCTGTTGT TGCTGCTGTT AAAAACACGC CAAACATTTG CATTTTGGTA      300
ATCATACTAT AATCACCTTT TTTCACACCT GCTTCAGCTT TGTGAACCAT CCAGATACCT      360
ATGGTATTTC TGATAACATC TAGTTGGACC ATCAATAAAG CTACTGGTAT ACCAATAGAG      420
ATTCCTACAG ATGGTTTTTG CCCAGTAGTA ATTGTAATAA AAGTTGCTAC GATACTTGCT      480
GTTTGATAGT CAGGTATTGA AGCTCCTCCc AAAATTACTT ATTCCCaAAG ATAGTAACTG      540
TAAAGTACCa CCGATAAATA AACCTGtTGA TATATCmCCt AGGATTAATC CTGTAACAAA      600
ACCCGCTGnT TACCGGTTGA TAAGTTCCAA ACaTTGTTGA GTTTTTTTCA AAGTTAATAA      660
AAAATCCATA CAGCACAATT AATACTAATT GAAAAATACT TGCTTCCATT AATTTTCCTC      720
CTTAATTACA TCAAATAATT TTGAAAATTT ACCTTTTTCA CTTGTTTTGA TAAAAATCGG      780
GATTTGTTGA ATGTCTGCTT CGACTACCAC TGTTTTTCCA CCATCAAAGT TTTCTCTTGT      840
CCATAGGTTT GTCCAACTTT CACCTTTTGG TAAATAAACT TCTCGACTAG TCTCTTTATA      900
ATTTTCAATA GGCGCAACAA GAATCGCATC TCCAAATAAA TAGGTATCTT CAATATTCCA      960
AGCATTCTGA TCATGTGGAT ATTCATAAAA TAATGTTCTC ATAAGTGGAT CACCAAACTC     1020
ATGGGCCTGT CTCATCAACT TACTAATATA AGGCTTCAAT TCTTCTCGTA tTTTTATATA     1080
TTTCGtTAGT ATTTGTTCGA CTTCTTTACC ATAAGACCAA ATTTCATTTG GAGCTCCGGT     1140
AACCATTGAT CCTCCTCCAT GATCAGATAA TGGTTTGCTG TGTGGTAAGC GATCACCATG     1200
CATTCGTAAA ATAGGCGAAA ATGTTGCATA TTGGAACCAG CGGACTAGTA ATTCTCTAAA     1260
CTCAGGATCT TTAGGATCTC CTCCATGAAA TCCTCCAATA TCTGTTGTCC ACCAAGGTAA     1320
CCCAGCCAAC CCAACGTTTA ATCCAGTATT TACTTGATTA CGAAAAGCTT CAAAGCTAGA     1380
ATCAATATCT CCTGACCATA ACAAAGCTCC ATAGCGTTGT GCCCCTGCCC ATCCACCTCT     1440
TACAAGTGTA ACAACTGATT GATTATTTCC TAAACCTTCA GAAATCATTG ATAAATAATC     1500
GATTGGATAC AAATTCCCAA TTTGTAAATC gGTTCCTTTT TTATATCGAT AATTATCGAA     1560
ATCATATACA GAATACCCAG GCTCAGCTAC ATCTAGCCAA AAATATTTAA CTCCTTTGTT     1620
AACATAATTC TTTTTAATGA GGTTCCAAAC ATATTCTCGT GCTTTCTCAT TAGTCATATC     1680
AACAAAAATT GTATTTCCTT GAATTTGCAT TGTTAATCTA ACGCCACGAT TTACTTTCAC     1740
TAAGTAACCA TTTTCTAGAT ATTCATTAAA ATTCATTGCA TCAGTTTGTA CAGTAGGCCA     1800
TATGGAAATA ATTGGCTCTA CATTATATTC ATCTCTTAAT GTcTTAAAAA TTGTTCAGGT     1860
TGAGGCCAAT ATnCTcATCA AATCGATATT CACCTTGTTT AGGCCAATGG AAATAGTCAA     1920
TAGCTATAGT TGAAAGTTTT ATTCCCAAAG CATAGTATTT TTTTACAACA TCCATTACTT     1980
CATTTGGAGA TCGATAGCGT AGTTTACTTT GCCATAATCC CAATAAGTTT TCAGGCATTT     2040
TAGGAACTTT ACCAGTTAAA TCACTATAAT TTCGaAGTAT CTCTTTTGGG GTATCACCTG     2100
CAATTACTAC ATAGTCAATA TAATTAGTAG AAGCCATTTT CCATCGAGTC ATGTTTTTAG     2160
```

CAAATGATAC CTCTCCTATA CCTGGATTAT TCCATAAAAA TCCATATCCT AAACTCGATA 2220

CATAGAATGG TATAGATACT TGTGAGTTTC TTTGAGCTAA TTCTAGAACT GTATTCTTCA 2280

AATCAAGATA AGATTGTTGA TATTGTCCCA TTCCAAAAAT CTTCTCACTT GATTCAGATT 2340

CAAATCTAGT TGTAACCTGA AAACCACCTT GTAAATTAGA GTGGTATTCT CTAGACTTCA 2400

ACTTCAGCGT TGAATTAAAA TCTTTAGTTA TTTCAATTGT ACCTGCATCT TCACTTCCGT 2460

CATCATGTTT GACAGCTCTC AATCTGATAA ACTCTTTTAA TAGCACTTCA TTTTTTTCAT 2520

TATAAAAAGT CAATCGGTCC CGATGGTCTA ACACAACTCT AATTTTCCCG TTTATTAATT 2580

CTGTCTTATT TTCGTCTTGA TTTATTATGA CTGTACCTAA AGCATCAGGC TCACCAATTA 2640

AAGAATTAAG TCTATTTTCA AAGTTCATAT CTACAAACGA TCTGACTCTA ACGGCATTTT 2700

CACCCCAAGC TTCTACAATC AATAATTCAT CATCAAAAGT TCTCcGAAAA AGATTTTGGT 2760

GTAACTTCAA TCACATCTTC CAACCCACTT TCAATTTCTT CTTAATTAGT TTGGTATATG 2820

ATAACAATAC TAAGTAATCG CAACAATACA AACGCTTTCC ATACGCACGC CTTTTGCTCT 2880

CTCTGTGTcm ACAAAAGGTT ATAGAtTAAT TTGTTTTCmC GAtAAACTAA TTAACTAAAA 2940

AAATAAACAT AAGATAATAA GCATCAATGA CTTGAAATTT ATTCGTTAGC ACGTTAAACT 3000

AATTAATAAA TAGACTATAA AACAGGCTCA GTAAAATGTC AACTGTAAAC GGCTATATTT 3060

TTTTTGGATA AATTAATTGG CTTTAGGCAA TAGGGAATCT TAACCTTAAT TnCACTAATT 3120

TATTCnAATT TAGGGGGGAA ATGGTTTCC 3149

(2) INFORMATION FOR SEQ ID NO: 127:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7019 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

ATTcCCCkGC tTCyTTTyCC ATAAAAATAC yCTTTtCTAG ATATTTTAAa GCATTTTaTr 60

GAAACGTTtC ACCAAAATAT GAACATCTTT TTCTTAAAAA TAATTCTGTT AAATAACAGA 120

AAAAAGTCTG GTCAATCCGA CCAGACTTTT TAAATATAGA TAGTAAAATG TGTAATTTCA 180

GGGATCACCC CAAAACGGGC ATGAATGGCG GTCGTTCCTA AGCCTGTATT AACATATAGT 240

TGCGTTTCTk GGTCCAACTG ATACAATCCT CTGGTATACT TTTCTGCTAA AACATTTGTG 300

ATTGGATAAA AAGGAATCCA AATTTGTCCG CCATGGCTAT GTCCTGCTAA AACTAACTGT 360

GTACCCGTAC CAACAAATGC ATCCGCGACA TCTGGTTCAT GGGTCATTAA TAACGCATAG 420

TCATACGTTT CTCGATAGGA TAATGTTTCT GAAACGGAAG GATTTCCCAA TAACGAGTCA 480

TCTAAACCAC CAATAAAAAG ACGGCGTCCA TCTGCTAAGG TTACTGTTTC TCCTTGGTTT 540

TTCATAACGG TAAAGCCACT GGCCTCCATG aCTTCrnCAT ACACCGCAGA AGCACCGCCA 600

CCATAATCAT GATTTCCCCA AACTGCATAT TTTCCAATTG TCGCTTGTAA CTGTTTTAAC 660

TTTTCAATCA TCGGTTCTTT TTGTTGCGGA TTTTTGGCAT AATTGTCAAA TAAATCACCT 720

GTAAAAACGA TAATATCTGG TTTTTCACGA TTGACTTTTT TTATGACTTT ATCTAACCGC 780

TTCGTCTCAT AAAATTCTGA AACTTGTATA TCCGAAAGTT GCACAACATT CAATGGTTCT 840

CGTTGATTTT TCTTTCCCAA AGTCACCTGg ATTAACGTGa ACTAAATTCG GCTCAATCTT 900

CCACGcATAG ATTGGGaGTG CGaTAATTAA GaAAAAAAAT AAGAGCATTA TACTTAAACC 960

AATTTTTTTA ACCACATCAC TCGCTCCTTT CTTTTCTTTG ATTTCCGAAG ACTTAGTATA 1020

GAGAAAAATC TTCTTAACCA CAATCTAAAA CAAAGGAAAT TAAGGCTTTG TTAAACAACC 1080

TAAAGAATTT CCACAATCTC TTATCCTTTT CTTAGCTTTT TTATAAAAAA AGCTAGATAA 1140

TTTCATCTGT CGATGAAATT ATCTAGCTAG AACATTCTTC TCTTAATGAT GATGTTTTCG 1200

AAACGTTTGA TAACCAATTA AGCAAAATAA AATCCCCAAT AAAATCATTC CAGTTACTGA 1260

AAATAAGCCT GCTAACATCT TTCGCATCCT CTCTGTTGCT AATCTTCCTT ATTTATAGTC 1320

ACTATTTCTT GTTGTAAAAT TGGTTGAAAT AGACGACGTT CTGCTAATTT TCTTGcTAAA 1380

ATTGCTTCAC TTTTTTTCGT AAAGATACCT AAGTTATGGC GCTTGCCTTT CAAAGAGATA 1440

TAAGCCAGCC ATTTCCCAGA AGCGGTCTGG CTCACCCCTT TGTGTCCACT GCTATTATTT 1500

TTGGGGCGTT TACTTTGTAA CTTCCCTGGA ATTACGCCAG CAACTCTTTG TCGGTAATCC 1560

TTCTTTTTAC GATGACCGCA ACTTTTGGTA TGGCCATTCA ACAAATTTTC ACAGCGAACC 1620

CAGCAAGATT GTCCGCACGA ACACAGACAA TACGCTTTCA TTGTCTGTTG AATACTTTCA 1680

ATCCGTACCA CGACTAATTC ATGATAGTTT TTTTTCAACA TTGTTTGATA GCGGTGGTGA 1740

GCACATTTCG TTCGATACCC ACGAATTAGT TGTCTTCTTG AAGCTTCTGC ATAGCTCCCA 1800

CAAGAACAGC GACAGGCCCA AACAATCAGC CCTTCTTTTC TTTTTGCCGT TGGTGCTACT 1860

ACTGTCAACT CACCAAACTT TTGGCCAGTT AAATCTTTAT AGTTACCTGC TTGTCTCAAT 1920

TGACTATCCT TCGTATTCTT GTTCTCGCCA CATCTCGATT GGGTCTCTCT CATAGTGCTC 1980

ATTTTTCTGA GTCTCTCCCT TTCTTAGTTC TCTGACCACC TCAGCAAATA CTTGCCAATC 2040

TTTTGTTTCC GGCAAATGAC GCCATGTCAA TTGACGACAC CGTACTTTCG AACTATAAAA 2100

ATCGGACAAA TACAAATCGG TTTCTAAGGT TAATTTGTGA TCAATCATGA CATCCATATA 2160

ATTGAAAGAA GCAATAATTT TTGCAATAAA GCGGTTATAA TCCTCCCCAC CAGAAAAATC 2220

AACAAATACA TGAATCGTTT TTTCAATTGC GCAAAAAGAC CGATCATTTA ACAAACAAAA 2280

CATAAAATCA TAGTATAGAT GCACGCGTTC ATACATTAAA AATCGTTTTT GATACAGCGA 2340

CTCTGTCTTT TGTATAAAAC CATCTAAGGC GCGATGAATT TGAGGAAACC GCTCTTCAAA 2400

GAAACTGCTC TGCTTGGCGG GAATAAATGC AGCCACCGAA AAGTATAGGC GGTCCCAGCG 2460

ATAACAAATC AACTTTAATT TTTCTTTTGT TTCCACCGTA ATTCGCTTTT CAAAGAAAGG 2520

AATAACACTA TTCATATGAT CAATAAATTC GTCTAACATG GTTAACTTAT CTGGAAAAAC 2580

AATCGGACTA TCAATGTATT CATTTAAACT TAAAAAAAGA AAAATATATG AAGTCTCTTT 2640

GCCAAAATCC TCTTTAAAAA GAGGGGTTTC GCTTGTATAA ATTTTTTCAA ACGCATGCTG 2700

TAAACAGCTA GGCACATGAA CTTCCGTAAA AGAGTCCACT ACATGATCAG AAAATACACG 2760

AAAACGTTGG ATCGTTAAAA GAATAAATAG TTGTGCGCTT TGATTAAGAC TAAACGTTCT 2820

TTGATACATC AGTTGTAAAT CGTTTATAAA TCGTTTAATG CCAGGACTAT TTTCCAGCAA 2880

AATGTTAGGC AGCCTTTCTT GATTACTTCC ATAAAAATAA TACAAGCATT GAAAACAAAA 2940

ACTTCGGATA TTTTTCTCTT CTCCAACCAA GCCTGTTGAC TTTAATTTAA TATGCCAATT 3000

AGCTAAAATC AGATTTAGTT TATACGAGAA GGCATAAATT TTGCTTTTGC TCATAAAATG 3060

AATTTGAGAA AATTCTTGAA TCGTGTGCTT TTGATAGAGG CCCACATCAA GAAGCAACTT 3120

GGCCATTAGT GATTGATCAA AATAATATTT TCTTAAGGCT TGATAATCAA TTTGCGATAA 3180

TTTTTGCGCA CTAATCAGTC CTTTAGCTGG AAAAAGAATT TGACTCCTTG AAGAAATCTG 3240

TGTACATTCA TAATTTATTT GACGGCACAA CGATTCTAAC TTTTGAGTGG ATAAATCAAA 3300

ATATCCATAA AATTCTTCTT TTAATAATGG ACCTTCACTA TTCATAAAAG TTTGATAAAT 3360

AAGTAATTTA TCTCGGAATT CTTGGTCTAA TAATTCATCT AACAAAGTAG TCACATTTTC 3420

AACTCCAAAC TATCACGCCT TAAGTATCTA TTTTTTTAAC AAGACGTGAG TGTCAATCTT 3480

GCTTATTTTT TCCACAAATT TAAAAATGAT TCTTTCAATA TTTAGGAATC TTCATAGAAA 3540

AGTATTATTT TAACCAAAAA AAGCCCCCTT GCGTCACCAA GGAGGCTTTT TTTGTTATTT 3600

TTTTACTTTA AATTTAAAGA AAAGCTGATT TGTTTGACTA ATTTTCTGTT TCTTTTCTAG 3660

CCTATTCGTC AAAAATGATT GAAACTTTTG GTTCGTCTTC TATACTATCT TCCTATAAAG 3720

TTCTTACTCT TTTAAGTATC AGCTTCTGTC AATAAAAGTC CTCATTTGGC ATTGCTAATC 3780

TGAATACAGT ATGACTACTT GTTGCCGCTA TTTTTTGCTG TTATTTTACT CTTTAATTGA 3840

AGCGCTAGAA ACGCTGTCAA CTGTTGGCTT TTCGGTCGCT GACTCTTCTT CATGGACTGA 3900

TGCACTTGAG ACACTATCTG CCGTTTCTGG CGCTTCTTTT TTGGAGGCAC CACTGACTGC 3960

ATCGACGTTT TTTCCACCGA CTAATTCTTG CCCAGTTTGT TTTAAATACC AATCAACAAT 4020

TGGTTTAAAG TCTAAAATAT ACTCATTAAT TCCCGCGTAA TTTCCATTAT CATCATACAT 4080

CGCTTGATAG TTATGCACTA CATAGGTATC TGGTCCGTGC ATTGGAACAT GAACGCGAAT 4140

AGCATCTGTT TTTCCTGACC GTAATTGTTG AATCACCCAT TCTACATTTT TCAATGCTTT 4200

TGGCGGATGG CAATTCGCTA ATGGATTACC AACTTGTCCT GGTTGCCGTT TAGCAAACAT 4260

TTCTTCAGCA GCCATTTTTT TATTGTAATA TAAAAATTGA TTGTTACTGT CTGCATAAGT 4320

CAATTCCATT GGCATTGATT TTAAGAAATA ATTTAGCTGA TCTACGGTTA AAATGCCTCG 4380

ATCTAATTTT ACATATgTGt CGCCTTCCAC AGCCTGAAcC GCtTCTGCTG CTTGtTCTAC 4440

CCAATCATCT GCTGCCATAT CAACCCCTTC GATCGTTGTT TCAATGGTTC CTTTTGCATG 4500

TAAATCCTCA GGTAAGTCTT TTGGCTTAGC TCCTTCTAAT TTTGAAATCA CATCAATGAA 4560

TTCCaCGAAT TTTTGTAATG TGCTTTCTAA AAATTGACGT GTTCCCtCCG CAATTAAGTT 4620

GCCGTTTTCA TCAAATGCTT CTTTAGCTTT CCCTAATAAA AACTCGTTCC CaGGCATCaC 4680

GATTGCATTG ACACCAGGTG CATCAAGAAT TTGACGGAGA TGTAATTGCG CACGCGAGGT 4740

TCCTTGATCA TAATAAGAGC AACCAACAAT CATTACTGGT TTATTTTCCA GCGGATGAAT 4800

TTTGAATGAT AACCATTCTA ACACACTTTT CAAACCGGCA GGAATGGTAT GGTTATGTTC 4860

GGGCGTCGCA ATAATTACGC CGTCTGCTTG TAAAATTTTG CGATTCATAT TCTGAATCAA 4920

GACACTGTTC GTTTGATCCT TACTTTGATT AAACATTGGA ATATCTTTAA TTTCTAAAAT 4980

TTCCAAGTCG AACATTTTAT AAAACTCTTT GCCAATATAT TGCAGTAGTA ACCGATTATA 5040

AGAAGAATCA GCGTTTGATC CGACGATCCC AATTAATTTC ATTTTATTCA CTCCATTTCG 5100

TTAAATCGTT TCCCATGAGA AATTTTCTGC TTCTTTTTTA TTTTGCGCAT GCGCATGTTT 5160

CAGCTGTTCT GTAATTTCAA TAAATACACA AAAATCTGCA AACAGACCAG CTAGTTTATC 5220

TATTTGTTCT GAATCTTTTA ACGCATTGTT TTCATCAAAA GCTTGTAAGG AATGCGCTAA 5280

TAGAAATTCA GAGCTTGGCA TAATTCGTGC TTTCAATTCA GGAGCATCTA AAATTTGACG 5340

CAGATGCGCT TGTGCTCTAG AGGAACCTAA CGTGCCATAG GATGCACCAG TAATCATAAC 5400

TGGTTTATCA ACAAACGGAA AAATCCCGTA AGATAACCAA CTTAGGGCAT TCATCAAAGA 5460

AGCCGGCACC GCATGGTCAT ATTCAGGCGT ACTGATAATG ACGCCATCTG CTGCTTCAAT 5520

TTTTTCAGCA ACCGTTTTAA CTATTTCTGG CAAAACTTTG TCTTCTGGCT TGTTAAACAT 5580

AGGAAAGTCA ACAATTTCAA CTAGTTCAAT ATCAGCCTCC TGCGCAAAAT AGTGTTGCAT 5640

AAATTGCAAA AGTTGACGAT TGGTTGATTG CTCTGAGTTC GTTCCTACTA AGCCGATAAT 5700

TTTTTTCATA ATGAGCTGTC CTTTCTTTTT CGGTTCTATG AATGGATAAT GTGAAGAGTA 5760

TCTGCTAACC CGCTGCTATA TAAAATTTGC TGGTCTTGTG TAATGATGAC TCCTTCAATC 5820

CCATCTAGTT GATTGAGGAT TTCAAGCGCT TCTGGAATTG GATAACCAAA TAAACGCGTT 5880

GTCCAAATCT CACCATCCAC TGAAGCATCC GAGATAATTG TGAGGCTGGC CATTTCAGTT 5940

TCCAATGGGT AACCGGTTGT CGGATCCAAC AGATGGTGAT AGACTCGCCC AGCTTCGGTT 6000

AAACTTCTTT CATAAATGCC CGACGTCACA ACCGATTGAT TAGCCACTTC GACAAGTAAG 6060

CTGATTGTTT CTCTGGATTC TTGCGGATTG CGAATACCAA TTCGCCATTT TTTGTTTTGA 6120

TGCGTAGAGG GcCCCAACGT CACAATATTG CCGCCCAGAT TAATCAAAGC AGACGTCACA 6180

TGGACTTCTT TCAAATAGGc AATTAGTAAA TCAGCGATAT AGCCTTTTGc TAAAGCCCCC 6240

AGATCTAATT TCATTCCtTC TTCTTCCAAA AAAACCGTTT GTTTTAAGGG ATTtAAGTGA 6300

ATCTTTTCtG GATTAATTTT TTTCAAACAG GCTTTAATTT CCTCTTCTGA CGGGACACGG 6360

GCATCTTTAA AACCGATTCG CCATGTTTGA ACTAACGGGC CAATGGTAAC ATTCAAATGA 6420

CTAGCTGGAT CACAACTATG TTTTTTCCCT AATGCAATTA ATTGATACAA CTCTGGATGA 6480

ACGCTGACTG GTTTTTGCCC AGCTTGTTGA TTGACAGCCA TCAGTTCCGA AGTTGAATCA 6540

TTTGCACTAA AGCGTTGTTC ATAAGTGATC AATCGTTGAT GCACTTCTTC TAGAATTTTT 6600

TCAGATTCTT CATGGTCTAC AAATACATCT ATTACTGTGC CCATCAAATA AATAGTTTGA 6660

CTTTGTTGCA CCTCAAGACT TCCTCTCCAC GAAAGATTGT GTTTTATTTC ACATTTATTC 6720

ACAAATGAAT CATACGCCTT TATTTGATGC CAGTCAAGTC GTTTATATAT TTTTTTATAT 6780

AAAATAACGA CCTTAAAATT TTCATTTAAG GTCGTAACTT GTCTTGAACT CTTTTAAAAA 6840

ACGCTTTTTT TATTATATTT TGAAAATAAA TATATAATTT TCTATATAAA CTAATTTCCC 6900

TTAGATAAAA TTGTTTCTTT CAATGTTTGA ATTGCTTCCA ACAAGATGCG ATTTTTATTG 6960

ATTTTTGCAT ACGCGCGGGC ATCTTGTAGT AGTTCTAAAA TTTGCTCTTG CGGCTTTTC 7019

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32768 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

TATTTATTCT TTTAACGAAA CACACAATGA AGGTAAAGAG aCTCTtGGTG GTAAAGGGGC 60

AAACTTGGCA GAAATGACAC GCTTAGGTTT GCCGATTCCA CAAGGATTCA CGATTACAAC 120

CCGTTGCTGT ATGGACTATT TGGCAGATGC TACTTTCTTT GAAGAACACT TACAATCAGA 180

AATTTTGAAA GCGGTTAAAA ATCTAGAAAC TGAAACAGGC AAATCTTTTA CCGCGGACAA 240

TGAGATTCTC TTAGTTTCCG TTCGCAGCGG TGCTGccTtC TCGATGCCTG GTATGATGGA 300

TACTATTTTA AATTTAGGTT TAAATGATCA ACGTGTGAAA AAATTTGCGA CCTTAACCTC 360

TCCTGGTTTT GCTTTTGACT GTTATCGCCG ACTTATTCAA ATGTTTGGAG ACGTCGTCTA 420

TCACATTCCC AAAGAATTAT TTGACCAACA AAAAGAGCGC TTGGAACAAG AGCTCAACAA 480

AAAAATTACC GCTTTTCAAG AAGAAGATCA TTTTGCCTTG ATTGTCCGCT ACCAAGAAGT 540

TTTTGAACAA CATCAGGTGG TCTTCCCGCA AGACCCTGTC GCACAACTTT TTGAAGCAAT 600

CAAAGCTGTT TTTGATTCAT GGAATAATCA ACGTGCGGTG GTTTATCGAA ACTTACATCA 660

TATTGCACAT GATTTAGGTA CAGCTGTCAA CATTCAAGAA ATGGTTTTCG GCAATCGTGG 720

ACTAGACAGT GGCACTGGCG TTGTTTTCAC TAGAAATCCG GTCACAGGGG AAAATCAACT 780

GTTTGGAGAG TTTCTTTTGA ACGCGCAAGG AGAAGATGTA GTGGCTGGTA TTCGGACGCC 840

AGAACCCATT CGTCGTTTAC GTTTGACTAT GCCAAAAGTT TATCAAGATT TTCGCCATTA 900

CGCAGAACTT TTAGAATATC ATTATCGTGA CATGCAAGAT ATTGAATTTA CCATTGAAAA 960

TGAAAAATTA TATATTTTAC AAACAAGAAA CGGGAAACGA ACCGCCGCAG CTACGGTTAA 1020

AATTGCTTTA GACTTAGCCA AAGAAAATCG CATTACTAAA CAAGAAGCAC TGTTACGTGT 1080

GACACCTGAT ACGATTGATC AATTAATTCA TCCCGTGTTT GATCAAGAGA AACGACAACA 1140

CATGGAACGT TTAGCTATGG GACTTCCTGC TAGTCCTGGT GCAGCAAGTG GCCAAATTGT 1200

TTTTACAGCA GAAAAAGCCA AAGAACTGAC CAATCTAGGA AAGAAAGTCA TTCTTGTTCG 1260

TCAAGAAACT TCACCAGAAG ATATTGAAGG CATGGTTGTA AGCGAACAAt TGTGACCAGT 1320

CGTGGCGGAA TGACTTCTCA TGCCGCTGTG GTCGCACGAG GAATGGGAAC TTGCTGTGTG 1380

ACTGGTTGTG AAAGTTTAAC CGTCAATGAA GAAACCAAAC AACTACACTG CGGGCCACAG 1440

GTCATCTTAG AAGGAACAAT CATTTCCGTC GATGGCTCAA CGGGGGAAAT TTATCTTGGC 1500

GAAATTCCTA CAATTTCTGC TGATAATAAC GATGmCCtTC ArGAGCTCTT GTCTTGGGCA 1560

GATGCGTaCG CTGATtTAcT GTTCGTGCCA ATGCAGAAAC TACaCAAGAT TTAGAGACTG 1620

CTATCCGTTT TGGCGCAGCG GGGATTGGTT TAGCGCGAAC AGAGCATATG TTTTTCGGTG 1680

AAGAGCGTGT CTTGGAAATG CGTCGCCTGA TTTTAGCAGA GTCTGAAAAG GAAGCAACCT 1740

ATGCGTTGGA ACAACTTCTT CATTTTCAAC AGGAAGACTT TTATCAAATG CTGAAAGTAG 1800

TTCAAGACAA GCCCATGGTG GTCCGCTTGT TAGATCCGCC GATGCATGAG TTTTTACCTC 1860

ATGAAAAGAA TGATATTCAA TTATTGGCGA AgCAACTACA ACgGTTCCCG GTGACCATCG 1920

CCAAACAAAT TGAACGGTTA CAAGAAACCA ATCCAATGTT GGGTCATCGT GGTTGTCGCT 1980

TAGGTGTGAC CCAACCACAA ATATACAAAA TGCAAGTTAC AGCCTTATTT ACTAGTGCCA 2040

TTCGATTAGT CAAAGAAGGA ATTACTGTTT ATCCAGAAGT GATGATTCCT CTAATCGCCG 2100

AAAAAGAAGA ACTTCTCTAT CTCAAACGCA TTTTAAAAGA AACGATTGAC GGTCTCTTTG 2160

AAGAACATAA AATGACTCCT TTCCCTTACG AAATTGGGAC AATGATTGAA CTACCTAGAG 2220

CTTGTCTGAT TGCGGATCAG TTAGCAGAAG AAGCAGACTT TTTCAGCTTT GGGACAAATG 2280

ATTTAACACA AATGACCTAT GGTTTTTCTC GTGATGATAT CGGTAAATTC ATTAATACGT 2340

ATCGGGAAAA GAAAATCATT ACCCAAGATC CGTTTCAATC ATTAGATCAA ACAGGTGTTG 2400

GTCAATTGAT TCAATTAGCT GTAGAGAAAG CGCGTCGCAC TAAACCAAAC ATGAGTATTG 2460

GCGTATGTGG AGAAGTCGGT GGTGATCCGC AATCCATCAC ATTTTTCCAA ACACTTGGCC 2520

TAAACTACAT CTCTTGCTCA CCTTATCGGG TACCTATTGC GAAACTCGCA GCTGCCCAAG 2580

CCAAAATTTT AACCGAACCG GCTGTGACTG AAGAACAAAT GGTCCTTCTT TAATTTAACG 2640

TGTGTAACAA ATGATTCATC CTTACAAGAT AAAAAGTGCT AGGrACATCA GTCGAAACGC 2700

TTGATGTTCC TAGCaCTAAA TTTTAAGATC AATGATTGTA CTCTGGTTAT TTTCTTATAA 2760

ACGCGGCATC TTTTGTCAAG GGGTCTGTTA CATCTGATAC ATATCTTTCT ACTTTCATAT 2820

GCAAATCATA TTTTTCTCTA AGTTCCGTAA TACGGGTCAT CATTGGAGAA GCGTGGTGTT 2880

TGTCTAATGC TTCCTGATCT GTCCAACTAT CAATCAGCAG AACTGTTTCT TCATCGTCCA 2940

TTGGAAAAAA GTATTCATAG CGACTATTCC CTTTTTCTGC ACGAATCGCC TCTACAATTC 3000

CGCTGGCAAT CATTTCTTTG GCAAATTCTC TTGCACCTCC ATTTGTCCCA CTATAATAAA 3060

TATTTATAGT AATCGCCATC TGTCTTCCCC CCTAAATTTA CTTGCCTTTT TACGTTCCCG 3120

ACTGGAATTG AACCAGCGAC CTACCGCTTA GGAGGCGGTk GCTCTATCCT ACTGaGCTAC 3180

GAGAACAATC TGTTTTAATT TTAGAAGTCT TCACTACATC TGTCAATAGA CAAAAAGAAA 3240

GTATCCCGCA TAAACAATTA TGCGAGATAC TTTCTTTCAG TGAAaCACGT CGTTCGTTTA 3300

TCTATTATTT TTTTTGCCGC CGATCAAGAA TCATTTTTCC TAAATAATAA AGTCCTGTCG 3360

CAATAGCTAT ATATTGCCCA ATTTTATTAT TGCGTGCAAC TTTTTCTTGT TTCGTTTTCG 3420

CCATTGTCCC CATCACCTTT CTAAAATAAT TCTTTTTACA TTTTTATTAT AGTAGGTTCT 3480

GTCATTTTTT GCGAATAAAC ACACTTTTAA AAATAATCTT CTAAAATTGG CATGCCTTTG 3540

ACTAAACGTT GGCCtAAACG TTGAGCTATC GTTGCGTCTC CTTGCAAGCG TTCATAAAAA 3600

GATAACGTCT CTGCTGAACG ATAGCCCAAG AATAGTTGCG TCCACGTTTG GATATCTGCT 3660

TTCAAAGCAG CAGTCCCTTC TTTTTCCGCA GCACCTTTAG TCACTGTTGC TTTTCCTTGT 3720

TCGTCAATTG TGATTGTCCA AATGCCTTCA TTCCATGGCC CATAGGAATC TTCAATCTCT 3780

AAAGAATACG TTTCTTTTTC TCCTGATTGA AATGGATACT TTTCCAGAAA TGTTTGTAAC 3840

TCAACGATTC GTGCCATCAT GTAAGGTAAA ATTTTTACAG AAGCTGCTGG CGTTGGCATT 3900

AAATCATTTA AATCTTTTCC AGCAAAACCA TTAATCCAAT GGAACGATTG AACGGAACCA 3960

CTATGTGAGC CAATAAACCC AGCTAAAGCT TTAAAAGCTG TATTAGTTAA ATAGTTCCAT 4020

TCTACGATTT CAAAGGTGCC AGCAGCTATC CGATAAATCA CATAGCCTTC TGCTTTCCCT 4080

TCTGACGAAT AATAAATCGC TTGATTATTT GGTTTACTTG CACGATTCAG CGTATAATCC 4140

AGCCACCAAG TTTCACGGAT CACACCACCA GAGTGCGCTC TTTGATTTTC aAGGTATACG 4200

TCTTTAATGA CCTCTTTACC GTCAGCCCAA GaAACTCGtT TAATGGTACC TGGAACTCgT 4260

TTAAcACGTG GCCaATCTTC CgTTTTAAtT GTATACTCTG CTTGTTCaAA TGTTTGTTCA 4320

TAGCCATATT GACGATAAAA TGGATATGAA AACGGTGCTA AGTAAGAAAG GGCAACCTTT 4380

TGTTTCGCTA AATCAGCCAA CATTTCTTTC ATAATAGCAG AGATGCCGCC TTCACCACGA 4440

TATTCTGGAT AAGAAGCCAC ATAGCCAATT CCAGCCATGG GATAGCGGAC ACCATGGAAA 4500

TTCACTTGAA AAGGTGTCGC CATCACTTGA CTTGTTAATT GCTCATCAAT TAGGAAGCCA 4560

TAGCTTTGCG TATGTGACAA CAATTTTTCA AAACGTTCTT GcGTTCCGCC GTGGGTTCCT 4620

GATTAAAAGC ATAGATAACA AGATCAAACA TCTCTTTCAT TTCTTCTTTG CCCATTTTTT 4680

TGACTCTTTT AGTTGTCATT TTCTTTCCTC sTCGCACATT GTATCAATTC ATTTTATCAC 4740

AGAAACGATT CTCTTTCTAA AAGCCGTTAT TTGCTGTCAT TTCTTTGAAC CAACGACCGC 4800

TTTTCTTAAT GGTCCGTTTC GCATCATCAT CTAAATCCAC AGCGATGAAG CCATAACGAT 4860

TTTTATAGGC ATTGAGCCAA GACCAGTTAT CCATACACGT CCACATATGA TATCCTTGAA 4920

CATTACTACC TTCTTGAATC GCTTGGTGCA CCCATTTCAA GTGATCTTGC ACGAATTCAA 4980

TCCGATAATC ATCTTCAATG ACTCCTTGTT CATTGACAAA GCGTTCTTCT CCtTCTACGC 5040

CCATGCCATT TTCTGAAATG TAGCAGCGAA TGTTGCCATA ATTTTCTTTT AAATTCATTA 5100

ATGTATCATA AATACCTTTT TCGTAGATTT CCCAGCCACG GTAAGGATTC ATTTTTTTGC 5160

CAGGCATATC GTAGACATCA TAAAAATCTT CTGGAAGTAA GCCTTCTGGT CGTGTTTCAA 5220

CAGGTGTTTC TTTTGCTTTG GCACGACGTG GTTGATAGTA GTTTACTCCT AATAAGTCCA 5280

CAGTATTTTC AGCAATTATT GCTAAATCTT CTGGCGTTGT TTCTGGCAAT AAGTCATTTG 5340

cTTTCGCCCA AGCAACCATA TCTTCTGGAA AATGACCTTT AACAGCCGGG TCTAAAAACG 5400

AACGGTTAAA GAAACCATCG ACAAATTTAG CCGCATTTAC ATCGGCTTCG TTGTGCTCAT 5460

CACGAGGATA AGTCGGCGTT AAGTTTAAAA TAATGCCAAT TTCACCTGAT AAGTTCATAG 5520

AGTGATAGAC GGCAATCGCT TTGGCACTAG CTAAAGCTTC ATGATAGAGT ACTTGAATCC 5580

CTTCTTTTAG ATTCACTTTA TCAGGATAAT GCCAGCCATA GAGATAACCA CCTTCAACTG 5640

GTACAATTGG TTCATTGTGC GTAAACCATT TTTTCACGCG ATCGCCAAAC AATTCAAAAC 5700

AGGTTTTGGC AAAGTCAACA TACAAATCAA CTACTTCACG GCTTTCCCAA CCGCCTTTTT 5760

CTTGAAGTAC CATCGGCATA TCAAAATGAT ACAAGTTCAT AAAGGGCTCA ATCCCGTGTT 5820

CTAATAAATC ATCAATGACT TGATTATAAA AATCAACTGC TGTTTGATTG ACTTTCCCTG 5880

TTGTAGGATC TGGAATTAAA CGGCTCCATT GAATGGATGT TCGAAAAGAG TTATGTCCCG 5940

TTTCTTTCAT TAACTGAATA TCTTCTTGGT ATTTTTTATA AAATTGTGAG GTCTTATCTG 6000

GACCGACTTG TTGAAAAAAT TTTTCTGGCG CTTCTTGATA CCAAAAATCC CAAATATTTT 6060

GACCTTTACC ATCACCTTCA AAAACGCCTT CTGTTTGAGG ACCGCTGGCT GCAGAACCCC 6120

ACCAAAAATT TTCTGGAAAT TGATATTTCA TGTTATTTGC TCCTTTACGA TTTAATCATT 6180

AAAAAAAGGA TTGAACCCTC CGTAAGAGTC CGGACGATTC AATCCTTCGA TTTTTTACTC 6240

GGCTGCTTCT GCGGCTGCTT CTTGTTTCAA GTATTGATCT GTTGCTACTT TTACGAAAGG 6300

CATGTAGATC AAAATAGAAA TCAATAAGTT AACAGCTGCT AAAACTCCGC CGGCGATACT 6360

CTTCGTTGCT AAAAATCCGC CAATAATTGG TGGTGTTACC CAAGGTGGCA TAAAGGTTGC 6420

AGCTGGTACT AAACCAGTTG CTGTCGCCAC ATACGCAACA GTTACAAGAA CCATTGGGAC 6480

TAAAATGAAT GGGATAAACA TAATTGGGTT TAAAACAATT GGTAAACCGA ACATGACTGG 6540

TTCATTAATA TTAAATAGTC CTGGTGCTAA ACTTAAGTTT GTAATGACTT TGTATGGTTT 6600

GTTACGACGA CCTACGATAA AGATAGCAAT TAACAGACCT AATGTTGCCC CTGTTCCACC 6660

TAAGTTTACA AATGAATCAA AGAAAGGTTT GTTTACAATA TAAGGAATTG CTTTACCTGC 6720

TTCTAAAGCT TTAATGTTCG CATCAATCGC TGGTACGTTG ATTGTTTGCA TTAATGGGTC 6780

AACCATGTTG GCACCATGTA AACCAAAGAA CCATAAGAAT GGTGTAATGA AAGCTAACAA 6840

TAATGCTGAT GGGTAGCTGT TTGCAAGTCC CmTAAATGGT TGTTGAACTG CTTCATAGAA 6900

TGAAATAACG ATATTGTCAA TTCCTAAAGC AAATAGAATT GATGTAATTA AACTAAATAA 6960

ACTGATkGTA ATCATTGCTG GTAATAATGC CgCAAATGAT TTTGCTACTG CTGGTGGTAC 7020

ACCATCAGGC ATTTTGATAA CTAATTTAGG ATTGCCAGAT AAACGTTGGA ATAGTTCTCC 7080

TGAAATAAGT GCAATAATCA AAGCAATAAA TAATCCAGTT GCGCCCATGC CTTCTAAGCC 7140

GCCAACTGCG AAGAATGAAG CTACAGAAAC TGTACCGGCT GCAATGCCGT CTTTTcCATA 7200

AGATTTAACT AAATTGTGTG CTACTAAAAA GGCTACTAAA ACGGCAAAGA TTGCGAATGT 7260

ACCGTTCCAA ATGTTACCAC CAAAGCTTTT CCAAATTGGT GCATCGTTAT ACATCATTGG 7320

GAAAATTTTG TTTAAAAAGT CCTCATATCC TGGAATTGGA ATATTATTAA TTAATACCGC 7380

CAAAGCCCCT AAAATCATTA ATGGCATCGT TACGACAAAC GCGTCACGAA TTGCTACCAA 7440

GTGCCGTTGC GCTGCGATAA CTGACGCTTT CGGCATAAAG TGTTTCTCCA TCCATTGAAC 7500

AAATCCGTCC ATTGTTATTC CTCCTAATTA ACGCTTTCTC AAGCGCGTTT TTCTTATAAG 7560

ACGCATTGAC AACGAGATTA CTCGTTATCA ATACGGCGAT AAAGATCGAT AATTTCTTTT 7620

GCTAAATCTG TAAATGCAAT CGCAGTCATT AAATGATCTT GACTGTGCAC TGTTAACAAT 7680

GTTACTTGCA TGTTATTGCC TTTTGCTTCT TCTGTTAACA TTCCTGTTTG CGAATGATGT 7740

GCATGTACAA GTGATTCTTC TGCATCTGCA ATTTTTTTGT CTGCTAATTC AAAATCCCCC 7800

GATTTTGCTG CGGCAATCGC TTCCATTGCA TCGCTTTTAG CATTTCCGCC ATACATGATT 7860

AAACCCATGA TTGCTTCTAA GTTCTGTTGT CCGTCCATTA CCTCTACCCC ATTTTCTGCT 7920

CTAGTTTGAT GATCCCATTA ATGAAATTGC TTGATCTAAA ACTTTTTCGC CATTCATCAT 7980

GCCATAGTCC GCCATGTTAA TTACATCTAA AGGAATCCCT TTTGGTTGTA ATTTTTGTTC 8040

AAATTGCCCT TTCATGAAAC GAACTTGTGG GCCTAATAAT AAAACATTCA CCTCTTTATT 8100

TTCCAAGTTT GTATCTGCTT CAGAAGCCGA TACTGCAAAG ATGTCTGCTT CCATACCACG 8160

ATCTTCTGCT GCTTTTTGCA TTTTTGTTAC TAATAAACTT GTGCTCATTC tGCGGAACAA 8220

ACTAACATAA TTGTTTTTTT TGCCATTTTC CTTCATCTCC TAAATTCTCT TTGTTTTGCT 8280

TTACACCTTT ATATAAGCAA GAAGCGTGCC AAGTTTTAAC ACATTAAAAA CCTTGATATG 8340

ACAACGTTTC TTGATATAGA AAGCGTGTAC ACACTAAAAA TAAAACGATT ACACACAGCT 8400

TTTTTGTGTG CGTTTCCAAA ATAATAGGGC AACAAAAGAA AGTGAGCGAT AACCCGAAAC 8460

GACTCATCGC TCACTTTCTA AAGCTGAAGA GAATCAACTG GACAAAAGCC GCTTTTTCCC 8520

TTTAAGCAGT TTCCATCATT TTAAGCATTT TATAAAACTG TTTCTTTTTA AAATAACTTA 8580

AAAAGATACC TAATACTGCA TCATTCAGTT TTTGTAACGC CCCATAAGAA ACCATTGTTT 8640

CTGTATAACG CAGTTCACAG CTTTTTTCAT CAATTGGCTG TAGCTCATAC CGCACCGAGA 8700

AATCATTTTT AGTTGTGGAT GTTCGAAATT GATAGGTCTT ATTTTCTACA GCTTCTTCAA 8760

TAACTATTTT CGCGCGACTA TTTTTACTAT ATTCTTTGAT ATATTCAAAC TTACGTAACT 8820

GCTTTCTTGT AAGAGTTTTG CCTGTTTGCT TACGGATATC AAAAATAACA GAATTCATTA 8880

ATTGCGTGTA AAATTCATCT GCTGAAATAG TTAATGTTTT CGTAATTTCC ATTACCGTTT 8940

CACCTCATCC TTTGTCGCCT CTTCTTCAGC AACTTTTTTT CCTTTAAAAA ACAGGTAACC 9000

ACCAAGTCCA ATGAGAAAAA TCCCCGTTGT CAGACTTAAT AAATCATAGT TTAAGGAGCT 9060

GGGATTCAAC GAATACAAAA CAATCAAGAA GCCAATACTA CATAAAAATA ATCCATTTCT 9120

AATCATTTTC ATCCATCCTA TCTTTTTAGA CTACCGTTGC TTGGAAGTAT AAAACGCCAC 9180

CCCAaGGTTG TTCCTTAATC AGCTGtTTTG CAACTTGTGC AGCAGTTTCA GGGTCTCCTT 9240

CAACTTCAAA AAATAACTTA ATACCATTTT TTTCTTTAAA AGTAAATGTT TCATTTTCCC 9300

CTTGATATGC CTCTAATAAA GTAATAATTT CCTCTTGCTT CATTGCACTT GCAATCGAAA 9360

TCATGCCGCT CGCTCCTTTC TATACAAGTA ATTGCTGTTT TTATTGCTTA TTGCTCTCTT 9420

TCGCCTAATG CAGGATCCAT TTTGTTAGCA GCGATAACAA ACGGTGTCCA AATCAAGAAA 9480

GTGACTACCA TACATACCAA CGCTACGATT GGTGCGCGCC AGTCGGCACC TGTCGCTAAG 9540

AAAGATAATA GCAACGGCGG TGTCACCCAT GCGACTTGTT GAGAAACTGG GGCGACTAGC 9600

CCAAGATACG TTGCAACCCA GCCAATTGTA GTAGCAACTA ATGGCGCAAT TACAAATGGA 9660

ATAAACATAA TTGGATtCAG AACAACTGGT AAACCAAACA TAATCGGTTC ATTGATATTA 9720

AAGATCCCAG GCCCTAATGA TAATTTACCA ACTGTTAAAT AATCTGCTCG TTTGGAGAAT 9780

AACAAGATTG CAATAATTAG GACGATTGTA CCGCCTGAAC CACCAAACCA AGCAAAAGCA 9840

TCAAACGAAC CACGGACCCA CATGTAGGCT TTACCGTCGT CAATAGCGGC TAAAACCGCA 9900

GCTGTTCCTG TTTTTCCTTC ATAACCTTTT TGGAAAATAT CAATATTGAT TAATTGTGCT 9960

TGTCCCCAAA TTCCTTCAAG AACTGGTGCC AAAACGTTTG GCCCATGAAT ACCAAAGAAC 10020

CAGAAAATTT GAACAAAAAC CGTTACGATT AACACAGCGC CTAGTCCTTG TGATAACCCA 10080

AGGAACGGTT CCGCGATATA TTTTTGAATT AAGTCAATTA ATAATTGACC GCCTGAAAGT 10140

TTTGACACCG TAAAGTTAAT AATTGCTACG ACATACAATG CGATTGTTGC TGGTAAAATT 10200

GCCGCAAACG CTTTTGAAAC TGCTGGTGGT ACAGAATCAG GCATTTTTAT TGTTAAATCC 10260

GCTTGCATTA ATTTACAAAA AATAATGGTT GATAAAGCGC CCATAATCAT CACCGTAAAG 10320

AAAGCGTTAC CATTTAGATG GTTGAGATTT AACCAACCCC ATGGATTGAC AGTCAATTGA 10380

TTATTAGCAA TTGACGCACC AATTTCTGCC CCATTAATTG CGTTGGCAAT CTTTTCTGGC 10440

ACTTTTACAG CTAATTCAGC TGTGTTATTT GGTGCAAAAG CAACACCTGA AATTAAGGTA 10500

GCTAAACTGA CAATCCCACC AGCTAGCTCA TTCACATTGT ATGCTTTCGC TAAGTTATAG 10560

CCCCAAGAAA AAGCGAAAAT CAAACCAGCG ATAsCcAATG TACCATTCCA AACAAAACCA 10620

TTAATCCCAA TGATGACATT CAATACTGCA AAAATACCTT TGTCTGCTTC CAAACTTGCC 10680

GCATAAGTTG AGACAAACTC TTGGGGAAGA TCTCGAATAA TTGCGTTAAT CATTACGGCG 10740

ACTGATCCAG CCATTGTCGC AGGTAATGTA CCAATAAAAG CATCCCTTAG TGCCACCAGA 10800

TGTTTCTGTG cACCAATTTT TGCGGCAGTT GGTAAGATAT ATTTTTCCAT CCATGCCATT 10860

AATCCGTTCA TTTTTATTCC TCCTGTATCC ATTTATATTT ATTAAAAAGG CAGTAGCCTC 10920

ATTAAACGAA TGAATCGAGC GTACAAAACA AAACCTTTGG TAACATGCAC TATCCTTATT 10980

TGCTATTCAC TTTTCCTAAA CGTTTTAATT CCTTATCCCA TTAATGAGAT TGCTTGATCT 11040

AAAACTTTTT CGCCATTCAT CATGCCATAA TCTGCCATGT TAATTACATC TAAAGGAATC 11100

```
CCTTTTGGTT GTAATTTTTG TTCAAATTGC CCTTTCATGA AACGAACTTG TGGACCTAAA   11160
AGTAAAACAT TCACCTCTTT ATTTTCCAAG TTTGTATCTG CTTCAGAAGC CGATACTGCA   11220
AAGATGTCTG CTTCCATGCC ACGATCTTCT GCTGCTTTTT GCATTTTTGT TACTAATAAA   11280
CTCGTGCTCA TTCCTGCGGA ACAAACTAAC ATAATTGTTT TTTTTGCCAT TTTTTCTTCA   11340
TCCCCTTAAA AGTTATGACG TTATACACTT TATTTATAGC AATAATTGTG CCAAGTTTTT   11400
TCCTCACTAA AATACTGATA TAACAGCTTT TGTCATCGAA AAATGCGTGC ACACACTAAA   11460
AGAAAACGCT TACACACATG AAATTGTGTG TAACTTACAA AAAGAAAACA TCAAGAAAAA   11520
AGAGAGCGTT TTCTCATTTC TTGATGTTTT TTTTATTTAT TATTGTTTAC TTTTAATTTC   11580
ATTTTCTAAA AACATTTGGG TAACATGCGC AACTTCCGCT TCTGGAATCG TCACACGATA   11640
CGCCCGTTCA ATTAACATAA AATTCGTGCG TACGACATCG ATTTCTAAAC GATGTGTTTT   11700
AACATATTCC GCTAAGTTTG GAAAATTTCT TGTTTCTTCT CCTTTAATTA AAGCATCTAC   11760
CAGAAAAGCC AAATGAATCA TGATTCCGGC ATCCACCCCT GGTTCAACAA TAATATGCAT   11820
GTCTGTTTGT ATTTGATGAA CTGTTTTTTG TAACATGAGA ATTAATTTTT GTAAGGAGTC   11880
CACAGAAGTA ATGACGCCAC TTAGTGAATG AACAATTGTA TCAATCGCCA CTTCATCACT   11940
AGCAATCCGT TTCAGTACAT TTAATTTTTC ATCATCAAAA ATATCATAGA CGGAAAAGAA   12000
GGGAATATTT TGATATTCAA CATCCaCAGT TCCCGCAATG GCCTTAATTT CATATTCTTC   12060
CATCAAGTTG TCGATATGTT TTTTAAATGT TTCTCGTTCA ATAAATTGCA TTTGAATCAG   12120
TTCCACTTTT GTTTCATCAA TAACTGGTAA GATTCGTTGG TATAGTTTTG CGGCAACCCC   12180
TTCTCCAGTA AAGCACGTCA CGATTACTGC TTTTTTTACG TTTTGACGTT TTTGCAGACT   12240
AGAACGAAAT TGTTCTCTCA CAACGCTTTC AAAAGAAAGT TGAATATTTT GGTAAATGTC   12300
CtCTAAACTt CGTCCCACGC TTGCCATTCG AATAGCTTCT AAGACAATCA TTGTGCTCGT   12360
CATTGTAATC GCCTTTGTTC GGATACCTGT TTCTTCAAAC AACATATTTC CAAACGAATT   12420
TAATGAACCC ATATCCGTCA GTAATArAAT CCCATTATTT AACGATTCTT TTTGCGTAAT   12480
TACTTGGTTT CGTAATTGTT cATACATTGT TTGAACTTCC ATCGTTAACG GCATATTCAT   12540
CGCTGTTCCG ATTGACGTTC CtAAAAGTTC TTGGGCGGTT TTCAACATAC TTGAGGCAGT   12600
AGAATCCCCa TGCATTAAGA CGAACACACC AACATTATTT TCTGGAGTTC TTTCTTCTTT   12660
ACCGACATTA ATAGATAAAA ACATACTAAT AAAGCCAATT TCATCGAAAG GAATCTCTAT   12720
GTTGTTCTCT TCTTCAATAA TAGAAGATAA ATCTAAGGCT ACTTGAAATT CTTTTTTCAT   12780
ATTTTTACGG ACATTATTTA AATCAGGATG GACAATCACG TGACCTTCTT TGACACGTTC   12840
AATCGTACTT TGTAAATGAA GCGCAAAGGC aAAGCGCgCa GTTTCGTTGT ACGTGTGTTC   12900
CaAACGCTCT TCTGCGaTAT CATaAAGACG ATTCGTTAAG TCCCAAATAG CGTCTGGTAA   12960
TAACTCCTTA TGTAAAGAAT GTTGGGTTAA TTCnTCGACA TAATTTTTAA AATACGCATT   13020
CATATCTTTG GAAATCAATT CTTCTAAATC AACATTTTCT AAGTCAGCTT CGGACAAGTC   13080
```

AGCGACTTTT TCTyCGATTT CaTTGTAAAC TTGCATGTCT CTAGCTGGAT CTTGTGACCA 13140

AACAATTTTG TTGGTGCCTG GCTGGAAAGA TAAATAATTA CTTTTACTAT CGACAAATTG 13200

ATCCAAGCGC TCAGGTACTT CTTTAATTTT TAATAATCCT TTTTGGACAG GTAACGACAA 13260

ATCCCGTTTA CGGATAACCA AATGTTCTAA ATTGTGCGTC CGATAATAAA GGAATGACTT 13320

AGCACAAACT AATTTTAAGT CCCGCTTCAC CTGACCAATA TTTCCTTCTG CATCGTATAA 13380

CATAAAAGCC AAAACTGCTT CGCGTTCGAT ATCAATTTGT TGGCCCAAAC GTGTTGCCTC 13440

TTGCTGAATA AATAACGAAA TGATGGCATA ACGCTCATCT AGCGTCCGTG CATCTAAAGA 13500

TGGTAATGTA ATCGACATCG GAATCCGCCG ATTAAAAGTG GTTAAAAAGT TTTCTGAGGA 13560

TTCCGTTGTC GCACCAATAA TTTGAACAGA GGCTTCATAG GTTTGACCAC TTTCTCCTAA 13620

TGGACGGTAC ACGCCTTTAT CAATAAACGT AAAGAGCATT TCTTGACCTT CTGGCGGTAA 13680

GCGATGAATT TCATCTAAAA AGAGAATCCC GCCATCTGCT TTTGCCATTA AGCCTGGACT 13740

ATCTTCATTC GCACCAGTAT AGGCACCTTT TTTTACGCCA AAGATATGAC CAAATAATAA 13800

TTGTGGGTTT TGTGCATAAT CAGCACAATT AAAAGAGACA AAAGGGGCAT CTTCATCTAA 13860

GGTTTTTGAA TCAATGGCAA AACGGTACAT ACATTCAGCA AAAAGAGATT TCCCTGTTCC 13920

TGTTTTCCCA AAAATAATCG TATGCAAACC TCTTGGTGGA TATAAAATAG CCGCCTTAGC 13980

CTGCTGAATA CTCACTTTTA AAGAATCATT GGCACCTACT AATGTGTCAA AACTAACAGA 14040

GGTGGACTCA TCTACATGGA CTTGTTCTCC TTCATTGATT GTTGGCACTG AATAGATAAC 14100

AGGACGTCCA GTACTTCTAG TAATTTTCTT TTCTTTGAAT AATTCATTTA AATAACGGCT 14160

GGCATTACTG CGATCAATCA TTAACTGTTC TGCTACCGCT CCTGCCGTTA AGCCAGTTTC 14220

ATTTTTTTGA AGTAACGCTA AGATTTCTTC CTTCCGAGAT TTCATGGAAT CCTCCTCGAT 14280

ACAACGTTCT GCTTGArGCT CGTAAAGCGT TATTTCTTCG CTTTACGAGC yCCTTTGTTT 14340

TTCGTGTTTT TCTTTTTGTT TTTCTTCTTT yTGCGTGTTT CTTGTGGTTC TTTGACAACA 14400

GAATTTACTG TTTCTGGTTT TCTTTCTTTC GTACTCTTTT TTTCTGAATA CGTATCTTTG 14460

ACTACGTTCT TTTTTTCTAG TTCATCTTTG GCTTTTGGCT GTTCATGAAG TGCACCATCA 14520

TATAAAAAGA TTTCCTGATA AGGAATGGCC AATTGTTTCA TCAAACGTTG ATAATCCCTT 14580

TTGGTGCCAT CGTTGATAAA CGTAATCACT GCCCCTTGTG CACCCATCCG GCCAACGCGA 14640

CCAGAACGAT GAATATAGCT CTCTTCGGTT AAAGGAACAT CTACGTTTAC GACGTAAGGG 14700

ACCCCAGTGA AATCTAAGCC CCGTGCCGCA ATATCCGTGG TTAATAACAA CTTCACTCGT 14760

TCTGCTTTAA ATTGATCGAT CGCCAATTTA CGCAACTGTT TACTTTGATC TGAAGCCAGA 14820

CCAATAGCCG GAACATTCTC ATAAACTAAT TTTTCTTCTG CTGCACCTAA ATCAGCAACT 14880

TGATTGAAAA AGACCATGCT TCTGAATGCT TCTGTATGGG CCAAACGACG TAAGTAATCA 14940

GCTTTTTTTC TTGGTGCTAA CCGAATAAAG TAGTGGGCGA CTTGGCCTGC TGAAGTATCT 15000

TCTTCCGTCA CATCAATCAC CTGTAATTTT TGGCTTAAAG ATTGCGCTTG ATTCACGACA 15060

CGATCAGCAG TCGCTGAATA AAAAACAAGT TGGTATTCGG TTGGTGTGTG TGTCAAAATT 15120

TGCTTGGTTA AACTTAGCTC TTCCTCTTGG AAAAGTTGAT CAACTTCATC CATTACGATT 15180

GTTTTTAACA ACTGAGCTTT GACTTTTTTA TTCTTCATTA ACTCTAAAAT TCGACCTGGT 15240

GTTCCGATTA ACACTTCGGG ACGTTTTTTT AACTTGTCAA TCTGGCGTGA AACATTTGCC 15300

CCGCCGATTA ACGTTTGAAC CTTTAATTGA AGTGGCTTAG CCCAAGTCCG AGCAACTTCA 15360

GCAATTTGCA TAGCTAATTC TTGAGACGGA GCAATAATCA ACAATTGATT TCCCTGTCCT 15420

TTTTCAACAG TTAAGAGTAA AGGCmGCmTG TAAGCAAGTG TTTTTcCmGT TCcTGTCGGC 15480

GAAATTCCTA AAACATTTTC TCCTtCGCGA AGGGGCGTAA AACTTTTTTC TTGAATTGTT 15540

GATGGTTCGT TAAAGCCAGA AGCTTGCCAC TGTTCTTGCC AAGCCTCTGG AAGTTGTTCT 15600

AAAAATCTCA TATTGTTCCT TTCCTATCGC TTAATTCTTA TTTATCAGCA TCAAAAATAA 15660

TGCCGGCACT ATGACGTAAA TCAGTGACTA CTTTATTGAC ATTTCGAGCA AGCTCAACCC 15720

ATTCTTCATA CTTGGTGCCC CAATACTTAT CTGTTGGGTT ATTTAAAACT GCTGCAAAAT 15780

CACGGGCTTC TTCAATCATT GGATTATCGG CTACTTGAAT ATCTAAGACA TCACGTTCTT 15840

CATGTGCGCG ATCATGGTAC TCAGCCGAAG AAATAGCATT GACACTATCT AAAATCAAGG 15900

TCCCGTTTTC AAAGTAAATT TCGGATGGCG CAAAAGAATC ACTAATTTTC CCAGTGTTTA 15960

ACGTCACGTC AAACTGATCA TATCGTAAAA TAATTGTTCC TAAGCCATCG ACACCTGTCG 16020

CAATTTTACT AGCAAAATAA TGGCTTTCGT TTGGTACGCC AAACCAGCCT AAGGCTGCAT 16080

AAACTAAATA aACGCCTAAA TCCGCTAAGG CTCCCCCAGA AAAATGAGgA rAAAAtATTT 16140

GGCTCTTTGC CTTCTAATAC TTGATCATAG CGAGAAGAAT ATTTCATATA CGTAAAATTC 16200

GCACCTAAAA TTTGGTTTTT TAATGGCAAT AATTCAGCAA TTTTTTGGAA ACTTTGCTCA 16260

TGAATATTAC GGGCCGCTTC AAAGAAATAC ACACGATTTT TATTCGCTAA TTCAATAATT 16320

TCTGCCATCT CATCAGGCGT TGAAAAAGCT GGTTTTTCAA CAATTACATT TTTTTTCGCT 16380

AAAATTGCTT GTTTCGCTTG TTCAAAATGT AATGAATTTG GTGAGGCAAT ATAAACCGTA 16440

TCCATATGTG CAATACCAAA AAACGTCTCT AAATCAATTG CATATTCCAC ATCCCCATAG 16500

CGAGAACCAA ATTCTTGTGC CGTTGCTAGT TTCCGAGAAT AAACAGCCGT TAAATCATAC 16560

GCACCCGTTT CTAAAGCTGC TGCAACAAAT TGATGACTAA TCCAATTGGT ACCAATTACG 16620

CCTAAATGTA TCATAATCAA AAACCTCCAT TTTTTGTTTA TTTACTCTCT TTATAATAGT 16680

AATCTCTCGT ACAAAAGTCA CGTAATTACC CTGCCTTAAA GCGCTAGGAG TTCTTCAAGT 16740

TATTTATTAG GATCATCAAA TGCCTGTAAT AAATCAAAGG TTACCTGTGA CTTAAGTTCA 16800

ATCGTTGCTG GTGTTACGTG GCAATTGTAA GCAAGGACTT GCACACCTTG ATCAATTGCA 16860

GCACCCACAT TTTCAGCAAA GGCGGGTTGC ATCTCTTGGT GAAGCGTTGC TTGCTTAATT 16920

TCTTCAAATT GGACCACAAA CAAAATGTAA CAGCGGTAAC CCGCTTTAGT TGCAGCCATT 16980

AGTTCAGTGA CATGCTTCAA GCCACGAAGT GTTGGCGCGT CAGGAAAAGC ACCGATGCCT 17040

```
TTATTTTCTA AAGTCATGCC TTTCACTTCG ACAAAAGCTT GTTCGTCCGC ATCCGTTTCC    17100
ACTAGAAAAT CAAATTTTGA ATGTGCAAAT CGTTGCTCAC GTTTAACAGT TTGGATCGTA    17160
CCAACTAAAC CAGGAAGCAC AATTTGACCA TTTTTTAACG CTTCATTGAC TAATGTATTC    17220
GGTACTTGGC TATCAATATT AATCCAAAAA GAACCTTTTT TTACTGCAAT CAAATCGTAA    17280
TCTGTTTTTC TTTTAGGCGA TGGTTGATAA CTAAGCGCCA CAACAGCACC TGGTAAAAAA    17340
ACTTCTTTCC CTCTGCCAGT ATTTTTGACA TGTGTGATTA CTTCTTCATT CGTTTCCATT    17400
AAACGACAGT GTGCAATAAA CCGATTGGGT CGTTCGATAA AATAAGCAAG TTGGACATTC    17460
GGGTACTTCA TACACAGACT TCCTTTCTTA CACACTATAC CATAAAAAAA CAAGGAAAAG    17520
TATAGACCAA AACCAACAAT TTTACTATCT TGTGTCAATC ATTAAATATT TGTCAATTCT    17580
CAGCGATTTA GTCGAATATT TTAGACAAGT TCTATATAAT AAAGGTAACG AAGACGAATA    17640
AATGAAAATG AAGGTTAGGA GGAATTTCTT TGAAAAAACA TTGGACTTTC TTGGTTGGTG    17700
CATCTTTACT TACTGGATTA GCAGGCTCTA TCCTCTGGAA AAAAAATAAA CAGCTCCCTT    17760
CCCCACAGGA GCTTCAACAC AACTATACTG GGCATTGGTG GTTTGTGGAC GTGAAGAAAG    17820
CCACGCAACA CACCTTAGAA ATCACCTCAT CTTTCAATGT TTTCATTGAT GGCAAACAAA    17880
CGCATGCCTT GCTTGTTGAA TTAACCCCTA AACGCTTAGT TATCTTAGAC AAGTTAGGTT    17940
ACCATTTAAT CATTCAATGT GAAAATGGTG TCCCTGCGAC GCTTTACGAT GAAGCCGACG    18000
ACGAAACCTA CACACTAGAA CGGGCCATTT ACAATTAATA AATCACCCAC GACACATAAA    18060
ATGAAGTATG TGCGGTGGGT GATTTATTTT TTTACATCCC TTGCAAAAGC ATTCAGAAAG    18120
TTTTCTTACA ACTTGAATAT TATTAAATAG TCAAATGCGC CTTTTTCTGA TAAACTTAGG    18180
ATGATGTTTT TTTGAAAGGA TGAGCCCCTA TGACACCAAA TCGCGAAGAC TATTTAAAAT    18240
TAATTTTTGA ATTAGGTGGC GACGAAGTTA AAGTGAATAA TAAACAAATT GTTTCTGGAC    18300
TCGATGTTTC GGCAGCTTCG GTTAGTGAGA TGATTTCAAA GTTAGTAAAA GAAGATTTGG    18360
TTGAGCATTC TCCTTATCaA GGGGTACAAT TAACTGAAAA AGGCTTAAAA AAAGCGAGTA    18420
CGTTAATTCG CAAACACCGA ATCTGGGAAG TCTTTTTAGT AGAGCACTTA AATTACACTT    18480
GGAATGATGT GCACGAAGAG GCAGAAGTTT TAGAACATGT TACTTCACAG ACGCTTGTGA    18540
ACCGTTTAGC GGATTATTTA AATCATCCAG AATTTTGTCC ACACGGTGGT GTTATTCCCG    18600
AAGATAATCA ACCCATTCAT GAGGAGAAAC GCCAAACGTT AACAGACTAC CCTGTTGGCA    18660
CAAAAATTCG GATTGCACGT GTCTTAGACG AAAAAGAATT ACTGGATTAT TTAGTTTCCA    18720
TTGATTTAAA TATTCAAGAA GAATATACGA TTAAAGAAAT TGCTGCATAT GAAGGACCGA    18780
TCACCATTTA TAATGAAAAC AAAGAATTAT CCGTCAGCTT TAAAGCAGCA AACACAATTT    18840
TTGTTGAGCC GTTGATTAGA GAAAGTGAGG AAAACTAATG AATGAAAAAG TAGCATTATT    18900
TACTATTTTC GGGGGGACAG GTGATCTTGC CAAACGTAAA TTGTATCCTT CCCTGTTCCG    18960
CTTATATAAA AAAGGATTAT TAGCTGAGCG TTTTGCGGTG ATTGGTACCG CACGTCGGGA    19020
```

ATGGACAGAC GAATACTATC GTGAAATCGT CCGTGAAACC ATTCAAGACT TAAATCCCAC 19080

GGCCGAAGAG GCAACTGAAT TTTCTAGTCA TTTCTATTAC CAATCACATA ATGTGAATGA 19140

TACGGAACAT TATAATACCT TAAAAGAATT ATCTGATCGC CTAAATGAAC AATACCACTT 19200

AGAAGGCAAT CATGTTTACT ATTTAGCTAT GGCCCCTCAA TTCTTTGGAA CCATTGTTAA 19260

TCACTTGAAA TCACAACATA TTATTTCGGA AGAAGGCTTT GATCGTCTAA TTATCGAAAA 19320

GCCATTTGGT TCTGATTATG AATCTGCCTA CGAATTAAAT GAAGAAATTC GTGCTGCATT 19380

TCCTGAACAA GATATTTTCC GAATTGACCA TTACTTAGGC AAAGAAATGA TCCAAAATAT 19440

TTCAGCCATT CGTTTTGCCA ATAATATTTT TGAATCACAA TGGAACAATC GTTACATTGA 19500

TAATGTCCAA ATCaCCTTCG CTGAAAGCTT AGGTGTCGAG GATCGTGGCG GCTATTATGA 19560

CCACAGTGGC GCACTAAAAG ATATGGTTCA AAATCATATT TTACAAGTCG TTGCATTGCT 19620

TTCTATGGAG CCTCCTGTTG CCTTTTCTGA AAAAGAAATT CGGACTGAAA AAGTAAAAGC 19680

TTTACGTGCC ATCCGCCTTT ATTCTGAAGA AGAAGCCCTT CAAAATTTCG TGCGTGGACA 19740

ATACGGCGAA GGTCAGCTAG CTGATCAAAC ATTTGCAGCC TATCGTGATG AACCAAATGT 19800

GGCAGAAACT TCATCCACTG AAACTTTTGT TGCTGGTAAA TTTTTAATTG ATAATTTCCG 19860

TTGGTCTGGC GTCCCTTTCT ATGTTCGCAC AGGTAAACGA TTAACTGAAA AAGGCACACG 19920

CATTAACATC GTCTTTAAAC AAGTACCTAT TAATGTCTTC AAAGATGATA CGTGTGAAGA 19980

ATGCGATAAA ACCGATTTAC CACCGAATGT TTTAACTATC TATATCCAAC CAACTGAAGG 20040

ATTTTCATTA ACCTTAAATG GAAAAGAAAT CGGTCAAGGG TTCAATACAA CTCCAGTGAA 20100

ACTTGATTTC CGTCAAAGTG CTGAAATGAC TGAAAATAGT CCAGAAGCAT ACGAAAAACT 20160

TCTATTAGAC TGCCTAAATG GCGATAGTAC AAACTTCTCT CATTGGGACG AAGTTGCCCA 20220

ATCTTGGCGC ATTGTTGACA TTATTCGACA TGCTTGGGAT AAAACAGACG TTTCTTTCCC 20280

TAATTATGCC GCTGGCACAA TGGGCCCGCA AGCTGCCTTC GACTTATTAG AAAAAGATGG 20340

CTTTACTTGG GAATGGCAAC CAGATAACTG GTACCGTGAC CGCGGTCAAC TTGATCAATA 20400

AATAAGCCAA AAAAGATCCG CTAGTTAACG ACTAGCGGAT CTTTTTTCTT GTTTATTCGA 20460

ATCCTTCAGC AACCGCTTCT GGATAATTTT CTTCAACAAT GTGTGCACAA CGGCCaCAAA 20520

GTGTTGGTAA TTTTTCATCT GAGCCGACAT CTTGACGAAC GGAACGACAA CGGTCACACG 20580

TTTCGCCTTC TGCTTTTTCA ACTAAAATAG CCATGTCTTC AAATTGAACA GCTTCACTAG 20640

GTGCTACTTC TTTTGATACT TCAAAGTCGG AAACAATTAG TAATTGAGCA ATATCTGCAT 20700

CAACAGCTGT CATTAATTGA CGAATTTGTT CATTTGGATA AACGGTCACT TTGGCTTCTA 20760

ACGATTTACC GATTAATTTA GAATGACGAG CTTCTTCCAA CGCTTTTAAG ACATTGTCAC 20820

GGAAATCCAT AAAGGCTGCC CATGTATCCA TCAATTCTTC TTCATTAGTA AACGTTTCGT 20880

AACCTGGGAA TTCAGCTAAT TGCACATACT CTTCTTCTTC TTGTAAGAAA CTCCAAATTT 20940

CTTCCGCTGT ATGTGGAATA ATTGGTGTCA ATAATTTTGT TAACGAAACC AATGTTTGAT 21000

AGAAAACAGT CTGCATACAA CGACGTTGAT AATCGTTTTC AGCTTCAATA TAAACGACAT 21060

CTTTCGCAAA GTCTAAATAG AAAGAAGATA AATCTACAGT TAAGAAGTTC ATAACTGTCC 21120

GATAAATATG CATGAAATTA TATTTTTCAT AACCATTTTC ACGGATTTCT TGAATGACTT 21180

GATTTAAACG AACCGTCATG TATTTATCTA CAGAACGTAA ATCGGCATAA GCTACAGTAT 21240

GCTCAGCTGG ATTAAAATCA CTCGTATTCG CTAACAAGAA ACGCATAGTA TTACGGATTT 21300

TCCGGTACAC TTCAGATACT TGGTTTAAGA TATCCATTGA GACACGCACA TCGGCTTCCG 21360

CGTCAACACT ACTTACCCAC AAGCGTAAGA TATCTGCCCC CATTTGATTG ATAACTTTTT 21420

CAGGTAAGAT AGTATTGCCT AACGATTTAC TCATCTTACG GCCTTCGCCA TCGAGAACCA 21480

TCCCTTGAGA AATAATTGAT TTATAAGGCG CTACTCCATT GATAGCGACA CTTGTTGTAA 21540

TACTTGAGTT AAACCAACCA CGATATTGGT CAGAACCTTC TAAATACATA TCTGCTGGGA 21600

AAGTTAACTC TTCTCTTTCT CGTAAAACTC CTTCATGTGA AGAACCTGAG TCAAACCAAA 21660

CGTCCATGAT ATCTGTTTCT TTGGTAAACT CGCCATTTGG TGAACCTGGA TGTGTAAAGC 21720

CTGCTGGTAG TAACTCTTTT GcTTCACGCA TAAACCAGAT ATTTGATCCA TGTTCTGCAA 21780

ATAAATTCGC TACATGTTCA ATTGTTTCTG GCGTAATGAT GGCTTCCCCA TTTTCAGCAT 21840

AGAAGATCGG TAATGGTACA CCCCACGCTC TTTGGCGAGA GATTACCCAA TCACCACGGT 21900

CACGAATCAT ATTATATAAA CGTGTTTTTC CCCAAGGAAT TAGCCAGTCA ACTTTCTCAA 21960

CTTCATCTAA AATATCTTGA CGGAATTTAG AGATTGAAGC AAACCATTGT GGCGTTGCAC 22020

GGTAGATAAC TGGTTTTTTG GTACGCCAAT CATGTGGATA ACTATGCGTA AAGAAATCTA 22080

ATTTTAATAA CGCGCCTTTT TCTTCTAATA ACTCAGTAAT CATTGGATTC GCTTTGTCAT 22140

AGAAAATTCC TTCAAAACCA GGTGCTTCAT CTGTAAAGAC CCCACGGCTG TCAACAGGAG 22200

AAATAACAGG TAAATCATAT TTACGGCTAA CGATATAGTC ATCTTCACCA TGTCCTGGTG 22260

CGGTATGAAC TAAGCCAGTC CCGGCATCTA GCGTCACGTG GTCGCCTAgC ATAACAAGTG 22320

aTGTTCGATC ATAGAATGGA TGTTGAGCTG TCATACGATC TAATTTTTCA CCAGAAAATT 22380

CTCGCAAGAC TTCTACTTCT TsCCAACCAA TCGCTTCTTT AACTGTTGTT AATAAATCTT 22440

TAGCAATTAC AAATTaCGAC CGTCTGCCTT CACTTCAACA TACGTAAAAT CAGGATTAAC 22500

TGAAATACCT AAGTTGGCTG GTAATGTCCA AGGGGTTGTC GTCCAGATGA CGAAGGCTGT 22560

CTCGTTATCT AAAAGACCTT TACCATCGGC CACATTAAAA GCTACGTAAA TAGAAGGTGA 22620

TTTTACATCT TTGTATTCAA TTTCTGCTTC TGCTAAAGAA GATTCACTTG AAGGAGACCA 22680

ATAAATTGGT TTTAAACCTT TATAAATATA GCCTTTTTCT GCCATCTTAC CAAAAACACG 22740

AATTTCTGCT GCTTCATACT CTGGATCTAA CGTAATGTAT GGATGTTCCC AATCGCCTGA 22800

CACACCTAAA CGTTTAAAAT CGTTACGTTG TTTATCGACT TGTGATAAGG CATACTCTTT 22860

ACATTTTTCG CGATACTCAG CGACAGTCAT TTCTTTACGT TTAACCCCTT TATTGGtTAA 22920

CACTTGCTCa ATTGGTAAAC CATGAGtATC CCAACCAGGc ACATAAGGAG AACGGAAACC 22980

AGACATTGAT TTTGaACGaA TAATAATATC TTTACTGATT TTATTCAAAG AATGTCCTAA 23040

ATGAATATTT CCGTTTGCAT ACGGAGGGCC ATCATGTAAA ACAAAGGTTG GTTTTCCTTC 23100

GTTTAATTTT TGACGTTGTT CATATAAGCC TTTTTCTTCC CAGTCTTTTT GCCaTTCTGC 23160

TTCACGGTTT GGCAAGTTCC CACGCATTGG AAAAGCTGTT TTTCCTAATT GCAATGTTTC 23220

TTTCATTTTC ATATGCATTT GCCTCGCTTT CATTTTATCG ATTTAATTGT GTTAGTTCAT 23280

ACTAAAGATG AACGATAGCT GATAAAAGTA GATAAAAAAA GAGTACTCCa TCCCaAGGGA 23340

CGAAGTAcTC GTGTTACCAC CCAAATTCcA AAGTAGAGTT GCCCCTACTT TCTCTGAAAT 23400

TGATAACGGT AATTATCCGG CAGTTCCTAC TCAAGTGTTC AGAACTGCAT CTAGGATAGA 23460

GGATCTGTTT AATGCCAGTG TCTGTCAAAC TTTCACTATC ATTGACTCGC TGCTAAGAAT 23520

AAACATTAAA CGTTCTGTTC TATCTATTTT GATTCTTCTT CAATTGTTTC TTCAGTGACT 23580

TCAATTACAG GCTTTTTTTC TACTACAGCA TCTGTCTTCT CTTCAGAGTT TACAACAGTT 23640

TCATTTTCGT TGTCAAGGTC TTGTTCATCA AGAATTTCCT TAACTGCTGT ATGTTTGTCT 23700

CCGACATAAC TAGAAAAAGG CTTTAAAATC TCTTCCCACT CTTCACTCTT TACTTGcTCT 23760

AATTGAGTTT CAAGCATTAA GCTTAAGCGT TGaTGGAAGA CACGTGTTTT CTTCTTCAAG 23820

TCTTCTGTTT CCCCAGCTAA TTGACGGGCA CGTTCAATTG CTTCTGCTAA GATTTGTGTT 23880

GATTTCGCTT CAGCATCTGC AATCATTGCA TTTGATTTAC GTTCAGCTTC TACCAAGGTT 23940

TCTTTGGCTT GGTTATCTGC CGAAGTGATA ATCATTTCTG ATTCTTTATT GGCACTTGAT 24000

TTCACTTTAT CAGCGGTATC TtGTGCAACA ATAATTGATT GATTCAATGC ATCTTTCAAT 24060

TCATtGAAGT ATtGTAATTT TTCTtCTGCG TGTTTTAAtG AwTTCTCAAG TTCACGATTT 24120

TTTTGTAATG CAtCTTCATA ATCACGTGTG ACTTGATCTA AAAAGTCATC TACATCGTCT 24180

TGGTTATAGC CTCTCATTTT TGTTGAGAAG tCTTTGTTTT GAATAwCTAA tGGAGTTAAT 24240

GCCATATTCC TTCACCTCTT TGGATTATTT TTGTAAAACC CCTTATGAAC ACATTCTCAT 24300

ACTAATCCTG AAATAGCAAA TATCTTACTT TTTTTATTAT ACACAACTTC GTCTGAATGA 24360

ACAGCCAAAA TGGTTTTCTT AAACAAAAGT TCGAGATTAT TTTCTCAATA CCCCTAAGAG 24420

GACACGAAAT TTTTCTTTCT TGGTTTTTCC TTCTAATTCT TGAATTTGAA GGCGCCCAAA 24480

TCCCCGTATC GAAACAATAT CTAATAAGTC TAAAGCAAAA TCTGGTCGTG TCGTTTCTGT 24540

CCAATTGACT TTTACTTTAC CAGATTCGAT CAATTGTTTt GATCGCTGCC TTGAAATATT 24600

AAATACGGCG GAAATCACAC TATCCAGCCG CAAAGAGCTC ATAGTCGTTC GCTCTTCCTG 24660

CCAACCATCC TTCGGTATCA AAATTTGTGT ATAATCCCGT TCTTCTAAGC GAACAGTAAT 24720

TTTACCAATT TTGTCGACTT GGTTAACAAC GAAGCCAGCT ACTTCTTGTG CTATAAAAAC 24780

CTGCCACTTA TCACCATCGG AAATAATATC TCCAAAGGCT TCCcTACGAA CCCCAACATT 24840

CATCAaGGTG CCTAAAATTT TCCCaTGaGA TAaTGTCGCA AaTTTAGATG GATAAATAaT 24900

kTCAAATAAT ACAACACCAA AATCTGTCTC TTCTGGTATG TAATACGCTG GATAAATCAA 24960

GCAGCGTTTT CGTTCTGCTT GCTCATAACC ACCATAAAAA GTAAACGATA AATCACTGTT 25020

TTGTCGAATC AACGTTTCCA AAATATACGC CTGTCTAGGA TCAAGAAAGT CTGTCAAATA 25080

AGGCGCGTAT TGCATCTCCA CTTGCTCTAG CCAATCCCCG ACAGTGTCAA TAAAAGGATG 25140

TTCATCTTTG CGGAAATGTT GATACACATT TGCGTTCATA GATAAATCAT CCTTTCTAAA 25200

AGTTATTAAT AAACAATAGC TCGTAAAATG ACTGCTAAGC CACCCGCTGC TAGGTTAAGG 25260

ACAATAATCC CTACCATTAC GTTAAATCCA ACCATTCCAA TCCTTAAATT CAAACGGTCA 25320

AACAAACTTA AGTATGGTTC ACAAATTCTA GAGAGCAAAC GCCCGAAAGC AGAATCATAT 25380

GCTCCTGGAA ACCATGAAAG TAGCGCGTAA ATCACTAGTA ATCCTGAATA GATTTGAACG 25440

CCTTTGTATA ATAGAGTAAT CAGTGTTGCG ATAATAACAC CTCCCTAAGG CTTAGAAATC 25500

AAAGAATTGT TTGTTGGCTA AAGATTGTGC TGTCGCACTA TCAATTTCCA TATTCGCTGG 25560

TGTGCATAAG AAGATTTCAT TGCCAACGCG CTGAATATCT CCGTCTAAAG CATACACAGT 25620

ACCAGTTAAG AAATCAACGA TGCGTCGTGC TTGGTCTTCT TCTACTAATG TAAAATTAAC 25680

TAGCACAGCT TCTTCAGCAA ATAAATGTTT CGCAATGCTC ATTGCTTCTG AATACACACG 25740

TGGTTCAATA ATAGTAATTT TATGTGTTTT TGCATTGGCG GTGCTTTCTT GTGCTCGTCT 25800

TGTATTCGTA GTCGCTGATT GATTAGAACT ACGCATAGAA ACCACTTTTT TCTCTGTGTA 25860

AGCAGCTTGT TGTGTGTCTC TTTCTTGATG CGCTTCAGCA GAACGATAGC GTGCAGAAGG 25920

TTTTTCAACT ACAGGCTCAC TTTGTTGTGT ATAAGCTGGC TGACTGTAAC TTTCCTGTTG 25980

TTGAACAGGG CGTTGTGGTT TAGGCCGAGC TGCACGACGC GGTGGTTCAT TAACTGCTTT 26040

GCGCTCTTCA TATTCTTCGT AGTTATCATA ATAATCATCT TCTTCGCCAG ATAAGCCAAA 26100

AAAGCTCGAT AACGCATCCT TGTTAAAAAT TGACATAAAT TCCTCTCCTT TACGCATCTC 26160

TGAATAACGC TGTTCCGATT CGTATAAAAG TTGCACCTTC TTCAACTGCG ATTGGGAAAT 26220

CGTtACTCAT TCCCATGCTA AGTTCTGTAC AAGGAGCAAA AGATAAATGT TTCTCATTTA 26280

CGGTCATTTG CAGCTCTTTT AATtCTCTAA ATGTTGCATG TAATACTTCT TCACTTGCCC 26340

CAAAAGGTGC CATAGTCATT AGACCGACGA TTTGAATTTT GTCCAAATCA GCAAGTTCTT 26400

GAATGAAAGG TAAGACGTCT TCTTTACGAA AGCCGTGCTT GCTTTCTTCC CCAGTTATAT 26460

TCACTTCAAC AAAACAACGA ATAACTTTCG TTGCTCTTTT TTGAATTTCT TTGGCTAGAC 26520

TCAGACTATC TAATGCATGA AAATAATCTA TCTCATTAAT AATTAATTTT ACCTTCCTAC 26580

GTTGTAAATT ACCAATTAAA TGCCACTGAA TATCATTAAA TTTTTCCAAT CCAGCTTTCT 26640

TTTCTAGTAA CTTATCTACT CGGTTTTCAG CTAAATTCTG GACACCAAGC TGAACAAGCT 26700

CTGCTGCTTC CTCAATTCCT ACTGATTTAG TGACGCCAAT CATTGTCACA TCAGCAACAG 26760

AGCGCTGTGA GCTTTTGCAT GCTAGCTGAA GTTGATTATT AATATTTGTT AAGTTTTCAG 26820

CTATCATTTC TAATCGATGC TCCTTATCGT TTTCTGCGGA AAAATGGTGG CGTACTTAAT 26880

TCGTCATCAT TGTGATTTGG TTCTTCACGG TGGAATGTGT CAAATTCTTT TTTGTCTACT 26940

TGTTCTAATG AAGATTCATC AACTTTTGGA CGTGTATTTT GTTCTCTGCG AATATCCCAA 27000

TCGCCAAAAG CACTTGCTTC TTCTTGTGAT TTTGGTTGAT CCATTTCCAC AGATTGAGTT 27060

GTTTGTTGCA TTGGTTGAAC CGCTTGTCTT GTTTGACGGT GCGGTTTTCT ATCCTTTTTT 27120

GATTCATCAA TACCTGTAGC GATTACTGTT ACACGAATTT CGTCGCCTAA ATCTTCGTTA 27180

ATTGATGTAC CTAGGATGAT GTTCACATCA CCTGAAGCAG CATTTGTAAC GATATCTGAA 27240

GCATCTTGCG CTTCAAACAA GGTCATATCT AAGCCACCAG TAATGTTTAA TAAGACTTGT 27300

TCCGCACCAT CGATTGATGT TTCTAACAAT GGTGAAGAAA TAGCTTTTTT CGTTGCTTCA 27360

ATTACACGTT CTTCGCCGCT CGCTACACCA ATTCCCATTA AAGCTGTTCC TTGGTTTTCC 27420

ATCACTGTTT TCACGTCAGC AAAGTCCAAG TTTACGTAAC CTGGTGCAGT GaTTAAATCT 27480

GAAATCCCTT GAACACCTTG aCGTAATACA TTATCAGCTT CTCTAAATGC TTCAAGCATT 27540

GGCGTTTTCT TGTCAACGAC TTCTAATAAG CGGTTATTTG AGATAATTAA TAGTGTATCA 27600

ACGTTTTCTT TTAATAAGGC AATTCCTTCA GCanGCAAAc GACCACGTTT TGGACCTTCA 27660

AAACTAAATG GACGAGTTAC TACACCAACT GTTAAAGCGC CTAATTCTTT AGCGATTTTT 27720

GCAACTACTG GCGCAGCACC TGTACCAGTT CCGCCACCCA TACCAGCAGT AATGAAAATC 27780

ATATCCGCGC CTTGTAATGA TTCTGAAATC ACTTGTTCAC TTTCTTCTGc AGCTTTTTGG 27840

CCAACTTCAG GTTGTGAACC GGCACCTAAA CCACGAGTGT ATTTAGGGCC TAATTGAATC 27900

ACTGTTTCTG CTTTTGAATG TTTTAATGCT TGAACGTCTG TATTGGCTGT GATAAATTCC 27960

ACGCCTTTAA CGTTTTCTTC GATCATACGG TTTACAGCGT TTCCGCCGCC GCCACCTACA 28020

CCGATAACTT TAATGACTGC ACCGTTGTTA ATGTTATTGT CTAATGAAAA TTCCATGATA 28080

TCGTATCCTC CTGTTTGTGT GATTAGTCGA AAATGTTCGA GAAGAAGTCT TTGATTTTGC 28140

CTGTCACTTT TTCACCAGAC TCACGTTCGT TGAATTCTTC GTATTCTTCT TGAGGCTGTT 28200

CAGCATATGT GTCATAGCGA ACTTCTTGTT GGACAGCGAC TGATTGTGCT GGTTTCGATT 28260

TTTCACCAGG GATTGCATAT TTGGCAATGT GATAAATATC GTTTAGTTGC GCTGAATATT 28320

CCACAATGCT AATGACATTA GCAAAGACTG GGTTACGTAA GCCCATGTGA TTTGGTACGT 28380

ATAGTTTTAC GTTGGCTTCA AAAATCTCTT GCGCCAAATC GACAATACCT GGCATACTAG 28440

CGGCGCCACC AGtTAACACA ACGCCTCCAG GTAATTCmAA CGCATCAATT TCATCCAGAA 28500

CTTCTTTTGA TTTTCTCAAA ATTTGTTCCA CACGCGCTTC AATAATTTCT GAAAGGTAAC 28560

GTTCGTCCAC TCTTACAGGC TCTGACTTAC CGATAACATC AACAGGAAAC TCTTCGTTTG 28620

CTGAAGTTCT TTCTGGATAA GCATCCCCAT AATTAATTTT TAAGGCTTCG GCGTTGTTAA 28680

AAGAAGTATT CAATACGATT GAGATATCTT TCGTGATGAA CTCGCCGCCT TCTTGGTTAA 28740

CATGAGTAAA TTTCAATTGT TTGTCGTGGg ATAACAGATG TCGTTGtTTG aCCGCCACCC 28800

aTATCaATGa CAATkGtTCC aAAGTCTTTT TCGcCATCTG tTaAAATTGT TTCCGTTAAA 28860

GCAAGCGGTG tAATCACTAA TTCGtTAATT CCTAAgCCAG CTTTTTCAAC ACATTTACGG 28920

```
ATGTTATGGA TGaTTGTTTT TGGtCCAGTA TAGACAACGC CAAACATCTC CATGCGAACG    28980
CCAAGCATTC CACGTGGGTC TTTGATGCCT TCAAAGCCGT CGACGGTAAA GTCTTGCGGT    29040
aAGATTGCAA CAATTTGACG TTCAGGAGGA GTTGAACGAA CCAAAGCTGC AGATGCGACG    29100
TTGCGCACAT CCTCATCTGT AATTTCTTTT GATTCACTGC TAACAGCTAT CATTCCTTGA    29160
CAGCTTTCTA CTTCTAATAA ATTCGCTGGT AATCCAACAT TTACGCTTTT AATTTGAATG    29220
CCGGCTTTTT CTTCGGCTTG TCGAACGGCC CGTTGGATTG CTTGAACTGT TTGATCAATA    29280
TCGACGACAA TCCCTCGATT CAAACCGTCT GATTTTGCAT TTCCAACGCC GATGATGTTC    29340
ATTTGACCTT CAATATATTC AGCGACAACT ACTTTGACAG AAGTTGTACC AATATCGAGG    29400
CCTACATACA TTCCTGTTTT TGCCATGAAC GGAATTCCTC CTCCTATTAA TTCCAATATC    29460
TAAGCGGACA AAGGCGTCCG TTTAACTATC TCTTTTTTAG TATTATTCAC TTCTTCCAAT    29520
TTTACCATAA AAATGGGGGT TTTAGAAAGC AGATTATACG AATTTGTTCG TTAATTTTTG    29580
AAACATTTTT GTAAAAAACT TATTCTGCTT GTTGCACTTC TTGATTTTCT CCATTTGTCG    29640
GCGTACCGTT ATTCGCTTCA TCCGTTGCTA CGCCTGCATT TGGGTCCACA ACTTCTTGAT    29700
TTTCAACCGC TGAATCTTCG GAAACTTCAC TACCCGTTTC TTCACTTTCT TTTTCATACG    29760
GATAAGAAAA AATGCCGACT TCCATATCAA CGATTCCAGG TTTGTCCATT TGCTCGGCTA    29820
CTTGTGAATA ATAAGCCATT TTTTCACTTA AATCAGAAAT ATTAACAATC ACTCGATTTC    29880
CATCATTCAT ATAGACATTA ATCAAGTCTT TATTGGTTTT GGATGGTGCA TATTTAATTT    29940
CTGAAATTCC TTGTTTAATT TCTTGAGGTA ATTTGTTGTA AGAGGCCATT AATTCAGGGA    30000
TTAATTTATC CTCTTTGAAA TTCTCAAAAA TTGGTTTGCC ACTTTCAGCA GCTTTGGTGG    30060
TTTCCGCCAA AGTTTTACCA TTCTCTAAAA TCGGATGATA GCCGCCTTTA GTCGCAGCAA    30120
GAGCGACAAT TTGATACTCT TGAATATCAA TTTTAAATGA ATTAATACCA TTTAACTTAA    30180
TGTTCGCCTT TTTTATAATT GGAAACTTTT TCTGAATATT TGCAGAAAAA TAATTTCTAT    30240
TTGAATACTG CTCCCATAAG CCACTGCCCG TTTCAAGTTT TGACTGCTGA ATAATCGCTT    30300
GGCTCTCCAC ACTTTTGTTA CCTGAAACTG TTACTTCACT CAATCGACTA AGAGGCGAAA    30360
CAAAATACAA TGCCACTAAA ATGGCTGTCC CTAAACAAGT TAAAATAAAG GCTAAACGCC    30420
TATATAAGCG CTTGTTCCGT GTTTTCTTAA CGTTCGGTAA ACGATCTGCA AAATGCTCCT    30480
CTTCTTCCAC TGTTTCTTGT GGAAGATGTT GTTGAGGCTC TTCTTGACTT TGTTCTTGTA    30540
AAGTGATGGT GACTTCTGCT TGTTCGTTTT CACCTTTTGA AGCAGCTTCT TCCGCTTGCT    30600
TCTTCAAATA TTCTTGGTTT TGTTTTTGCC AAGGTGTCAG TTCAGGTTGA GACTCTTCCG    30660
ATTTTTTTGG AGGTAATGGT TTTCTTTTTT TGAATATATC GTTCTCGTTA CTAATCTTCC    30720
ACACCTCCTA TCTCAGACTA ACGTTTTCAC AACTTGATAC AAACGATCAC TTGCATCTGG    30780
GATGCCTTCC CCTTTGGACG CCGTTGCCAT CTGTTGTCTT TTTTCATTGT TTAATAAAAT    30840
ATCATCTATT GCCGCAACTA AACGAGCGCC GGTCAATTCT GCATCTGGAA TCATTTCAAC    30900
```

AGCACCAACT TTTACTAAGC TTTGCGCGTT TTTCGTTTGA TGATCATTCG TGACGTAAGG 30960

ACTCGGAATT AAAATTGCTG GTAATCCTAA TGCGGTGAAC TCCGCAATTG AGGTTGCTCC 31020

TGCACGTCCC ACCATTAAAT CGGTATTAGC CATCACTTCA ACCATTTTAT CAATATACGG 31080

TTGCACACTT ATGTTTGTCA ATTTCTTTTC AGAAAACTTT AGACTTTCTT GAAGTTCCTG 31140

ATAATATCTT TCTCCTGAAG CGTATAAAAC TTGATACTCT CTCTCTTCAA ACAAAGGAAA 31200

GGCCTGTTCA AACGCTTGAT TGATTTTCAA GGCGCCACGA CTTCCGCCAA ATAAAACCAC 31260

CGTTTTCTTT GCTGGATCCA AGCCAAACTC TGATAAAATA GCTGACTTTT CCACCGTTAC 31320

GACTTCTTGT CCCCGCGGGT TCCCTGTTAA AATTGTTTTT TCTTTCGGGA AAAAACTAGC 31380

GACATCCGGA AAACAAATCG CAATTTTATC CACATAACGA CTTAAAAATT TATTTGTCAT 31440

TCCGGGGATA CTATTTTGTT CATGAATAAT TGTCGGTATT TTTAATTGGT GGGCCGCATA 31500

AACCACTGCA CCAGAAACAT AACCGCCTGT ACCAATGACC ACATCTGGTT GAAATTCACG 31560

AATAATTTTT TTTGCTTTAT TAATGCTTGT TAAAAATAAA TAAATTGTTT TAAAATTTTG 31620

AGGACTTAGC GAACGCTTGA ATCCTTGAAT TTTAATCGTT TTAAATGGAA TTTTCGCCTT 31680

TGGCACAATT TGACTTTCTA GACCATTtTC TGTTCCCACA TACAAGAATT CTGTTGCAGG 31740

TGCTTCTTTT TTGACGTGTT CGACGAAGGA TAACGCAGGA TAGATATGCC CACCTGTGCC 31800

ACCACCAGTT ACTAGTATCT TCATTTTCTC ACATTTCACT TTCTTTTAAT TGTTGAACAG 31860

CTTGCATAAA GGCTTCCCCG CGTACTTCAA AATTCGGGTA TTGGTCCCAA CTTGCGCAAG 31920

CAGGTGATAG TAAAATAGTA TCATCTTTTT CCGAATAATC AAAGGCAATG GTAACCGCCG 31980

TTTGAACATT TTCAGCAAAT AAAATTGTTT CAATGTTCGC TTTTTTAGCA GCTTCCGCCA 32040

ATTTTTCTTT GGTTTCTCCA AACAAAACAA TTGCTTTGAG TCCCAACAGG GCAGGAACCA 32100

ATTCATCAAA TGAGTTACCG CGATCCAAGC CACCCGCAAG TAAAAGTAGC TTTTGGTTGT 32160

CAAACCCACT TAACGCCATC TCTGTAGCTA AAATATTGGT TGCTTTTGAA TCGTTATAAA 32220

AACGTCTTTG CTGAACTTCG CCAACAAATT GCGTTCGATG AGGAACGCCT GAAAAGTTTG 32280

TCAAAGTTTG TCTAATCTGA GCATTCGATA CATTTTTTAA TTTAGCTACA CAAATCGCTG 32340

CGAGTGCATT TTCAATATTG TGACTGCCAG GAATCCCTAG CTCATCGGCG GGCATAATAT 32400

ATTCTTCATT GAAATATAAT TTCCCATCTA AAAGATAAGC CCCTTCTACT GCTTCTTTCG 32460

TTGAAAAAGG CAATACGTTG GCAGCTGTGG TTTTGGCTAA CGTTTGAAGC TCTACTTGAT 32520

TCCAATTTAA AATCAAGGTG TCCTCTGCGG TCATGTTTTT TTGAATGGCC CATTTTGCAG 32580

CAACATATTC TTTCCGCGAA CCATGATAAT CCAAGTGTGC CTCAAAAATA TTTGTAATTA 32640

CTGCAATTTG TGGGTGAAAC GTCTCAATTC CCATTAACTG AAAACTAGAA AGTTCCATCA 32700

CAAGATTATC CTTGGCCGTT GCTTCTTGAG CCACCGTACT CGCTGGAAAA CCAATATTTC 32760

CCGCCAAA 32768

(2) INFORMATION FOR SEQ ID NO: 129:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6259 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

GGCTGGATTT TCCTAGTAAG AGCTTGTTGG AmCAGCTTAA TAAAGACACA AGCCTTTTTT 60

GACGTCGTCG CTCAGGATTT TGCTAAAATG GCCCAGTTTA CGTATCATTT ACAAGAAGCA 120

AAAAAAGTCA CAGAGTATAC AAATTCTTTA AGATTAAAGG ATTTACAAGG AAAATAAGCT 180

GAAAACAACA TAAAAATAGT GTATAATTGA AAAAATTAAT CTGCGAGAGA AAGGGAACCT 240

TTTTTGATTG TAGATCACCG CAGTTGTTTG TTGCCTAGTC TCTAAGAAAT AAACCTAAAA 300

TTATTGATAT AAAGGATGAA CAAATGAAAA AAGAAGAAAT GCAAATGCGT AATACACGTC 360

GTCAAAAATC AGGAAAAAAT AATAAAAAGA AAGTAATTAT TACTTCTTTG GTTGGACTAG 420

CTCTGGTTGC TGGGGGCAGT TATGTTTATT TTCAAAGTCA CTTTTTmCCA ACCACAAAAG 480

TAAATGGAGT TTCTGTAGGC TGGTTAAATG TAAATGCTGC AGAAGAAAAA TTAGCGCAAG 540

TTAATCAAAC CGAAGAAGTT GTGGTTCAAA CGGGGACAAA AGAAGAAAAA ATTCAACTTC 600

CTAAAAAATA CCAATTGGAT CAAAAATTTT TAAAAGACCA TTTACACAGT AGCAAGGTGA 660

AGCTACCGTT AAACGAGGCA TTCAAAAAAG AACTAGAAGC CAAATTAGCA ACTTTGAGTT 720

TTCCAGAGGG GAAACCAAGC AAAAATGCGA GTATCCGTCG AGGCAATGGC ACTTTTGAAA 780

TTGTTCCCGA AGAACAAGGC ACAGTAGTGG ACACACAGCG CTTAAACCAG CAGATTATTG 840

CGGATGTTGA AGCGGGAAAA GGCAACTATC AATATAATGC CAAAGATTTT TATAAAGCCC 900

CTGAAATTAC AAAAGAGGAT CAAACGTTAA AGGCAACATT GACAACGCTC AATAACAAGT 960

TAAATAAAAC AATTACAGTT GATATTAATG GTGAAAAAGT AGCCTTTGAT AAAACACAAA 1020

TTCAAAACGT GCTGAATGAT GATGGCACAA TCAACAAAGA AAAACTAACT ACTTGGGTGA 1080

CACAATTAGA AACAACATAT GGTTCTGCTA ATCAACCAGT TTTATTTACA GATGTTCACG 1140

GCACGACACG TCGTTTTAAA AACAACGGAA GTTATGGCTG GTCGATTGAT GGGGCCAAAA 1200

CGCAAGAACT ACTAGTAAAC GCGCTGAATA GCCAAGAACA AACGAATGCA ATCACTGCTC 1260

CGTTGGTTGG TGATACCAAA GAAAATAGTA AAATTGCCAA TAATTACATT GAAATTGATT 1320

TAAAAGATCA AAAAATGTAT TGTTTCATTG ATGGCAAAAA AATAGTCACC ACAGATGTCA 1380

TTACTGGCAG ATATAACAAA GGAACCGCAA CAGTACCAGG ATTCCATACA ATTTTATATC 1440

GGACAACCGA TGTGAATTTA GAAGGTCAAA TGCTTGATGG TTCTCGATAC AGTGTGCCAG 1500

TAAAATATTG GATGCCGTTA TTAAGTCAAG GGGGCGTTGT CACACAAATC GGGATTCATG 1560

ACTCCGACCA TAAATTGGAT AAGTATGGCG ATAAAGAAGC CTTTAAAACC GATGCTGGTA 1620

GTAATGGCTG TATCAATACG CCAGGAACAG AAGTTTCAAA AATCTTTGAT GTATCCTATG 1680

ACGGAATGCC GGTAATTATT TATGGACATA TCTATGATGA TGCACCAGGT GAATTTGATA 1740

```
AACCTGTAGA TTACGGCGAA GAAGTATAAA AAGAGATATT AGGcAAGACC GTACTTTGGT    1800
CCAGCTGTGT ATCTGAAAAG TAAAAATTGG AGACGTCAGC GTAACGACTG ACGTCCCCAA    1860
TTTTTTATTG ACAAGCGTCA TAACCCGCTC TATGATACAA ACAGAACAAG CAATTTAAAT    1920
TGGGGGAACG AAGATGCCAT TTGTACATGT AGAATTAATT GAAGGCCGCA CAGAAGAACA    1980
GTTAACTAAT ATGGTCAAAG ATATTACAGA AGCTGTGTCA AAAAACGCTG GTGCACCAAA    2040
AGAGAATATC CaTGTGATTG TCAATGaATT GAAAAAgACC GCTATGCACA AGGCGGCGAA    2100
TGGAAAAAAT AAGTTTTCAG AGAATTTCTG TTGAATCACT TGCTTTTTCT GAAAATAAGT    2160
GGTTAAATAA GTGCATAAAT GATTGATGAG AAAGACAACT GAACTTGCCC GCGATCTATT    2220
AGAGAGGAAA ATCTTGGCTG AAAATTTTCC AGATTGGCAC AAGGGATGAC CACTTTCGAA    2280
ACCTTTGTTG AAACAAAGGC GGTCGCACCG TTAACACGCT AGAGGAACAC AGATTATTTG    2340
TGTTCAAAAT TAGGTGGAAC CACGATATTT TCGTCCTAGT ATTCGTTAGA ATACTAGGAC    2400
TTTTTGTTTT TTAACAAGGG GAACAGGTGG CTGTTTCGTT TCATTGTGAA AAACAGGAAC    2460
ACAATGAAAC GCTTCAAATC AAAAAAATTA AAAGGAGAGA ATTTTATTAT GTCTATTCAC    2520
ATTACTTTTC CAGATGGCGC TGTTAAACCG TTTGATTCTG GAATTACAAC ATTTGATGTT    2580
GCTAAAAGTA TTAGCAACAG TTTAGCCAAA AAAGCTTTAG CTGGTAAATT CAATGGTGTT    2640
TTAATCGATT TAGATCGTCC TATCGTAGAA GATGGTTCGC TTGAAATCGT GACACCTGAT    2700
CATGAAGATG CTTTAGGAAT TTTACGTCAT TCATCAGCTC ATTTAATGGC TAATGCCTTA    2760
CGCCGTCTTT TCCCTAACAT TAAATTTGGC GTAGGTCCTG CGATTGATTC TGGGTTCTAT    2820
TATGATACAG ATAATGGAGA ATCCCCTGTG ACAGCGGAAG ATTTACCTGC AATTGAAGCC    2880
GAAATGATGA AAATTGTGAA GGAAAATAAC CCAATCGTTC GTAAAGAAAT CTCACGTGCA    2940
GAAGCGTTAG AATTATTTGC TGATGATCCT TACAAAGTTG AATTAATTAC AGATTTGCCA    3000
GAAGATGAAA TCATCACTGT CTATGATCAA GGCGATTTTG TTGATTTATG TCGTGGTGTT    3060
CACGTCCCTT CAACAGGACG GATTCAAGTC TTTAAATTAC TTTCAGTAGC TGGTGCTTAT    3120
TGGCGCGGAA ACTCTGACAA TCATATGATG CAACGGATTT ATGGCACTGC CTTTTTTGAT    3180
AAAAAAGATT TAAAAGAGTT TATCAAAATG CGCGAGGAAG CCAAAGAACG TGACCACCGT    3240
AAATTAGGaA AgAATTAGAT TTATTTATGG TTTCACAAGA AGTTGGTTCA GGGTTACCTT    3300
TCTGGTTACC AAAAGGCGCA ACCATTCGTC GTACAATTGA ACGTTATATT GTGGACAAAG    3360
AAATTAGCTT AGGTTACCAA CATGTGTATA CACCAATTAT GGCAGATGTG GAATTATACA    3420
AAACATCTGG TCACTGGGAT CATTACCATG AAGATATGTT CCCACCAATG GATATGGGTG    3480
ATGGCGAAAT GCTGGTATTA CGTCCAATGA ACTGTCCACA CCATATGATG GTTTATAAAA    3540
ATGACATTCA TAGTTACCGC GAATTGCCAA TTCGAATCGC TGAATTAGGG ATGATGCACC    3600
GCTATGAAAA ATCTGGCGCA TTATCAGGGT TACAACGTGT TCGTGAAATG ACTTTAAACG    3660
ATGGCCATAC TTTTGTTCGT CCTGACCAAA TTAAAGACGA ATTTAAACGT ACTTTGGAGT    3720
```

TAATGGTGGC AGTCTATGCT GACTTTAACA TTACGGATTA TCGTTTCCGC TTAAGCTATC 3780

GTGATCCAAA TAATACAGAC AAATATTTTG ATGATGATGC GATGTGGGAA AAAGCGCAAG 3840

CGATGTTAAA AGCTGCCATG GATGAATTAG AATTAGATTA CTTTGAAGCA GAAGGCGAAG 3900

CTGCCTTTTA CGGTCCGAAG TTAGATGTTC AAGTAAAAAC AGCTTTAGGA ATGGAAGAAA 3960

CATTATCAAC CATCCAATTA GACTTCTTAT TACCAGAACG TTTTGACTTA ACTTATGTTG 4020

GCGAAGATGG TGAAAATACA CATCGCCCAG TTGTTATCCA CCGTGGTATT GTCTCAACAA 4080

TGGAACGATT TGTGGCTTAC TTAACAGAAG TTTACAAAGG CGCTTTCCCT ACTTGGTTAG 4140

CACCAATTCA AGCAACTATT ATCCCAGTTT CTGTAGAAGC GCATTCTGAG TATGCTTATG 4200

AAATCAAAGA ACGTTTACAA GCACAAGGCT TACGTGTTGA AGTCGATGAT CGTAACGAAA 4260

AAATGGGCTA CAAAATTCGG GCCTCTCAAA CACAAAAAGT ACCTTATCAA TTAGTGGTCG 4320

GGGACAAAGA AATGGAAGAC GCAACGGTGA ACGTCCGTCG TTATGGAAGC AAAGAAACGT 4380

CTGTCGAAGA TTTATCAATT TTCATTGACA GCATGGCTGC TGAAGTTCAC AATTACAGCC 4440

GTTAAAAAAA GAGAnAAAGT CAAAAGGACT TCTGAATTCA GAAGTCCTTT TGACTTTTTT 4500

CAACAACAAA ATAATTAAAT ATTTCCTTAG AGCTGGGAAT AATCCTGATT TTTTGACTAT 4560

TTTTTGTTAC AATGAGAATG CGAAACTTTT TAGTGAAGGG ACTAGAGGTA TGAAGAAAAA 4620

CTCATTTATA GnTrAaATAA AAAAAaCAGA TGCCGCAGCA AAGAGTCAAC GAGGTAGAAA 4680

GCCACATATT CCTTTTCGAT TGAACTTATT ATTTTTTGTG ATATTCACTT TATTTGTTTC 4740

ATTAATTGTC CGTTTAGGGT ATTTACAGAT TGTAGAAGGA GAAGAGTTTA ACAAAAAAAT 4800

CACTGCCAAT TCTTCACTAC AAATCACGAC ACCATCCCCT CGAGGACAAA TTTATGATTC 4860

ACAAGGAAAA GTCTTGGTTT CTAACAAAGC CAATTTAGCA ATCACATACA CACGAGGAAA 4920

AAATATTGAA GGAAAAGATA TTTTACCGAT TGCGAATAAA GTTAACGAAC TTATTAACGT 4980

TCCAGTTGAT CCAAACTTAA CGGATCGTGA CAAAAAAGAT TATTGGTTAG CGAATCCTGA 5040

AAATTTAAAA GCAGCCCAAT CCCGTTTAAC CGATCAAGAT AAAGAAGATG AAAAAGGCAA 5100

CAAAATTACG GACGAAGGCA CATTATATGC CAAAGCAGTT GAAAAAGTCA CCCCTGAAGA 5160

AATTGCGTTT GATGACCGAA CCTTACAAGC AGTAACGATT TTTAAACGAA TGAATGCAGC 5220

ATCACAAATG AACACGGTCT TCATCAAGAA TGAAGGGGTA ACAGAAGGTG AAATTGCGAC 5280

AATTGGTGAA CACACAGCGG AAATATCTGG TGTTTCAACA GGGACTGATT GGACGCGTGA 5340

TTATTCTCAA AGTGGCGCAT TACGTAGCCT GTTGGGCACc GTTTCAACTG AAAAACAAGG 5400

ATTgCCAGCC GAAGAaGTCG ATGanTATTt GAAAAAaGGC tATGCaCGAA aTGACCGTGT 5460

CGGGACTAGC TATTTAGAAA AACAATATGA AGACGTTCTA CAAGGTAAAA AAGCAAAATC 5520

AGAAGTCGTT TTAGACAATA ACGGAAAGAT TGTTTCACAA ACCCCAATTT CCAAaGGGGA 5580

AAAAGGCTCC AACTTAAAAT TAACGATTGA TTCAAATTTC CAAAACAAAG TTGATGAAAT 5640

TTTACAACGA AACTATTCAC AAATAGTCAA AACCATTGGA CCATACTCAG AAAATGCATA 5700

TGTCGTTGCT ATGAACCCTC AAACAGGAGC TATTTTAGCG ATGTCCGGAT TTCATCATGA 5760

TTTAGCAACT GGAGAAGTCA CACCTAATCC TTTAGCACCT ATTTTGAATT CTGAAGTTCC 5820

TGGATCAGTT GTGAAAGCAG GTACTTTAAC AGCTGGTTAT GAAACAGGCG TCATTAAAGG 5880

CAATGACGTT TTAACAGATG AAGCTATTTT ACTAGCAGGT AGTAATCCAA AAGCTTCTTG 5940

GTGGAATGCT TCTGGCGGAA CAACCATGCA ATTAACTGCG GAgCAAGCGT TAGAATATTC 6000

TTCGAATGCT TATATGATGA AATTAGTCTT TAAAATGATG GGTGTGAATT ATTATCCCAA 6060

TATGATTTTC CCTTATGAAG TAGGAGATGA CACTGTCTTT AAAGAATTAA GaAAAGCTTT 6120

TGCAGAATAT GGTATGGGAA CGAAAACTGG AATTGATATA CCAGGAGAAA CAACAGGAAT 6180

TCAArACAAG GACTTTAAGG ATTCGTCATC tGCGCCaCAA GGCGGTAATC TTCTTGACCT 6240

ATCTTTTGGA CAATATGAT 6259

(2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9767 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

CGCTGACTCA ATTAGCGTAA TGGAGTTTGT TTTAGAACTT GAAGATGAAT TTGGAACAGA 60

GATTTCAGAT GAAGATGCAG AAAAAATTGA AACAGTTGGT GCCGCTGTAG ATTATATTGT 120

GAGTAATTCA TAAAAAATAA TAAGACAAAG TCGATTAACA TTCCTGTTAA TCGACTTTTT 180

TTGCGTACAT AATGGTTATT TTATTAAGCA ATAAGGCTAT TTTTACACAT ATTTTTGTAA 240

AAAAAATCAT ACATGTAAAA AAATAACTGC GAAAAAAGTT ATTTTATCTT GCTTTTTATG 300

GAATAATTCT ATAAAATTAT GAAAAAACAA TGAGCTGTAA ATTAACATTG ATAAAAATTA 360

GAAAGTGCTA TCATTAAATA TTAGTTAATT TATATAATGT TTTTGAAAGG ACGTAAGGGG 420

TGCGTATGGA TGCAAGGAAA TGAACAATTA CTAGAGGTTT CAGGATTAGA AACAGCATTT 480

AGAATCAAAG ATGATTATTA TAATGCCGTC GATGATGTTT CATTTGAATT GAAGAAAAAT 540

GAGATTCTGG CAATCGTCGG AGAATCTGGT TGTGGAAAAA GTACCTTAGC AACAACGATT 600

ATGGGCTTAC ATGATCCCAA TCATACCAAA ATTACTGGGG AAATTTTATA CAACAATTTA 660

AACTTAACTA CGTTTAATGA AACATTGTAC AACAAAGTCA GAGGAAACGA TATTGGGATG 720

ATTTTCCAAG ATCCGTTATC TGCACTGAAT CCGCTCATGC GAATTGAAGA CCAAATTAAA 780

GAAAGTCTCA CGTATCATAC CAAAATGACC GCACAACAAA AACAAGCACG GGCCATTGAA 840

CTGTTAGATC AAGTGGGGAT TCCCAACCCA GAGCGTGTCG GGAAACAGTA TCCTCaTGaA 900

TTGTCTGGcG GcATGCGCcA ACGGGTGaTT aTTGcCATTG CAATCGCGTG tAAACCACCG 960

aTTTTaATTG cmGrTGaACC AACAACGGCC TTAGACGTTA CCATTCAGGC ACAAATTCTC 1020

GATTTGTTAA AAGATTTACm AGaAGAAAcA GGGAGcGGCA TTATTTTAAt TACCCATGAC 1080

TTAGGTGTAG TTGCGGAAAT GGCTGATCGT GTAGCGGTTA TGTATGCGGG TCAGTTTGTA 1140

GAAGTAGCCA CTGCTGAAGA ATTGTTTAGT CATCCGAAAC ATCCATATAC ACGTTCCTTA 1200

CTTCAGTCCA TTCCCCAAGA GAATTCAGAT GAAGAAGAAC TTCATGTGAT TGAAGGGATT 1260

GTTCCTTCGT TAACCAAATT ACCTAGAAAA GGTTGTCGGT TTGCCCCACG AATTCCTTGG 1320

ATTGATGCAT CGGAACATGA AGAAAATCCA ACATTACATG AAGTTGCACC GAATCATTAT 1380

GTTCGTTGTA CATGTTATAA ACATTTTCAT TTTAGAGACG GGGAAGGGGA AGCGTAATGA 1440

CAGAAATTAT TCAAATTAAA GATTTAAAAG TTCATTATCC AATTCGCAGC GGCTTCTTCA 1500

ATCGAATCAC TGATCATGTT TTGGCGGTTG ATGGTGTGGA TTTCATGATT GAACAAGGTA 1560

AAACGTATGG GTTAGTTGGT GAATCGGGTT CTGGAAAATC AACGACAGGA AAAGCAATTA 1620

TTGGTTTAGA AAAAATTACA AATGGTGAAA TTATTTATCA AGGACAAGAT GTAACCAAGC 1680

CGCGAAGTCG AAAAGCGATT GGATACAACA AAGATGTCCA AATGATTTTC CAAGATTCGA 1740

TGTCAAGTTT AAACCCTAAA AAACGAGTAC TAGATATTAT CGCCGAACCA ATTCGAAATT 1800

TTGAACGGTT AAGTGACCAA GAAGAAAAGA AAAAAGTCAA AAGTTTGTTG GATATTGTGG 1860

GGATGCCAGA AGATGCGTTG TACAAATATC CTCATGAATT TTCTGGTGGA CAACGCCAAC 1920

GTTTAGGCGT CGCCCGAGCG GTGGCTACCA GCCCCAAATT AATTATTGCA GATGAACCTG 1980

TTTCAGCACT GGACTTATCT GTGCAAGCAC AGGTATTGAA TTTTATGAAA AACATTCAAC 2040

AGGAATATGG GTTAAGTTAC TTATTTATTT CCCATGATTT GGGCGTAGTT AAACATATGT 2100

GTGACAACAT TGCCATTATG TACAAAGGCC GTTTTGTTGA AATCGGTACA CGCCAAGATA 2160

TTTACACTAA TCCACAACAC ATTTATACCA AACGCTTACT TTCAGCAATT CCTAAAATTG 2220

ACGTAGCGAA CCGAGAAGCT CACAAAGAAG AACGAAGAAG AGTTGAGCAA GAGTATCGGG 2280

AAAATCATAA AGACTATTAT GATGAAAATG GGCGGGTTTA TAACTTACAT GCGATTAGTC 2340

CAACCCACCA AGTAGCGTTA AAAAATGGAG GTGCTGAATA ATGTGGAAAA CAATTTTACG 2400

CCGTTTGTTA TTGATGATTC CGCAAGTGAT TATTTTAAGT GTCTTGATTT TCCTGTTGGC 2460

TAAAATGATG CCTGGTGATC CGTTTACAGG ATTGATTAAT CCGAACCAAG ACCCGAAAGT 2520

GATTGAAGCC ATGCGCGAAg cTGCCGGGTT AAATGATCCT TGGTACGAAC AATATTTCCG 2580

TTGGATTGGC AATGCCTTAC ATGGAGATTT TGGGCATAGT TTTATCTTTA AATTACCTGT 2640

GTCCACGCTC ATTGCTGGTC GGGTGGGGAA TACGATTGCA TTGGCCGCTG TGTCAGTTAT 2700

AATCACGTAT TTAATCGCCA TTCCTTTCGG TTTAATTGCT GGCCGTTACC AAAATTCATG 2760

GTTTGACAAA ATGGTGGTTA TCTACAACTT CTTCAGTTTT GCCGTGCCAT TATTCATCTT 2820

TGCCTTGATT ATGCTCTTTA TTTTTGGGTA TCGCTTAGAC TGGTTCCCAA CGAGTGGAAC 2880

CGTGACTGTT GGTTTGGCTG AGGGAACGTG GCCGTATTAT TTAGATAAAC TGAAGCATTT 2940

AATTTTACCA GGGGTCACTC AAGCGTTGCT AGGGACAGCG GTCACTATTC AATATTTGCG 3000

CAGTGAAGTG ATTGATGTTA AAAACATGGA CTTTGTTCGT ACTGCTCGTT CAAAAGGAGT 3060

GCCAACGAAT AAAATTTTTA ATCGACACAT TTTTAGAAAT GCAGCATTAC CAATTGCCTC 3120

TCAATTAGGT TATGAAATTA CTGCTTTGAT TGCCGGATCA GTCGTTATTG AAAAGATTTT 3180

TGCTTTCCCA GGAATTGGAA AATTATTTAT TGATAGTATC ATTCAACGTG ACTATTCCGT 3240

TATTACCGCT CTCGTGTTGA TTTTGGGATT GGCAACCCTT ATCGGAACCT TAATTTCTGA 3300

TATTGTAATG AGTATTGTTG ATCCAAGAAT TCGAATTCAG TAGAAAAAAC ACGAAGAAAG 3360

GGTGATAAAA ATGGAATTGA CAGAAGAAAA AAGAGAAGAA GTCTTACAAG AAAGTATTCC 3420

ACCAATGGGC TTTCGGATGA TTGCGCGAGA ATTCGTCAAA GAAAAAATGG CGATGTTTTC 3480

TTTAATCCTT TTGGTCATTA TCTTATTAGC CGTTTTTATC GGTTCGTTAG TGTTAGACCA 3540

AAGTGCCGTT ATGCATGTAA GTATTTTAGA TAAATATGCG GAACCAGGTA CCGTTACAAT 3600

GAATGGCACC AAGTTCATTT TAGGTGCCGA TGAAGGGGGT CGTGATGTTT TAGGACAGTT 3660

GATTATCGGG GCTAGGAACT CGATTTTAAT CGGTTTTGCC ATTACGATTA TTACGTCTAT 3720

TATCGGGGTC GGTTTAGGGA TTATTTCCGG TTATTATGGA GGCATGATTG ATAACATTTT 3780

GATGCGTATC GTGGATTTTA TTATGATTTT GCCAATTATG TTAATTATTA TTGTGTTCGT 3840

TTCCGTGATT CCTAAATATA GTATCTGGTC CTTTATTTGG ATTATGTGTG CGTTTTATTG 3900

GGTTGGAAAA GCGCGGTTGT TCCGCAGTAA AACGTTATCT GAAGGGCGCC GAGATTACGT 3960

TAGTGCCTCT AAGACAATGG GAACAAGCGA TTTTAAAATT ATGTTTCGAG AAATTATGCC 4020

GAATTTAAGT TCATTAATTA TTACGAACTT AACGATTAAT TTCGCGGCCA ATATCGGGAT 4080

CGAAACAACG TTAACTTTCT TAGGCTTCGG GTTACCACAA AGTGTCCCAA GTTTAGGAAC 4140

CTTAAtTGgc TACGCCAGCA GTGGGGATGT ATTAGTCAAT AAACAATGGG TTTGGTTACC 4200

TGCATCAATT CTAATTTTAG TCTTGATGTT AAGTATAAAC TATGTTGGAC AAGCATTTAA 4260

GCGCTCAGCA GATGCACGAC AACGATTAGG CTAATTAGCC AAAAAATGAG GGAGGAAAAG 4320

AGATGAACAA GAAACGGATT TTAGGTGCAA TCACGTTAGC TTCTGTGTTA GTATTCGGGT 4380

TAGCTGCATG TGGTGGCGGC AATAAAGGCG GGGGCAATAA AGCAACGGAA ACAGAAGACA 4440

TTTCAAAAAT GCCAATCGCT GTTAAAAATG ATAAAAAAGC AATTGATGGC GGTACATTAG 4500

ATGTCGCTGT AGTTATGGAT ACACAATTCC AAGGACTTTT CCAGCAAGAA TTTTATCAAG 4560

ACAACTATGA TGCACAATAC ATGCTTCCAA CGGTACAGCC ATTATTTAAC AATGATGCAG 4620

ACTTTAAGAT TGTCGATGGG GGTCcTGCGG ATCTGAAATT AGATGAAGAT GCCAATACAG 4680

CAACCATTAA ATTACGTGAC AATTTGAAAT GGTCTGACGG TAAAGATGTG ACAGCCGATG 4740

ACGTGATTTT CTCTTATGAA GTCATTGGTC ATAAAGACTA TACAGGGATT CGTTATGATG 4800

ATAACTTTAC GAATATTGTT GGCATGGAAG ACTACCATGA TGGTAAATCG CCAACCATTT 4860

CTGGCATAGA AAAAGTCAAT GATAAAGAAG TTAAAATCAC TTATAAAGAA GTTCACCCAG 4920

GAATGCAACA ATTAGGTGGC GGTGTTTGGG GCTCAGTTTT ACCAAAACAT GCCTTTGAAG 4980

```
GAATTGCTGT TAAAGACATG GAATCAAGCG ATGCAGTTCG TAAAAACCCT GTGACTATTG     5040
GACCATACTA CATGAGTAAT ATTGTGACAG GTGAATCTGT TGAATACCTA CCAAATGAGC     5100
ATTACTACGG TGGTAAACCT AAATTAGATA AATTAGTGTT CAAATCTGTT CCTTCTGCGA     5160
GCATTGTAGA AGCGATGAAA GCGAAACAAT ACGATATTGC ATTATCAATG CCAACAGATA     5220
CGTATCCAAC ATACAAAGAT ACTGAAGGGT ATCAAATCTT AGGACGTCCC GAACAAGCCT     5280
ACACGTATAT TGGCTTTAAA ATGGGTACGT TTGACAAAGA AACAAATACA GTGAAATACA     5340
ATCCAAAAGC TAAAATGGCA GATAAAAGCT TACGTCAAGC CATGGGCTAT GCAATTGACA     5400
ATGATGCAGT CGGCCAAAAA TTCTACAACG GCTTACGAAC AGGGGCAACA ACGTTAATCC     5460
CACCAGTCTT CAAGAGCTTG CATGATAGCG AAgcGaAAGG CTATACGCTT GATTTAGACA     5520
AAGCGAAAAA ATTATTAGAC GATGCTGGTT ATAAAGACGT AGACGGCGAT GGCATTCGCG     5580
AAGACAAAGA AGGCAAACCA CTAGAAATCA AGTTTGCTTC AATGTCAGGC GGCGAAACTG     5640
CACAACCACT TGCTGATTAC TATGTCCAAC AATGGAAAGA AATTGGCTTA AACGTAACGT     5700
ATACAACAGG ACGCTTAATT GATTTCCAAG CATTCTATGA TAAATTGAAA AATGATGACC     5760
CAGAAGTAGA TATCTATCAA GGCGCGTGGG GCACAGGTTC AGATCCTTCA CCAACCGGCT     5820
TATATGGTCC AAACTCAGCC TTTAACTATA CACGTTTTGA GTCAGAAGAA AATACTAAAT     5880
TACTTGATGC GATTGATTCA AAAGCATCAT TTGATGAAGA AAAACGTAAA AAAGCCTTCT     5940
ACGATTGGCA AGAGTATGCC ATTGATGAAG CGTTTGTAAT CCCAACGCTT TACAGAAATG     6000
AAGTCTTGCC TGTCAACGAC CGTGTAGTTG ACTTTACTTG GGCAGTTGAT ACGAAAGATA     6060
ATCCATGGGC AACGGTGGGT GTCACAGCAG ACTCACGGAA ATAAAAAAAC AGATTCTAGA     6120
AGACATAAAG CTCTTCTAGA ATCTGTTTTT TTGGATAGGA TAGTTTCTTA GGCCATCTTA     6180
TTTTATGTGA AAATTTTTGT TACTTCAAGC AATTACAATT TATATGTATG CGTTATTTTT     6240
TTGTTATAGA TGAATTGTTT TTTATGCAGT ATGGTTTTTG ATGGGAGGTT CTGTTTTTTT     6300
ATTTTATTTA GTGTGTATAT CTACATGTTT GTTATGAAAT GTTATTTTTT TTATTATATA     6360
TATAATACTC CCTTTACGTA TAAAGAACGA TAAGTTACAA TTAGATTAAG TTTTAGTTAG     6420
TTTTTGAAGG TGTGTTTGGT ATTGTTTATA ACTAAAACTT ATATTTAAAG AAAGGAGTGA     6480
ATGGATATGC CACTTTATAT GGAGAATTAT TTGCTTGCAT TGCAATCAAA ATATTCACAA     6540
GAAATAACTG TAAAATGGGT TTGGGCAGTT GTAATATTAC TTTCAGTGAT TGGCTGTATT     6600
GGTTATGCAG TTTACTGTAG TTGGGTAGGT GGAACATTTG CTGGGAGTAT TAAAATTGGT     6660
GTTCCCGACT TGGTACACGT TACCTTCAAT TGTAAACGTT GATATTTAAG GAGGAAGGAG     6720
CCTCTATTTT ATGGAGGCTT CTTGTTAAGA AACTTATGAT AAGAATTGAA AATTTAACAA     6780
AAAAATTTGA TAGGAAAGTT TTAGATAATA TAACTATCTC ATTGCCTAAA AACCGAGTTA     6840
GTGTGAnTGT AGGTATAAAT GGAAGTGGGA AAACTACTTT ATTAGATTGT AATGTAGGAT     6900
TGAAAGATGC AACGTCTGGA AAGGTATTTA TTGAGAGCTA TTCTAATGAT TCTGAAAAGT     6960
```

TTAAAGAATG TATATTCTAT ATACCATCGG AATTTTATCT ACCTAATTAT ATGACGGGTA 7020

AAGAGTACTT AAACTTTGTT TTATCAAGAT ATCGCTGTAG TGATATAGAA CGGATAGATG 7080

ATTTTTTAGA ACTATTTGAT TTAAAATTTG CTGGAACTAA TTTAATAGAA TCGTACTCAT 7140

TTGGTATGAA GAAAAAAATA CAAATAGTtG CAGCAGCACT AGCTAACACT GATTATATAT 7200

TAGGTGATGA AATATTTAAT GGACTTGACT TTGAGACAAC ATTGCTTACG TTAGAGCTGT 7260

TTGAAAATTT ATCTAGAGAA GTAGGGATAG TGATTATTTC TCATAATAAA TTAATTATTG 7320

AAAGGTTTTC AGAGAATATT TTATTGATGT CCAATGGCAA TTTAACTCCA TTTTTAGGGG 7380

CGTCAGAGAA TTTAGAAAAA GAAGTGATAA GTACGGAGAA AATTCATGAA AAAATTAAAT 7440

ACATCAAAGG ATATCATCCT ATTAATCGAG TTATTTATTG ATAATGCTTT ACGAGGTTTG 7500

CTCAAGCACT CTATTTTGTC AAAAAAGAGT ACTCGCATCG TAATAGGTGT GTGTGTATTT 7560

TTTATCTATT TTGCATACTT TTTTTTTAAT ATGAGTGAAC TAGCTAGAAT TGTACCAGAT 7620

TCAGAAAAAA TTAGTCATAT TTTGATAGAA CAAGGTCGAA AAATAACTTT TTATAGTTAT 7680

TTTAGTAATA CATTTGTTTT AGGTATAATA GCATATATTT TAGTTGaTAA TACTGTTGCa 7740

TTAGATAAAA ACTCCTTATT TTTTGTGAAA ACATTGCCtT TTAGAAAGAA AGACATCAGC 7800

TTATCTTTTA TGCTATTCAA ATTTGTAATT ATGATATTGC TTTATGAATT AGTTATGATC 7860

ATTTCTACGC CAGCTATTAA ACTTGTGACG ACTGTTCCTA TTGAATATGT CATTTTTTTT 7920

ATTGTTCAGC ATCTTTTTTA TCTAGTGGTA ATTGGGGTGA TTGAGTTTAT ACATTGTTTG 7980

TTTACATTCT TTTTGAAACG AAGAGTGAAA CAAATTAATA GTCTCAAAGT GATAGGTGAT 8040

AGTTTATTAA TGATTTTTGC TACTTTTTAT TTTTTTGATT TTCGTTATGG TTTAGAATTA 8100

TTTCTAGCAA ACCAATTGTG GAGTATCTCC TATATGATTC CTATCACTTT TTTTCTAATA 8160

GTGGTATGTT TAATGCTTAT TAGTATAGGC CTATTAAGAC TAATTAGTAT TCTTGAAAAT 8220

CAAATGTCTC AACATTCAAA ATATATATAC ATTCCATTCT TAAATAAAGT tATATTTCGT 8280

TATAAAGCGA ATTGGTATTT TATTAGTTTT ACTACAGTTG TTATGTTGGT TATTTTTTTC 8340

CAAAGTGGTT TAAGAACAAT GTTATTCATT TTGACAACTG TTATGGCCTT TTCAGGCGTT 8400

TTATTATTAA GCTATGGAGA TATTACTGCT GATTTCAGAA AACAATATGA TTTATTGAGA 8460

ATAAAAAATTA GAAATGAATG GTTGAGTCmA CTATTGCkAG TGATAATGTT GGCAATGCCA 8520

TtGtTACTAC TTgtTTTTGG GGATTtGGGG AGTTAGCTCA ACTAATTACA GCGTTGTCGT 8580

TATCATTAGT AGCTATAATT TTGGGATATG TATTTCCAAA ATCACAAGGA AGTTTAAATG 8640

AAACGACtTC ATTATTGTTA TTATTTATTG TATTTGTTCT AGTTAGTTTA TTGACAAATC 8700

GCTCTTTTGG TTGGTTGATT CTAATTGTTT TAGTAnTCTT GCATTTGTTA GTAATTAAGA 8760

AGGTGAGAAA TGAAAAAGAA TAGTAAATAT TTAGGAAGTA TAACAGTTAT CTATTTGTTA 8820

GTTCTTGTGA TCGAAACATT ATGGGAGCTA TTCCATACCA GTCGCACAGA ATCAACGAAA 8880

GTCACTACCT TATTAGGAAT TACGATTGAT AATCGTATAA GTAAACATGA AATTTCAACT 8940

ACTTTTGGTT TGACAATTAA GGTATTAGTA CTCTATTTGT TGTTATTGTT AGTTGTCTAT 9000

ATTTTGACTG GGTTATTTGG AAAAAAGAAA ATATGATTAT TAAGTTATCA ATATTGTATG 9060

TGTTGTCAAT ATATGTCGTG TTACTTGTAC ATGAGTGGTT ACATTTTATT TTAGCCmAAT 9120

TCTTCAAATA CAATGcTTAT ATAAAGAGAG TTGGTCTATA CCCATTTAAA GTGGTGTATA 9180

CGAACAGAAA TAATCCtCTA GATAATTTAT TGATTTCAGC GATATCTCCG CTTTTCTTGG 9240

TTATAGTAGG AATAATTTTG CCATTAAATT ATTATACTGT GATTTTAAAA GTATCTTGTA 9300

TTAGTAATAT TTTTAATCTC TTACCGTTTA CTGCAGATGG AGAAATAATT TTGCTTTCAA 9360

TTTTTCAAAT TTTTAGGAGG AGAAATAAAT GAAAAAAAGA AAAGTGTTCA TTATTTTTCT 9420

TGTTAGTTTG ATTATATTTG TTTTTGGTTA TTACTTAGCG TTTAATCAAA ATAAGTATGT 9480

ATCATATAAT GACGAGACAA AACAGTTTGT ACATACTGAT ACTTCAAAAA TAATGGAAAA 9540

TCTGGTGAAC AGACAACAAG GAATTTATTA TTTTGGCTTT CCGACTTGTC CTTGGTGTCT 9600

AGAATTATTA CCAATACTAG ATAAAGAATT AGAGAAGGAA AGTATGAATG CTTATGCTGT 9660

AAATACTCGT GGTGATGACT ATACaGAAAA TGATGAkGAA TTATTACaAA AATTTkATCa 9720

AAAATATACT GGGGATGAGT CGCTTTCTGT ACCTTTTATT GTTGCTA 9767

(2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7900 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

AAAATCGTTT TGGATTTTTT CTCGGnTnCT TTCTATTTAA GTTATTTTAG TTGGATATCG 60

TTATTTAATA AGGCTATTAT ATGGTATAAT GATAATTAAC TGAAGTGTGG AGGAAATAAA 120

ATGTTATCAA TTGTTGTCCC TTGTTATAAT GAGGAAGCAG CGATTCCGCg TTTTTTgAAG 180

AAGTAGAAAA AATTAGCCAG AAACTGTCTC ATTCAGTAGA ATATATATTT GTGAATGATG 240

GCTCAAAAGA CAATACGTTA GCTGTTTTAA GACAGCTTTA TCGCGAACAT CCAGAGAAGG 300

TCCGTTATCT TTCTTTTTCA CGAAATTTTG GGAAAGAAGC GGGCTTATAC GCTGGTTTAA 360

AAGAAGCAAC AGGTGACTTA GTGACAGTCA TGGACGTTGA TTTACAAGAC CCGCCTGAGT 420

TATTACCTCA AATGATTGAG ATGATTGAAA CCAGTACAGA TTTAGATTGT GTTGGCACAC 480

GACGCATAAC AAGAGACGGC GAACCGCCCA TTCGTAGCTT TTTTGCACGG ATGTTCTATA 540

AACTCATTAA TCGAATTGCT GAGACGGAAA TGGTGGATGG TGCCCGTGAC TTTCGTGTCA 600

TGACCCGGCA AATGGTCGAT GCGATTTTAG AGTTATCTGA GTATAATCGT TTCTCAAAAG 660

GGATTTTTAG TTGGGTCGGC TTTAAAACAG AATATATTGA ATTTAAAAAT CGAGAACGAA 720

TTGCTGGAGA AACCTCTTGG TCTTTCTGGA GTCTATTAAG TTATTCAATT GACGGGATTG 780

TCAATTTTTC AGAAACACCA TTGAATATTG CTTCATACGT TGGCGCCTTT TCGTGTATTG 840

GTTCTGCACT AGCAATGTTA GTGATTATCT TTCGAACATT AGTTAATGGT GATCCAACGA 900

GCGGCTGGCC TTCAATGGTT TGTATTGTAC TTTTTGTAGG TGGTTTGCAA CTATTATGTT 960

TAGGAATTAT TGGCAAATAT ATCGGTAAAA TTTTCTTAGA AACGAAAAAA CGACCAATTT 1020

ATATTGTAAA GGAATCAGAA AAAGATTCTA AAAAATAACC AAAAAAGCAG AGAAAAATTC 1080

TCTGTTTTTT TGTTGTGCAT TAAAAAAAGT GTGAATAGAC TTTTCAGAAA AACAACATTT 1140

TTTACATAAA AAATACATTA ACTTTACCTA TTTTTATCTT AATGAAACAG CTTTGCTCTC 1200

GCGGTTGCAA GAAATAAATG TTAAAATCAA AAGGTATGCT ATTAAATGTT TTTTTGGAGG 1260

TTGGGTcATG ATAGAAGTTA TCGACTTGAA AAAAGTATTC GATAATGGGT TTGAAGCGTT 1320

AAAATCGGTC AATTTTACGA TTGAGCAAGG GGACTTGGTT TGTTTGTTAG GACCAAGTGG 1380

GTGTGGGAAA TCCACAATTT TAAACTTAAT TGCAGGTTTA TTGCATCCAA GTGATGGAGA 1440

TATTCAGTTT CGACAGCAAT CGGTTGTCAA AACAGCACCA AAAGATCGGA ACATCGGGTT 1500

CGTTTTTCAA AACTATGCCT TATATCCACA TATGACTGTT CTAGAAAACG TAATGTTTCC 1560

TTTGACGGTA GGAAGTAAGA AAGTTCCCAA GGCAGAGGCA CAAGCTATTG CCGAAGAATA 1620

TATGAAACTA ACGAATATTG AAGAGTTAAG CCATAAAAAG CCAGGCACAC TTTCTGGTGG 1680

TCAACAACAG CGTGTTGCGA TTACCCGGGC GTTAGTTCAA AAACCAGATG TTTTATTATT 1740

AGATGAGCCT TTAAGTAACT TGGATGCACG TTTGCGTTTG AAAATTCGTG AAGAAATCCG 1800

CCGCTTGGTG AAAGAAGTAG GGATTACAAC AATCTTTGTA ACGCACGACC AAGAGGAAGC 1860

GCTATCAATC AGTGATAAAA TTATTTTATT AAACGAAGGG GTTATTCAAC AAAATGATGA 1920

ACCTCAAAAC CTTTACTTAG AACCAAATAA CTTATTTGTT GCTCAGTTCA TAGGTAACCC 1980

AATTATTAAT TTATTGTCTG TTGAAGTGAA AGACGGCAAA ATGTACCACA AAAGTTTTGA 2040

AATTCCGCTT GAGCGTTTTG AGCAAGCACG CTTTAAAATG CCAATGACTG ATGGGAAATA 2100

TACGTTTGCT TCTCGTCCgG AAGATGTGGT ACCAGCTGAA ACAGGTCTCT TTACCACAAC 2160

GACAGATTTA GTGGAATTGA TTGGTCGAGA ACGTATTTTA CGATTTACAT TAGGAAATGA 2220

ACAAGTGAAA TCAATTGTAA GTGTAGAAGA AGCGATTGAA GAAGGAGACA CTTTATCTTT 2280

TGATTTTTCA TATAAAAAAG TATTTATCTT TAACGAAGCG GGAGACCGGG TTTACTAATG 2340

AAAAAATACA ATCCAGAAAA TCAACCAAAA GCATGGCTCT TCCTTCTGCC ATCATTAGGA 2400

ATCATTTTAC TGTTTAGCGT TTATCCACTG TTTCGTTCTT TGTGGATGAG TTTCCAAAAA 2460

GGTTCCTTAA TTAACCAACG cTACGCTGGT TTAGAAAATT ATCAACGTGT ATTGAATGAT 2520

CCAATCTTCT ATAAAGCGTT AAAAAATACA GCGCTCTATG CATTTGCTGT GGTACCGATT 2580

GCCTTGATAA TTTCCCTAGC AATTGCTTGG ATTATCTTTG AAAAAGTTAA ACATAAGAGT 2640

TTCTTTGAAA CGATTTTCTT TATGCCTTAT GTAACGAGTA CGATTGCCAT TGGGATTGTC 2700

TTCCGTTACT TCTTTAATGG CGATTATGGA ATTGTCAATT ACGTCTTAGG CTTTTTCGGC 2760

ATTCCTTCTG TCAACTGGCT AGATAATGTT CAAATGAGTA TGCCAACATT GATTATTTTT 2820

GGGGTTTGGA CGAGTTTAGC ATTTAATATT ATTATTTTGT TGGCTGGGTT GAGAAATATT 2880

GATGAAGAAC ATTTTAAAAT TGCAAAAATG TTTGGCGCCT CAGACGGCGA AATTTTCCGA 2940

CGTATTACGT TTCCGCAGCT AGTCCCGACC ATTGCCTTTT TATTAACGGT CAACTTAATT 3000

GGCGCCTTTA AAGTTTATAC ACAAGTTTAT GCTTTATTCG GTGGCCGTGC GGGGATTGCC 3060

AATAGTGCCA CAACCGCAGT GTACTATATT TATGACAAAT TCCACATTGC AGGACGTCCC 3120

GGAATTGCGA TGGCCGCCAC AGTGATTTTA TTTGTGATTA TTTTAGTGGT CACTTTCTTA 3180

CAAAATAAAC TCTTAAAGAA AGTGGGGCAA TAAGCCAATG AAAAAAGTTT TAACGATTAT 3240

CGCATTTGTT TTCTTAGGCA TTTTAGCCGT TATTACGTTA TTTCCATTTG TCTATATGAT 3300

TTTAGCTGGT TTAATGAGTT ACTCTGAAGC GACCAGCATG CCGCCAACCA TGTTTCCGAA 3360

ACAACCACAA TGGCAAAACT ACACAGAGGT CTTTCAAAAA GCACCCTTTC TTCGGTATTT 3420

CTTGAACACA GTCTTTGTTT CAGGCGTTAC AACGATTGCT ACGGTAGTCA CCGCCGTGTT 3480

AGCTTCATTC GCATTAACGA GTTTGAAGTT CCGTTTTAAA AATGTGGTCA TTGCTTTGAT 3540

GATTTCGTTG TTAATGGTTC CGTATGAATC TATTATTTTT ACCAATTATC AAACGATTGC 3600

GCAATTAGGT TTGTTAAATA CTTACAGCGC CTTAATTATT CCATTTTTAA CTAGTATTTT 3660

TTATATTTAC TATTTAAACG GTTATTTAAA AGGTATTCCT GATACTTTTT ACAAGGCTGC 3720

CAAAATTGAT GGCGCTAGTG ATTTAGAATA TATTTGGCGT ATTTTAGTAC CAATGTCAAA 3780

ACCAGCGTTG GTAACAGTAG GGATTTTAAC ATTTATTTCT AGTTGGAATT CTTTCTTGTG 3840

GCCATTGTTG GTGACGAACG AGAAAAAATA CCGTCTATTG AATAATGGAC TATCGGCCTT 3900

TGCGACAGAG AGCGGTAGTG ACGTACATTT ACAAATGGCT GCTGCAACAT TAACCGTTAT 3960

TCCAATTTTA ATTATTTACT TGATTTTCAG AAAAGAAATT ATCAGAGGAG TTGCAAAAAA 4020

TGGAATCAAA GGCTAAAACA ACTGTGACTT TTCACAGTGG CATTTTAACA ATTGGAGGAA 4080

CCGTGATTGA AGTGGCGTAT AAAGATGCCC ACATCTTCTT TGATTTTGGT ACAGAATTTC 4140

GACCAGAATT AGATTTGCCT GATGATCACA TTGAAACCTT AATTAATAAT CGTTTAGTGC 4200

CAGAATTAAA AGACTTATAT GATCCACGTT TAGGTTATGA ATATCATGGT GCAGAAGACA 4260

AAGACTATCA ACACACCGCT GTTTTCTTAT CTCATGCCCA TTTAGATCAT TCACGCATGA 4320

TTAATTATTT AGACCCAGCT GTTCCGTTGT ACACCCTAAA AGAAACGAAA ATGATTCTAA 4380

ATAGTTTAAA TCGAAAAGGT GATTTTTTGA TTCCATCACC GTTTGAAGAA AAGAATTTTA 4440

CCCGAGAAAT GATTGGTCTT AATAAAAACG ATGTGATTAA AGTTGGCGAA ATTTCAGTGG 4500

AAATCGTGCC TGTTGATCAT GATGCGTACG GTGCTTCGGC ATTACTGATT CGCACGCCTG 4560

ACCATTTCAT CACATACACA GGTGATTTGC GCTTACATGG CCATAATCGT GAAGAGACTT 4620

TAGCTTTTTG TGAAAAAGCC AAACATACTG AATTATTAAT GATGGAAGGC GTAAGCATTA 4680

GCTTCCCAGA ACGTGAACCA GATCCAGCCC AAATAGCGGT TGTCAGTGAA GAAGATCTTG 4740

TTCAGCACTT GGTTCGTTTA GAACTAGAAA ACCCGAATCG ACAAATTACC TTTAATGGTT 4800

```
ATCCAGCAAA CGTGGaACGT TTTGCTAAGA TTATTGAAAA GTCACCACGT ACAGTCGTTT   4860
TAGAAGCAAA TATGGcTGCG TTGTTACTTG AAGTATTTGG AATAGAAGTT CGTTATTATT   4920
ATGCTGAATC TGGTAAAATA CCAGAATTGA ATCCAGCGTT AGAAATCCCg TATGACACGT   4980
TACTAAAGGA CAAAACAGAC TATTTGTGGc AAGTTGTGAA CCAGTTTGAC AACCTCCAAG   5040
AAGGTAGTTT ATACATTCAC AGTGATGCAC AACCGTTAGG GGACTTTGAT CCACAGTATC   5100
GCGTGTTTTT AGATTTGTTG GCTAAAAAAG ACATTACTTT TGTCCGCTTA GCTTGTTCAG   5160
GACATGCAAT TCCAGAAGAT CTGGATAAAA TTATTGCATT GATTGAACCT CAAGTATTGG   5220
TTCCAATCCA TACGTTAAAA CCAGAAAAAC TGGAAAACCC GTATGGTGAA AGAATATTGC   5280
CAGAACGTGG CGAGCAAATT GTTTTATAAA TTTTTTTAAG GGAGAGAAAA AAATGAAGTT   5340
CAAAACTCTA GCAACAACAG TGTTAGCAAC CGCAGCTATT TTCGCATTGG GGGCTTGTGG   5400
TAACGGTAAT GGGGCCAAAG AATCAAACGA TATTGTGAAA GAAGTGAAGG AAGATACGAC   5460
AATCACTTTC TGGCATGCAA TGAATGGGGT TCAAGAAGAA GCGTTAACAA AATTAACGAA   5520
AGACTTCATG AAAGAAAATC CAAAAATTAA AGTGGAATTA CAAAATCAAT CTGCTTACCC   5580
TGATTTACAA GCCAAAATCA ATTCGACTTT AACTTCACCA AAAGATTTAC CAACAATTAC   5640
GCAAGCGTAC CCAGGCTGGT TATGGAATGC TGCACAAGAT GAAATGTTAG TGGACTTAAA   5700
ACCATATATG GATGATGACA CAATCGGCTG GAAAGATGCA GAGCCAATTC GTGAAGTATT   5760
GTTAGACGGC GCCAAAATCG ACGGCAAACA ATACGGCATT CCATTTAATA AATCGACAGA   5820
AATGTTATTC TATAATGCTG ATTTGTTGAA AGAATATGGT GTTGAAGTAC CGAAAACATT   5880
AGAGGAATTA AAAGAAGCTT CTAAAACAAT TTACGAAAAA TCCAACAAAG AAGTCGTTGG   5940
TGCTGGTTTT GACTCGTTAA ATAACTATTA CGCAATTGGA ATGAAAAACA AAGGCGTTGA   6000
TTTTAATAAA GACTTAGATT TAACAAGCAA AGATTCACAA GAAGTCGTGG ACTATTACCG   6060
TGATGGTATC GAAGCAGGTT ACTTCCGCAC AGCTGGTTCA GATAAATATT TATCTGGCCC   6120
ATTTGCAAAC AAAAAGGTAG CAATGTTTGT CGGTAGTATT GCTGGTGCTG GTTTTGTTCA   6180
AAAAGATGCT GAAGCTGGTG GCTATGAATA CGGTGTTGCA CCACGTCCTG AAAAAATCAA   6240
CTTACAACAA GGAACAGATA TTTATATGTT CGATAGTGcT ACGCCAGAAC AACGGACAGC   6300
GGCATTTGAA TTCATGAAAT TCTTAGCTAC TCCTGATTCA CAATTGTACT GGGCACAACA   6360
AACAGGTTAT ATGcCAATTT TAGAATCTGT TTTACACAGT GATGAGTACA AAAATTCTAA   6420
GACAACCAAA GTACCTGCAC AACTTGAAAA CGCAGTAAAA GATTTATTCG CTATCCCAGT   6480
AGAAGAAAAT GCTGATTCAG CCTATAATGA AATGCGGACA ATTATGGAAA GTATTTTTGC   6540
TTCATCAAAT AAAGACACGA GAAAATTATT GAAAGATGCA ACATCACAAT TTGAACAAGC   6600
ATGGAACCAA TAATCAATAA ATAAACAGAC GTGGGAGGTT GGGATCTGAG TCGTTATGGC   6660
TCGTGTTCTA ACCTTCCTGT TCTTTTGGGC GGAAGAAGGA GAGAAACAAA ATGCAAATAA   6720
AAATTTTAGC AACAAGTGAT ATGCATGGGT ACATTATGCC AACAAGTTAT AGTGAAAAAA   6780
```

AGATGGATTT ACCTTTTGGA ACCGCAAAAG CAGCAACCAT GCTGAAAAAG TTACGGGCGT 6840

CTGCCAAAGG ACCAGTTTTT CAAATTGAAA ATGGGGATTT TATTCAAGGT TCGCCGCTTA 6900

GTTACTACGT AAGAAAAGCA GAAACACATT CAGTGGCTGC CATCACTAAA ATCATTAATC 6960

AAATGAATTA TGATGTTAGC ATTTTAGGTA ACCATGAATT TAATTACGGT TTAGATTATT 7020

TAAAGGAAAC GATTGCCAGT TATCAACAGC CAGTTTTAGC TGCCAATATT CTTGGTAAAG 7080

ATGGACAACC TTATTTCGGT CAGCCGTATG TCATCATAGA AAAACAAGGG GTTAAAGTAG 7140

CCATCCTTGG CGTGACGACT CAATATATTC CTCATTGGGA ACAGCCGGCA ACAGTCAAAG 7200

ATTTAACGTT TAAAAGTGTC GTTGAAACAG CGGCGGAATA CGTGCCAAAA CTGCGGGAAG 7260

AAGCGGATCT AGTTGTTGTC GCTTATCATG GCGGTTTTGA AAAAGATTTA GAAACAGGGG 7320

AACCCACGGA ATTACTAACA GGAGAAAATG AAGGCTACGA TTTATTGGAA AAAGTTCCTG 7380

GGATTGATGC CTTAGTCACT GGTCACCAAC ATCGAGAAAT TGCTACAAAA CTGAATGGAA 7440

TACCTGTGAT TCAACCAGGT TTTCGAGGAG CTTTCGTTGG AGAAATTACT TTAGAAATTG 7500

AACCTATGGC TAAAGGCTAT CACGTTATTG GAAGTGACGC AGCTATTCAT CCTGTTGGAA 7560

ACGAACAACC AGATACAGAA GTTTTAGCTT TGACAACAGC GTTACATGAC GAGGTAGAAG 7620

AGTGGcTTGA TCAACCAGTG GGAAATGTGG AAGGGGATAT GACGATTCAA AATCCGAATG 7680

CCGTACGCCT CAAAGAACAT CCATATATAG aATTTATTAA TAACGTCCAA ATGGCTTCAA 7740

GCGGAACGGA TATTTCAGGG ACTGCACTTT TTAACAATGA AGGCAAAGGG TTTAATAACC 7800

AAATTACGAT GCGCGATATC ATTACCAACT ACATTTATCC AAATACCTTG GCTGTATTGC 7860

GTGTCACTGG ACAAGACCTC CGTGAAGCGT TAGAACAAAG 7900

(2) INFORMATION FOR SEQ ID NO: 132:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5480 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

TCCTTACATT TCATGTTATA AAATGTAACA TAGAAATTAA GAACAAAAGA TTAAGATCAG 60

TTGAAGTCCG AAACAGCAAA AGTGGCGAAA ATATGTTCGG TATGTTAAAC TAATAAGTGG 120

ACATTCTGCC AGAATGGGCC GATAATTTAA GAGAAATTAG ACACTTTGAG GGGGAGTTAT 180

AAATGATCGA AAGAGAAACT TCCAATACCT TTCATTTGGC TTCTAAATAT GAGCCTGCCG 240

GAGATCAGCC AGCCGCTATT GCTGAACTAG TTGACGGTGT CAAAGGTGGA GAAAAAGCTC 300

AAATTTTACT TGGAGCGACC GGAACAGGTA AAACATTTAC AATCTCTAAT GTGATTCAAG 360

AAGTGAACAA ACCAACATTA GTGATTGCAC ATAATAAAAC ACTAGCTGGA CAATTATACG 420

GGGAATTTAA AGAATTTTTC CCAGATAATG CGGTCGAGTA TTTTGTCAGT TACTATGATT 480

ATTATCAACC AGAAGCATAT GTACCCTCAA GTGATACATA TATTGAAAAA GATTCCACTA 540

TTAATGATGA AATTGATAAA CTACGCCATT CAGCGACAAG TTCTTTATTG GAACGAAACG 600

ATGTCATCGT GGTAGCATCA GTTTCTTGTA TTTTTGGTTT AGGGGATCCG CGGGAATACA 660

GTCAACAAGT TGTTTCATTA CGTGTGGGAA TGGAAATGGA TCGTAATGAG TTACTAAAAA 720

GCTTAGTCGA TATTCAATTC GAACGAAATG ATATTGATTT TCAACGGGGG CGTTTTCGTG 780

TCCGCGGTGA TGTGGTTGAG ATATTCCCAG CTTCACGGGA TGAGCATGCG CTACGTGTTG 840

AATTTTTCGG AGACGAAATT GATCGAATTC GTGAAGTAGA TGCGTTAACT GGTGAAATTG 900

TCGGCGAAAC AGAGCATGTC GCTATTTTTC CGGCAACACA CTTCGTGACG AACGAAGAAC 960

ATATGGAACA CGCCATATCG CAGATCCAAG AAGAATTAGA GGCACGTTTG AAAGTGCTTC 1020

GCAGTGAAAA TAAATTATTG GAAGCTCAAC GTTTGGAACA GCGAACCAAC TATGATATTG 1080

AGATGATGAG AGAGATGGGC TATACGTCAG GGATTGAGAA TTACTCTCGA CATATGGATG 1140

GTCGCCAAGA AGGAGAGCCG CCATATACGC TGTTGGACTT TTTCCCAGAT GACTTCTTAC 1200

TTGTAATCGA TGAGTCTCAT GTAACAATGC CGCAAATTAG AGGAATGTAT AATGGCGACC 1260

GAGCACGGAA ACAAATGTTA GTGGACTATG GTTTCCGGTT ACCAAGTGCA CTTGATAACC 1320

GACCGCTTCG TTTAGAAGAA TTTGAACAAC ATGTAAACCA AATTGTTTAT GTATCGGCAA 1380

CACCAGGCCC TTATGAAATG GAACAAACAG AGACTGTTGT TCAACAAATC ATTCGACCAA 1440

CAGGTTTACT AGATCCAGAA GTAGAAATCC GCCCAATTAT GGGTCAAATt GATGACTTGG 1500

TGGGCGAAAT TCATGAaCGG ATCGAAAAAG ATCAGCGGGT GTTTGTCACA ACCTTAACGA 1560

AGAAAATGGC CGAAGATTTA ACCGACTATT TCAAAGAATT AGGCTTAAAA GTTAAATATT 1620

TGCATAGTGA TATAAAAACA CTGGAAAGAA CGGAGATTAT TCGCGATTTA CGATTAGGCG 1680

AATTTGATAT TCTAATAGGT ATTAACTTAT TACGTGAAGG AATCGATGTG CCAGAAGTTT 1740

CCTTAATTGC GATTCTAGAT GCGGATAAAG AAGGCTTCTT ACGTAGTGAA CGCTCGTTAG 1800

TGCAAACGAT GGGACGGGCT GCCCGGAACG CTGAAGGAAA AGTCATTATG TATGCAGATA 1860

AAATCACCGA TTCGATGCAA CGAGCTATGG ATGAAACGGC ACGACGTCGG GCAATCCAAG 1920

AAGCATATAA TGAAGAACAT GGTATTGAAC CAAAAACAAT TATTAAAGAA ATTCGTGATT 1980

TGATTTCTAT TTCCAAAACA GCTGATAAAG ATGAAACAGT GGTTCAATTG GATAAATCAT 2040

ATAAAGATTT ATCAAGACAA GAAAAAGCTG ATTTATTAAT GAAACTTGAA AGAGAAATGA 2100

AGGATGCGGC CAAAGCATTA GACTTCGAAA CTGCAGCCAC ATTACGAGAT ACCATTCTTG 2160

AATTGAAAGC TGCCAAATAA GTCAACCAAC GGATTAAGTA AAAGATACGA TGAACCATCG 2220

TATCTTTTAC TAATCTTTTC TCTATAGAAT GACTTGAAAT GATTTGACGT AAGAGATTTA 2280

GTAAAAGGAG AACATATATG GCAAATGATA AAATTGTGAT TCATGGTGCA CGCGCCCATA 2340

ACTTAAAAAA TATTGATGTC ACGATTCCTC GTGACAAAAT GGTTGTCGTA ACTGGACTAT 2400

CTGGTTCTGG CAAAAGTTCA TTAGCGTTTG ATACGCTTTA CGCAGAAGGG CAACGACGTT 2460

```
ATGTAGAGAG TCTCTCTGCT TATGCGCGAC AATTTTTGGG ACAGATGGAT AAACCAGATG   2520
TTGACAGCAT CGATGGTTTA AGTCCAGCTA TTTCAATTGA TCAAAAGACG ACAAGTAAAA   2580
ACCCACGCTC AACGGTGGGG ACGGTCACAG AAATCAATGA TTATTTACGA TTGTTATTTG   2640
CTCGGGTAGG TCATCCAATT TGTCCTAACG ATCATATTGA AATTACCAGT CAGTCTGTTG   2700
AACAGATGGT TGATAAAGTC TTAGAGCTAC CAGAGCGAAC AAAAATTCAA ATTTTGGCCC   2760
CTGTTGTGGT CAAGAAAAAA GGCCAACATA AAAAAGTGTT TGAAATGATC CAACGTGAAG   2820
GGTATGTCAG AATGCGGGTG GATGGTGAAA CCTATGATGT TAGTGAGGCA CCAGAACTTG   2880
AAAAAAATAA AAAACATGAT ATTGCGATTG TGATTGACCG TATCGTTGTG AAAGAAGGCA   2940
TTCGCTCTCG CTTGTTTGAT TCGTTTGAAG CGGCTTTGCG TTTAGCTGAA GGTTACGCAA   3000
TTGTTGATGT GATTGGTCAA GAAGAAATGT TGTTTAGTGA ACATTATGCT TGTCCTTATT   3060
GTGGGTTCAC AGTCGGTGAA TTGGAACCGC GCTTGTTTTC ATTTAATGCA CCATTTGGCG   3120
CTTGTCCCGA TTGTGATGGT TTAGGTGTTA AATTAGAAGT GGACAAAGAC TTAGTGATTC   3180
CAGATCCCAC TAAAACCTTG AGAGAAGGGG CCATTGTTCC GTGGAACCCT ATTAGTTCCC   3240
AATATTATCC ACAAATGTTG GAGCAAGCAG CCACCAGTTT TGGGATTGAT ATGGATACGC   3300
CGTTTGAAGA ATTACCTGCA GACCAACAAG AAATTATTTT AAATGGTTCT GGCGAGAAAA   3360
ACTTTCATTT TCATTATGAA AATGACTTTG GTGGTGTTCG TGATGTGGAA GTGCCATTTG   3420
AAGGGATTTT AAAAAATATT AAACGACGTT ATCATGAAAC GAATAGCGAT TTTACACGAG   3480
ATCAAATGCG GTTATACATG ACAGAATTGA CTTGTCGAAG CTGTCAAGGG TATCGTTTGA   3540
ATCCGCAAGC TTTAGCTGTA AAAATcAATG GCACGCACAT TGGTGAAGTC AGTGAATTAG   3600
CAATAAAAAA TGCGGTCCAA TTTTTTGAAG GTGTGTCTTT ATCTGAACAA GAAACAACGA   3660
TTGCTCGGCC AATTTTAAAA GAAGTCGAAG ATCGGTTAAC CTTCTTAAAA AATGTTGGGT   3720
TAGATTACTT AACCTTAAGT CGAGCTGCTG GTACACTTTC TGGCGGAGAA GCACAAAGAA   3780
TTCGCTTAGC AACACAAATT GGATCAAACT TATCAGGCGT TCTTTACATT TTAGATGAAC   3840
CATCAATTGG TTTGCATCAA CGAGATAACG ACCGCTTGAT TGACTCATTG AAAAAAATGC   3900
GGGACTTAGG CAATACATTA ATTGTGGTGG AACATGATGA AGATACAATG ATGGCTTCGG   3960
ATTACTTGAT TGATGTGGGA CCTGGCGCTG GACATCTGGG CGGTGAAATT GTAGCCGCTG   4020
GTACGCCGGA AGAGGTTGCT AAAAATCCGC ATTCATTGAC CGGGCAATAT CTTTCTGGGA   4080
AAAAAGTGAT TCCTGTACCA AAAGAACGTC GCAAAGGTAA TGGCAAAGCG ATCAAAGTCA   4140
CTGGTGCCAG TGAAAATAAT TTAAAAAATG TCAGTGTTGA ATTTCCGCTG GGTGAATTTG   4200
TTGCTGTCAC AGGCGTTTCA GGTTCAGGGA AAAGTACTTT AGTCAATCAA ATTTTAAAAA   4260
AAGCGCTGGC CCAAAAATTA AATCGTAATT CCAATAAACC CGGGAAACAC AAAAGCATCA   4320
CTGGTTATGA AGcCATTGAA AAGATTGTTG ACATTGACCA AAGTCCAATT GGTCGAACAC   4380
CAAGAAGTAA TCCAGCGACT TATACTAGTG TTTTTGATGA CATTCGCGAT TTGTTTGCCC   4440
```

AAACAAATGA AGCCAAGGTT CGTGGCTATA AAAAAGGTCG CTTTAGTTTT AACGTCAAAG 4500

GTGGCCGTTG TGAAGCTTGT CGCGGGGATG GTATCATCAA AATTGAAATG CACTTTTTAC 4560

CTGATGTCTA TGTGCCTTGC GAAGTTTGTC ATGGCAAACG TTATAATTCT GAGACGTTAG 4620

AAGTTCATTA TAAAGGCAAA AACATTTCTG ATATTTTAGA TATGACGGTG GAAGATGCCG 4680

TAGAATTTTT CAAACATATT CCTAAAATTC ATCGTAAATT GCAAACAATT GTGGATGTAG 4740

GCTTAGGCTA TGTGACATTG GGCAGCCGG CAACTACTTT ATCTGGTGGG GAAGCACAGC 4800

GGATGAAGCT AGCGAGTGAA CTTCATAAAA ATTCTAATGG GAAAAACTTT TATATTCTGG 4860

ATGAACCAAC AACAGGCTTG CATACCGATG ATATTGCGCG ATTGCTTCTA GTATTAGAGC 4920

GTTTAGTGGA AGCGGGCAAT ACCGTTTTAG TGATTGAACA TAACTTGGAT GTCATCAAGT 4980

CTGCGGATCA TGTGATTGAT CTAGGGCCAG AAGGTGGCGA CGGTGGTGGG ACGATCGTGG 5040

CGACCGGAAC ACCAGAAGAA GTTGCTAAAG TCAAAGACAG TTACACAGGA TATTATTTAA 5100

AACGTGTTCT GAAATAAACA TTAATGATTA GATAAAGAAC CTTGTTTTTA TGGGAAACAA 5160

GGTTTTTTAT TATAAAATAA CAAATATTTT TTAATAAAAA TTAAAAATAA ATGTTGACTT 5220

AATGTGTAAT TAAAAGATAT ACTATGTTTG TTAACAAAAT AACAATGAAT ACGAAGGGAA 5280

GTCTTAAATG AAAAACAAAA AAATAACAAC AAGTTTATTA ACAGCAGGAA TGGTCTTAGC 5340

TATCGGAATC ACACCTTTTG TACATAGTAA CACATGGAAG CAGCGACCTT AATTTCACAA 5400

TCAGACCATG AAAAGTAATG AGTTTTACCA GAAAGACTAA AAGAAGCACG GGCATCAATT 5460

GGGCCGGCGC AGTCCAGCGT 5480

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10391 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

CAGGCTTTTT AATGACGAAT GTCGTTCAAC AATCCATGCG CGATTTGCGT CGCGATATTG 60

AAGAaAAAAT CAATCGTTTG CCAGTTTCTT ATTTTGATAA GAACCAACAA GGAAATATTT 120

TGTCACGGGT GACGAACGAT GTGGATGCTG TCAGCAATGC GATGCAACAA AGTTTTATCA 180

ATATTGTTTC AGCAGTCTTA GGTATTGTGA TGGCGGTAGT GATGATGTTC TTAATCAATC 240

CGCTGATGGC GATTTTTTCA GTGATTATGA TTCCGTTGTC TCTGATTATT TCCAGAACAA 300

TTGTTAAAAT CTCCCAGAAA TATTTCCAAG GAATGCAAAA TTCTTTAGGA GACTTAAATG 360

GTTATGTCCA AGAAAATATG ACTGGGTTCA GTGTCTTAAA ACTATATGGT CGGGAAAAAG 420

AAACCCTTGA AGGCTTTAAA CAAGTCAATC ATCGTTTAAA TGGTTTTGGT TTCAAAGCAT 480

CCTTTATCTC AGGATTAATG TTGCCATTGG TTCAGATGAC CGCTTATGGG ACCTATATCG 540

GGGTAGCTGT CCTTGGTAGT TACTATGTGG TTGCTGGTGT GATCGTAGTG GGGCAATTAC 600

AAGCGTTTAT TCAATATATT TGGCAAATTA GCCAACCAAT GGGGAATATT ACGCAGTTGT 660

CTGCAGCTTT ACAAAGCGCT TCAGCTTCGA CCATGCGGAT TTTTGAAATC CTAGATGAAC 720

CAGAAGAAGA ACTTAACGAA CAAGATGTTC CTTTGCCAGA ACCTATTTTA GGCTCTGTTG 780

AATTTGAAAA TGTCAGCTTT AGTTATGACC CAGAAAAACC GTTAATTCGT AATTTGAACT 840

TTAAAGTTGA TGCGGGCCAA ATGGTTGCGA TTGTGGGACC AACTGGCGCT GGGAAAACAA 900

CCTTAATCAA CTTACTGATG CGTTTTTATG ATGTAACAGA AGGCGCCATT AAAATTGATG 960

GCATTGACAC GAAAAAAATG AACCGTAGTG ATGTCCGATC TGTATTTGGA ATGGTATTGC 1020

AAGATGCTTG GTTGTATAAA GGTACCATTG CAGATAACAT TCGTTTTGGG aAGTTAGATG 1080

CCACGGaTTA TGaAGTTGTC GATGCaGCGA AAACGGcCAa TGTGGATCAC TTCaTTCGGA 1140

CAATGCCAGA CGGGTATGaA ATGGAaATCa ATTCTGAGGG AGATAACGTT TCCCTTGGtC 1200

aAAAACAATT GTTGACCATT GCCCgAGCGG TAATTTCTGA TCCGAAAATT TTGATTTTAG 1260

ATGAGGCGAC TAGTTCAGTC GATACACGCT TGGAAGCCTT AATTCAAAAA GCAATGGATC 1320

GTGTTATGGA AGGACGAACG AGTTTCGTTA TTGCCCACCG CCTATCAACT ATTCGTGAAG 1380

CTGATTTAAT TCTTGTTATG AAACAAGGAG AAATCATTGA AAAAGGTACG CATCATGAGT 1440

TGCTGGAACA AGGTGGCTTC TATGAAAAAC TATACAATAG TCAATTTGCT GAAGAAGGCG 1500

ACTATGAGGA ATAAAAAGAG GTCGGGACAG AAGCGTTTAA CTCCGAGAAA TAAGAAGGAA 1560

TTTCCGAAAA TTGTTCTGTA ATTTTTGGAG AATTCCGGCT TATTTCCGAA GGAGTTGCTT 1620

TTGTTCCCGC CGTTTATCAG TTTTTGAGCG TGGAGCAAAA ATCCAAAGTG ATTTTTGTCC 1680

CACGCTCTTT TCTTGTCTTT TTTACTATGA AAGTGAGGAA TTCTTGTCTA TACTTAATGA 1740

AAAGAGGTGA CAGAGTTTGA AAAAAGAATA TTTTATTGCT GCTAGTGCAG ATGTCTATGG 1800

AAACGTACAA CTAGCAAAGG ATAGTAATAT TTGGTTTCAG TCGGTCTTAA GAGGTGACAG 1860

TAACACAATT ACCATTGGTG AAGGTAGTAA TATTCAAGAT GGCACGATTA TTCATGTTGA 1920

TGAAGATGCC CCGACAATTG TAGGCAAGTA TGTAACGGTG GGCCATCAAT GCATGTTACA 1980

CGGTTGTAAA ATCGGTGATG GTGCTTTGAT TGGCATGGGT TCAACCATTC TAAATAATGC 2040

TATTATTGGA GAAAACAGCC TGATTGGTGC GGGCTCTTTA GTGACAGAAG GCAAAATTAT 2100

CCCGCCAAAT GTTTTAGCAT TTGGTCGACC CGCGAAGGTT ATCCGTCCGT TAACGGCAGC 2160

AGAAATAGCC AGTAATCATA CGAACGCCTT ACATTATATT CAACGAGCAG AAGAATTTCG 2220

CCAACAAAAA TACCCGAAAC TAACATAGAA TTAGGAGGAA AATGATGAAG AAAATTAAAG 2280

TAATGACGGT TTTTGGAACA CGGCCAGAAG CAATCAAAAT GGCTCCCTTA ATCAAAGTTT 2340

TAGAAGAGCA AAGTCAAGGA TTTGACTCTG TAGTAGTCGT AACAGCGCAA CATCGCCAAA 2400

TGTTGGATCA AGTGTTAGAA GATTTTCAAA TTACACCGAA TCATGATTTA AATATCATGA 2460

AAGATGGGCA AACCTTAACA GATATTACAA GTCGTGTTTT GAATTTATTA ACTGAGGTTC 2520

TTGTGACAGA ACAGCCGGAT ATTGTCTTGG TTCATGGCGA TACAACCACG TCATTTGCGG 2580

CTGCATTGGC TGCTTTTTAT CAACAGATTC CAGTGGGACA TGTGGAAGCA GGATTGCGGA 2640

CCTGGCAAAA ATATTCTCCG TTTCCGGAAG AAATGAATCG TCAACTGGTA GACGTGTTAA 2700

CTGATATTTA TTTTGCCCCA ACCACTCAAA GTAAGGACAA TTTAATCAAA GAAAATCATC 2760

CAGAAGAACA TATTTATGTG ACTGGCAATA CAGCAATTGA CGCGATGGCG TACACCGTTG 2820

ATGCCCACTA TCAAAATGAT TTATTGGAAA AAATCCCTAC AGACCAACGC ATCGTGCTAA 2880

TTACAATGCA TCGTCGTGAA AATTTAGGCC TGCCAATGGC AAATGTGTTT AAAGCTGTTC 2940

GTCGACTTGT AATGGAACAT CCAGAAATTG AAGTCATTTT CCCTATGCAC AAAAATCCCA 3000

AAGTCCGTGA AATAGTAGCG GAACACTTAG GTGAATTAGC TCGTGTGCAT TTAATTGAAC 3060

CATTGGATGT CAAAGATTTT CAAAATTTTG CAGCGAAAAG TGCTTTAATT TTAACTGATT 3120

CTGGCGGTGT GCAAGAAGAA GCTCCTTCCT TAGGTGTGCC AGTTTTAGTT TTACGAGATA 3180

CAACGGAACG TCCAGAAGGC GTAgCGGCAG GAACATTGAA ATTGGTTGGT ACAGACGAAC 3240

AAGTAGTCTA TCAAGAGGCG AAAACATTAT TGACTGATTC TGATGCTTAC CACGCGATGG 3300

CGCATGCACA AAATCCGTAT GGGGATGGTC AAGCTAGCCG CCGAATCGTT GAAGCCATTG 3360

CCTATGAAAT GCAGAAATCA GATAAAAAAC CAGATACGTT TACAGCAAAA TAAATAAAAA 3420

GGCAGAGAAG CGCTGAAAAA TGTGCTTCTT TTGCGTATCT TATTTTGAAG AATCAATTAG 3480

ACAAAAAATG CTAGAAAGTA GTGAAGTGAA GATGTATCAA ACGATTTTAT TTGATTTAGA 3540

TGGCACAATT ACGGATTCTG GTAGCGGCAT TATGCGTTCT ATTTTATACG CAACAGAACA 3600

ATTAGGCTGG CCAGCTCCTA GCGAGGAGAC ATTACGTTCT TTCATAGGGC CGCCTTTATA 3660

TGAGTCGTTT TTGCACATGG CGCCTTCAGC TGAAGCAGCG CAACAAGCAG TGGGTCATTA 3720

TCGAGCGTAT TATCAGAGAA AAGGTATGTT TGAAAATCAT GTGTATCCAG GGATCCCAGA 3780

GGTACTAATA AGGTTAAAGG AAGCGGGAGC AAAGTTATAC ATTGCGACTT CCAAGCCAGA 3840

AGAATTTGCT AAAAAGATTA TTACTCATTT TGATTTAGAT CGTTATTTTA CAGGCATATA 3900

TGGAGCGAGT ATGGACGGTC ATCGTTCTAA AAAAGCAGAT GTGATTCAGT ATGCATTAAC 3960

GGAAGCTCAA CTAGACCCAA CCAAAGAAGC AATTATCATG GTTGGTGATC GTAATCATGA 4020

CATTCTTGGT GCTCAACAAA ATGGCTTAGA TAGTATAGGT GTCTTGTATG GATTTGGTGA 4080

AGAAACAGAG TTACAAGAGG CGGGAGCGAC ATTCCTAGTA CACTCGCCTA AAGACTTAGG 4140

TGCTATTTTG CTACGAAATT CTTAATAGAA AATGGATTTT CGATAAATTT AGGAGGATTT 4200

TAAGATTTAT TATCGTTTCA ATAGAAAAGC GCCATCATAG GCTTTAAACG TTATCAATGA 4260

TGAATTTTTT ATCGGGGAAT GACGGTCTTT TTATTAAAAA CATTTAATTT TAAAGAGGTT 4320

TCATGTATAA TAGTTTCAGT CATAAACGAT AGAATAGTAA GGAGGGTTAG TTTTGGCTAA 4380

TGATAACCAA AATAACCAAG ATCCCAAATC TTCGCTTAGA GATCAAGTAA CTGGGTCTTT 4440

AAATGGTAGA AATGATGGCG ATCAACCTGA TTCCTCTGAA AAAAATGATA GGTCACCCCA 4500

ACCTTCATCT GATGAATCAC AAGAAACAGC GAGCCGCACA ACACAAACAC GTGCAGGTTC 4560

TAGAGCTGCT CGAAGACGTG GGAAAGACAA GACAACACAG ACAGTCGAAG AGCCAACCCC 4620

GATTGAAACA GACGAAAAGC CAACAAATAC AAAAAAACAA ACACGTAAAA AAGAAGATCG 4680

TTTAGTCGGT CGTATCGTGT TAATTGTAGT TTCTGTATTA GTTTTAATGA TGGCAATCTT 4740

TGGGTTTACT TTTTATAAAT ATGTAGATGC TGGGCTACAG CCACTAGACA AGAACAATAA 4800

GAAGCTTGTT CAAGTGCATA TCCCAGAAGG ATCTTCTAAT AAGCAAATTG CTGCGGTTTT 4860

AGAAGAAAGC AACGTCATTA AAAGCGGAAT GGTTTTTAAC TACTACGTAA AATTTAAAAA 4920

CTTAACAGAT TTTCAAGCGG GCTATTATCA AATGTCTCCA AGCATGACAC TCGATGAAAT 4980

TGGTGAAATG CTAAAAGAAG GAGGTACGCC AGAACCAACT AAAATTGCAG ATGGCAAAGT 5040

AACGATTCCT GAAGGCTACG ATATAGATAA AATTGGCGAA GCTATTGAAA AAAATACAGA 5100

TTTCAAGAAA GCAGATTTCA TTGCATTAAT GAAAAACGAA GACTTCTTCA ATCAAATGAA 5160

AGCTAAGTAT CCTGATTTAC TTGAAAGTGC TGCAACTGCA GAAGGTGTTC GTTATCGCTT 5220

AGAAGGCTAT TTGTTCCCGG CAACTTATGA TTACTACAAA AAAGCAACAC TTCCTGAATT 5280

TGTCGAACAA ATGATTGCTA AAATGAACAC AGTAATGGAA CAATATACGC CGACAATTCA 5340

TGCAAAGAAT TTAACCAATC AGCAAGTCTT AACGTTAGCT TCATTGGTTG AAAAAGAAGG 5400

AGTCAAAGAA GCAGACCGCA AACAAATTGC GCAAGTATTC TTTAACCGCT TGGCTGCCGA 5460

TATGCCAATC CAATCAGATA TTTCTATTTT GTATGCTCTA GGAGAACACA AAGAAACTGT 5520

AACGTATGCT GACTTGGAAG TGGACTCGTC TTATAACTTG TATAAAAATA CTGGTTATGG 5580

ACCAGGACCA TTAGATAGCC CAAGTGAGGA GTCAATTAAA GCAGTCTTGA ATCCGACACC 5640

AAGTGACTAT TTATACTTTg TAGCAGATAT TTCTACAGGA AAAGTCTACT TCTCAAAAAC 5700

ATATGAAGAA CATCAAGTAT TAGTAGATCA GTATGTCAAT AATTCAAGTA GTGAATAAAA 5760

TGAAAACATA AAAGAGCTTG GTGAATCAAT TTTGTATGAT TCACCAAGCT TTTAGATACT 5820

CAAAAAGTAA TCTGTTTAAC AGAGGGAAAG AGTAGCAACT AGAAATTTAT ATAAATAAAA 5880

TTTAGCAAAA TCACAAAAAT TCAATAATTA TATTTACAAT TTAGAAAAAG CGTGGTACTC 5940

TTTTGTGGAT TGAAAGCGGA ATTTTTTTCA ATAACCAACC AAGAACTACG TATCGTTAAG 6000

CGAAAAAGGA GAATTAACAT GGTAGAAAAA GTATATCCAA TGACTCTTGA AGGAAAAGAA 6060

AAATTAGAAC AAGAGCTAGA AGAATTAAAA ACAATTAAAC GTAAAGAAAT CGTAGAGCGT 6120

ATTAAAATCG CTCGTAGCTT TGGGGATCTT TCAGAGAACT CAGAGTATGA ATCAGCTAAA 6180

GATGAACAAG CTTTCGTTGA AGGACGTATT ACAACGTTAG AAAATATGAT TCGTTTTGCA 6240

GAAATCATCG ATAATGATGG CGTGGACCAA GACGAAnTTT CCATCGGTAA AACAGTTACT 6300

TTCCAAGAAT TGCCTGATGG AGAAGAAGAG GAATACACAA TCGTAGGAAG TGCCGAAGCC 6360

GATCCATTTT CTGGTAAAAT TTCTAACGAT TCACCGATTG CGCAACCnTT AATTGGTAAA 6420

CAATTAGGCG ACCAAGTGAT TATTGCAACA CCTGGTGGCG ACATGACAGT TAAAATTACA 6480

AAAGTTGAAG CAAACTAAAT AATTGAGTAG GGCAGGCAGC TTCCAAGGCC tGTCCTTTCG 6540

CAATCTATGA CATTTAGGGC TAAAGTCCTA TGGTAACCTG aGGCGCAAAC CACTATGATT 6600

AAATAAAATA AACGAGTAAA ATAAATGACA GAGGGAGGAG GCAAATCTAT GAATAACCCT 6660

TGGCGCTTTT TTATCGTCGC AGAAGCATTA CTTTTTATTC TGGCGTTATG GCAAATTGTA 6720

CATAATCCTG GaTTAGCTGT TTtATTAACA ATTGGCGTTT TACTTGTGGC CTACGTTTCC 6780

AGGAAAGCAT CTAAAACACA TTTTAACAAC TTTCAATTCG TCCTCGGCGT TGTTTTTATT 6840

GTCATTGGTG CAATGAATAG CACGGCTGTT TGGTTTATGT TGATTTTTGG CGTACTCTTT 6900

ATCGGCTTAA AAGGCTTTGA GATTTCAGGC GTGGATATAG CTGAGCGAGC ACCTTGGCGA 6960

AAAAAACAAA TGATTATGGT GGAGACGGCG GCAAAAGAAC CTAAAAATGG CAAACGGTTT 7020

AAACGCCGCT GGTTTGCCAA CGAACGCATT GGTAACAATA TCTATGAATG GGACGATATC 7080

AATATTGATT TAATCTCTGG GGACACCATT ATTGATTTAG GTAATACGCT ACTACCGAAA 7140

GAAGACAATA TTATTATTAT TCGTAAAGGT TTTGGCCGCA CGCGAATTCT AGTGCCGTTA 7200

GgTGTAGCTA TTTTGTTAGA ACATTCAACT TTTTACGGAA CGGTACGTTT TGAAGAAGAA 7260

AAATATCAAT TGAAAAACGA ATCATTAAAA ATTTACAGCA ATGATTATGA TACCAATCTT 7320

CGTCGTTTGA AAATTATGAC GAACACTTTA GTAGGAGATG TTGAGGTGAT CCGTGTATGA 7380

CCGATCGGAT TTCAAGACGC ATGATTTCAT TATATGCGTC CCTTAGCACC TTTATTGTTA 7440

TCTTAATTAC ATTGTTTTCA TATTTTCATT CGATTAAACA AAACCGGTGG TTATTAGAGC 7500

TTCTTCAGAG AAAAGTCTTT TATTTACCAC TAATTGTGCA CATTGTTCTC ATATCCTTAC 7560

TAATAGGCTT ATTGACCTTT TTACTGATTT CATTGGTTCA AAAAGGGCAA TATGGACGGA 7620

TTGAAGAAAA ACTTCGGTTA TTGGCCAACG GTAATTATGA AAGTCCAGTC TTAAACAAAC 7680

CAACGACCAG TGAAAATCAA GACCATTATC TAACCGAAGT CGAACAAGAT ATTTGGTCGA 7740

TTAAAAATAA ATTATTAGAG ATGTCTAAAG AATTGCAATT ATTAAACAGT CGACCGCAAT 7800

TAATGGATGG GCAAACAAAA GAAGAAATTT TAGAGAACGA GCGGCATCGT TTGGCGCGGG 7860

AGTTGCATGA TTCAGTCAGT CAACAACTTT TTGCAGCCAT GATGATGTTG TCTGCATTAA 7920

ATGAACAAGC ACAACGAACA GAAACCCCGG AACCATATCG TAAACAACTA GCCATGGTGG 7980

CAGAAATCAT TAATGCCTCC CAATCGGAAA TGCGCGCGCT ACTATTGCAC TTGCGTCCTA 8040

TCAGTCTAGA AGGAAAAAGT TTGCGTAAAG GTATTGAACA ATTACTGAAA GAACTACAAA 8100

CAAAAATTAA AATTGAATTG ATTTGGGATG TTGAAGATGT TCATTTAAAT AGCAGCATTG 8160

AGGATCATCT TTTCCGAATT GTGCAGGAGT TACTTTCAAA TACCTTAAGA CATGCCAAAG 8220

CAAAGGAATT AGAGGTATAC TTACACCAAG TCGATAAAAA CGTGTTATTG CGTATTGTTG 8280

ATGATGGTGT CGGCTTTGAT ATGAAGGAAC AAAGTAATAA AGCCGGTAGT TATGGCTTAA 8340

ATAATATTCG AGAACGTGTT GTCGGCATGG GCGGTACAGT TAAAATTATT AGTTTTAAAG 8400

GGCAGGGAAC CAGCGTTGAA ATTAAAGTTC CTGTCATAAA GGAGGAAACT GCAAGTGATC 8460

AAAGTAATGT TAGTGGATGA CCATGAAATG GTCCGTTTAG GCGTTTCATC ATATTTATCT 8520

ATTCAAGAGG ATATAGAAGT CGTAGGCGAA gcAGAAAACG GTAAGATTGG CTATGAAAAA 8580

GCATTGGAAC TACGTCCAGA TGTTATTTTG ATGGATTTGG TAATGGAAGA AATGGACGGC 8640

ATTGATTCAA CAAAAGCGAT CTTGAAAGAT TGGCCAGAAG CCAAGATTAT TATTGTGACG 8700

AGTTTTATTG ATGATGAAAA AGTGTATCCG GCGATTGAAG CTGGTGCAGC GGGCTACCTA 8760

TTAAAGACAT CAACAGCACA TGAGATTGCT GATGCAATTC GGGCGACTTA TCGCGGAGAG 8820

CGTGTGTTGG AACCTGAAGT GACGCATAAG ATGATGGAAC GGTTAACAAA AAAACAAGAG 8880

CCGGTGTTGC ACGAAGATTT GACAAACCGG GAACACGAAA TTTTAATGTT GATTGCACAA 8940

GGTAAAAGTA ATCAGGAAAT AGCTGATGAA CTCTTTATCA CTTTGAAAAC AGTTAAAACA 9000

CATGTTTCAA ACATTTTAGC AAAACTAGAT GTGGATGATC GGACCCAAGC GGCGATTTAT 9060

GCTTTTCAAC ATGGTTTAGC CAAATAAATA TCAGATAAAT CCATAAATAA ATTCTAGGAA 9120

ATCCTTCTAG GAATTTATTT ATGGATTTGT TATACTAATT TTAACGAACA GACAGGAGCG 9180

GTCAGAAGAA ATGAAACAAA ATTTTGCAAT TATTGGTTTA GGACGTTTTG GCGGCAGCAT 9240

ATGCCAAACT TTGGTGGAAG CTGGTCAAGA AGTTTTGGCC ATTGACAGCA GTGAAGACCG 9300

TGTCAATGAA TATATGAATA TTGCAACGCA TGCAGTGGTG gcTAACGCCC AAGATGAAAT 9360

GACGTTGCGT TCTTTAGGGA TTCGGAATTT TGATCACGTA GTGkTGGCAA TTGGTGAAGA 9420

TATTCAAGCG AGTATTTTAG TGACACTAAT GGTTAAAGAG ATGGGGGTTC CCAATGTCTT 9480

GGCCAAAGCA GTCAATGAAT ATCATGCCCG TGTCTTAGAT AAAATTGGGG CAGACATGGT 9540

CGTTCACCCA GAGCGGGATA TGGGCATTCG AATTGCTCAC AAACTTGTTT CACGCAATAT 9600

TTTAGATTAT ATTGAATTAT CTAGCGAATT TTCATTAGCT GAGGTTCGGG TCACCAATCC 9660

TAAATTTTAC AATAAAACCT TAGCAGAATT AAATTTCCGT CAGCGATTTG GTTTAACTGT 9720

GGTAGCGATT CGCCGCTCGA AAACAGAAGT AATTGCTTCG CCAGATGCTT CTGAAATTGT 9780

TCGAGAAAAC GACAATTTAT TAGTTATCGG TGATACCGCA GACGTGGATT TATTAGATGA 9840

AAAAATGAAT GGCTAAGCAC TTTATTCTCT AAAAGCACTG CGCTGTGCTA CAATAAATGG 9900

CGACAGAGAT GTGAAGGAGG AACAAAGAAT GAAATTAGAA GTGACACCGA AAGCCCAACA 9960

ATGGTTTAAA GAAGAAACCG GGGTTCAGCC TGAGACAGGC ATTCGTTTTT TCGGAAAAAT 10020

TTACGGAAAA ACACCTGTTC ATGATGGTTT TTCCATTGCG ATGTCTGTAG AAGCGCCAGA 10080

TGAGCCAATG ATTAAAGAGA ACCTAGATGG GATTACCTAT TTTATTGaAG AAACAGATGA 10140

TTGGTTCTTT AAAGGGTATG ATTTGATTGT TGATTATGAT GAAAAGAATG ATGAACCGCG 10200

TTATCmATTT GAAGCAAACC mAGrAGaTTT AACmAAATAA TTTTAGTGCT AAACAAAAGA 10260

GGCTGAAGTA GTGCGTAGGA CTATTTCTGT TGCTTTCGTT TAACCGCTAG TCAnnGTGAG 10320

AAAAAATCGT TTTGGATTTT TTCTCAGGTC TTTCTATTTA AGTATTTTAG TTGGATATCG 10380

CTATTTAATA A 10391

(2) INFORMATION FOR SEQ ID NO: 134:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3852 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

AATGTTAAAA GGAnAAATAA TAGAGTTAAA GTTGAAAAAG tGgAGTTAGA TTATGTAGGA 60

AGTATTAmTG TkGATAkGGA TTTATTAGAA GcAGsAGGaA TATATGAATA TGAaAnAGTA 120

CAAATAGTTG ATGTTAATAA TGGAAATAGA TTTGAAACTT ATACAATAGC TGGAGAGAGA 180

GGAACAGGGA TGATTTGCTT GAATGGAGCT GCAGCTAGAT GTGTATCAAC TGGAGATAAA 240

ATAATTTTAA TGGCATATTG TGAATTAAAT GAAAATGAGG TTAAGGATCA TAGACCTAAG 300

GTTGTTTTTG TCGATGATAA TAATAAAGTT GAAAGAGTAA CATCATATGA AAAACATGGA 360

AGGTTATCAG aTATAATTCT TTAGGTTTTG TAATGTTATT TATATATAGC TAGACTATTA 420

TTGAAATATT TATTGATTAC TAATTGGATA GATGTCAATA AAGTAGACAC AAAAAGAGAA 480

GATACTTAAA GATTTTCAAA GTAAAGAGCC TCCATTTCAT TTGGTGtAAG CATATCTCAT 540

GTTCCATGAG GACGTTTGGT GTTATAGAAG CCTTCAATGT AGGCAAAACA AGCCAACCGT 600

ACTTCTTCAA TTGAATGGAA TGAACGGCGA TTTAATTCTT CTTTTTTCAT ATATTAAAAA 660

AAAGCTTCGG TAACGGCGTT ATCCCACGGA TAACCTGGTT TTGAAAGTGA GTGGACAATC 720

GGATGATCCT CTATAAATTT TCTAAATTCA AAAGAGGTAT ATTGAGACCC TTGATCTGTA 780

TGAAAAAGAA CAGAACTCGT TAAtGcgAAT GGCAACAGTT TTTTTGACAA ATTTTATAAG 840

GTGCAGAACT TCTTTCCGTA TACTATTCCG ATTGTCCTAG ACAATGAGCC TCCTCGGATG 900

TGACGATCTT CGGCAGGAGC TAACAGTTAA AGTTAGACTG CCTCGCTAGC GTTAgcACAT 960

CCGAGCTCAT TATCAAGGAT TCGcTGCGCC AATCTCTGGT TACGGTAACC AACCGGTGTC 1020

TCGTAGCCAA GTGTACCGTG CAACCGAAGt GGnTTCCACC AATTGACATA GTCAAATAAC 1080

TCCAAATCCA ATTGTTGTAA GGTTTCAAAT GTGTATTGAT AGACAAATTC TACTTTCAAC 1140

GACTTATAAG TTGATTCAGC TACGGCATTA TCAAAAGGAC AGCCTTTATG ACTCAATGAT 1200

CGATTGATGT CAAAAGTCGT TAATAATTCA TCAATAGCTT GGTTATCAAA CTCTTTTCCA 1260

CGATCAGTAT GAAAAATCTC AACCTCTGTC AGAGGTTGTC TGATACGGCT AAATGCTTTT 1320

TTTACTAGAA CGGCATCTTT ATGTTCTCCA CAAGAATAGC CGAGAATTTC TCGATTGAAC 1380

AGATCCAAAA TGAAACAGAC ATAATTCCAT TTTTTCCCGA CTCGTACATA AGTCAAGTCT 1440

GTTACGATCG CTTCTAATGG GTTGTCTCTT AAGAATTTAC GATTCAATAC GTTTGTCGTT 1500

TTGGCTTCAT TGCAAGTAGA ATGATGTACT TTAAAATAAG CAACAGTATA GCTCGATTTT 1560

AATCCTCTAT TTTTCATGAT TCTACTAATT TTTCGTCGGC TGATCTGAAT GCCTCGTTTT 1620

GATAAGGCTT TTTTTATTTT TCTTGAGCCG TAGgCtTTCG GCTGAGGATA AATTCTTCAG 1680

```
CGACTACTTC TTCAAGTTCT GATTCGTCTT TCTTTGGTTT TGATTGATAA TAATAGGACT     1740
GACGTGATAG ACCTAATATT CTGCACATCG CTGATATAGG GTAAAGATGC TTATTCGCAT     1800
CGATTACTTG TCTCTTCGTC CGAATATCAG CGCTGCTTGC TTTAAAATAT CATTTTCCAT     1860
TTCTAATTGC TGGTTTCTTT TACGTAGTTC TAACAATTCT TTTTGTTCAG GCGTAAGATT     1920
ATCTTTTTCT TTGAATGAAC CACTCGTTTT AGATTGCTTT ACCCATTTGT CAAATGCTGA     1980
AGCCGTTAGT TCATATTCTC GAATGATTTC TACACGTGGC TTTCCAGCTA AGTAAATATT     2040
GACTATTTGT TGCTTGAATT CTTGTGAAAA AGTTCTTCGT GTTCTCTTAG ATATAAAAAT     2100
TCCTCCTGGT ATGtTTTCTT CTAGTCTACA CACCTTAATT TTTCTGTCTA GTTAATTGTA     2160
GCCTATCCAG TGCTTCTGCC GGAGAAAATC AGATCCGAGG ATGCTCATTG TCAAGGGCAA     2220
TCGGAGCCAT AACACCGCAG CATTCACAAC ACCTTATAAT TTTTGTCAAA AAAAGTGTTG     2280
CCATTCCnAT AGATGAAATT CTTCAATTGA GAGTGTAAAA TATTTTGTGT AAATAGAAAA     2340
AAGGAAGTCC CTTCTGTAGA ATAGAGTTAC CACAACACAT TCACAGAAAA GAGGACTTCC     2400
ATATGAACGA TTTTACTACA GAAATTCTAA AGACyCTAGC GAACnAAAGG CGATTTGaTG     2460
AATTATTAGT TGCTTTAATT ATTTATAAAT AACTATAATG CTTTTATTAT AGTATCGTAA     2520
ATAGTTTATA TGAATACATT AATTGTAAAA ATGGTGGATG GCTCTGAACG ATTAACAATA     2580
TTTTTTTGTT TTGACTATCA TTTTATTGTT CAGAAATAAT TTATTGCACT TTTTGAGGTA     2640
AATAGTTTAG AATTATTTTA ATATAAGTAC TTCAGATTTA ATAAAGGAGA AATATCTAAT     2700
GAAAAAAATT GATTTAATAA ATATGATAGG TATGTTAATA GGGATATTAG TAAATATAGT     2760
AATTTTCACA GATTGGCTCG GGGTGTTATT TTCTAATCTT ATTCCCATAC TTATTATAGG     2820
CATTTGTGGA ATTATACTTT CAATATTAGA ACTTTTTGAA AGCAGAAATA CAATGAATAG     2880
AATATTTGCt TGTATTATAT TAATTGTTAA TCTATTGCCT ATGGTCTATT TTACTTTTTT     2940
ATATTTTGCA CTGGGATGAA ATAAATTAAC CAAAAGTTTC TTTTTCTTAT AATGTGTATC     3000
AATATTAAAA CTGAAAAAGA TATGGATAAT AGGATCAAAG CCCTCATTGG AACACTCAAT     3060
ATATTATCAG ATTTAAATTT CAAATAGATA ATTTCATTTT TTAGTAATAT CGAAAGATAC     3120
TTTTTTAATG ATGAGAGTGT AATGTAAACC GTGTAAGTAA CCATTAATCC TAGGCAGGTG     3180
GTTGcTTCGC GGTTTTTCTT TTTCTACAAT AAGATCAAGA GGTGATTACA CGATGGCTAG     3240
AAAGAAAAGA AATCCTGATG CCGAAAAGTT AGCTGAATCC ATTCTGAATG CCTATCAACC     3300
TGAATCTGTC GATGACATGC AAGATGCTTT GAAAGATGTG TTTGGGCCCC TTTTTGAAAA     3360
AATGCTTCAA GGAGAATTGA ATAATCATTT AGGTTATGAT GCCCATTCTA AAGAGCCTAA     3420
GGAACACGAT AACCGTCGAA ATGGCTATGG AACTAAAACG CTTAAAACCA GTTTTGGTGA     3480
AGTAGCTATT GATGTTCCTA GAGACCGGGA AGCTTCCTTT GAACCAGAGT TAATTCCTAA     3540
GAGAAAACGA GATGTCTCCG ACATCGAAGG GAAGGTTCTT TCCATGTATG CACGAGGAAT     3600
GAGTCAACGA GATATTGCCG CAACCGTCGA AGCTATCTAT GGCTTTGATA TTTCCCATGA     3660
```

AATGATTTCA GATATCACTG ACGCTGTCCT TCCTGAATTG GAAGAATGGC AAGCCCGCCC 3720

CCTAGCAAAG TGTTATGCCT TTCTATTCGT TGATTGTATG TATGTTACTT TACGAGAAAA 3780

TTATGAAGCC AAAGAATATG CTGTATACAC CATTCTTGGC TATGATCTCA AAGGAATAAG 3840

GAGATTTTAG GA 3852

(2) INFORMATION FOR SEQ ID NO: 135:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8805 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

ATTGGCTAAG CGTTTAAACA AATGCGTCTT CTAAATAAya GGTTTTAATG ACATGACTTA 60

nTTCyCCTTT TtGtCCATAA TTTCTTGTAA tCGTgCACGA TTTTGTTCAT TTCTTCTTGA 120

TTGACAGAAC GAATGTCAAA TTCAATGATG CCaTTGTTGG CTTGATATTC TCGCGTATAA 180

ATGGCTGGAC TTTCTGCTTT TAATGCTTGG ATGACTTCCT TAGCGGTTTT TCTGCCGTCC 240

ACTTTGACAC TGGCGCGATA AATGTCTCGG CCAGCGCCAT CTTGGACGAT TTTGGCGGTT 300

AAATCTGATA AGTTGTTGAT GGCTTCAACA AACGGTTTTA AGCGCTCTTG CATCGAAGCC 360

CCTGATTCAC TGCCATGTGC CAAGTATTCT TCCACTGCTT GCGTAAAGCC CAAGATATTG 420

TCTTTGCCAA TTTTCATTGC TCGACCAATG CCTTTGCCTT GTAGGCGCAC CCAGTCAATG 480

TATTCTTTTT TCCCGACCAC TAAACCGGCA CTTGGCCCTT CAATTGCTTT GGCGCCGCTG 540

TAAATCACCA AATCAGCGCC TGCTTCAGTG TATTTGAAAA GATCTTCTTC TGcTGCCGCG 600

TCAACAATTA GCGGCACCTT GTGGCGTTGC GCCACTTTGG CCGCTTCGGC CACCGTCAGC 660

ATACTTTTTT GCACGGTATG GTGGCTTTTG ATGTATAGAA TGGCCGCTGT TTTTTCAGAA 720

ATCATCATTT CCACGTGTTC TGGACTGCAC ATGTTGgCGT AGCCAGCTTC CACCACTTGG 780

CCGCCACCTT GCGCTACCAT GACTTCTACA GGCGTGCCGT AGTCCACGTT GTGCCCTTTA 840

GGTAAAACAA TCTCCCGCTG TTCGATTTTT TCTGTATAGG GATGATACGC GTGGTACAAG 900

CTGCCTTTTC CAATCAAGGC AGCCACCGAT TGGGCAATCC CCGCCGAAGC CGAGGAAACG 960

ATTTGGGCAT CTTCTACCTT CAACAAGTTG GCTAAAAAGG CCCCTGTTTG CACACTAAGT 1020

TCACTCATTT CAAAGAAATG TTCGCCGCCG AATCGTTGAG CTGCTAGGAC TGCCTCCGAA 1080

ACTTTGGAGA CACCTAGAAT GGTCATTTTG CCAGAAGCAT TAATGACTTC TTTTAAGTGG 1140

AATTTTtCGT AACTAATTGT CATAGATTTG TCCTCCAATA ATTGTTTTGA TTGGGCGAAT 1200

TTGCTCTTTG GCTACTCGTG TCAATCCATT AGAATCAGTC AATGTTTTTT CTTCTGCTTG 1260

AATCGTGAAA ATCGTTAAAT CAGCATCTTT ACCTATTTCC AACGTGCCTT TTTGCGTTAA 1320

ATGGAAATTC TCAGCAGGAG CCTTGGTTAC CTTTTCAATA ATTTCTGGCC AATCATAGCC 1380

GACCACACGG AGTTTTTCCA TCGTTGTGGC TAAATCATAA ACAGGACCAT TTTCTCGGTT 1440

ACGGATGTAG ATATCTGTGC TAATTGATGC AGCTTTCATT CCCTCGCGCA ACGCTGTTTC 1500

AGCTACGTGA AAATTGAAAC TATCCGTACC ATGCCCGATA TCAAATACGA CTCCTTTGTT 1560

GTACGCTTGC CAAGCGAAAT CTTTTATTTT ATCCGTTGCT TGGTCTAAGA TACCATTTTC 1620

TTTCCCATTA AAACAATGGG TTAGGACATC CCCTTTTTCC ATCAAGGCTA AAATTTCATC 1680

TAAATGAGGC GGTGCTGATC CAATGTGAAC CATTAAGGGA ATTTCTTGGT TCTCTTGCTG 1740

AATCTGTTTG GCCAATTCTA AAGGCGTAAT GCCGTTATCG CCAATAACGG TCCTACTCAT 1800

ACGGGCTTTA ATCCCTACAA CAAAATCTGG TAATTCTTGA ATCGCTTTTT TGACTAAACT 1860

CGCTTGTACT TTACTTAAAT CTGCGAGTTC GTCCTGAGCA ACGATGCCCC ATTTAGAAAT 1920

ATTGACTAAT CCAAAAACAT TTGTTTTTGC TTGcTGCGCT AAGTCATAAA ATTCATGAAT 1980

GTTTTCAGCA CCTGTTGTCC CAGCATCAAT CACTGTCGTA ACACCCTTTT TGACCCCAAT 2040

TTCATCTGGA TAATCATAAT AAAGAGCCAT TTTTTCAAAA CAATGAACGT GATCATCTAT 2100

CCAGCCTGCG GATACATAAG TACCTGGTTC TAAGTGGATA GTTTCTTTTG CAGAACCTGA 2160

AATAGTTGCT GCAACAGCAG CTATTTTTTT CTCTTTAATT GCAATTTCAA CAGGCATACC 2220

ATTTACTGTC TGTCCATTTT TTATAAGTAA ATCGTAGTCC ATATTCTTCC CTACTTTTCT 2280

ATAAATCTAA ATTTTGTCCA AGGCTTAATG AATTCTAATA ATACAAGTTC TTCTTCTATA 2340

ATACGCCCTA CTCGGTTTTT GCGTGAGTCT TGATGTGGTT CTAAAACGAC TTGTAATTCA 2400

TTTTTGTATT TTCCAAATGC ATCATTTCCC ACAACAACAT CTCCAGGTTG TAGCATTGCT 2460

TGATTATCAT GTGGGGGATT AACTTCATTT TTATATTTTT TACGCACTTC TGTGGAACGA 2520

ATCATCTGCT GAGTGATATC TCCACGGCGA AAATGTTGAT TTTTTGTTAC GATTTCTTTT 2580

TCAATTTCAG AAGTCGCTTC TTCAAAAACA ATGGTCAATT CTGTTTGATA ACGATTCAAC 2640

TGACCCAATG CTTCTAATTC TTCGTTGGAA GCATAAGCAT TTCCAATAAC AACATCATCA 2700

ATTAAATTTG TTGCAAATAA ATGTTTAGCT TGAACAGTGA CTGGCAACTG TCTATGCATT 2760

TCTAAAGTAG GTAGCCCATC ATTCACATCC CAAGGACCAA TTTCACCTAC CTGTGATGAA 2820

ATAAAAGCAG CTGTCCGAAT ACCTTCCTTT TTAAAACGGA TGCTACATTT TTCAAAGAAA 2880

TCATACGGCA ATGCAGTTCC CGCTTGTGGA TAAAAGTTAT GACAACCATA TAAGAAAGGT 2940

CGATTCGCTT CATACGTCAA AATGTTATCA AGATAAGCGA CATCATTACT CATATTCAAC 3000

TCAATAGCCA CGCCGAATGG ATTAAACGTC AACATGGCTT CTTTATTTCC ATCAAACCCT 3060

GTATCCAACC GAATCCCATC AGCTCCTAAT TCGCTGAAAA AAGTCAAATC ATCATATGAA 3120

ATTTCCAATT CATCAAAGAT ACTTGGAACG ACATCTAAAA TTGTTTCAAA TCCTAAATTT 3180

TTCGCATAAC TAATTAATTC TTTAAATTTT TTCTGAACCG CTTCTTTGCC ATCGGTCACT 3240

TCTAACATAC TCATAAAAAT TCGAGTAAAC CCACATTCGC TGGCTTTTTT TAAATACGCT 3300

TTATCTTGAT TGATATCACT ATGATCTGGA TAAACTGATA CACCTAATGC TCTTTTCATC 3360

TTCGTTCTCA CACTCCGCTA TTTTCAAATT TATTTCGAAA TTTCCAAGCA ATCACCGAGC 3420

TACACAAGAT TGTTGTAGAG CCAACGAGCA ATACTGGTTG ATTTAGAAAT CCGCCAATTA 3480

TCATGGAAAG AGTAATACAA ACTGCCGTAA AAACCAACGT CTTTACCAAA CCAAATTGAT 3540

AAACAACAGT AGGCAACTGC CCTTTTTCTA TAATAATTTG ACTAACGATA AAAAAGAAAA 3600

TGAACCAAAT AATCATCTTA TACGCTCCTG ACTTTCTTCT CTGACTTAGA AGGGGCAGGA 3660

TATTCATATC AGAACACCCG CCCTAACATT TTTTAATTTT TTGCTGCTTG TTCAGCTGCT 3720

AATTCTTCTG CTTCTTTATC CAATTGTTGT TTTTCATAAA CTTTGAAAAA TGGATAGTAG 3780

ATAGCCATTG TAATTAAGAA GTTAACAATC ACTAATACTG CGGCCGAGAA ACTCCAATTG 3840

GTACTCATCC AAGCAGCAAT TGGTGCTGGA ATTGCAAACG CTAAGCGTGC AGCCATCATC 3900

GGAATTATAT TAGCGACTGT TAAGAAATAA GATAACGTTG TGGTAATTAA CGGTGCCACG 3960

ATAAACGGAA TACCTAAAAT AGGATTCATT ACGATAGGCG TTCCAAAAAT AACTGGTTCA 4020

TTAATATTAA ACAAGCCTGG TAAsAACsAT AAACGGCCAA GACTTTTCAG ATACGTTGAT 4080

TTAGAAAACA TCATTAAAAC AACTAAAGAA AGCGTAGTTC CAGCGCCACC AATCCAGATA 4140

AACCATTGTA AAAACTGTTC TGTAAAAATA TTAGGTAACT GATGAACGTT TGCGCCATTG 4200

GTAAAAGCTT CAAGGTTTTC TGCAATCGAC ATATCCCAGA AAGGACGAAT AACTGGGCCC 4260

ATAATCGCTG GACCATGAAT CCCTAAAACC CAGAAGAAAC AGATTAAGAA GACTGTTAGT 4320

AAGCCGCCAA ATAAACTGTT ACCAGCAAGG ATACCTTTAA GTGGCATTAA TAACGTACTC 4380

AAGAAACCAT TCAAATCAAA ACCAATTACG TGGCGAATAA CCCAGAATAA TAAAAGAATA 4440

ACTGCTCCTG GAATTAACGC AATAAATGAA TTTGACACTT CAGGCGGTAC GCCATCTGGC 4500

ATCTTAATCA TAATATCTTT TTCAATAAAG AAGCGATAAA TTTCAACAGA TAACAGCGCT 4560

GTTACAATCG CACCAAATAA AGAAGCTGAA CCTAAATTTG CTAAATTAAT ATAGCGCCCT 4620

GCTGTAATGA CGTTGTCAAC ATCTTCAAAG ACACGTGTTG GTGGTGCCGC TGTTACTAAA 4680

AAGGCCATTA ATGCTAAAAT TCCGCACGTC AAAGAGTCAA GCTTATAACT TTTGGCCAAT 4740

GATGAAGCAA TCCCAAATGT TGCATAAAGT GCAAGTAACC CCATCGTATA GCGGGAAGGA 4800

ATATCTAAAA TTGCTTGATA AGGTTCAATC AATTTTGTGT ATGCCTCAAT TGGAATATTT 4860

TGAAAAAATTG TAAAGAATGA GCCCACAATT GTTAAAGGCA TCGTTGCAAT TAATCCTTTT 4920

CTAATCGCTG TCATATGGCG TTGCGAACCA AATTTATTCG CAATCGGCAT TAACTTAACT 4980

TCCATCCACT GTACAAAACC ATTCATTTGT TAATTCCTCC GTTCAACTAA TAAArATAAT 5040

AATTATTTCG CATTTATAAT AACATAGTAA cATTATAATG TAAaCGCTTA TTCTTATTTG 5100

TGACAGAGAA AGTGCAAAAC AAAAGGCTTT TTCTTATTAT AAAGAAACTG GGaATATTGC 5160

ACCAAAAATC ATTGCGCCTA AAATTGCGCC CCCAGCGATT GGTTTTTTCC AAATGTAAAA 5220

ACCTAAAGCT CCTAATGTAG CACCAATCCC AATAGGAATC GAAGCACCCG CCGCACTCAA 5280

AATAATTAAC GGGCCTAAGA AGCGTCCTGA GGCATTACCA GCCCCCATCA TCACATCCGC 5340

ACCAAAAGTG GAATTTCCTT GATTGATAGT AAATTTCCGA ATCAACACAA TTACGCCACC 5400

```
AACTGCTAAC CCTAGTAGCG CACCTGTTAA AAGTGCTAAG CCAAAGTTTT CTACAGGTGC    5460
TTTAATACCC ATGCTGAGGA AAATTGCAGG TACGCCAATC CCGATCCCCG TTAAAATTGA    5520
GCCACCTAAA TCTAAAATCC CGACTAAAGA ACCTTCTAAA ATTCGTGCAA ACAGGAAACT    5580
TGCACCAAAT GCTGCGGCAG cGCCATaAGA ACCACCATCT AAACCAGCTT TTAGCATCGC    5640
AACAATTGAA ACTTCATTAA AGGCGCCCAC ACCATAAACT ACATACATAT GTGTTCCTGC    5700
AAAAATTGCC GCTGACATAA TCCCGACTAA AATGGGAAAG CCCCAATCAG CATACCAAAA    5760
ACTTTTGCCT TTTAAACCTT CTTGAATTTC TTGTTCTTGT TGTGCCCATT CATTTTTTTC    5820
TTGCTGTTCC ATCTATGAGA CCTCCTTTTT TCTATTTAAC GGTGAAGACA TTGTGTAAAT    5880
TATTTAACCA TTGAGGAATA GACAAGCGGA AACTTTCGAT CATTTCTAAG TCGAACCCTC    5940
GGAAGAAGCC ACTTAAAATA AATAATAAGA TAATGACACC AAGCATGCCG CGTGTCATTC    6000
GGGTCCACCC TAAATCATCG ACACCTTTAC CGATTAAGAT CCCTAATACA ACTCCTGGGA    6060
CAGCATTGCC CATGACTAAT TGAGATAATC CGCCAAATAA TGTTCCCCAG AAACCCGTCC    6120
GTTTTCCTGC ATCTAAAGCC GCTAGCCAGA ATAAAACAGG CATCACAGTA TTGATTAAAA    6180
TCGTTGCTGC AGGTACCAAC ACCGCAACCG CTGTTACTTG TAGCGATTCT GGAATTGCTG    6240
CAGCTGTACT ATTTAAGAAA GCCACAGTCG CCATTCCTAC AAAGACCCCG ACAATCCCCA    6300
TTTTTTTGGG ATTATGCATC GTTTTAGTGA CATCTTTATC CTTTACCAAT AAAATTGCTG    6360
CAGCCCAGTT AGGCACAATC CGGTGGGTCA CATCTTGTGT AAAGGCCCCT GCTCCGACGG    6420
CCGAAGCCCA AGCATTAAAG AAAAAACCTA AACCAAAAGA AAAGTGTGAT GCTGGATCTC    6480
CTTGTGCCGC GTTCATTTCT CCTAATGTTC TAAATGCCCC TAACCCTTGC GTTTGTGGTG    6540
CATGAAACAT CCGAGCAGCG CCTGCGCCAG CTGCAAAACC TAGTAAACCA CCAATAAGCA    6600
ATGACTTCAG CAAAATGATT ACAAATTCCA TTCCTTCTTC CTCCCTTTAT ACGCCTCTTT    6660
TTGAAAGAAA CGGAACTACT ACTGCGTTCG GTTCTTTAAT TGTTTCCGTT AAAAAATCAA    6720
CTTCTTCGAC TGCAATGGCT TGGATTTCCA CTTCTACAGC TAGTTCGACA TGATAAATCA    6780
CTCTCTTTCT AGGAAAGAAG AAGAACAGAA ATTTTTCTGT ATAAACTTGT TCTTTTGCCT    6840
GACGAATAGT TACTTCTTTC GGTTCAATCC GGATGAGGAC CTCTGTTGTT TCTTTTAATA    6900
CTTTTTTTGA AACATCATTC AATGCACTTG CAAAAGCTTC ATTCTTCTTA TTACCTTTTC    6960
CAGAAACAAT CACGGTTTGT GTTGACCGTT TTGTGGCATC CATAATTACC GCTCCTTTCT    7020
AAGTTACTAC TTTCTATGCT CCGTGTTTCT TAATAAATGC TTCAGTTACT CGTTTACCAA    7080
GTTCTTCCTG ATCCATAAAA CCAAATCCAA GAACAGTTTT CCCATCGTTA ATTGCTGTTA    7140
CCCCTTCTTC GATTGATCGC ATGCCATGTC TTTCGGGATA CCCGTATTTA GTAGCAGCCG    7200
TTAACGCTCC TGCACCACCA CTTCCACAGA AAGAAATGCC TAAATCAGCC TTTTCTTGTT    7260
GCATCACATC TCCTAAACGC ATATCAGCGC CCATTCCTGG AACAACAACC GCGCGACCAC    7320
CGGCTTGCTC CACTCCTTTT GCAACATTTT GTCCTTTTCC TAAACGATCT GCGATAACTA    7380
```

CTGTAATCAT AACTTTTCCT CCTATTAATT GTTTTGTTTA GCAACTTCAA AATGAACAGC 7440

TAAAACATAC CATTCATCTT CTGATAACTG ACCAATTGCT TCTGTTGTTT CTTCTGCAAT 7500

TTTCAACGTT TCTGGCGTTA GTTCCGTAAA TAACGTCCGA TCAACACCAG CTAATGTCGT 7560

TCCTTCTCTT GAACGATTCA GCATCTCATT GAGATGGTTA ATCAAAATCG TCCACTGTAA 7620

TTCCGTTGGT TCAATTTGTG CTTCTGCTAA TCGCGCACTG GTTTGTTTGA TAATCGATTG 7680

TAGCTTTTCT GGTTCAGCAC TTTGTCGAAT GATCTCCTGT GCTTCCATTG CTAGTTCCAT 7740

CCTCTTCCTC CTTTGGTTGT TCTAAATAAT TAAGTAATGC TGAAATTTCA GCATGATTGA 7800

CAGGTAACTC ATTTTTAGCA AATAAACATT CAATTTGCCG AATTTCTTCT TGCCTTTCAT 7860

GAGATACTTC TCTATTACAA GGAATCTCTT CGTATTGTTT ATTAAACGTA ATCCGATGTG 7920

TCATTAATAA AACGTGAAGT AAAAAGGCTT CTTGATATTC GCTCAGTAAC TTTGTTTGAA 7980

ACATTCGTTT TAACTCTTCA TAAATAATAT ACGCTTTTAC TAATAAATCT TGGCTTTGTA 8040

TCCGTTCATT TTTTTCAATT ACATCAGATT TCGTGGAGCG AGCGGTACCT TTCCCTTGAC 8100

GTTTCCCTAA TAAAATTTTA TCGACCTCTT CCCGAATCAA ACTTAAGTCT GTGTCCGTCA 8160

ACAATGGACT AACTTTAATA AATGGGTATG TCATTTCTTC AATTGGGAAA ATACTCAATA 8220

CCAAATCGGG CTTTTCTGCT TGAATCACTT CAGTTGCATT TAAAACAGAA GCAAATCCTG 8280

CGAGTTCAAC GTTTGGTACT TCCTCTAAAA TTTTTTGTTC AATTAAACTG GTCACACCTA 8340

AGCCCGTCGA GCAAACATAC ACAATCCTAA CAACTTTAGC TTCCTGCTGT TGTCGTTCCT 8400

GAGCTGCCAA AAAATGCAAC GCAATATATG CGCAGAAAGA ATCGTTGACT AACAAAACAG 8460

AACGTTGAAT TAATTGTTTA CAGGCACTTT GAATAGCTAA AAACAACGCT GGGTGTTTTC 8520

CTTTGATTTC TTCAGAAAAC GGATTGTATT CATTAATAAA TAAATGTTTC TTTGTAAAAC 8580

GTAAGGACAG GTGTGCAAAC AAGTTCGTAA ATAATTGGCG ATCATTTACA AAAGGATAAT 8640

TCAGCTCTTT TGAAACGCTT TCAATCACTT GCTTTGTTAA TTGAACCATA TCTTGCCGTT 8700

GATGGGTTCC GAATAAATCA CTATAAATAT AAAAATATTC ATCCGCTAAT AAAGGCAAGT 8760

CATACCGTTC AAACACCTCT TGTAAAAGTA AAAGTGGTAt TCGTT 8805

(2) INFORMATION FOR SEQ ID NO: 136:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2738 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

TATTTTTTGC TTAATATAAT TAATGACATC TTTCTTTTGT TCTTTAGCCT CTGATATATA 60

TTCCTTTAGT TCTTTATCCG TAATTGTTGA ACTGTTAATT CCTTGTTCAT ACAATTGTCG 120

TCTAGTTTTT GAAACATCAT CTGTTGGATT CAATTCTCTA TTCTTTAAAT TCTCGGTTGT 180

TTCAATTGTT ATAGATGATG TAGAAGTTGG TGTTGTGTCC GGATTACTAC TATCTTTTTT 240

AGAAGACTGG GTAGTTTGCT GCTTAGTTGA TACTTCAGTA GATGAAGCGT CGCTATTTTT 300

TCCCCCAGTA TTATTACTAC ATGCCGTTAA AAATAAGGCT AACAATAGTA CAATAAGTTT 360

TTTTTTCATT ATATCCTCCT ATAAAATAAC GAAACATGGT CGGCCTTCTG TCCAATCAAT 420

TGGATCATTT GACGGGGTAT TCCAAAGTGC TTGTGTTGAC ACCCAACCTG AGTAAAACTG 480

GCGACCATAC GGATCATAGG TTAAAGTGTA ACCATTTTTG TACTCATGAA CGACAATAGC 540

ATGTCCAATC CCAACATTTT GATATAAAAT AACGGGATAT CCTTCAGAAA GATAGGCGTT 600

TAATGTAGCT ACATCATTAA TAACTTTGAT CGAACGGCCA TATGATTGAG CTGTTATCAC 660

TAAATCAGTT GCAGAAGCGC CAAAGTATCT TTGATTTAAC CCGCCATAAG ATGCCATTCT 720

TGTTGCAACA CTGACAGGAT TTACATTATA ACCGTAAGAG CCTCTTAAAA TCATAGATAA 780

ACTTGTTGGA ACACATCCTG cTGCGCCAAT GGTATAGTTT CCGATATAAG TATTAATCCA 840

TTGTGGGTCA CGTTGAGAAT AATAATAAGG TTTTTTCTTT AATCTTTCTG CACGTAATCt 900

GTTTGCTATC TCTGCTGGTG TGcGAGCTTC ATCGCCAACT AAAAAGGmTT TTCCgATcCT 960

TTAGGAGCAG TTTCrCCyTT TCTAACTAyT TTGATTTGrA TrGCTTCyAA ATGTTTAGAk 1020

GCAGCTTGAC TTCCAGCTGr GGCrCCATTA GCTGCCCAAT TyAACCAACC TTTATCTTGG 1080

ATATGGACAC GGTAATAGAC ATCATAATAT CTTGAAACAT CACCGGmCAG TTTAATTTTA 1140

ATGGCTTCTA ATTGTAAATT TTTTCCAGTT GTGCCAGAyA ATGCCCCATT rGCTTTATAA 1200

kCTTGCCATC CGATATTTTG GATGTGCGTA GAGTATTCTA TrTTTCCAGC TAGAGCAGCA 1260

TCACTCAGGT TGATTTTGAT TGCTTCCAGT TGTAAATTTT TCCCACTTGT ACCAGCTATT 1320

TCrCCATTCT TGACAACCCC TTGCCAACCA ATATTTTGCA CGTGTGTTTG ATArTTGACA 1380

kTTGGTTTGA TTTCATCTGG ACGTTTTGCT TCATTTCCAA CTAAAAAGGA TTTTCCaTTT 1440

CCTTTAGGAG CAGTTTCGCC CTTTCTAACT ACTTTGATTT GGATGGCTTC TAAATGTTTA 1500

GATGCAGCTT GACTTCCAGC TGGGGCGCCA TTAGCTGCCC AATTTAACCA ACCTTTATCT 1560

TGGATATGGA CACGGTAATA GACATCaTAA TATsTkGAAA CATCACCGGA CAGTTTAATT 1620

TTAATGGCTT CTAATTGTAA ATTTTTTCCA GTTGTGCCAG ATAATGCCCC ATTGGCTTTA 1680

TAATCTTGCC ATcCGATATT TTGGATGTGC GTAGAGTATT CTATGTTTCC AGCTAGAGCA 1740

GCATCACTCa GGTTGATTTT GATTGCTTCC AGTTGTAAAT TTTTCCCACT TGTACCAGCT 1800

ATTTCGCCAT tCTTGACAAC CCCTTGCCAA CCAATATTTT GCACGTGTGT TTGATAAwTG 1860

ACAGkTGGTT tGATtTCATC TGGACGTTTT GCTTCATTTC CAATTAAAAA GGCTTTTCCA 1920

TTTCCTTCAG GTGCCGcTTC GCCCTTCTTC ACTAATTTAA TTTGAATGGC TTCTAATCTT 1980

TTAGCGGCGC TTTGACTACt GCGGACTCAC TGTTTTTTGC CCAATTCAAC CAACCTTTAT 2040

CTTCTATATG CACACGATAA TAAACATCAT ATGCGTTAGc tATTTCTCCA GTTAAGCGAA 2100

TTTGAATCGC TTCGAGCCGC AAAGACTTTC CGCTTGTACC AGACAATTGA TTATCTTTCA 2160

CATATCCTTG CCACCCAATT TCTTGGACAT GTGTTCTATA TTCTACAGAT CCAGcCAAAT 2220

CTGAATTTGA AATATTCATT TTAATTCCTT CAAGACGTAA AGAGCGTCTG CTAgkcCCAG 2280

AAATTTCCCc GTTTTTGaCA AcCCCTtGGC cAwCCGATGT CTTGGGATAT GTGtCtGATA 2340

ATTAATtCAG gAcTTTTyCT cTGCtTTGCC gGTGCTATAG GTGTGtCCTC ATTACTTkGT 2400

GATGTTyCCA AAGTAGTTtC AGTAcGTCGC TACTATyTTC TGTATCATTA ACAGTTGTTA 2460

CTTCTGATGT AGAAACTTCA GTGGTGGAAT CGGAGGTTGA GGAAGTGGAC GTTGTTTCTG 2520

TTTCCTCTTG AGAATTTATG ATGCTTGGGT TCATTTCTTC AGTGACACCG CTACTCATTT 2580

GTCGCTCCGT TAGAGATGTT TCTGAAGCCA CCGTTTCTTC AGTAGTTCCC TTAGAGCGTT 2640

CCAAAGCGTA AGCAGAAATA GGGGTAACGA GAAATCCCGA CAACAAAACC GCTAACATAG 2700

AACTAGTTAT AACACTTTTT TTCATTAATT TACTCCTT 2738

(2) INFORMATION FOR SEQ ID NO: 137:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5550 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

GATTGGACGA AAGTAACTAA TAAATCTGTT TTTATAACTG GTGCTTCAGG CATGATTGGG 60

ACCTTTTTAA TTGATGTGCT TATGTACCAA AATAAAACGA GAAATGCGAA TATTTCGATC 120

TGGGCGATGG GACGAACATT GTCGCGTTTA GAAGAGCGAT TTACTAGTTA TCTAGAAGAT 180

CCGTTATTCC ATATAGTTAT CGGAGATGTC ACAGAAGAAA TTCAGATAGC AGAAGCTTGT 240

GACTATGTGT TGCATTGTGC TAGTAATACA CATCCTAAAT CATATGCCAG TGATCCAATT 300

GGAACGATTA TGACCAATAT TGCTGGTACG CAGCATATTT TAGATTATGC GGTCAAAGCG 360

AATAGTGAAA AAGTGTTATT TTTATCAACT GTTGAAATCT ATGGAGAAAA TCGTGGAGAT 420

CTTGATAAAT TTACGGAAGA CTATTGCGGC TATATTGACT GTAACACGCT ACGGGCCGGT 480

TATCCCGAAG GAAAAAGAGC AAGTGAATCA TTATGTCAAG CCTATATTCA GAAATATGGT 540

ATTGATGTTG TTATTCCAAG AATTAGTCGT ACTTTCGGGC CTACGATGTt ACTTTCGGAT 600

AGTAAAGcTT CTTCTCAATT TATTATGAAT GCAGTCCATA AAGAAAACAT TGTGTTAAAA 660

AGTGCAGGGA CCCAACTATA TTCTTATGCA TATGTTGCAG ATATTGTATC TGCTCTTCTG 720

TTTTTATTAG TTAAAGGGCA AAAAGGGGAA GCTTACAATG TTTCAAACGA ACATTGCGAT 780

ATTACGTTAC GTACTTTTGC GGAAACATTA GCCAATGTGG CTGGAACAAA AGTGATTCAT 840

GGAGAAGCCA CCGTCCAAGA AAAACAAGGG TTTTCTAAAG CAACAAAAGC ACTACTAGAC 900

AATCAAAAAA TCTACGCATT AGGTTGGCGG CCACTGTACG ACAGTATGGA AGAACCATTA 960

AAACATACAA TTAAAATTTT AGAAGGTAGC GGAAATTGAT ATGCCCAAAA TTAGTATTAT 1020

TGTTCCTGTA TACAATGTAG AAAAATATTT AGAAAAATGT GTACGCTCTA TTTTAGCTCA 1080

AACGTTTACT GACTTTGAAT TAATTCTGGT GGACGATGGC TCTCCAGACA GTTCTGGAGC 1140

AATGTGTGAT CAGTTTGCTG AACAAGATCA ACGGGTAAAA GTTATCCATA AAGAAAATGG 1200

TGGGCTAAGC GATGCTCGTA ATGCTGGAAT TGAAATAGCA ACAGGTGAGT ATTTAGGTTT 1260

CGTAGATAGT GATGATTACA TTGCAGATGA TATGTATGAA CTATTATATA CAAATATCGT 1320

AAAAGAAGAT GCCGATTTAT CAATCTGTGG TATTTATGAT GTCTATGAGG GGAAAGAGCC 1380

AATTGTAAAA AGTTTAATAC AGGGAACTTT TTCCAGAGAA GAAGCATTAT TGTTAATTTT 1440

ACAAGGAAAT ATTATCTCCG TACATGCTGT AAATAAGTTA TATAAACGAA AACTCTTTGC 1500

GGATTTAAGA TATCCTAAAG GAAAGTATCA TGAAGATTCA TTTATTATAG TAGATTTGTT 1560

GTCTGAATGT CAGAAAGTCT CTATTGATTC AACACAAAAA TATTACTATT ATCATCGAAT 1620

GGGCAGCATT AATACAGAAA CATTTTCAGA TAAACAATTT GAGTTTATAG AAGCTTGGGA 1680

GAAGAATGAA CTTAAATTGA AAGGTAAGGG AGCAGTTATA GAAGAAGCTG CCCATCAAAG 1740

AGTTTGTTTT GCGAATTTTT TAGTATTAGA TAAAATTTTG ATTAGTAATG CTCCGAAGAA 1800

AAAAGAAACG AAACAAATTG TTCGCTATCT TCGGGAAAAC TTTATTTTTA TTATGAAAAA 1860

TAAAGTTTTT ACAAAAAGTC GTAAACTATC AATGATTTTA CTTATGTTTG GTCTTCCTTT 1920

TTATAaAATA CCAATTAAAT TGAAAAGAAA ATAGTCAAGT CCAGACTCCT GTGTAAAATG 1980

CTATACAATG TTTTTACCAT TTCTACTTAT CAAAATTGAT GTATTTTCTT GAAGAATAAA 2040

TCCATTCATC ATGTAGGTCC ATAAGAACGG CTCCAATTAA GCGATTGGCT GATGTTTGAT 2100

TGGGGAAGAT GCGAATAATC TTTTCTCTTC TGCGTACTTC TTGATTCAGT CGTTCAATTA 2160

GATTGGTACT CTTTAGTCGA TTGTGGGAAT TTCCTTGTAC GGTATATTGA AAGGCGTCTT 2220

CGAATCCATC ATCCAATGAT GCGCAAgCTT TTGAATATTT TGGTTGATCG ATATAATCAT 2280

GAATCAATCG ATTTTTAGCC TCACGCGCTA AGTTAATATC TGTGAACTTA AAAATTCCTT 2340

TAACAGCTTC TCTGAAAGAT TTTGAATTTT TTTTAGGAAT GGTGGTAAAG ATATTTCTTA 2400

GGAAGTGAAC TTGGCATCTT TGCCAACTTA CGTTGGTGAA GGATTTTCTA ATGGCAGAGA 2460

CTAATCCTTT GTGCGCATCA GAAATAACGA GTTCCGTACC TTGTAAACCG CGTTCTTTTA 2520

GGTATTCAAA AATGTTGTC CAGGTCTCTT CGCTTTCGCC ACTTTGAATC ATGAAGCCGA 2580

TAATTTCACG GTCGCCATCT TTGGTTATTC CAATCGCTAT ATGACAGCTT TTTGAGAGTA 2640

CTCGATTTTC TTCTCGTACT TTTATATAGA GTACATCGGT CATTAAGTAA GGATAATTTT 2700

TTTCTGATAA TAAACGATTC TGCCACTCGT TAACCATAGG TTCTAGCTGT TCTGTTAAGC 2760

TAGAAACGAA GGACTTAGAG ACGGATTTAC CACAAAGTTC TTCCACAATT TTTGATACTT 2820

TACGAGTTGA AACGCCTGAT ACATACATTT CCAACATTGA AGCCATGAGG GCTTTTTCGT 2880

TTCGTTGATA ACGTTCAAAC ACTGTGGGTG AAAAATGGCC ATCACGTGTT CTGGGTACTT 2940

TTAATTCTAG CGTGCCTACA CGTGTCGTAA AGcTGCGCTC ATAATAGCCA TTTCGTTGAC 3000

TTTGTCGGTT TTCTGTTCGT TCATATTCTT TTGCTTGAAT ATATTCTGTT CGTTGATTTT 3060

```
CCATTAGTTG ATTAAATACC GTTGTTAAAA TATTTTTAGA AACGTCATCC TTTACAGAAT    3120
ATTCAATAAT GCTTTGAATC TCTTCGCTTT TCAGTGTAAA ATGTACTTGG GTCATGTAAA    3180
AGTCCTCCTG GGTATGTTTT TGTCGTTAAA AACATTGTAC CGTAAAAGGA CTGTTATATG    3240
GCCTTTTTAC TTTTACACAA TTATACGGAC TTTATCGACA AACTGATAAA TTAAAAGATT    3300
CAGTAAATAA TTCCATCGAA GATATTATCA ATGAAAAGAG TTTAACTTTA AGTAATCAAG    3360
TAAAAGAATT AGAAAAAGAA TATGAAAGCA AAATACAAGA AGCAGAAACA ATAAAAGATG    3420
AGAGTATATT AAAAGCAGCA ACTCACCACT ATGAAAAACA GAAACAAAAA ATTGAGGAGT    3480
CTTTTACAGA AGAAATTAAC AAAGAAGTTG AAAAAACGAT CAATACTGTT GTGGAAGAAC    3540
AAATAGAAAA AGTGGAAGAA AAGAAAAAGA AGACAACAGA AGATGATGTT CGAGATCATT    3600
TACGAGGATT TGCCCGTACA ATTCCTGCAT TCTTAATGGC ATACGGAGTT GAAGATACAA    3660
GATTAGAAAA TTTTGAGGTT AATATTGATG AAGCCACATT CGAAGATTTG ACGAGTATTA    3720
CAATAGAGGA ATTCAAGAGG CTCCGTGATG GATTTGAGTA TCCAGATAAT AATGGGGAAA    3780
CTAAAGTTAT CCCAGGATTA TTCAATGAAG TTGTATTTAA TGCTAGTATC CAGGAATTTT    3840
TGGAGACAAA AAACCGTTTG GCAAATTATT TTGATGATAG CTTGCAAGAG GACATTTTTG    3900
ATTATATCCC ACCACAGAAA ACAAATCAGA TATTTACACC AAGACGTGTT GTAAAATCGA    3960
TGGTAGATTC CTTAGAAAAA AGAATTCCAG GAATTTTTTC GGATAAAACA AGGAAATTTG    4020
CAGACTTATA TGTTAAAAGC GGTATGTACA TTACAGAAAT AGTCAAGAGA TTAAATCATG    4080
GATTAAAAGA TCAGATTCCA AATCAACAGG AACGTATTAA ATGGATTCTT GAAAATCAAG    4140
TTTATGCATG TGCTCCAAGT AATATTATTT ATAATATTGC CAAGAACTAC ATATTGGGAG    4200
AAATGGAGCA TGTTTCTAAT ACAAATATTA TTGAACTTGA TACACTGGCT TTGGCAAAGG    4260
AAGCAAAGTT ATCAGAATGT ATTTATAAGG AATTCGGAGA TGTTAAGATG AAATTTGATG    4320
TAATTATAGG AAATCCTCCA TATCAAGAAG AAAATACTGA AAGGAATAGA GATGACGCTA    4380
TCTATCACTT ATTCTTAGAT GAGGCGTACA AGATCGCAGA GAAGGTTGTT ATGATTACGC    4440
CAGCAAGATT TTTATTTAAT GTTGGATCGA CTCCAAAAAT TTGGAATGAA AAAATGCTAA    4500
ATGATAAACA CTTATCTGTT GAAATGtATG AACGGAAAAG TGaAAAAATC TTTCCTAATA    4560
CaTCgATTAT TGGTGGAATC GCAGTCACTT ATCgTGATGC CTCAAAAATC CTAGGACCTA    4620
TTGGAACTTT TACACCGTAT CAAGAATTAA ATTCAATATT GAGTAAAGTA AAAGAAATTA    4680
ACCAAGAAAG TAACTTTAGT TCACTGATGT TCGTACAAAA TAAATTTAAT TTAGATAAGC    4740
TATATCGTGA TTTTCCAGAT TTTGAAACAC GTCTTGGAGG AAATGGACGT GAACGCCGTT    4800
TAACATCTTC AATATTTAAT GTATTACCTG AAATTTTTTC AGATAAAAAA ACAAAAGAAC    4860
ATAATCTGGC AATTTACGGT AGAGAAAaTA ATAAGCGAAC aTATAAATAT GTCTCTAAAG    4920
AATATATTGA GGAACATCCA AATATGTATA AATGGAAGGT TTTTGTTCCA GCAGCAAATG    4980
GTGCTTCGGG TACAATAGGA GATAGTCCAG CAAGAATTAT TAGTTTCCCG GTCTTAGGTG    5040
```

AACCTAATTC TGGTCATACC CAAACCTTTA TAAGTTTAGG AGCTTTTGAT ACGAAGTTTG 5100

AAGGTGAAGC GTTGTTAAAA TACATCAAGT CAAAATTCGC CCGTGCGATG CTAGGTGTAT 5160

TAAAAGTAAC GCAAAATAAT AAGACTCAAG AAACATGGTC AAAAGTGCCT GTATTAGATT 5220

TCACAAAAGA TTCTGATATC GATTGGGAGA AACCCTTGCC TGAAATTGAT CAGAGTATAT 5280

ATAAAAAATT TAGTTTGAAT CAAAATGAGA TTGATTTTAT AGAAAACCAT ATAACGTACC 5340

TTGATGAAGA AAGTTTAAAT AATTTATAAG TCAAGATAGA TACTTGTGAT TAAAAAAGTA 5400

ATTGTATTCA ACTGGCATCT ATTTTTGGTT CAAATTGGTT CAAGTAATAT CAATATCAAT 5460

CTATTTCCAT ACATTTTTAT CTATATAACA TTTCCCAAAA CGTTGATATA AATGGATTAG 5520

GATGTATTCA ACACACGTCG TATTTTTATA 5550

(2) INFORMATION FOR SEQ ID NO: 138:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28690 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

AGAAGAAAAT AnCATCAAAT ACGCTTTCTC TAGATGGGTT AAAATGCTTG GAGTACACAT 60

TACGATACGA TTGTTGCTTC CATTGTGATT CCTTTTGAAG AAAAAGATGG CTnTGTGACA 120

GAACTATCTG CAGGCCAAGC CTTTAACCGC ATGATGGGAG ATCCTGACAA AGAGGTTCGT 180

CAACGTTTGT TTACAGCCTG GGAGAAAGCT TGGAGTGGCA AAGCTTCAAT TTTAGCAGAC 240

ACCCTAAATC ACTTGGATGG GTTCCGTTTA TCTGATTACA AATTGCATGG CACTACAGAT 300

TATTTAGAAA ATCCCTTAAA CTATAATCGA ATGAGTCAAG AAACATTAAC GGTTATGTGG 360

GaAACCATCC AAAAAAATAA ACAACCATTC GTCGACTTTC TAACGCGCAA AGcGCAAtTA 420

TTTGGCAAAG AAAAAATGGA TTGGCAAGAC CAAGATGCGC CAATTATTTT AGGTGATTTA 480

GAAGAAAAAA CGTATTCCTT TGATCAAGCC GCTGCCTTTA TTTTAGAAAA TTTCCGTAAA 540

TTCAGTCCAA AAATGGCTGA TTTTGCCCAG ATGGCTTTTG AAAAAAGTTG GATTGAAGCT 600

GAAGATCGTC CAGGCAAACG TCCTGGGGGT TATTGTACAG AATTACCTGA AACACAAGAG 660

TCGCGTATTT TCATGACCTT TGGCAATTCA ATCAACGAAG TAGCTACTTT AGCCCATGAA 720

TTAGGTCACG CTTTCCATAG TAGCGTGATG TGGGATTTAC CTTCATTGAA TCGAGAGTAT 780

GCCATGAACG TAGCAGAAAC GGCTAGTACC TTTGCAGAAT TGATTGTTGC TGACGCCACC 840

TTAAAAGAGG CCAAATCAGA TGTTGAAAAG ATTAACCTTT TAGACACAAA AATGCaAAAT 900

GCCaTTGcCA TGTTTATGAA CATTCaTGCG CGtTTTATCT TTGAAAATAA TTTTTaTCaA 960

GCGCGCCAAA AAGGTTTAGT AGCGGAAGAA GATATTACTG AATTAATGTT AGCGGCGCAA 1020

AAAGAAAGCT ACAAAGATAG TCTTGGCAGT TATCATCCTC ATTTCTGGGC GAGCAAATTG 1080

CACTTCTTTA TTGATGATGT GCCTTTCTAT AATTTCCCAT ATACTTTTGG CTATCTATTC 1140

AGCATGGGTA TCTATGCGTA TGCCAATCAA CAAGGCAGTT CTTTTGAAGA CCAATACATT 1200

GCGTTATTAA GAGACACTGC TTCGATGACA AGCGAAGAAT TAGCGAAAAA ACATTTAAAT 1260

GTTGACTTAA CAAAACCTGA TTTCTGGCAA GCCGGCATTG ATCAAGTCCT AAAAGATGTG 1320

GAACAATTCA TGACTTTAAC TGAAAACTAT GTAAACTAAG CAAAAAAGAA GCAGTAACCA 1380

ACATCAGGTT ACTGCTTCTC TTTTTTTCAA ATAGTTTCGA TACGCAAGGT AGCCCACAAG 1440

CATTAGTCCA AACGTAACAA ATTCAACTAC TGGCATGGCT AACCAAACAC CGGTTAATCC 1500

TAGTAATTTC ACCAAAATAA ATAAAACTGG CACAATTAAA AAGAACCCAC GCAAAGAGGA 1560

AATCATTAGC GCCATTTTCG GTTTACGAAC TGCCGAAAAG TAATAAATCC CCATAAAATT 1620

CAATCCTGTA AACAAGAAGG AAATCGCATA TAAGCGGATG GCTGTCCTAG CAATTTGCGC 1680

CAACTGGTCG TTCCCTTCGC TATTAAAGAC CTCAATAATT TGTCCCGTAA ACACTTGACT 1740

GATTAAAAAG AACAGGAAGC CGGCGACCAA GTACGTAATA AAGCTAAGTT TAACAAATTT 1800

ACGCAACACC TTCGTTTCCC CTAAACCATA GTATCGGCTG AGTAATGGTT GCGCCCCTTG 1860

CCCAATCCCT GTAAAAATAG CAATCACAAT AATATTGAGA TTGGCAACAA TCGCATAAGC 1920

TGAAATCGCC ACATGACCAA CTAAGTTTAA CAAAACAATA TTAAATAAAA ACATGACAAA 1980

GGCCGAAGAA AATTCATTTA AAAAGGAGGA AAAACCAATC GACATAATTT CTCGAACATC 2040

CTGAAATTTC ACTTTAAACT TTTCAAAACG TAAAACTCGT TGCGGCGACC ACTTATGCAA 2100

CGTAATCAGG ATTAAGCTTG TTAACGGTGA CATCACCGTT GCAATGGCCG CACCTGCCAT 2160

ACCCCACCCA AATTGATAAA TAAAGACATA ATCCAAAATA ATGTTGACAA TTCCCCCACT 2220

GACAAAAGCA ATCGTTGCCA AACGAGGATT ATGGTCATTT CTTAAAAAAG TAATCAGGCA 2280

ATTGTTTAAA ATAAATAATA ATGAACAGCT AAGGTAAATC ATATAATATT CCTTTGCCAA 2340

ACCAAACAAG ACACCGGTCC CCTGAAGCCC CCGAATAATT TGATCGGAAA AAATCAAGCT 2400

CACTAAAGTA AAGAGACTCC CTATAACGAA AACTAACCCA ATCGTTAAAC TAAAATACTG 2460

ATTGGCTTTT TTGATTTCTT TTTGCGCAAC CGTGGTCGAA AACAACGTAG CGCCGCCAAC 2520

GCCAAACATC CAACCTAATC CATTAATAAT ATTAACGAGT GGTAAGACAA TGTTAAGCCC 2580

CGCCAACGCC TCGGTGCCGA TCCCATTTGC AATGAAAAAT GTGTCCGCCA AAATAAAAAT 2640

TGATTGCGCC GCCATACTTA AAATTCCTAA CAGAACATTT TTATAAAACA ACGCAGTAAA 2700

ATGGTTtTGC TGTGtTTGCT CCATACCCAA ATCTCCTTCT AACAAAAAAA GGACTCAtTC 2760

TTTTTGCTTC AAGGTTACGC AAAAATCAGA ATCCTTCAAT TATTCAACCC AGCGATTAAA 2820

TAATCTGTCT ATTTATTGAT ACTTCTAGGG TAGCATTCTC TCTTCGAAAA AACAACCATC 2880

CTTTTATTTT TCTCATTTTA AACTTTCTTA AGCAGTTGCA GTTGTTTTTG AACCACTTGC 2940

CAATCTTCTT TCACGACTTC TTTTAAACGA TAATTATAAA AGACAGCCGC TGGATGGAAG 3000

AGTGGAATGA TTTGATAGGT TTTCTGCGTC CAATGATAGT TTTGTGATTG CTCATCAAAT 3060

GTTTGAATGG GCGACTGAAT GACCGTTCCG TGCACTGCAG TGATTGTTGG CTTTGGTCCT 3120

AAAAGTCGTT GCAACCCAAT ATTCCCCAGT GTCAAAATTA GTTCAGGTTG GCAAACAGCG 3180

ATTTCCCAAT CAAGCAAAGG TGCAAAAATT TTGATTTCTT CTTTGGTTGG CGTTCGATTA 3240

GGAAATTTGG TTACTCTTTC TCCCGTTTTT GCAATCTGCT TCGTTTTTAC AGAAAAAGGC 3300

CGACTCCTAA CCGCACTGGT GATATACACC GTTTCACGAC TTAAGCCAAC TTGTTGTAAT 3360

TGTTCCATCA ACAATTGACC TGAACGCCCT ATAAATGGTA CGTTTTCTTT AATTTCTTCT 3420

TTTCCAGGGG CTTCCCCAAT CAGCATCACT TTTGGCTGTT GCGGACCTTG ACCTTCCACA 3480

AAACCTTCTA ATGGTAAGTC TTTCGCTTTA GCTAAAACTG CTTTTTTCAG ATGTTCAGGA 3540

TAGTGCATTT ACCATCACGC CTCTCTTCTG GTGTTCTTCC TTCGCTATAT TGATTATAAA 3600

TAAGGCCCAA AAGAAACTCA AAAAAAGGAG CCTATTCATT TAAGCGCTGG TTTTTCTACA 3660

CACAAAGCTT TGTCTTTAGC TAAATAAAAA GGCCGTAAGT AACCAACTAA ATTAGACATT 3720

GGTAAAACCT CCGATGGTAA ATCGAGCTTT AATGTCTCCT TTAAATAGGC TTTGCGTGTC 3780

GCAATCCGTT GCCACATGTC TGGATAAGTT TGCTGAATCT CTTTTTGCAG TGTTTCATCA 3840

GCCAAAGCCA CACATTCTTC TGCACTCACC CCTGTATAAC CAGGTACAGA AGGAATAATA 3900

TCCACTTGAA AGAGCATGCC ACTTTCCAAG CGAATGGCTG AATCTGGATA AATAGGCGAC 3960

GACATCCATT CATCATCCGA AACTAAATGC CCAGGATTCA AATGCCAGTG ATAGATTTCT 4020

TTTGGTAACT GTTCTTCAAT CGCTTGGTAC ATTTCTCGCC CTAACAGTCC GATTCGAATG 4080

GTTTCAAGCC AGTGAACGAC AGCTTGAAAA TAGGGTTTTG ctACGCGCTC TAAATAATCA 4140

CGTTGCGCTT CTGGCAATTG CTGCTCGTTT TCAATAACAA AACCTGTTCG GCTAGACAAG 4200

CCGCCTTTGT AGCCTGTAGT TAATGATAAG GCATCACCTA ACTGAATTTC TTTGGCAGTC 4260

GGATACATGT TTGCATATTC AAAACGTTGA CCAGTTGCCG CAATGGTTAC TACCGTCGGC 4320

GTTTGACCTT CATCATTTAA CAAAGCCCCA ATGTCCGTTT CACGCTGTCC TATTTCAATT 4380

GCATTTAAGG CCTTTAACAT ACTGCGAGAA GCCAAATTCG CCCCATATTC ATAGTGCGCC 4440

AATTCGTTAG CATTATTTGT TGCGCGAGCC CCCTTTGCTC CAATCAATAA GTGGGTCCCA 4500

TTTAGTAAAC GAGCTTCTTT CGGCAGCGCT TTTTTCAGTG CTTCGACAAT AAAATGTGGC 4560

ACATCAAACA TAGTAGCAGG TTCTTGCTGT TGTGTCGTAA ACATTTTCCA CCCAACGATA 4620

CCTATTTTTT GAGCAGTTTC ATCGAGAAGC GTTTCAAAAA TCTGCGACAA CTTTTGTTCC 4680

CCAGCCATTG GTTGATTAGG TAGTGAAAAA GCTAGATAAT GAATTAAATC GGCTGAGATG 4740

CGGGCATGTT GGCACAGTTT TAAATTTTCA TTACCTAAAA TAAGTGTCGC CGCACCATCC 4800

TTATTTAACA CAAGAAGTCC TTCTTCAAAC CTCGGAATAA AGCCAGTTAA GTATTCAAAA 4860

TTACTGCCAT GCTCTTTGTC TGCATAGAAG ACCAATTGAT CGAATTTTTC TGTTTCCATT 4920

CGTTGTAATA CTTTCTGCTT GCGCGCCATC ATCGTTTCAT TCGTCAAATA CGTTGGAAAA 4980

ACATTCTCAA AAACAGTAGG TTCAGCAATA TGTGTATAAT TAATTTTTCG CTTCATTATT 5040

ATCCTCCATT ATTCATCTGC GACAATCAAC GTTACGCCTG CTTGCGCTAA CTGCTCACAG 5100

AATTTTTCAT CAGGTCGTGT ATCTGTCACT AAACAATCAA TTTCTGTTAA TTCGCAAATT 5160

TTAAATAAAC CCGTTTTGTT TAATTTCGTT TGGTCTGCTA AGGCAATGAC TTGTTTGCCT 5220

TGTTGAATCA TTCGCCGCTT GACTTGACCT TCTTCTTCTG TGTGAATAAA TAATCCTTGT 5280

TCAGATAACG CAAACACACC TAATAATGTA TAGTCGGGGG CAAATTCCCC AACCTTAGCG 5340

ACTGTTTCTG CCCCAAACAA AAAGAGTTGT TGTTTATGCA GTTGCCCAGG TAATACTGTA 5400

ATTTTTGTCT GATCCATTTG CGCTAGAAGT AGCGCATGAC TGAGAGAGTT CGTTATCGCG 5460

TGAATCGGAC TATTTTCCAT CGCTCGAATA GCTGCTTCCA CCGTGGTTGA TGAATCAAAA 5520

AAAATGGACG CTTGTTTGGG AATATTTGCC CAAACTTTTT GGGCAATCGC TTGTTTTTTC 5580

TCTGGATGCT TCACTATTCG TTGAGCATAA TCGTAAATAG TTGCTTCAGA CTGCACTGAG 5640

ATAGCGCCAC CGTGCGTTCT CTTTAAATGT CCTTCCTTTT CTAAACGAAT TAAATCTCGC 5700

CGAATAGTAT CTTTTGAAAC GGAGACTTGC TnCCGTAACG CTTCCATGGA AATACTTTTT 5760

TGTTCATTCA CTATTTTTTG AATCTTGGCT AAACGTTCTT CAATCAACAT GCTTCCTCAC 5820

CTCTTCTAAG TAAAGATAGC ACCTTTTGCA TCATTTTGCA AACAAATGCA AAACAATGCA 5880

AAAAAAGACA AAAAAACTAG TCCTTCATTT TCTAACGAAG GGCTAGTTTG AGAGTGGTTT 5940

TAATAGGTAA ATTGAATAAT CATGCCGCTA GGATCTTCTA ACGCAAGTAA ATGATCTTTG 6000

GCAATAAAAT GTAACTTTGC TTCTTCCGCA TTGGCTTGGA CTATCTCTAA ATCTGCCGCA 6060

GTTGGTAAAC GGAAGGTGTA ATTTTCTAAG CCAAATTGAT TTTCTTGAAT CGTTGGCAGC 6120

TGTTTGCCGC TCCATGTATT TGTCCCAATA TGATGATGGT ACTGACCCGC TGCAAAAAAC 6180

TTCGCTTGTT GACCAAAGTT AGACTTTAAG CCAAACCCTA AGTTTTTATA GAACGTTTCA 6240

GTGGCATCTA AATCGGCCAC TTTCAAATGA ATGTGCCCAA TTTTAGAACT TCGGAAATTC 6300

CTAGCCAGTG ACCGTCAGCT TCTGCAGCTA AACTATCACC GTCTAATTCT TCTGTCACAC 6360

CGATAATTTC ACCATTTGGA CGAATATCCC AGACTTCTTC TGGTTTATCC CAATAAATTT 6420

CGATGCCATT ATTTTCTGGA TCACTTAAGT AAAGGGCTTC ACTATAGCCA TGATCTGctG 6480

CGCCAATTTG CACATTTTCC TGCAAAAGCC ACAGCAACAT ATTGCCTAAA TCTTTGCGGG 6540

TTGGTAGCAG GAAAGCAATA TGGTACAAGC CAGTGGTACC CTGATTTTCT TCCCCAGCTG 6600

TTAGTTCTTT TAAGACAAGA ATTGCATCGT TGGTGCCTTG TGGCGCTAGC CAAGCACGTC 6660

CCTCTCCTTG TTTTAAAAGT TTTAAGCCAA TTGTTTTCGT ATAAAAGTTT ACCATTTTAT 6720

CTAAATCAGC CACTTTCAAA GTGACTTGTT GAATCCCAAT ATTACTTGGT AATGCTTGAA 6780

CCATTTGCTT TCACTCCTTT CCTCTTCGTT TTACTAAAGC AATGATTTAG TTGAATTCTT 6840

TAAAACGAGC ACTATCTTCC ATGTATTCTT TTTCGcCAGc tGGCGCTTCA ATGGAACCAA 6900

TGACCATTTG TGCACGTAAA TTCCAGTTTG CAGGAATTGA CCATTCTGCA GCAACCGCTT 6960

CGTCAATAAC TGGATTGTAG TGTTGTAGGT TTGCGCCGAT ATTTTCTTGC GCTAAGGCAG 7020

TCCAAACGTT CGCTTGTGTT AAACCAGTTG CTTGTTCAGA CCAAACAGGG AAGTTATCCG 7080

CATATAATGG GAAGCTATCT TGTAAGTTTT TCACGATATC TGTATCTTCA AAGAATAAAA 7140

TTGTGCCCGC ACCAGCTGCA AAACCTTGTA ATTTAGCTTG TGTGTTTGGA AATGCTTCTG 7200

CTGGAGTTAA TGGTTTTAAG GCGTCTTCAG TCATTGCCCA TAATTTTTTG TGAGCATCGC 7260

CGAATAAGAA CACAACACGT TGTGTTTGAG AGTTAAAAGC TGTTGGACTT TCTCTAACAA 7320

CTTCTTTTAC TAAAGCAGTT ACTTCTTCTT TTGAAATTGG TAAGTTTTCT CCTAAAGCAT 7380

AGATTGAACG ACGGTTTTTC ATCATTTCAG TAAATTTtGA CATAATAtGT TCaTCCTTtC 7440

GAATTtCCTC AGTAAATyCA TTaATTnTGT kTACATGCTT ACTATACTAC CTTACTAAAA 7500

GTAAGTAAAy CTTTTTGTCT TACTTTTTGT AAGTTTTTTT TATTTTTTAG TGAAAAaCCT 7560

TATTCTACCA AAGAAAAAAC AACCAGCTTT TTTCTAAGAA AAAGAAAAAT CTGGTTGTCG 7620

ATTAATAAAT GTAGAAGTCA ATTGAATCGT TTTCGGCCGT TTTTCTTGAA ATAAATGTGT 7680

GGCTTTTTCC CCTAATTCCA TTAATGGATG ACAGATTGTT GAAAAATTTA AAATTTCACT 7740

TAATAAAGTT TGTTCCTGTC CCACTAAAAA GAGTGGTCTT TCAATAACGG CTTGAATCCC 7800

TGCGGCAATT TCATCACCGT TGGTCACAAC CGCTTGGAGC GTTGGATCTT GCCCACAGTG 7860

ACGTCCTACT TCACACCCAT CTTCATAGGT TAAACAGTCC TTAAATAACC GATTTTCAGG 7920

AATCGCTCCA TAAATTTCCT GATATGCGTC AACTAAATAG CGTGCGCTTG TACTTTTTTC 7980

TGGTGCACGG TTAATCGTAA TTGCCACTTC ACTACATCCT TGTTGTTcAA TGCTnGGAAG 8040

ACCCGCCGAT AAACGGTTAA ACGATCGATA TAGATACTAG CATAACTAGG GTCTTCTATT 8100

TTnTCACAAA AAACAACTTG CTCTTTATAT GGAGCAAACa CaCTTAArGG ATTTGCCTTT 8160

GaCAAAAcAA TCATGCCaTC aAACGCACCC nGTGGCCcAT TTTTTCCmAG tAACTTTTTT 8220

nCAACTTCTT TATCaTAATC CgTTGGTAAT AAACTGACCA TCATCCCTTG TTGAAATGCT 8280

TGGTGGATAA TCCCACTTAA AATAGCATCA AAATAAGGTT GATTCGTATA AGGAACCATG 8340

ACACCTAATG TATTACTTTT CCCAAAGCTA AGATTTCTCG CTAATATATk GGGCTTATAG 8400

CCCaACTCAT CGATAATTGC TTGGATTyTC AGTCGTTTTT CTTcAGAAAC ATAAGGATGa 8460

TGATTAATCA CACGCGAAAC AGTCGTAACT GAAGATCCTG AAAGGCGCGC AATTTCTCGA 8520

ATTGTTGTCA TACTTTCACC TCATTACTTT ACAAAATTGA GTTAAATAAG GCACAATCCC 8580

TGTCCCGTAA TCAACAAAAC GATAAAAGGT CTCCTGCCAA GCAAAAGAAA TTTCCTGAAT 8640

ATCCTCCTGC TGTGTCAAGT GcCAtCCGTT ATCTGATGGT nACGGrATTT CTTGATGCCA 8700

AATCTTATAG TATTTTTTGG GACTATTTAA CTGCAGACTT TTGGCCGTCG TAGTCACsGr 8760

GACAAAATAG TAATCCATAC CAAGTCTATC GCTTATCTCC ACTATAACAT TCATTTTCAT 8820

GCCCTCCAAT ATTTTGATAC ACCTAGCTTA CCTTATGAAA CGCGTACCAG AGCAAGAGGA 8880

AAAATAAAAA AAGTGAGACC ATGTTATACT AGATGTATCA TATAAAAGCA AAGGATGAGT 8940

GCAATGGAAA CTGTTACAGT AGCAACAAAA AGCGACTTAC TAACAAATTT GAACAAGCAA 9000

ACAGATGAAA TCATTATCTC GGGGGCACTT GCACAAGAAA TTTACCACTT AAAAAAACAA 9060

CAATTAACCG AAACGGAAGA AATGGGCTTT TTAGTGGGCA GCCAAGGTGC TGGCTCATTT 9120

TTAACCTATT TAATCGAGAA GTTATCCTAT ATAAATTCGG CGCTATCGCC AGAAGAACAA 9180

AAAATTGAAC ATGCGATTTT GCGTTTGTAT GTTATTAAAA AATTAACTCC AGAAACTACT 9240

TTATTACGTC TAAAACAGTT AGATTATTAA AATTTTATTT AATCAAGATA GACCAATTCC 9300

CCGCGTACAA CGAGTCCCTT TCATTAATCA CTTGATATAA TTACCTTTAT TCAAGTTCCC 9360

ACCATTGTCA TTATCGATTA ACCAATGGTA TATTAGAAGT AACAGATAAG TCATTTTATT 9420

TGAAAAAGTG TTTGGGAAAT GTTTGGGGAG GAATGTTTAT GTGTCAACCG CAGGAAAAAA 9480

AGAAAGAAAT TTCTGTTGCT AAGTTACGTT ATCGTTTTGA AGTACCAATG ATTATCATTG 9540

GTTTTGTTTT ATTGTTTTTC GCATTGGGGT TAGTTGCCTT ATTGATATTG AGTGATTTGC 9600

CGATTCCTGA CTGGTTATTA ATTGTCTTAA TTGCACCAAT TGCGCCAGTC ATTGCTTTTT 9660

TTACCATTCG TTATATGTAC ATGGATACTG CTTCTAACGG AGTGGaAATT ACAGAAAATC 9720

AGTTACCAGA AATTTATTCC TTGTATATTG ATTTGGCAAA AGAAATGGGC TTTGGCACTA 9780

AAAAAATTAA GGATGCCACG TCTTTATTTA ATCAACGGGA ATGGTGTCTT AAACGCTTTT 9840

GCGGCAAAAT GCTCTTTACA CAGACGCTAT GTTGTCATTC ACAGTGATTT GTTGGATATC 9900

GCGTACAATA CAGGCGATTT TTCGCTGATT CGCTTTATTT TGGCACACGA GCTAGGACAC 9960

CATAAATGTG GGCACACTAA TTTATGGCGA CTAATGTTAT CCATTATTTT GAAACCAGTT 10020

GCGTTAGATA AAAGTTTTAC AAGAACGCAA GAATATACCG CTGATCGCGC AGGTTTATAC 10080

TACGCCGAAG AAGGTGCTTT AAGTATGATT TATTTATTTT CAGGAAAATA CATGGGCAGT 10140

CGTGTGGATT TAGAAGAATA CTTTCACAGT ATTGATTTAC ATGATGATAC CATCTGGCTT 10200

AAATTGAGTA ACTTCTTATC CGATCATCCT GTTGGCTTCC GTCGAATGCA AACATTGAAA 10260

AAAGCCAAAG ATACTGGCAA TTGGGACGTT CATGGGAAAT TTTTCTAGCA AGAGCCAACC 10320

ATCATTAAAT AAAnGTGCTG GGACAAAAGA TTCTTTAACT TTTGTCCCAG CACGTTTATT 10380

AATCCACCAA TAAATGAACA AATTCTTCTT TGGTGATATT TACAAAGTAG TCTTCAAAGT 10440

TAAACGCTGC TAATGCTTGC CAAATTGATG AATACTGATA GTCCACACCT AATAAAGCTG 10500

CAACGATTTC CGTTGTTTCT TTTTTGGCAA AATAATCACC ATAAATCGTT AACTCAATTA 10560

TTTTACCTTT TTCAACAGTC AAACGTGCAT CCACGATGCC GCCTTTAAAC TTTTCTTCTT 10620

TTTTGATCGT AAATTTAGGC GCTTCACCAA AAATCCAAGC TTCATTGCCA TACACTTCAG 10680

CGACCAATTT CCTAATTTCC TGTTGATCTG CTTTGGTTAG GACATATTCT TTCTCCGCAA 10740

TTTCAGTTAG AGATTCGACA TCAAAGAGCT CCATCAGTAA ACGATTGCGA AATTCCTCAA 10800

TGGTTAATTG TTGATATTTT TCGTCTAAGT ACGGCCGTAA ATTGGTTACG CGTCCCCGAA 10860

CAGATTTtGT CCCTTTTGAT TCAATTTTTT TCTTAGAAAC AGTTAAAACA CGTTGAACTT 10920

CTGCCAAATC AACATCATAC ATCAAGGTAC CATGAGTATA CATCTTCCCC TTTTTTGTAT 10980

```
ACATGGCATT GCCAGAAAAT TTTTTCCCAT CAATTAATAA ATCATTGCGA CCGCTAATCT    11040
CGGCCCCCGT TGCGCCCATT TTATGTAATG CTTCAATAAt TGGTTTAGTG AACGCTTTAA    11100
AATCCCCAAA AGCTTGATGT CCTTCCTGAA CAACAAAACT AAAGCTGACA TTCCCGAGAT    11160
CATCATAAAC AGCACCACCG CCCGATAAAC GGCGAGTAAT AACGATTCCC TTCTCTCTAG    11220
CATAAGCTAA GTCAATTTCT TCATAAGCAT TTTGATTTCG GCCTAAAATA ACACAGGGTT    11280
CCTGAATATA AAACAACACT AAGGGCTCTT CAAAAGAGAG AGTGTTTAAT AGATAATCTT    11340
CCGTTGCTAA ATTTTCACGA ATATCTCGTG ACTGCATTAT GACATAACGC ATAAGGTCGC    11400
CCCACTTCTT TTTCTTCTAG TGTATCACTT TCTTTCAAAA TTGGTTCACT AAACGTCTTC    11460
TCTGGCCATT CTCAAAAGCG GATGATCAAC AAAAAAACGA ACGTGGGACA ACAATCACTT    11520
TGGATGTTTG TTCCACsCTC AAAATCTAAT AAACAGCAAT TGCAGAACTG TTTAGCTTTT    11580
TAATCTAACC AAGTTACTAC TTCTGAAACA TCTCGTCGGA CTGTTTCAAA GCCGGGTGTT    11640
TGAGCTTTGC CAATCGGAAT AATCAGTCCG TTCACTAAAT GATCCTCAAT ATTAAATAAC    11700
TTGCGAACAG CTTCTGTTTC AAAACCAGAC ATCGCCACAG TATCATAGCC TCGATCTTTT    11760
GCAGAAAGCA TCAGCTGCAT CGCAACTAAG CCTTGGTCTA ACATTAGCCA CTCTCTTTTT    11820
TGTTCTTCGC TAACATCATA GTACCAAGAA ACATTTTCAT TAATTCCTTG ACGAATTTCT    11880
TCCGTCATCG TTCCTTGTGC GACAGCTTTA GCACTGATTT CATCTGCATT GGCCACTGTA    11940
TAAGCAGTCA AATCACCTAA TAAAATAATC ACCGCCGCTG CCCCTTTAAC TGCTGGTTGT    12000
TGCATCGAAA CTTCGTAAAG GGCCTCTTTT TGTTCTTTTT CTGTGACAAT TAAAAAACGC    12060
CATGAATTTA AATTGGTACT ATTTGGCGCG AACATCGCTT CCTCAATTAT GGCCATTAAT    12120
TCCTCTTTAG GAATCTCAAC ATCTGGCATA AAATCCCGAA CTGAGCGGCG GCTGCGAACA    12180
ACATCTTGAT ACATCATATT TCTTCTTCCT TCCTCTTGCC TTACTTGACA TAGTTATTAT    12240
TACTCACTAT GATGGAGAAA AAAAGTACGC ACTTTTTATT CACCTAGGAA GAAAAAAGTA    12300
ACCATTACGT GTCACAAGGA GAAACTATGA CAACTGATAA ACAAACCAGT ATCAACTTAG    12360
CCCTTTCCAC CATTAACGGC AAATGGAAAT TATCATTAAT GGATGAATTA TTCCAAGGAA    12420
CCAAgCGCAA CGGTGAATTG ATGCGTGCGC TCGACGGAAT CACTCAGCGC GTCCTGACAG    12480
ATCGCTTGAG AGAAATGGAA AAAGATGGCC TAGTCCATCG AGAATCCTTT AACGAACTAC    12540
CCCCGCGTGT GGAATATACG CTCACTCCGG AAGGATATGC TCTTTACGAT GCCCTCTCTT    12600
CACTTTGCCA TTGGGGCGAA ACATTTGCAC AAAAAAAGGC GAGACTTAAC AAATAAGTCT    12660
CGTCTTTTTT TATCTATTTT GGGTATTATG GGTAAACTTA AGATATTTTT GAAGTCCTTC    12720
TTGAATTGAA TAAACGGGCG AAAAACCTAA GCCTTTCAAA GCAGAACTAT CTGAAAGAGA    12780
ATATTTTATA TCGCCATCTC GTTCCTTTTG ATAAGAAACT GGCAAAGAAA GTCCCACTAA    12840
TTGGTTCATT TCTTCAATTA ATTGATTAAG GGAAATGGCA ACTTCGGTAC CCACATTATA    12900
AACATGTCCC AACGCTTCAC TTTTGGTTGC TACTAAATTT AATGCAGTTA GGACATCCTC    12960
```

AATGTAGATA AAGTCACGTG TTTGCTGACC ATCTCCAAAT ATTTGAAATT GGCTCACTTG 13020

CCCTTGCTCT AATTGGATAT AACGATCCAT CAAAATGGAC AGGACACCAG AATAGGGTGA 13080

AGCAGGATTT TGATTGAGAC CATAGACGTT AAAGAAGCGA ACTGCGCTCG TGGGTACCCC 13140

GTGTAAATGG TATTCATTGA GAACATAACG TTCAGCAGCA AACTTATCAA TAGCATAGGG 13200

ACTCAAAGGG CAAATCACCG AAGTTTCTCT TTTAGGCAAC GTTGGCTCTG CACCGTAAAC 13260

GGCTGCTGAC GAAGCAAAAA CAAAACGTTT GAGCTTTTTT TGATATTTTT TTATTGATTC 13320

CAGTATTTTC AATACACTTA AAAAATTGAC CTCATGGGTC TCTAGAGGCT GTGCCACAGA 13380

ACTTGCCACG CTTGCCACGG CCGCAAGATG AAAGATATAC TCAAACGAAT GCTTTGAAAG 13440

TACTTCGTCC AATAATTGCT GATCCGTCAC ACTTCCTTTA ATAAAAACCA ACTGCTCAGT 13500

TTGTTGCAAA TTTGACACTC GCCCCATGGA CAAATCATCA ATCACAAAAA CTTGATACTG 13560

TTGACTATAA AAATTAGCCA AATTAGAACC AATAAAGCCA GCTCCTCCAG TAATTAATAG 13620

TTTCTTCACC TCTTTGCTTC CTCTCTTTTT TTTTAATTTA ATTCCACTTT TTTCGTGTTA 13680

TTTTCTCTTT GGAATAAAAA AATGCAATAT CGAATGAACA CATAACTTAC AATAGCCACA 13740

GCTAATAAGA GTTCAAAAGT AGACAAAATA TTTAACAGAC TCGTTAGGCC TACCTTTTCT 13800

GCAAATAAAG GTTTAACTTT GACTAATGCT GTCGCAGAAA TTACTAATGG GAATGTAAAG 13860

GCAGCATAAC TTGGATAGAA TGTCACTTGC ATAATTTTAT ATAGATTCAC TACTACAATC 13920

CCATAAATAC ATTGAGATAA AATTAGCAAG AAAACAACTA ATGGCATATA CAAATGACCA 13980

CCCATTGAAA TATAGCCAGC TAAACATAAC GAACCTGGCG CTGTCATAAT CGTCAATAAA 14040

GGAATGGTCG ATTCAGGCAT TTCTTTCGAC TTCGTAATTC TCTTGATTAA AATTGGTAAC 14100

AAGAGTAAAT AAAAGAACAA CGCTACCCAA AAAACAATTT TTCCCAACTC ATTGATAAAC 14160

AATTGAGAAG TATTAGGAAT AACGCCAATG CCGACAAACG TAATAAACCA GCTTGGATAA 14220

ATATGTTCCA GTTCGATTTT TGTTGGAAAA ATATGCACCG CTACAAAATA AATCATTAAA 14280

ATAaAATGTA AAATCAcTGC GCCAATCCAC AACACGTTAA TCCAAGCAGC ATTGGTAAAT 14340

ATCCGATCTA AGAAAACACA GATAACCATT AAAGCCATTG TAAATGTTGG TGATACCGAA 14400

GCAATAATTG GATCATCTAA GGTGCTTAAT GCATGTTTCA TTGTAAAAAT AATTTTTAAA 14460

ATAACTAAAA TCATTAAGAG ACTGCCAAGT ACACAATAAA TGGTCCCAAT TGTTGCATAT 14520

CCCAAGGAAT AAAGTAAATT CCCCAAAGAA ACCATCCCTA AAATTAGTCC GCAAATGGGA 14580

ATCGGGACTT TCCGCAAGAA TTTTTTCGCA CTTTCTACCA TTATGTAGTC ACACTCCGTC 14640

TTTATTTATT ATTATTGATA TTAATCAAAA GTATGTTATT ATACTAAATG AAATATTTAT 14700

ATTTGTAAAA TAAATATTTG CAATACCAAC TATTAAGATA GTAAATAGAA AGAGGCGATT 14760

TTATGTTCGA ATGGATCAAA ACATTTATTG CTGTGTATGA AaCCAAAAAT TTTAGTACGG 14820

CGGCCAAACA ATTGTATATT TCTCAGCCAA CTGTTTCTTT GCAAATCAAG AAATTAGAGC 14880

AACACTTTTC GATCAAGCTG TTCTATCGTA ATGGAAAACA ATCGGTTATT CCTACCAAAG 14940

AAGCTGATTT TTTATATCCA AAAATGCTCT CGATTATTGA AAGCTTAACT ACTTCTTTTG 15000

CGCAAGCTGC AGAAAAAGAG AATTTTAAAG AAGACTGCAT TATAGCCTCT TCAAATACAG 15060

TTGCTATTTA TTTACTGCCA GATATTATGG AGAGTTTGGT TACCGCTTTT CCACTCATCA 15120

ACTTTTCGAT TCAACTGATG AATTCAAACG AAGTCGTGGA TATGGTTCAA AATAATCATG 15180

CCCACATTGG ACTAATTGAA AAACCCATTG AAACCAAAAA TCTACACAAA GAAATTGTTT 15240

ATGAAGATCA ACTAGTTCTA GCAGGAGATG CCGCTTCAAA ATTTTGGCTC TTAAGAGAAA 15300

AGAACTCTGG CTTACGTTTT TTCAATGAGC TTTATCTCAA CGAACACAGC ATCAACTTAC 15360

CAATTATCGA ATTAAACAAC AACGAAGTCC TGTTACAATT GCTAAAAAAT AACATTGGGC 15420

AATCAATTGT TTCTCGATTA TCTATTTCCG ATGAGATCCC TTGGCAACCA ATTGATCAGT 15480

CCTTTGCTTC TCGAAAAATT TTCATTGTCC AGACCGATCA CCATGCCAAT ACTTCCTTCT 15540

CTGAAGTTTA TAACCGAATC GTTGAAAAAA TTCGTGAACG ACAAACCTAA AAAACAGGAG 15600

AACGACTTAG CTTTCGCTAA ATCGTTTTCC TGTTTTTTAT TCATCAGTTT CCTCTACCAA 15660

TTCAAATTCA AAAGGTTCTA ACTCAAATAG TTGGTGTAAA ATTTGCTGGG TTTCTTGGAA 15720

CACTTGTTGT GGCTCTTTAC ACAAAAGAGT GCTTTGAATG GTCTGTTTTT TTGATTGAAA 15780

AAGAATTCTC GTTTCCTCTT CCTTAAATAC AACTTGATGA GTTcCCTTTC AAACACCGTA 15840

TATTCTTCGA ACATTTCCAT AACTAAtTTT TGACGAAACC CTTCTTTGAG TGCTAATAAA 15900

GTTACTAAAT CTGCAACGGt TAGCGkTTCT AACGCAAnTG kTTCaTCyTC TACATGAACc 15960

AAGATAATnC GtACACTTCA CATtCGGACA GGCATTGCAG CCGCCATTAA TCAAGTCTGA 16020

GGTAGTTGTT AyTTTCACCT GTTTCATCTT AATCCTTCTT TCCTACTTTT CGAGAAAAAC 16080

ACAAGGGGTC TCCTCAAAAA TATTCAATAA AATATGATTT ACTTCTTGAA AGACTTCTTC 16140

GTGATTCCGG CATTGTTTTT CCATTACTAA ACGCAGATCA TTCACAACAA AAGTCAACGT 16200

CGTCGCTTGT TGGGAAACAT GTATACAACG GTTGCCTTTA ACAAAACATT CTGTTTGTTC 16260

TTCAATGAAT GGTTCTTTTT GATAGCCATA TGCTGAAGCG ACACATTCCA GGATTTCGTA 16320

GATTTCTATT TTTGGAAGTT CATAGCTACA GCTACCGATA ACTAGTTGAA AACGGTTCGG 16380

CTGTTTTTCA TTATTGCTTG CATAATACAC TGTTGTCGTT GTCATTTTGC CTCCTTATCT 16440

CCTCTATTTT CCTACTCTTA TCATAACAAA AAAAGCACGC TTCCCTTATT TAAATTGTTA 16500

ATAAAGACTA TTACAGATTG GTATAACCAC ACCTAAAAAG TTCGTGATAT AATTCACTAA 16560

GAAAGAAATA TTACAGAAAG AAGAGATGAA CGATGAAAGT ATTTAACAAA AAAGTCGCAA 16620

TTATTGGTAC TGGTTTTGTT GGCACAAGTA TTGCCTATTC CATGATCAAC CAAGGGATTG 16680

CGAATGAATT AATCTTAGTT GATATTGACA AAGCCAAATC TGAAGGCGAn CAATTGACTT 16740

ATTAGATGGT GTGkCyTgGG GkCmAGAAAA TGTAAACGTC TGGGCTGGCG ACTATCAAGA 16800

CTGCCAAGAT GCCGATATCG TCGTGATTAC AGCTGGCGCT AaTCAAAAAC CTGGGCAAAG 16860

TCGTCTAGAT TTGGTTTCAA TCAATGCAGa AATTATGAAA ACAATTGTTA ACAATATCAT 16920

```
GAAATCTGGT TTTGATGGAA TTTTAGTGAT TGCCTCAAAT CCTGTCGATG TACTGACTTA    16980
TGTGGCCTGG CAAGCTTCTG GTTTACCTGT TTCAAGAGTA ATTGGAACTG GTACAACTTT    17040
GGACACAACT CGTTTCCGCA AAGAACTGTC TCAACGTTTA GCGATTGATC CACGCAATGT    17100
TCACGGCTAT ATTATTGGCG AACACGGGGA TTCTGAAGTG GCTGTCTGGT CTCATACCAT    17160
GATTGGTACC AAACCTATTT TAGAAATTGT GGATACGACA GAGCGCTTAA CTAGTGACGA    17220
TTTACCAATC ATTTCTGATA AAGTGAAAAA TACAGCTTAT GAAATTATCG ATCGCAAACA    17280
AGCGACCTAT TATGGGATTG GTATGAGTAC TGCACGCATT GTTAAAGCCA TTTTAAATAA    17340
TGAACAAGCT ATTTTACCTG TCTCAGCTTA CTTAGATGGG CAATATGGGC AACAAGATGT    17400
ATTTACAGGG ATTCCTGCAG TCGTTGGCAA TCAAGGTGTG ACTGACATTA TTGAATTGAA    17460
TCTGAATGCC GCTGAAAAAG AACTCTTCCA AAAATCAGTG ACACAATTAA AACAAGTGAT    17520
GGCATCGTTA CAACCGAACG CTTAGTAATT TTTAACTAAA AATAACACTT CAACTAAATT    17580
ATTCTCTACC AAAATAGATT CTAATTCCCC TTTATCGTTC TTTTTGGTAC CATGAAAGAA    17640
GAACTTCAAC TAAAGCAGAA ATTAGGAGGA CAAAAAATGG TAATCCAAGG AGATACGTTA    17700
GAAAATAGCG CACGTCGTTT ATTGCAAGAA CGTGGCGTAA CAATGAATGA TTTAGCTGAA    17760
CTTGTGTTAT TTTTACAAAA AGATTATATT GACAATTTAA CATTAGATAC TTGTTTAGAA    17820
AGTGTTGATG CCGTGTTAAC AAAGCGTGAA GTACACAACG CGATTATTAC AGGTGTCCAA    17880
CTGGATATTC TGGCAGAAGA AAACAAGCTA CTCTCACCGT TACAAGAAAT TTTAACGGAA    17940
GATGAAGGGC TTTATGGCAT TGATGAAATT ATGGCGCTCT CCATTGTAAA TGTTTATGGA    18000
TCTATCGGTT TTACGAACTA TGGGTATATC GACAAGGTAA AACCAGGCAT TTTAAAAGAG    18060
TTAAATGCGC ATGATGGCGA ACATGTCCAC ACATTCTTGG ATGATATCGT GGGCGCAATT    18120
GCGGCTGCGG CTGCCAGTCG GTTGGCCCAC TCACAACCAG ATTTAAGTGA TATTACTAAA    18180
TAAAAAACTT GCCTCTTCAT TTTTATGAGA AGAGGCAAGT TTTTTATTTG CTGGACTTCT    18240
CACAAGCTTT CTACTTAACA GTAATTATTT GCCGTGTTAT GTTACAGTCA AGAATAAACA    18300
GCTAAAGGAG GTTATCTAGA AATGAAAAAG TTTTTACGCG TATTTTTAAT TTTTATTGGG    18360
GTATTCTTTC TTGCAAGTTG CGGTTCAAAA ATAGAGACAA AAACCTTTGT AGGTTCGCCA    18420
CAAGATGGGA TTGATTCCAC CTTAACCTAT ACGTACCAAG GAGATAAAGT CTTAACGCAA    18480
ACGGCCAAAA ACATTGTTTC TTATGACAAA TTGGGGATTA CAAAAGAAGA GGCAAAAACA    18540
GCGTTGGAAC CCGTTTCTAA ACAATATGAG GACATCAAAG GGCTTGATTA CAAACTGACC    18600
TATGAGGACA AACAAGCCAT CGAAAAATTG ACGATTAATT ATGAAAAACT TGATTATGAT    18660
AAAGCCAAAA AAGTGGATGG TATCCAGATT GATGGAGATT CTTCTAAGGG AATCAGCATG    18720
AAGAAAAGCC AAGAATTGGT GGAGTCACAA GGTTATACAG AACAAAAATA ATAAAAGAAA    18780
GCGTGCGACA ATGGGCATTG TCGCACGCTT TCTTTTTATT AATTACATTG ATGTTGCCAT    18840
CTTCACTGCT TGAATTGGGT CAAAATGATG GACAGAGTAA ACGACTGTAC CGGCTTGCCA    18900
```

AACTACTGAG TTTCGATTAT AGTCAGTGGT TCCTGCCACG TTCAGTGTGA TTTGACCAAC 18960

GGTATTCGGT GCTGGTAATG TTTCTTGTTC CAAAATGTTG ACAACATTTT TCGCTAAATC 19020

ATCAGGCGAT TCACCATTGA TATTTGAGGC CCGTACGACT AAACTCCAAT TGCCTTCTTG 19080

CCAAGATAAG TAACTAGAAC CTGCCGCCCC TTGTTTATAC CCAGTAATAT TGTAACCTAA 19140

ATCTACTGGT TGACCGCCAT TGTCAATGAT TTGGTTCACT GCATTTTTTG CTTCGGCATC 19200

AGAGCCATAT GTTTTTTTCT CAAAAGAACT AATTGGCGTT AACTGATTGA CACGTGCATC 19260

ATTCACTGGA ATCGCTTCTT TTTCTGCATA ATATAAAACA CGGAAATTGT TTTGATCCGC 19320

TTGTGAAGTT GTCGCTGCAG TTAAAAAGTT ATTCGTTTGA CTCGTTAAAA TTGCCTGCGG 19380

TAAGCCTTGA CCAGGAAACA TGTTTGCTAG CTGGTTAAGC ACTTGGTTGG CTTGGTCATC 19440

TGGATTCGCA TTTTTCGCTG GTGATTGCTT CTCTGCTGAG CCATTATTTG CCGCTTCCTT 19500

CGTAGAAGAA GCTGTTGCTT TCGAATCAGC TGTTGTTGTT CCTGACGGTT TCGCATCACT 19560

AGATTTTGTT GTTGCTTTGC TACTTGATTC TTTTGTAGCA GATGCTTTTG TACTCTCGAC 19620

TGTGGTACTT TCCGCAGCAC TGGAACGACT ATCCGTATTT TCTGTACTTT TTCCGCCGCC 19680

GCATGCTGCT AAAGCAAGGG TTAAAAGAAA AGCACTAGAA ATAAGCCCTA TCTTTTTCAT 19740

AAGAATTCCT CCTTGTTTTC CCTACCCTTA TTGTACCTTA ATCATGTGTA TCCGTCACTT 19800

GTTTGTGTTT CGTTCTAAGT TGGTTTAAAT AACAGAAAGT GGACGGTAGG ATTCGAACCT 19860

ACACCTCAAC ATTGAGCAAG AATAAACTAC TGTTAGTCTC TATGAGTTTA ACACCACGTT 19920

CATCCACCTA TTTGATTGAT CCTTTGTATC ATAACTTAGA ATAATTAAAT TGCCTTTAAG 19980

GAGCTTCAAA GATTATATGA AGAATTGTAA AAAATCTGTG ACGAAAAAAC GATTGAACCA 20040

ATCAGTAGAT ACACTGAGGT TCAATCGTAA CCAATTGTTT AATTATCTGA AAAAACTCAC 20100

AATTCAACGT TAGTCCCTGG TGCTTTTTTG TTTTCCCCTT GTTGTTTTTT TAAGAATATC 20160

CCCAAAATGA CGGTGATAAT GGGACAAAGC AATGGGAAAA AGGCATATGG TAAGTATTCT 20220

AGTGCACCAA CTTTTAACGT TCCCATAATG AAGGTACCAC TAACTCCCCA AGGAATTAAC 20280

GAGTTGACTG CCGCCCCAGC ATCTGCCAAA GTACGAGTTA AATATTTTTT ATCAATACCT 20340

AAACGAGTAA ATGAGGATTT AAATGTTTCC CCTGGTAAAA TAATTGACAA ATATTGTTCG 20400

CCAACTAGTA AGTTAATTCC CACAGAACTT AAGGCAGTTA AAGCGATTAA TTTGGCTGGA 20460

TTATTCACAT AGCCTTTAAT TTTATCAATT AAAGTAGCCA CGATGTTAAA CTTAATCAAC 20520

AAGCCGCCCA AACCTAATGC CAAAATAATC AACGCAGCAG AACCTAACAT GCTTTCAATC 20580

CCACCACGAG ATAAAAGTGT ATCAATACTT TGATCCCCTG TATCTGCCAC ATAACCACTC 20640

ATTAAGATGG TAGAAACTTT CGCTAAACTT AGGTGATGAT CATTTATAAA AGCTAAGATG 20700

ACCGCCACTG TAGAGCCGAC TAGTAACGTT GGAATAGCCG GTACCTTTTT CCAGGCGAAT 20760

AAGAATAAAA CCGCTACTGG CAACAATGTA ATCGGTGAAa TCCAGAATCC TTGATGTAAC 20820

GTTTGGACCA TTGCATCAAC CGATTGTAAA TCCGCTGAAC CAGATTGATG CCCCAAGAAG 20880

ATATAGCCGA CAATCGAAAT AATAAACGCC GGAATAACAG TGTACATCAT GTTTAGTATA 20940

TGTTCAAAAA GGTTCACGCC ACCAATTCCA GCAGCTAAGT TTGTTGTATC AGATAACGGT 21000

GAAATGTTGT TCCCTAAAAA GGCACCTGAA ACAACTGCAC CAGCTGTCAT CGCATTGTCG 21060

AATCCAAGAA TATGTCCAAT ACCTAGAAAG GCAATCCCCA TAGTTGAAAC AGTCGTAAAG 21120

GAGCTCCCTA CCGTCACACC AACTAGTGCA CAGACCACAA AAACGGTTGG TAAGAAGAAC 21180

TTCACTGAGA TAATTTTAAA ACCATAGACC ATAATTGTTG GAATCGTCCC AGATAAGATC 21240

CAAGTAGCAA CTAAGACACC TATCAGCAAG AAAATAATAA TCGGAATGAT CCCTGGTTTG 21300

ATCCCTTCCA CAATGCCCTC GTGGATCTCA TCCCAAGAAA ATCCTTTCCA TTTGCCATAA 21360

AGCAACAACA ACATGAATAC AAAAAGAATC GGTATGTGAG GGGTTAATTT TTGACCAATA 21420

ATTAAGTACC CGAGAATCGC TAACAAGGCA ATCAAAATGC CTAAAcTTTC AGCGAACGAA 21480

ACAGTGGGTT TCGTTGATTT CATATGAAAA CCTCCTTTTA TGGAAAGCGA AAACTGAAGC 21540

AATAATCGAA CTTGCTTCAG TTnCCAAACG ACAATTTACT ATAGTCATCA GGTGAAAAGG 21600

CACTCAGCAA AATAATTGCT AGCCTTTTTA TTCACCTAAA TCTTCTTTTA CTTCCGTTGC 21660

CCAACTTGCT TTTTTAAAGC GATTAmTGAT TAACGGAATC ACAAACATAA CAATAGTGAT 21720

TAATGCAACT AAGACTGCAT CAGTCACGGT TGTTGTACCT AATGTGGCGA AAATAACAAT 21780

GATAACACTA AAAATTAACA TTGCTGCCAC TAAATAAGCT AAGCCATTAC CTTTTTTACC 21840

AATTCGGTAA GGACGTAGTG TATCTGGTTG AGATTTACGT AATTTAATTA GGGAAATCGC 21900

AATCAATACA TACACAATAC TATAAATAAT TGTTGTTGCA TTTGTTAAAG TTAAGAATAC 21960

AGCGTTGACA TCATCCATTA CCCCATAAAG TAAAGCGAAT AATGAGATAA CCACAGATTG 22020

TGtTAACACA ACGTTACGAG AAACGCCATA TTTATTTTCG CGATGGAAAC CAAATTTTGG 22080

TGGCAATAAG CCATCACGAG CAACTTGAAT AATTGTTTTT GATGGTCCAG TAACCCAAGC 22140

AGATAATTGT AATAACACAC CTAAGAAGAC CATGAAACTA AAGATATTAC CAATAATTGT 22200

TGGTAAACCT AAAACGTCAC AATAAAGCAA AATTGGTTGT GTAATGTTAG ACAATTCCAT 22260

TTTGCCATTT GGAACAATAT TAGCAACAAA CATTGCATTA ATAACGTTTA ATAAAACTAA 22320

ACCAATTAAT GAAATGAAAA CGCCTTTTGT ATAATTTTTT GTTGCATCTT TCAAACGAGG 22380

AATATAAACA GAAGACATTT CAATCCCAAC AAAGATAAAT GCAATCCCTG CTAGATATTT 22440

AAATGTGTCC ATATTTTCTA AGTTTGGTAA AACTTTTGAC CAAGAGAAGG CACCTAAATA 22500

GCCGCCAGGT GTTAAGCCTA CTTTAATCAT TGAAAGAAGA CCTAAAACAA ACATAACTAA 22560

TACTGGGATG TAAACACCTA GCCAAACACC AATATTTCCG CCAACTTTCG CCATGTCAAA 22620

TTTCAAGTTC AAGATTGTAA TAATCCAGTA AATCACTAAG ATACAACCTA ATACGAACCA 22680

ATGATTTTCA CCTAAATGAA CATTaCCAAA TGTATTaCCT AGCAAAGGGC CTAATGTAGA 22740

AGCAACCATT ACCATTCCTG GGAACATTTG CACCCATAGT AACCAAGCGA CAACAAAACC 22800

CCATTTGCTA CCTAAGGCAG TTTTAACCCA AAGTTGAGGG CCACCTTCTT GAGGAAGCAT 22860

CGTAGATAAT TCACCAGATA TCAATGCGAT TGGACCAGCA AAGAaTAAGA CCGCGAAAAT 22920
TGTGTAAGAA ATTAATGTCC AACCAACAGC AGCAAGTGTT GGGATACTAC GTACAGTTGC 22980
ACACAATGCC ATTGTTAAAC CGATAAAAGT ACCTAATTCA ATTTTTTTAA CTTGTTGTTT 23040
ATCCAATTTT GTACACTCCT TATAAGATTT TGGATAAGGA GCATGAATGA CTGTTATACG 23100
CTGAGTAAAA CAGTCATTCG ACTGCCTCTT ACCCAGACAA AGAATTGGTT AAGAGGTTAG 23160
CCATCTGATT GATTTATCTC TTAAATAAAT CAATCAGACG AACGTTCCCT CATTCGGTTA 23220
CCATTTCTTT TTTCCGTTTA AGATTCAATT ATTTTACGTC GTAAATTTGT TCTAATTTTT 23280
CTTGTAAAGC AGCTTGAATC TTAGGAGCGT AAACATCAAA TTCTTCTTTA TCGTTCATGT 23340
ATGGAGTCAT TACAGCAGCA CGTAACACAG TTACTTTACC TGCACGGTTC CATTCTTCGT 23400
CACTAAATCC CAAGCTGTTT ACAAATTTTA ATGGGcTGTT GCCATAATCT GGAATAGCAA 23460
AGTCAGTATG TGAAGTAATG AATTCGTTGT TGTAAATATT TCCTTTAACG TATGATGCAT 23520
AGTCATATAC ATCATGGTTT AATTTGTTCA TTGCTACTAA ATCATCGTTG CCTTTTTCTT 23580
TAAACACATA GTCAACCATG TTGAAGTCTG GATGAGTTAA TGTGTGGACT TCGATTTCTT 23640
TGTCGCCAAC TTTGAATGTT AAATCATTCA AGAAGTTATA GAAATGATGT GATCCTTCAA 23700
TTGAAGCACC AATTAATTTA CCATAACCTG CAACGTTTAA AGGTAAAACG TGATGTGCAG 23760
CCCATACAGA AGCAGCTGTT GCACCAGCTT TTGAACCTTC AAGGATATAA GCACCTAGTA 23820
ACGCAGGAAT ATCGGCACCT TTTTCAAATA CGTAAGTTGC GAAGTAAGAA ATAACGTCGC 23880
GCATACGGAT ATCTTGGATA ACGATACCAC CAGCTGAATA AGGGATATAA CCCATTTTAT 23940
GAGGGTCAAT TGTTACTGAT TCTGCTAATT CGATTGCTTT AtATGCATCA TACACTTCTC 24000
TTGAAATGTG TTCTTTTTTC TCTTTGAAGA CACCGTATTC TTCGTGAACA TCTTGTAAAT 24060
CTTCGTAAGG GATGAAGTTG TTGTCTTCGT CTAAGAAGAT GGCACGTCCA TAACCACCAT 24120
AAGCAGCATC AACGTGTACA TAATAGTAAA TACCGTCTTT CATTAATTCA TCGCGTAAAG 24180
CAATAATTTT ATCGATTGAG TCAACGGCAC CTTCTTCAGT AGAACCAACT ACACCGACAA 24240
CACCTAGCAC TGGAATTTGT TCTTCTGCTA AGCCACGAAC GATTTTTTCT AATTCGTTGA 24300
TATCCATACG GTAGTTATGG TCAACTGGTA CTGGGATAAC TTGATCTAAA CCGATACCGA 24360
TAATATCAGC AGCTTTTAAC CAAGAATAGT GTTTAGTTTG TGGAACTAAC CATTTACCGA 24420
TTGCTTGTAA ATGTTTACCT GAACGAGCTG AATGAGCTTT GATTTCATCA ATTTCATCTT 24480
CAGCTGATTC TAATAAGTCC ATAATTTCTT TTGTTGGCAT GTTCAATAGT TCCCAATCTG 24540
ATTTGCCAGC AACTAATTCT GGTTTTACTT CTTTCATAGC AAATGGTAAT GATTTAATGT 24600
TACGGGCATA CCATAAGCCT TCTAAGTTAG CTAAAGAACC ATCAGCAACG ATGTGTCCCC 24660
AACCATTTTT GTAGCTCATT AAGTGAGCAA ATTCATGTCC TACTTCTTCT TCCATTTGAG 24720
AAGTTGCTGG AGAAGATTCA TAGGCAACGT TGTTCCCGTT CCATAGCATT GCAAAGTTGT 24780
AAGCTAATAG AGAAGGCATT AATGTTTCTG AGTTCATATG TCCCCAATAT CTACCTGCTG 24840

TATGCCATGG AACTGAATGT GTACGCATGC GAGAAGAGAT TTCATTCAAT ACATCTTTCA 24900

TATGGTTGAC AGTTTTTTCG TAGCTTTCAG ATGTGCGTTC TTGAGAAGAG ATAACTGGCA 24960

TGTCTTGTGG CATGTAGTTT TGACGCCATC CTAAATGTTC ATCTACTAAG TCGATCAACA 25020

AGTCTTTATA TAATTGTCCG TTTTCGGCTT TGTCCCCAAT GAATAGTGCA TTAAGGTTCA 25080

TTTCGCCTTT TGCTAATTTT TCGTTTkTCA TACGAAAATT CCTCCTTACA AATTTTAATG 25140

AATTAGATTT TGCATTCACA ACTTACACCC AAACGGCTAA AGTGTAAAGA AGAaAAATAA 25200

CAATGGtTAT CATAGCTCTT TAATTTTATA ATTTTTTCAC AAGGTGTCTT GTCaGaACCA 25260

TTATATaAAA AAATAAGAGT GGGaTTgTGT GtACAATCCC ATCTTATTAG GTGTATTAAC 25320

GGTTTCACCA CTACCTGTAA GAGnAAATAG AATTTTAACT TTGATCAAAT AATCACACGT 25380

CTTCTCTTCG TGAAATTATT ATCAATTTCT AATCCATTAC TTTTGCTAAA AAGTAATTTT 25440

TCTtCCmTTT CCGTACAACA ACGAATTTtC CATCGAAATG TTGTGTTGGA TCGACTGCAA 25500

AATCTAAATC GGTAACTCGG TCACCGTTAA TACTTATAGC TCCTCCGCTC ACATCTTCGC 25560

GAGCTTGACG TTTTGATGGT TCGATTTCTG TGTCAACTAA CCAATCTACG ATGTTTTTAG 25620

CATCTTTGGT AATTTCAACA GTTGGCACAT GTTCTAACCC TTGCTCGATT TCCTCAATCG 25680

TTAAGTCTTT AATGTTGCCT GAGAAAAGAG CTTCTGAAAT TTTTTGCGCT TCTTCTAATG 25740

CTGCATCATC GTGAACAAAT CGTGTCACTT CTTCAGCTAA GCGTCTTTGT GCTTCACGTT 25800

TACCTGGTTC TTTTTCAACT TTTTcAGTAA CGCATCGATT TCTTCTTTAT CTAAGAAAGT 25860

AAAGAATTTC AAGTATTTGA TGACATCACG ATCATCTTGG TTTAACCAGA ATTGGTAGAA 25920

TTCAAATGGT GAGGTTTTCT TAGGATCTAA CCAGATAGCG CCACCCGCTG TTTTCCCAAA 25980

TTTTGTACCA TCTGCTTTTA GCATTAAAGG AATGGTTAAC CCGAATACTT TGGCTTCTGG 26040

TCCTTCTTTT TTACGAATCA AATCTAATCC TGCAGTGATA TTGCCCCATT GATCAGCGCC 26100

ACCAATTTGC AATTGAATAT TATGTTTTTT ATGCAATGTG TAAAAGTCAA TTGATTGAAG 26160

AATTTGGTAT GTGAATTCTG TAAAGGAAAT GCCACTTTCT AAACGACTAG CGACAATGTC 26220

TTTTGCCAAC ATCGTGTTGA CATTAAAGTT TTTCCCGTAA TCTCTTAAAA AATCTAATAA 26280

AGATAGTTCT GATAACCAAT CGTAGTTGTT CACCATTGTT ACCTCAGCAT CTTTACCAAA 26340

TAATTTTTTC ATTTGGTTTG AAAGACTGTC CACATTATGT TGCACAGCTT CCATCGTTTG 26400

TAAGACACGT TCGGTTGTTC GTCCACTTGG GTCACCAATT GTTCCAGTTC CGCCACCAAT 26460

TAAAATGTAT GGGTGATGAC CTGCTAATTG GAATCGTTTC ATCATCATAA AAGGAATTAA 26520

ATGTCCAATA TGCATGCTAT CTCCAGTTGG ATCGACACCG CAATATAGCG AAATGCTCGT 26580

ATTTTCTGTA AGTTCTCTTA GTCCTTCTTC GTTTGTTTGT TGATTGATTG CATCACGCCA 26640

TGCTAGCTCG TCAATGATAT TCATGTTATC ACCCTCCATA TTTATAATTA TCTCTTTATA 26700

GGAAATCGTT CATTTAACGA AACAAACGAT TAACACACTT GTTTTTACAC GATATCCCGT 26760

GTTTATTTCT ATGTATACAC GGGTTCACGA AATTTTCTTT GGTCTGCCAC GTATTCAAAA 26820

AAAGAATACA AAAAATCCCT AAGTCGTCGA AACGACTTAG GGACGAATTT CTCGTGGTAC 26880

CACCCAAATT TATACATAGT TTAGTTTATT TAAAATACTA TATACCTCTT AACAAATAAC 26940

GCTGTTACCC GTGCCCTGTA GGCAGAAGCT CAAGAATGTA ATTCGATAAA TAACGTGTAC 27000

TAGTTTCCAG CGCCACTAGC TTTCTGTAAC AGTGGTTATT TCTCTACTGC TTTTCTGTCA 27060

TTGCTTTGTA TTTAACTTGT ACATGTAGCA TATTCCTTTT CGTTTTTTTT gTCaACACTC 27120

TTkGCTTAAt TTTTkTTACT TTTATAAAAG ATGTCCCTTT TTCACAAAAA TAAAATAACG 27180

AATTCAAAAT AATATTTATT TGTTTTAAAA GCAACTACAG ATAAAAAAAT AACGAAAAAA 27240

TTTAAAAAAT ATTCTTTTAA ATTTTTTTGA CTTTTTTTGk GATTTTCTAT ACTTTTATAA 27300

AATTCCCGAr GAAAAATGAC GAACTTCGGA TCAGAAGTTC GTCaTTTTGT GAAAaaCAGT 27360

TTTAAAGAAT TGGTGAAAAT AATCGTGAAA TTTGTTGTTT AAAAATCAAC CATTTTGACA 27420

TGTCGCGAAC AACTTCTGGA GTCATTGTTG TACATTTTTT TCGATCTTCT AAGAAAATTG 27480

CTTCTAATTG ATCAATGATT TTTTTATCAT ATAACACAGC ATTTGCTTCA AAGTTTAATT 27540

TGTAACTTCG GATATCTTGA TTGGCTGAAC CAACTGTGCA GACTTCATCA TCCATGATCA 27600

TTGTTTTCGC ATGTAAGAAA CCACCGTTAT AAATTAAAAT TTCAATATTT TCTTTCATCA 27660

ATAAGCGGCC GTAATACTGT GTTGCTCGAT AAATAAACGG ATGGTCGGGC TTATCTGGAA 27720

TCATAATTTT AACATCTACT CCTGAAGCCG CAGCAACCTT TAAAGCAGCC AAGACACTAT 27780

CATCAGGAAC TAAATACGGC GTTTGAATCC AGACTCTTTT CTTGGCAGAA GTAATCAATT 27840

TAATAAAGGC CAATTTAATT TGTTCACGGT CACTGTTAGG ACCACTGGCA ATCATCTGAA 27900

TGGATGTATT TGCCTCTGGC ACCAATGCTT CAAGTTTAAA GAAATAATCC AATTGATACG 27960

CCACACGATT TTTCTCGGGG GAAGAGACGT TCCAATCCAT TAAAAAACGC ATTTGCAGTA 28020

ATGAGGCTGC TGGCCCTTGA ATCCGTAAAT GTGTATCCCG CCAATAGCCA AACTTTTTAG 28080

TCGTTCCGGC ATATTGATCG GCAACATTAA AGCCGCCAAT GTAGCCTACT TTTCCGTCAA 28140

TAACAACAAT TTTCCGGTGA TCATGATAAT TCAAACGAAA CTTCAAAAGT GCTTTTTGAG 28200

AAGTAATAAA CGTTTGGACA AAACCACCGT TTTTTTTCAA TTCATTCAAA TGATGAACCT 28260

TGGTGCCCTT TGAGCCAAAG GCATCGTAAA GCAACCGCAC TTCAACCCCT TCAGCCGCCT 28320

TTTCTTCTAA TAAGTTCAAA ATTTTGGTGC CGATATGATC CGTTACAAAG GCATAATATT 28380

CAATATGAAT GGAGTGCTGA GCTTTTTTTA TATCCGCCAT CAACGCATTG AATTTTTCCG 28440

TACCATCCGT AAAAATTTCA ACATCATTTT TCTTAGTCAG AGGCATCCTA TTTAACGAGG 28500

AAAAAAAGTC TGTAAACTGT TGTTGCTCCT CCGTCGGCAT CTTTTGACTA TAAAGTTCAA 28560

AACTTTCTTC TTGGAAATTC TGGAAATTCT CTAATTCCCG TAAATCACTT TGTTGCAAAT 28620

AAAATTTCTT TTkGTCGGkT AAACCTCGAC CAAAGrATAA ATATAAGAAA AAACCAAAAA 28680

CGGGTAGTGG 28690

(2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6146 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

AAATTATTGA AAAATCACCG AAGCAAACTC ATACAAAATA TACAAAATGG CAAACAGAGT 60

CCATTTATAA ACAAAAAATA AAAAGTCGTT TAGACTTTTT ATTAGAACAA AGTTCTTCAA 120

TTAATGACTT CCTAGAAAAA GCAGAAGCTT TGAATTTATC GGCAGACTTT TCTAATAAGT 180

GGACGACCTA TCGATTGCTT GACGAACCAC AAATAAAAAA TACTCGTAGT CGAAGCTTGT 240

CAAAAAGTGA CCCCACTAGA TATAnTTATG AAAAGATTAT TGAACGATTG AAAGAAAATA 300

AAAACGTTCT AACAGTGGAA GAATGTGTCA AAAGtACGAg AAAAAAATGA ACAAGAAAAA 360

AATAATTTTG ACTATCAaTT TACmATAGAA CCaTGGCAAC TGTCTCACAA AACAGAaAGA 420

GGATATTATA TAAATATCGA TTATGGCTAT GGCAATTCTG GAAAGCTCTT TGTAGGAGGG 480

TATAAGGTTG ATCCATTAGC CAACGGTAAT TACAATATCT ATATTAAAAG AAATGAGTAT 540

TTTTATTTTA TGAGCGACAA GGATGCTACT AAAAGTAAAT ACTTAACAGG GGCTTCGTTG 600

ATTAAACAAT TACGACTGTA TAATGGACAA ACACCACTGA AAAGAGAACC TGTGATGCAA 660

ACCATTGATG AATTAGTTAG TGCCATTAAC TTTTTAGCGG CGAATGAAAT CGAGGACACA 720

AGACAACTAA AATTATTAGA AGAAAAATTA GAAGTTGCTT TTCTTGAAGC AGAAAAAACT 780

TTAGAGACTC TTGACGAGAA GATGTTAGAG CTGCACCAAT TAAGTAACCT CTTATTAGAA 840

AATGAACTGC AGGGAGATCC AGAGTTAATT CAGAAAAAAT TGAAAACATT GTTACCGGAC 900

GCAACGCTTG CGGAATTTTC TTATGAAGAT GTTCGAGGAG AAATAGAAGC AATTAAAACC 960

AGTCAAACTT TGTTAGAGAG TAAACTGGAA CGAACTCGAA ATGAAATTAA TCAGCTTCAT 1020

GAAATTCAAG CCGTACAGGA AAAAGAGCCA GAGAAGCAAA CCGACATTAA GCCCAAACTA 1080

TAAACTATCT ATTGGTGAAC GGAGAGGAGA ATTTTATGTT AGTTTTTAAA CGTATTTTGT 1140

CTGTAATACT ATTTTTACCT TTTTTTATTC TTCATTTACT TTTCCGATAT ATTGGAATTT 1200

TAACATTGCT TACTAGTGGA TTTACCACCA CTTTATTATT GATTATTGCC CGAGTTTTGG 1260

CTTtTCTTTT TTTAATAACA CCAATTGTGA TCTTAGCAAC AGAAGAACTT AGAATCGAAT 1320

GGGCAAGTTG GGTTGGTTGG CTTTTATTTG GATTACTTGT TCTGCTTTTT CATTATCTAG 1380

CTGCCTTGAA TGAAGAATTA GGAGGCTGGT TTTTAGATTT TTCTGATTGG TTAAGCAAGG 1440

TTCCTAGCAA ACTGTGGGCT GAATAATTGC TCCTATGCTA AAACGCCTTA ACTTAAAAAA 1500

GCCAAGGCGT TTTATGGTAT AATATACGAG ATAGTAAAGG TTTATGACGG TGGCTAATCT 1560

CTGAAAAAGG GGTGATTGCC TATGTTGATT TTTGGCTATA CTCCTAGAAA GGACTAGCTT 1620

TGACTATTTT TGAAGCGTTA ATGTTTGCGA TTACTTTTGC GACATTAATT CTAAAAATTA 1680

GTGATAAAAA CAACAaAAAA TAACCGTCAA CTTTAGCCGA GTTTGGTTAT TTTTtGTAAA 1740

```
AATATTTAAC TATCGGTCAC CGTCTTAAAC GGTGCTATCA TGAGGGCGTA TGACTAGTAC   1800
GTCCTCTTTA CATGTTTATT ATAACACATT CTCTACAAGA AGAAAGGCGG AAGCCTTTCT   1860
TTTTTTTCAG CGCGCCTTAA CGGCGGTGGC GGTGGGAGTT TGGcGCGCTG AaAAAaACGC   1920
TCtTTTCTTG ATTTCTAATC AAGAAGATAC TCGCTAAAAG CGAGTATCAG GTTTGTCTTA   1980
TATTATGATG ATAAGGCGTG CTGTAAATTC TCTTGCCATT GTCTGACTGC TTGTTTTATC   2040
TTTTCTGGTA CTTsCTGACG GTTACATwtn CCTTGTATTT GTATAACCTG TTGATTTCGA   2100
TATTCTAATG TGAAGTAAGG CATATCTGGC TTTTCTTTTC TTCGGATAAA AACAATCGTT   2160
GTTTTTCCTT GCGTATGCCG TTCGATATAG TGTTGACTGC CCACACAATG GCGTAGAATA   2220
CTTCCTTCTT GAATGATTTC TTGTAATGAA TGTGGTGTGA GAAAAAGCAA ATCGTCAATT   2280
TCAGCTTCTA ACGCTTTAAT TTGGTTTTTC CGTTCTTGAT ACTGCTTTTC TTCAATCTCA   2340
CTTTTCAAAA GGTTTAGCGT GTTCATAGCG TTATCGTGAG CTACTTGTAA GTTTTCAGGA   2400
TACAATACAC TATCGTTATT AAGTGGGGTG TTCAACTCTT CTAACATGTG TAAATAGTCC   2460
ATATAATAGT CAAAACGCTC CCCTTTTCGT ATAAACCATT TTTGAAACTT TGTTAAGTTC   2520
ACTTCTTTCG GTAGTTGCTT AATTTGCGAA TAATGCACGT ACTTTTCTAT TTCGTTTATC   2580
ATAGGCGCAT TTCGCTCTTT TAACGTTTTT TCTAACATGA ATTCATGAAA GGTTGGATTT   2640
CTTGTTTTTA AtTTTTGtTT ATtCGCTTTT AGCCATTTCT TAGTAATTAC TCTCATATCT   2700
ACTTTATGAA AATACAGTCC AAAAACGTTG CGATAGTCGT CAAAAATGAT TTCATTCGCT   2760
AAGGTTGTGG CACCTATCTT TTGTAAAAAC TCAATTTCGT TTCGGTATTT GTAAATATGA   2820
GCTAGTTCTT GACGTTCTAT TTCAGGTAAC TGTAAGTACT TTAATTCCGA TACAGTGGTT   2880
AGTTTTTCTT GCCAGTTATT TGGATAGAAT ACTGTATTTG TATAAGCACT ACTCATAAAA   2940
CCATAATTTG CTAACAAGCC CTCATGCCAA TGTCCGTTCT CATAATGGGC TTTAACATGG   3000
GTATCGTTTG AAAACCGTTC AAGGTTTGAC GGCTCATAGA TAAGGTTTTC TATTCCATCT   3060
GTAATCGTCT GCCAATAACT ATGGGTTTGA ATTTCAATTC GTTTTGCACG AACGAGAATG   3120
ATTCCAAAGG AGTAAAATTT CGTTGGAAAC GATAATCGTG AGTATTTTGT TAATCGTTTT   3180
TTTATGATAG GGcAATTCTT GCGAGAACTC GCAAGAATTG TCTCCTTTTT GTTTTGCCAT   3240
TCATACGTGG GTATTTGTGC AGTACACCAG TTAAAAAAGG CTTTGGGTGG GGTTAATGGT   3300
TTTGGTGTTT TTTTATTCAT ACGCACTCGA ACAAATCAAT TTGTTCGTCC ACCTCTTTTC   3360
TTTCTGATTT TGTTTCTGAT TTTTTGGTTG GTTTCTTTTT TGTTGTGGTC GTTACTTTAC   3420
CAACTGCTTT GGTTTCTGGT AATTTCTCTA ATAAGAAATA CTCTTTTATC CAACTGAAAA   3480
CCACTTCGTC CGCCACCATG GCGCTTGATT GGTTTTTGGC TTGTTCTTGG GCTTTTCCTA   3540
TGCAGTAAGT CATAGCACCT TTTAACGTTC TATCTTyTTT GAGAATACCT AGAAAAAGCG   3600
CTTCGTCTGT TTGTCGGCAT AACCAGTTAT GAACAACTTC TAGTGTACTG TCATGGTCGG   3660
TTTCTAGTTC TTTTArCATT TTCTCAAAAG CTTGTTCTTG CGTGCTTTCT TGAGCTGTAA   3720
```

AGAAAAATCG TTCGTAAGAA wkTTTGTCGC TAATCACTAG CCAyTkrTAk TTrTTTCCTT 3780

GGTCTATyGT TTCGTCAATy GTTAGAAAmA GTGAyGArTA GTkwCCGTtG GggTCTTTwA 3840

TCGyAACrAT yGCTTGTTCw GGTAATTTTT CAAGGTAAGy CArTAACGTT TCGACTGGAT 3900

AAAGAGCGAC CTCATAATTA GGATAAAAAC ATAAAAACTG AACGTTTTTT CCGTCTTTTT 3960

CAATGTGTTC TTTTCTGAAA GCGACACTGA TATAGCCTTG TATGTCGTCA AAAATTGTCG 4020

TTGCTAGTGT TGGTTTGTAT TGTTTGAGTG CGTGTGTTAA TTCATTTACA GTTTGTATAT 4080

TTTGAAAGTC CATAACGTTC CCTCGTTTCT TTTAAGCTTC CCCTCTTACA TATGTAAGAG 4140

GGGAGCAAAT TTTCGCTTGT CAAAGGTCAG TCGGTACGTT ATACTTGAAC TACGGTTAGT 4200

AATTCAAGTA AAGTCCTATT AGAAATCTTT TGTCGTTAGC AAAAGGTTTC TTTTTTTGTC 4260

TTTTTTTCGT GGAAATACTT TGATAATGGT TCAAAATCTT TCATATCTCC ATTGCAATTC 4320

ATAGTACTTA CTAGCCACTC TCCAGACTGA CGTTGTGCAT ATAAATAGCC GTTGTTATCA 4380

GAGCCACGAT AAGGTGTAAA GGTTACTTTC CCTTCAACGA GTTCCGTCCA ACCGTCAATT 4440

CGGTCAAGGT CAGTTAGCGT GAGGTATTGT TGTTCTCTTA TTTCTTCTAA TGATTGTGCA 4500

AGATACATTT CTGTTTCTGG TCGAACATAT AGACAAAACT CGTGCCAACC ACAACAAGTT 4560

AGTTCCCGAC AGCCGACATT GTCGTATTCG TAAAGTACGT TTTCTTTCAC ACCTAAACAG 4620

GACAGGCCTT TTTTCTGATT GATAACTTTT TGAATTTTTT CAGGGTCTTG GTAATGTTCA 4680

TATAAAACAG CACCTACTCC CTCTATATAA CCGTCCCAGT GACAATATAT GCCTTGATAC 4740

GTTCCGTCTG CTTGTTCTTT AAAAATTGCT GTTCGTGTTC CCATAAAAGC TCTCCTCTTT 4800

CGTTGTTTTA AATGATAAAC TCTGTTGCTT CGTGATTATC TTCTGTCAAG ATAATTTCTT 4860

GGTCGTAATA CATGTTAGAA ACCAATTCAT GTGCCTGTTC TTCGCTTTCA GCAAATACAG 4920

AAACAGCTAC CTCTAAGGTT TCAACAATTG TTACACAGTA TTTTTTCTTT TTTTCTGCCA 4980

TATTTATACG TCCTTCCATT CAATGTGATT TAAAAAGAAT GCTAACTGAT TCAAACGATT 5040

ATTGGcATTT TTCAGAcAAC CATAGTTTAG TAACAATGTA GGCGTATCAT TTTCAAGAAT 5100

GGCTTTATTT AACTTCTGAT ATTCTTTACG TGCTTGCTCT AGCTGATTTT TTAGCTGTGG 5160

GTCGTTCATT TTTACTACTT CCTTTCTGGT ATTCTATTAT TTTTGGTAAT CTCTATCAGT 5220

GTTTTTTGCG CCTCTAACGC AGTTAAAATT GGTAAGCAAT CTCCTAAAGT CATACCGTTA 5280

CCGACAATTA CAAAAGGTTG AAGTAACTGC CCTACCGTCA TTTTTTCGTC TGAATGAAAA 5340

AGAGTAAATT CTTTTTCtAA AATATTTTCG CCTTGTAACA AACATTCAAT ATACACTTTG 5400

CTAAAGTCGT AAAAACTTTC TATTGAAAAT AATTTTTGAA GATAATATTC TACAGTTTCA 5460

AAATCAATGC AATCAAAATA CAAATTATTT TCATCAATCA AAAAACATTG AGAGCGTAAT 5520

GTGTTAAATA AATCTAAACA TTCTTCTAAA TCTAGCGATT TTAAATCGTT ATTTAAATAA 5580

CGATTAATGG ATAAGCTATT AATGAAATTT TCTTTTTTGT TCATTGGTTC TCTCCTTTAT 5640

TTTTTTGTTT TTGAGCAACC ATTTTTGTCT GGCGTGAGGG AGGGAAGCTG TTCGCTTCTT 5700

TGGCTCATGC TAGGACGTGG GACACTATCC TCCTTTCGTT TTTAGATTTG ATTGTTCATT 5760

TTCGCCAGTA TGTTTCGGTC GGTTGCTGTG CCGAGAAAGA ATCATAGACA CCAAGAAAAG 5820

ACAAGGGCGA ACGGAGTGAG TTCATATACC CTTGTCTTTT CGACTGGCTA TGATACTGTA 5880

ACAAACAGCG ACTGACCGGG ACATACGACA ATGAGCGGTC ATAAAACGAT TGAGAGGACA 5940

CACGTCCGTA AAAgCATGAG CCAAAGAGGC GAAATTTTAA TGCACTTAAA AGAACTAGAC 6000

GCAAAGCGGC TAGTTCTTTC TTTTGTTTTG CCTTGCTTTA GCAgGCATTT CTTTCGAGCG 6060

GCGGTGGCGG TTGGAGCGwG ATAGwAATAA ACACTAAAcA aAGTAATnAG GTATGAAAAA 6120

TGGAAAGACG ATTTTTCAGT GAATAG 6146

(2) INFORMATION FOR SEQ ID NO: 140:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 656 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

TTTTTGTCAT GATTTGTCAA TACTATGCTT CAGTCGCTTC GCAAGGTTTT GGGACGGAGC 60

TAAGAAATCA ATTAATGAAG AAAATTAATC AGCTTTCACA CAAAGAATTG AATAGTTTTG 120

GTACAGATAC CCTCATCACC CGGATCACAA ACGATATCAA CCAGCTTCAA TTAGCTTTAG 180

CGATGTTTAT TCGGTTGGTC ATTCGGACAC CTTTTTTAAG TATCGGTTCT GTGGTGATGG 240

CTTTTTACAT TGACGTGCAG ATGGGCTTTC TTTTCCTATT ACTTTTACCA ATTTTTAGCC 300

TTATTCTCTT TATTATCATT AAAGTGACTG TGCCTTTATA TCAAAAAGTC CAAGAATATT 360

TGGATCGGTT AAACCGTCAA ATCAGTCAAA ACTTAAGCGG TGTCCGTGTG ATCCGTGCGT 420

TTGCTAGAAA GGAAACAGAG CAACGACATG TTGATAAAGC TTCAGATGAT TTAGGCGATA 480

TTTACATTCG TGTATCGAAT GTsTCTGCTT TATTAACGCC TTTAACCACT TTGATTATGA 540

ATGTTGGAAT TTTATTTTTA CTTTATTTTA GTGGCTTAAA AGTTTCTTTT GGTTCCTTAC 600

AACAAGGGGA AGTTTTAGCA TTGATCAATT ATATGAATCA AATGATGCTC GCTTAA 656

(2) INFORMATION FOR SEQ ID NO: 141:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2776 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

CGGAAGATTG ACAAAAGATC CTGACTTACG TTACACCGCT AGTGGTTCAG CTGTTGCGAC 60

CTTTACACTT GCTGTAAACC GTAACTTTAC GAATCAAAAT GGTGATCGTG AAGCAGACTT 120

TATCAACTGT GTGATTTGGC GTAAACCCGC AGAAACAATG GCTAATTATG CCCGTAAAGG 180

TACATTATTA GGTGTTGTCG GAAGAATTCA AACTCGTAAC TACGAGAACC AACAAGGTCA 240

ACGTGTCTAC GTAACTGAAG TGGTTTGTGA AAACTTCCAA TTGTTAGAAT CTCGTTCTGC 300

TTCAGAACAA AGAGGAACTG GCGGCGGTAG CTTTAATAAC AACGAAAATG GTTATCAATC 360

ACAAAATCGT AGCTTTGGTA ATAACAATGC CAGTTCTGGA TTTAATAACA ACAACAATAG 420

TTTCAATCCA TCATCTTCTC AGTCGCAAAA CAATAACGGT ATGCCTGATT TCgATAAAGA 480

TTCTGATCCA TTTGGTGGCT CAGGTTCATC TATCGACATT TCAGATGATG ATTTACCATT 540

CTAAACGAAT TTGATAGGAG GGAACAAGAA TGGCACAACA AAGAAGAGGC GGACGTAAAC 600

GTCGGAAAGT CGACTATATC GCAGCAAACC ACATCGAATA CATCGACTAT AAAGATACTG 660

AATTGTTAAA ACGATTCATT TCAGAACGTG GCAAAATTTT ACCACGTCGT GTTACAGGTA 720

CTGGCGCTAA AAACCAACGT AAATTAACGA TTGCTATTAA ACGCGCACGT ATCATGGGAT 780

TATTACCATT CGTTAGTGAC GAACAATAAT TTTATGACAA TTTCCnGAAC CAAAGAGCTT 840

GCTCTTTGGT TCTTTTTTAT CTTGACTAGT GTTTCACGTG AAACATAACA AGGTTCGCAG 900

AATGAAGTGT TTTGTGATAA AATAAGGGAT GAAATAACTA GTGAATCACC AGAGCAATAT 960

GCTCGTTGGA ACAAGTCGCC TCAAAGGTCT GCAGAAAGAT TGGAGGCAAA AAAATGCAAA 1020

AGAAGAGAAT TCAAAAAAAC GGTTTCTTAA TTGTTGTGGG TCTTCTCTTA GTAGAATTTC 1080

TCCTCTATTT CTTACTAACA AATAAATGGC TGCTATTGGC GGTAATTATC GCATTAGATA 1140

TCTTTCTCTT AGTGGTTATT CGGCTGTTGA TTAGAGATGT AGAAATTACG AACGTAGAAA 1200

AGATTCAAGA AGCAAGTTCC ATTGCTGAAC AATCGTTGGA TTATGTTGTA AATGAAGTAC 1260

CTGTGGGAAT TATTACGTAT AACGGGGAAA CACGCGCGGT AGAATGGCTT AATCCTTATG 1320

CTGCTTCTAT TTTTAATAAA GACAATCAGC TAACGTTAAC CGCTAGCCAA GTTACGTCTT 1380

ATTTAGAATT AGCAGAACGA AACCAAGATA TTTTTACGAT TGACGAAAAT ACCTATCGCT 1440

TTAGCGTCAA TAAAGAACAA CATACAATTA CTTTTGAAGA TATCACTAAA GAAAGTAATT 1500

TGTATCAAGA AAAAGTCGAA ATGCAAACGG CTATTGGCAT TGTGTCTGTC GATAATTATG 1560

ATGATGTCAC CGATACAATG GATGAGAAAG AAATTTCTTA TTTGAATAGT TTCATTACGA 1620

CGATGGTTTC TGATTGGATG GACCAATACA AAGTTTTTTA TAAGCGAATC AACGCAGAAC 1680

GTTATTTTTT CATTGCCCAA TGGGAAGATA TTCAAAAAAT GATGGACGAA AAATTTTCTA 1740

TTTTGGATAC GATCCGTAAG GAATCAGCTA ACCATGAAGT AGCCATTACG TTAAGTATGG 1800

GGATTGCTTA TGGGGGCCCA ACCTTAGATC AAACCGGGAC CACGGCTCAA ACAAACCTAG 1860

ATACAGCTTT AGTTCGTGGT GGTGATCAAG TGGTTGTAAA AGAAGCCAAA GATGAAGCGA 1920

AgCCGTTATT TTTTGGTGGA AAAACGGCAG TAACGACGAA ACGTTCCCAA GTACGTTCTC 1980

GCGCAATGAG CATGGCAATT AAGGGAATTA TTGCGGAATC TGCTGACATC TATATTATGG 2040

GCCATCGTTA TCCAGATATG GATGCGTTAG GTTCAGCATT TGGTGTTGCT CGTTTAGCCT 2100

CGTTTAATAA TCGAAAAGCG TGGATTGTTT TAGATGAAAA TGAAATCATT CCCGATGTCA 2160

AAAGAGTGTT AGAGGCGATT AAAGAGTACC CAGAATTAGA AGAGCGCATT ATTAGTCCTA 2220

AAGAGGCCAT GAAGCGCAAG AAAGAAAGTA GCTTATTAGT TATGGTAGAT TACCATAAAC 2280

CGTCTCTATC GATCTCACAA GAGCTCTATG AGCGTTTTGA TAAAGTAGTA ATCATTGATC 2340

ACCATCGACG AGGAGACGAA TTTCCAGCAA AACCCTTGCT TTCTTATATT GAATCTTCTG 2400

CCTCTTCTGC TTCAGAATTA GTCACAGAAT TGATCGAATA TCAAAGTAAT AGCGCAAATA 2460

AACTGCAGGC CTTTGAAGCA ACCATGATGT TGGCGGGAAT TGTGGTTGAT ACGAAAAGTT 2520

TCAATACACG AACGACGGCG CGAACATTTG ATGTGGCTAG TTATTTACGA ACTTGTGGAG 2580

CAGACTCATC TTTAGTACAA TATCTATTAA GTTCTGATCT TACAAGCTAT CTGGAAATGA 2640

ACAATTTAAT CTCTAAAAGC GAATATGTCA CAAAAGATAC CGTCGTTGTT GCAGGGAGTG 2700

AAGACAAAGA ATATGATAGT GTCACAGCTG CCAAAACAGC GGACACATTA CTTTCTATGG 2760

CAGGGATTAA TGCAGC 2776

(2) INFORMATION FOR SEQ ID NO: 142:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7337 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

TCCATTCCAG CTATTTCTTT CTACATCTGT AATATGTCTA ATGTTATCTT GATCATGCGC 60

ATCAAACTCA GATTTAGTTG CCTGCTTTTG ATTTAGTACA TTTCCtwAAC CaACCTGakT 120

AGCTGTTACC GCatGaGGAT TACTTATATT TGATTTATGG GCATTAAAAT CTTCAATATG 180

TGTAGCACTA TCTTGCAATT CTTTTAAAGC TCCGTGGGTT CTGCTAAAGA ACCAATTAAA 240

ATAATCAGCT GGTGGTTTTT GTGAAGCTTT CCAACCATCG CTAGTTAAGC TTTCGGGTGG 300

TTTAATACCT GGTGCTAACC AAACAGGCAA TTCCTTTGTA AATTTCATTT GCATTCTCCT 360

CTCTTTCTAC AGCGGCAGTT TATAATCATT TTCTGGAATA AAGACTCCTC CCAAAGTTCC 420

TCCGATATTA CCATCAATAT CAGCGAATCC TTCTTTACTA ATCTCTATAT CAGTCGTACT 480

TGAAAATGAA AAAGTTCCCT CCAAATTAAC GTAGGCTACT CTAATGCCTC CAGCACTGAT 540

ACTCTCGACA ATTTGTAAAA ATTGACTAAT ACTTAGACCC GAACTATTCA AGTAGTCAAG 600

AGGAGCTTTT TTTATGATTA CACAAGCAGG TTCTTTTTCA TCTACCGTCT CATTAGCTGA 660

TATAATATGG ATATCACTAG GATGACAATT TAAAGATGTC GCAATTGCAT GTAACATTTT 720

GTCGATTGAT CCATCGCTTG TAtTTCTAGC TACCTTTCCA CGAATCAATA CTCTATATAT 780

TTCATCAGTC GTTTTtCCTC TCGCTTGTCC AACATTAGCA CCTAATTCAT CTAGTCCTTT 840

CCCTCGTGCT TCATCTATCG ATCGCCAGTT TGCTACTTTA TTCAGCAAAT CAGTCAAGTC 900

TTTAATTTCA AACTCAATAA TTTCAAGAAG TTTTGAAATA TTTGAATTTT CGCGATTGAA 960

```
TAGATCtGGT AAATAATCTC TAAGCTTTTT GGTCACTAGT AATCACCACA TTCTCTGTTT    1020
TACATTCAGC TGACTCATTT GGATTCAGAT TAATGTCTTT GACTTCGGTA GTTTCAGTAG    1080
ATAATCCAAT TTTTACATCA GCGACAACAA CGCCTGGnAT TTGATAAATT AATGGaTAAA    1140
TATAAGAaAA TCTGACTACT TCACCCaTAG tCaGATkGTw AATGTAATTA TTTACTATAA    1200
CTTTTATTtC TTCTTGTCCG kTTTTTTCAA ATTTTGAATC TACTTGAATA GAAATATTCA    1260
CAAAAATAGG AACAGATTTT GCATAATCAA ATTTAACAAT ATGACTAAAT CCTCCTAAGT    1320
CTTTTACTTC AACTTCCTGA TTACCAACAG TGTCAATCCC AGCGGCAACA CTTTTAAAAA    1380
TTGCCTGCCC AATATCGTCT TTCACTCCAC CTAAAATATG CACATGTACC GACTTGGGTG    1440
GATTACCATA CGAATCAGTT TCCATTGTTT TATTTTCAAC AACACTAGCG GTCCGTACCC    1500
CACTAACTTC TAGCAAAGCA GTTAGAATTC CGTTTATTGG CGGTCCAGGA TTCCCACGAA    1560
CAGAAATACC AATTCGGTCG CGATAGGCTT TATCTGTTTC ACGTTCTGCC CCACCCTCAG    1620
CCCTAGCAGG GTTATTAACA GATGATATTT CTTCAGTAGG TTCCACTTGT ACTATAATAG    1680
TATTAGCTGG TACGTTTGAA CTAGCATTCT CTTCTAAAGA AATTGCGCGG CCTTTCCCAA    1740
ATCCATTATC ATCAATTTTG ACTATATCAA TCATCTGAAA CATAACTTTA TTTTCTGTTG    1800
AAAAACGTAC GCCTTCATTA ATGATGTAAC CAGGTTTTCC TGAGAACTCC AGCTCAACCA    1860
TTGCAACTGT AGCTGGATTT CTATAGATAC CACTATTCGC ACCTAAACGA TCCAATGAAA    1920
CCCCTGTAGC TTGACTTATG AAGCCACTAT AATAAACTCT TTCAGTTAGT TCATGAGATA    1980
TAGATAAAAA CCATGCAACG ATACGAATAA TGATACCCAG AACAGAATGT CCTGAAACAT    2040
TAACATCCGA ACCAAACAAG GCTTTTGCCT TTTCAGACAT ATCATAGAGT ATTTCATCAT    2100
ATGTTTTTCT TTTAAATCCA TTTTCATCAA GCACGTATTA CCACCTCCAC TTCATCACCT    2160
AATGTTGATA GCATCTCCAC TGTAATATTC AACTGTCTAT TGTTTCTTAC AATTTCTATA    2220
TTTTCAATAC TATTAATTCT AGGTTCTTGA TCTAAAATCG CTTCTGTAAT ATCTTGTTTC    2280
AAGTAATCTT CATTGTAATT TTTTCCAAAC ATATTATCAC TTTCTAAACC AACGGATGTA    2340
TCTAATTTAA ATTCTTCTAG TCGTATTGAT AAGATCATGA ATACACTTTG TGCAAGTTCT    2400
AAATCTCCTT CAACTAGTAA AATTCCATAG TCAATAAATG ATAAGTCTCC ATTGACTATT    2460
TTCAAATCCC TCATTAAGCA ATCACCCCCA TAATAACAGC GTCATTTTGA CTATGCATTC    2520
TATCAGATGA TAATGAATAA TCAGCAGAAC TTCTATAATT ATCTATGTCA CGATCACAAA    2580
AAACTACTAC TACAACCGCT CCCTGTGAGA TATCAGACTG ACAATGCTTA GTGATCAACG    2640
CGTTTAGTAT TAATGCTCTT TTGCCTCCAT CTGATTTTAA TGCCATTGGC TGAACATCAG    2700
CTCTTGCACC TTTTACAGTG ACAACTCGAC ATAGTTGCAT AACATTAATT TCTTTTAAAA    2760
TTCTATTTTT AAATGATCGA AAAAAAGCTA AGTCAGTTTC TTTCATTAGG TTACCACCTC    2820
ACACGTTGTA ACAAATCGAC TACCATCATA GGAGTGCTGG cCATTTTTTA CATAAAAATT    2880
TCCyTTTACA TTTTTtGAAT CGATATAGAC TGCTGTACCT GTAGTGATTT TATGTTGCAA    2940
```

AAGACACTCA ACAGACCAAC CAGTATAATC ATGACTTTCA ACTTTTGTTG GTTGGTTAAT 3000

AAGTCCCGTT TCAGGACTTA ATCTGTATCT CTCTTTATTT CCATCTCGAA AATTTTTAAT 3060

TACAAGTTGA CCTCTTTTAA AATATAGTGA TGTGTCACAA GCTTTGGAAA TCTCTTCTAA 3120

AACCATCATT GCTTGGCCAT CAGCTGTGTA GCCAGACCCA TATATTTTAT TATTTTTTAA 3180

CTTAATTTCT GACAAAGGAA TATTTGCTTC CCTAGCAACT CGATTtATAA TAGTAkGrGC 3240

ATCCGTyCCA TTATTAAAAG TAATATTTAC TTCTTTTkGT yCCGAGTAAc TTTACCTyCT 3300

AAAAAAGTAA ATGTTGTTGC TCTATCTACc CGTTTAGAAC AGAGGGTTTT ACTTCTGCTA 3360

TCGTTCCTGA CGTAATCACT CCATTTGAAG TACCTGCGTA TCCTGCGTGG ATATAGACTG 3420

GATTTCCTTT CTTAATAAAA TCAATACTTT TTTTATTTAA ATTGTATATT GTCACACTCA 3480

CTTCAGATGG ATTAGGTGAG TCAGAAAAAG GTGCTGTAAA GTGAATTTCT AACCGATCCA 3540

ACCTGCCTGA ATCAGCTCTT AGTAGAACTC GATTCTTCCC ATTCTTGTCG TGTATTTCAA 3600

TTTGTAATAA TCGTTGCCAT TGCGTATTTC CCATTTAATT ATCTTCCCTT TCCAAACTTG 3660

GAACATTATA ATTCGGAGGT AAATCATCTA TATAAAGGAA AACCTGAATA CCAAAATTTT 3720

CAAAAGTTAT TTCTTTTGAA GCTCCCGACT CATCCATCGG CACCAAATCT GCTGAAGGTA 3780

AACGAGTATC TACAATGTCT TGCCAAAGTC TTTCATCGAT TACCATACGC TCACCAATAG 3840

CTATCGGTGT ATGATCTATG TCATATAGAT CTACAGTAAA AAATTTTTCT GTTTGATTAT 3900

AGTCCACCTC AAAAATGTAA TTTGTATTAC CTAAAGGTAT TTCGAATTTT TCAGGTAATG 3960

AATATTTATC AATAGGAATA TATGCTCTTA AAGACATTTA GTCACCTCAT TTCACACGCG 4020

CACGAGCGCC TATGGGTATA AATCTATCTG GCCACTTGTT CCAATCTCTT AATTGCTGAA 4080

TAGACGTACC ATATTGTTGC CACCAACCCC AATAAGTATT TCCTGGCTGA ACTGTCACAT 4140

ATACCGCATT ACTTGGTTTC GAAGGTTGTT TTGCAACTGG CGGTTGTGGA ATCTTCTCCC 4200

AAATAGTCTT AGCTACACGT ATGGGTTGTA ACGTAATTGT TATAGTGAAA CCATTTTCCA 4260

CTGTGTCATA CCCTTTACTG ATATCTTGTA TAACTGCATT TTTAAAATAT GATCTGCCTC 4320

GATAAACAAT CCAACGAACA TCTTTTTGCC ATGCAACTAA TGTATTGTAA GCTTGCTCTG 4380

CAGCGTtACC ATTTTTAGCG AGAATCCAGC CACTAATTGT GACTGGACCG CCTGTATACA 4440

TCATATTGTC AGTGATTGGT GCTCCTGaTT CAACAGGATA TTGAGATACA TTAGCAGCAC 4500

TGGTTACAGT CTCACTGaCG TTTACAATCT CAACTTTAGA TTTACCACTC TGaATGTATC 4560

CCATTAAACA ACCTCTCCTG TTCCTAGAAT ATTAAGTAAT TTTGCATACT CATCTTCTAA 4620

TGCTTGTCTT ACAGCCTCTT TAATATCACC AACAACAGAT TTATCAGAAT TATTCCCAAC 4680

ATTGATAGTA ATATTAGGAC TAAAATTAAC AGATGGAGAT CCTTGATTAA AGAGCGCTTT 4740

AGTTTTTTCA TGAGGATGTA CAGTTCCAGC TGTATCTGCT TCGAACAATT CTGGTCCATT 4800

CTCTCCTACT AAAACTGTTT CACCTTTTGA AGGGCGGCCT CCCTTTGCAT AAGCCCTTAA 4860

TCTATGACCT GTTGGTCCCC ATCCAGAATG TCCATACGGC AAATCTGTTC TCCAATTAGA 4920

ATTATTGAAA AATGCCATTA GTTGATGGAA ACCATTCATT ATATTTTCAT ATCCACGTAC 4980

TTTATATGCA TCAAAAGTTT GAGGAATGTA CTGAAGCAAT CCTCGAGCTG GGTTTCCACT 5040

AGCTGTATTT ACATCCCAAA CAGCAGAACT TTGAATAATA CTTTGGTTAC CACTAGACTC 5100

CCTCTGTATT TGAGCTAAAA TGCCGTTTAC TTCTGCATCA CTTACTTGTT GATTCATTTG 5160

TTTTGCAGCT TTTCTAACTT GAGGACCCCA TCCTGCAGCA CCTACCGCCA CACCACCAGC 5220

AGCTCCGCCA TAAGTATCAG GATTTACGTG TTGGCCATTT GGTCCGCCTT TTCTAAGTtC 5280

ATAATGAACA TGAGGACCAG TTGACCAACC AGrAGrACCT ACATCACCGA CGATTTGACC 5340

AGCTTTAACT TTATCGCCCA TCTTAGCTCT TATGCGGCTC ATATGTCCAT ACATAGcCCA 5400

TAGaTTATCm GcAACTTTaA TTCCTACGTG CTCACCTAAT CCAaTGGnAA GAAGATTGAA 5460

CCCaATCAAC CAAACCaGGA TATTGAGCTG GaATAGGCGT cCAGTTGGAG CAGCATAATC 5520

AATACCGTGT GAAAATCCCC ATAGAGGCCT GGTCGCTTAC CATAATCAGA AGTACGAACA 5580

AATGGTGATC CAAAATGTGG AGCAAAAGCA CCTGAGCCAA AAGAGACACC ATCTTCCGAA 5640

TCAAAAAAAC CTTTGAAAGC TTCAAGTTTA TTTTTCACCC ATTCGGCACT GGTGTCTTTT 5700

AGTTTGTTCA TGACACCAtG TCCCAAACCT tGGATATTTT TACCCTTTTT ATAAGkGTTA 5760

TGcTTATTAA ATAAGcCAGT CACTTTTCCa ATCGGGTCTG ATAGCCAATC TTTTGcTGTT 5820

TCAGCTATAT CTTTAGTTTT TTCTATAGCT TTCGTTCCTA CTTCTTTTGC TTTATCTACT 5880

GCATTAGATC CAAAGTCTTT TAATTTATCA ACAGTATTTG CACCGAAATC CTTAGCACTA 5940

TCAACAAAAT CAGAAAGCCC ATTTCCTTTA CCTTTATGGA AACCTGGTAA AACTGTTCCT 6000

TGCCCCATAC CACCATTTAA TACAGCTTTT GTATCTGAAT GGTTTAAAAT CCGTTCACCC 6060

GATGTTACAT TAGTAACTTC TGGACCATTT GATCCTAATA AACGCATTGA AGAAGTGCTC 6120

TTGTTATATG CTAATTCAAC ACCTTCTTCA CCAACTAGAG CTTGACCACT GTATGATGCA 6180

CCACGAGATC CTGTGCTAAA AGAGCGACCA TGCTTGTTTT TAGTAGAATT AGGTGTATTC 6240

GGTTTCCATT CTGGTATTTT AGGAATACTA AAGAAATCTA ATACGACATT AATACCACCT 6300

GTAACAGCAT TTACTTTATC GGCGAGTCCT GTTTTAAAAT TATCCCAAGT TTCTAATGCT 6360

CCAGACACAA CTTTATCCAT AGCTTCACCA AAGGCCTTAC CAAAACTTTT AGCGCCTTCA 6420

ACTGTGCCGT CCCAAATTTC TTTAAATTTC TTTTTGGAAT CCTTTTTAAA GACTTCCCAT 6480

TTAGATAATG TTTGTCCAGT TTCCCAGTCC ACTTGTTCAA GATGACCTTC CGCTTGCtTT 6540

TTAGCCTGCG TTACAACATC TTCATGTTGT TTTTCTGCTT CTTTGACTAC ACCGTCTCTT 6600

TGTTCTTGGG CTTTTTTTAA TGCATCATCA TATTGTTGTT GAGTAATCGT ACCATTGACA 6660

TATCTTTCTT CATCTAAAAT GCGCTTTGTT TCTTCATATT TTTCATTTGC AGATTTAATT 6720

GTACCTTCTT TTGCTTTATA TGAAGCTGCT ACAGCATCAG CAGCCTGTTG CGCGCTAATC 6780

TCACCTGATG CATTTTTCAA ATTACCCATA ATGATTTTTT GCTCTTTAGC TGAATCTGAT 6840

AATGCAACAA CAGCTTGATT TTTCATATCT TCATGAATCG ACTTATTAGC AGCAGCATGT 6900

TTTTCTTCAA CAGCACGTAC TGCAGCAGCT GAGGTTTTTT CTAACGACTC AATTTCTTTT 6960

TTATCGTTTT CTGATAGTTC TCGATTTTCT AATCTTGCTT TTTCTTTGAT TTCATTGATA 7020

CGATTTGTGT AATACTCTGT AGCAGTAATA GATTCATCAT ATTCTTGTTT TTCTAATTTC 7080

TTGAAATCTT GAACTTTTTC AGAGAACATA TTTGTTCGTA CTTTTGCTAA TTCATCAGCA 7140

GCTGCCTTGG CTCCTTGTAC ATCCTTTtCA TTAAGAATTC CCATACCAGC TAATTTATCA 7200

AGATTGCCAT TAGATTTATC TkTCTkTTTA TkCAGCTGTT TAGTAACTTG ATTCTGCATG 7260

TCTTCTAAAG CTTTATATCG TTTAGTAACT TCTTCCATTG GTAAGAACTC GTCCTTCAAT 7320

TTTACCTGTA GTGTAnT 7337

(2) INFORMATION FOR SEQ ID NO: 143:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1160 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

ATAAGGAGGA AATCTGATAT ATGAAAACAA AACGCTTTCA GCGCACGGCG GCGATGCTGC 60

TGTGCATTTT TATCTGCGCG TGGACAATGT TCCCCCTGAT GGGTGCGGCC TTTGCCkCsG 120

AGAATGGGCT GCATGACACC GCTGAACCGG AAATCACCTG CTTCTGTGAT ATTCACTGCA 180

CGGCGGATGC CCTCAATGCA GACTGTCCGG TTTGTACACA GGATGTGAAC GCCTGCACCG 240

GCACGGAACC GGAAGCGCCC ACCGAAACAG AAACACCCAC CGAGCCGGAA ATCACCTGCT 300

CCTGCAATAT CCGCTGCACG GCAGATACCC TCAACGCAGA CTGCCCGGTT TGTACGCAGG 360

ATGTAGCTGC CTGCACCGGC GCAGAGCCGG AAATTGCCTG TTTCTGCGAT ATTCGCTGCA 420

CGGCGGACAC CCTCAACGCA GACTGCCCCA TCTGTGTGCA GGAGGCAGCC GCCTGCATTG 480

GCGCGGACCG GAACCCGTCT GTACTTGTGA AACGCGGTGT GCGGCAGACG CACCTGACAC 540

GGATTGTCCC ATCTGTACGC AGGATGTAAC CGGCTGTACC GGCAAAGAGC CGGAACCGCC 600

TGCCAAACTG GTCTGTATCT GCGCCGCCCA CTGTGAACCC GGCGCGGTCA ACGCCGACTG 660

CCCCGCCTGC CTGTTGAACC TGGCAGACTG TATTGTGGAA GCCCCCACGC CCGTCCCGGT 720

CTGCGTCTGC GAAAACCGCT GTGAACTTGG TGCGGTCAAC GCCGACTGCC CGATcTGCAG 780

AACCGACCTG GCCGGATGCA CCGGcAAAGC CCCCGcGCCA ATyCCTGCAT CCAATCTCTC 840

CATCAAAATT ACACCGCCCT CCGGCTGGGC AACCGGCAGT GCTGCCGCCG AGTTTCGCAT 900

TACCGATGAA GCCGGGAACG GCTTTGCGCT GGTGCAGGTG AAAATCGAGA AAAACGGGCA 960

ATGGCGGGAT GTGACGGACA GCTTGAAGCA GAACGAAAAC CGCTGGTATG GGGAAATCGA 1020

CCTTTCGGAG AACTGCACCG TCTATGTCTG CGCcAACGGC CATGATGGGA AgTGTATGAG 1080

AAtCCCGCTA TATTGAGTGT TTCGACCGmA CggcCTGCTT ACGCGcCGGC ATkGAcGGcG 1140

GtTsTcCGTG CGGAACAAGC 1160

(2) INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12860 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

TTTACnTGAA CCAATTTCCA ATGTGGATGG GATGATTAnG CCAAACGTTC TGGCCATTnC 60

CCTGGCCATT TTTCAGGACC TTCTACCACA TATTGTGGAT ATTTTGCCAT ATATTCCGCT 120

GCTCCGGGTC TATTTCTTAC ACGCATATTT AAAATTCCTT TCTTTTAAAA ACAGACTGGT 180

GAAATTCGCC AGCCTGTGTG GATGTTAAAT GAATTAGTTT AAATACTAAA CATTCGTTTT 240

AATTGTAAAA TTTCTGCAGT CATTGCTCTT CTTTCGCCCT CTTGGTGATG ACGAACAATC 300

TCTTGTAAAA AGCTAAATAA ACCATACCAA TAAATTCGTT GAACCGTACT TTCATTCGGG 360

ATTAAACCAT AAGATAGTAA CCATTGATTC CAACCGGAAC GAGGGACATA ATGACTCAAA 420

ATCATTCCTA AATCTACCGC TGGGTCCGCG ACCATCACAG AATCCCAATC TACTAAATAG 480

AGATAATTAT TCGAGACTAA CCAATTCCGA TGATTCACAT CGCCGTGAAC AGCTACATAA 540

CTAGTTTCTG GATATTCTGG TAAATGATTA TATAAATAAG TGATTACTTC TTTTAAATAA 600

CTGTTTCGTT TTAATTCTTT AGGTAGTTGA CGCTTATACT CTTTCAACAT TGTTAGCGGG 660

GTACAAACTT CGCCACCAAT TTTTTTCAAC ATGTCTTTCA GCAAATTGGA ATGATGCAAT 720

TGGTATAACA CGTCAATCAC GTCATTGCGT TTGCCAATTT CGTCCGCATG TAATAAACGT 780

CCTTCCAGCC ATTCTTGGGC TGTTAGTACG TCGCCATTCG CTGTCCGCTT TGTCCAAATC 840

ATCTTCGGGG CTAACCCTTC TTTAGACAAT GCTGCAAGAA GTGGAGAAGT GTTTCGTTTA 900

ATGAACACTT yCTCGGTTTC TTkTATTCCC ATGTATGTTT GACCAGTTGC GCCCTTTATC 960

GGCTGCAGAC GCCAGTCTTT ATCCAACTGA AATTCCATGG AAATTATCTC CTATCTCAAC 1020

ACAATTTTAT TGCAATTTTT ACAGCAATTA CCATTTTAGT CGCCAAACCT TGCTTCGTCA 1080

AGCTTCTATG TGGTTCTTTT GTAAATTTTG CTGGGTTTcT CATCTTTATA ACTTATTCAA 1140

GCGCATTGGA ACATACATTT TAGTGAAAAT CACAACTTCT ACAATTAAGG CACCCAGAAT 1200

GATCAAGCTA TTTTGCAAAC TGCCTAAACG AAGGCAGACA ATCCCACCAA AAATAATGGC 1260

CGCAACAGAT AATGCGCCGC TAATTAACCG GCGAATCGCT GCTTGTTTTT TTTCTATTGA 1320

AATTGGATAT AGCTGTGTCA TCACCATATA ATCAAATTGC GTGTACATGG GAATCAACTG 1380

GAACCCGATT AAATAGACAA ACAATAAAGC AACTACCAAC GCAATCCAGA ATTCTTGTAA 1440

AAAGAAGAGT AGCACGCCAC CAACCAAAAT CAAACGAAAC AATAAGCCGC TATATTCGCT 1500

ACCTCGTAAA AAGCTCCGAA TGAATAAGTA AGCATACGTG TTCTCAGATG TTTTTTTGAC 1560

TACTCCGAGG AGAGGATCTA AGTATTTACG GCGTTTCACT GTACTACTCA CTTCAGGAAT 1620

ATCTGTAAAC AGCTGAATGA ATTTATAAAT TCTATGCATT CTGTTTTGTT CTTTTTGGAT 1680

CATTTTTTCC CAATCCAAGA AACTCTGTTT CATCTTATTT TGCCAAATAC TCGATAAACA 1740

AACGGCCAGA ATGATGCCAC CAATCACACC AACCCAAGGC GTCCAGTAAA TAGCTCCCGT 1800

CATAATCAAG AGACTACTGA TCAACCAAAC AAACCACCAA CTACGGTAAG TGGTTGACGA 1860

AAGATGATAG AGCCCATAGC TTTGTAAACT AAGATGTGTA TAAATCATAC AAACTAGCAT 1920

GACCAAAAAT AAAAAGAAGG TTTGAAACGA CCAACCTGTT GTGACTACAA TCAACGGCAT 1980

CAATAGGCCA CTTCCTAGAA AGCTGACTAC AATCGGTAAA AGTAAAGAGT AGCGCAgCGC 2040

TCGTTTCAGA TAAGCAGCGA ATTGTTTTTC TTTTGGTAAT AAAAACACCT TGTCTGGCTC 2100

TTCTGTCAGA GTAGCCAACT TACCAACTTG AATGAGTAGC AGCCAGCCCA ATAGAATTAA 2160

CGGTCTGCCC CAAACAAAAT CACGAGGTAA TTCTTTTAAC AACTGGGAAT AGTACAGACC 2220

TAATCCGCCT AATAAAAACA TACAGACTAA AACAAAATGA TCATTTAAAA TATAGCGCAA 2280

GTACTTGGAC ATTTTTTTGA AATGCCGTGA CAAACGTTGA CTAAAAAATT CGCCCATGTT 2340

ACCCCACCTT TTCTTCCTTC GTCAAGGCAA GATAAATATC ATCTAAAGAA GACTCTGGTA 2400

AATTAAATTC TGCACGTAAT TCAGCCATTG TACCGTTTGC CCGTAATTTC CCTTCGTGTA 2460

GCACCACAAA ACGATCACAA TATTTTTCAG CTGTTGCTAA AATATGCGTG GACATTAAAA 2520

TCGCTGCCCC TTGCTTACGC ATCGTATCCA TTAATTCTAA TAAAGCATGA ATTGCTAAAG 2580

GATCTAAGCC TAGAAAAGGT TCATCAATAA TATATAAACT CGGTTCAATT AAGAATGCAC 2640

AAAGTACCAT GACTTTTTGT TTCATGCCTT TTGAAAAATT AGCAGGAAAC CATTCCAATT 2700

TATTGTCTAA ACGGAACGTT TTTAGTAATG GTTCTGCTCG TTTGAACGCT TCTTCTAATG 2760

GAATATCATA AGCCAAGGCC GTTACTTCGA TATGCTCTtT CAATGTTAAT TCTTCATATA 2820

AAGAAGGTGT TTCCGGAATA TACCCGATTT TTTTACGATA TTCTTCAGGG GCTTGTTGCA 2880

ATGTTTCCCC GTCAATCATA ATTTTGCCTT TTTGCGGCGT AAGTAACCCA ATAATATTTT 2940

TGATGGTGGT ACTTTTCCCC GCACCATTCA AGCCAATTAA ACCAACCATT TCACCAGACT 3000

TGACGTCAAA ATTAATATCT TTTAAGACAG GAATATGGCC GTAACCTCCT GTTAAATGTT 3060

CAATTGTTAA GCTCATTTCT TTTGCCTCCA ACCTTTTCTC TCTACCATTA TTAAACCCTA 3120

AGCGAACCAA GAATTCAAAA ATTACTCCCT TAATTAAATG GATTTTTTTC TTTCTCTTAT 3180

TATACACGAA ATGAAACGAA CTTTTTAGCG AACTCACGTA CTTTTTCATC CaCAGAGAAA 3240

TCAACTTGTG GTAAAGTAAA GGAAAGAACC TAGAGAAAAG AGGAATTTTT CATGAATGAT 3300

TGTATTTTTT GTAAAATTAT CAACGGCGAA ATCCCAAGTT ATAAAGTCTA TGAAGACGAG 3360

AAAGTCTACG CGTTTTTAGA CATTACACAA GTGACCAAAG GGCATACCTT AATGATTCCA 3420

AAACAACATG TTGCTGATAT TTTTGAATAT AATGACGTTT TAGCAAGCGA CGTCTTTGCT 3480

CGTATTCCTA AAGTAGCACG AGCATTAGAA AAAGCTTTTC CAGAAATGGA AGGATTAAAC 3540

ATTTTAAATA ACAATAAGGA AGTGGCTTAT CAATCCGTTT TCCATTCCCA CGTTCATTTG 3600

ATTCCCCGTT ATTCAAAAGA AGATGATTTC TCTATTCATT TTGGCAACCA TCAAGAGGAC 3660

TATTCCGCCG AAGCGATGCA AGAAATTGCT GAGACCATTG CAAAGCAGGT GAACTAAATG 3720

AAAAAATTTA TTAAAGGTCT CTTTTTTGGC GCAgCTGCTG GTACCATTGG CGGCTTGTTG 3780

GCTGCTCCAA GAAGCGGGAA AGAAACACGC CAACATTTAA TCAATGAGCT AGAGGACTAT 3840

CGTTCACTGA AAAATCAAGT CACGAATGAC TGGGATCAGG TGCAACGAAA TTTGGCAGTT 3900

GTCGAAGAAA ACGTTCCACT GGCAACTGAA TTTTCAAAAG ACCTGCAACA AGAAATCACT 3960

GATTTCAAAT TCCAAGCGGA ACCTCGGATT GCTCAAATTA AAGAACAAAT TGCTAAAATC 4020

ACGGCTGAAT TACCTGACAC ACAAACCAAC AAGCAAAAGT AAAGTCTTTT ATGTTAATTA 4080

TTTCTTTTGT TATTAAAAAA ATAAAAAGTG TGACGTAAAT ACGAATTTTG TTTGAAACTA 4140

TGCCTGAGAC TCTTTCTTTT TCAAAATGAG GTAtGGTATA GTTATAACAG AwATAAAACT 4200

ArAAAAAACA GGAGTGCATA AGAGAATGAa GAaAAAACTA ATCTTAGCTG CAGCGGGCGC 4260

AATGGCCGTT TTTAGTTTAG CAGCGTGTTC AAGCGGTTCA AAAGATATCG CAACAATGAA 4320

AGGTTCAACA ATTACTGTTG ATGATTTTTA TAACCAAATT AAAGAACAAA GCACTAGCCA 4380

ACAAGCGTTT AGCCAAATGG TTATTTATAA AGTCTTTGAA GAAAAATATG GCGACAAAGT 4440

AACTGACAAA GAyaTTCAAA AAAACTTTGA CGAAGCCAAA GAACAAGTAG AAGCACAAGG 4500

CGGAAAGTTC TCTGATGCAT TAAAACAAGC TGGTTTAACT GAAAAAACAT TCAAGAAACA 4560

GTTAAAACAA AGAGCAGCCT aTGATGCAGG TCTAAAAGCC CACTTAAAAA TTACAGATGA 4620

AGACTTAAAA ACAGCTTGGG CAAGTTTCCA TCCAGAAGTA GAAGCACAAA TTATCCaAGT 4680

TGCTTCAGAA GATGATGCCA AAGCTGTCAA GAAAGAAATC ACTGACGGCG GCGATTTCAC 4740

AAAAATTGCT AAAGAAAAAT CAACAGATAC TGCTACGAAA AAAGATGGCG GTAAAATTAA 4800

ATTTGATTCA CAAGCAACAA CTGTTCCTGC CGAAGTTAAA GAAGCTGCCT TCAAATTAAA 4860

AGATGGCGAA GTGTCAGAAC CAATTGCTGC AACAAATATG CAAACCTACC AAACAACCTA 4920

CTATGTAGTG AAAATGACGA AAAACAAAGC AAAAGGCAAT GACATGAAAC CTTATGAAAA 4980

AGAGATCAAG AAAATTGCTG AAGAAACAAA ATTAGCCGAT CAAACATTTG TTTCGAAAGT 5040

CATTAGTGAC GAATTAAAAG CGGCCAATGT GAAAATTAAA GATGATGCCT TCAAGAACGC 5100

TTTAGCAGGC TACATGCAAA CTGAATCTTC AAGCGCTTCT TCAGAGAAAA AAGAATCAAA 5160

ATCAAGTGAT TCTAAAACAA GCGATACCAA AACAAGCGAC TCTGAAAAAG CAACAGATTC 5220

TTCAAGCAAA ACAACAGAAT CTTCTTCTAA ATAAGCATAA AAAAGAGAGT GGTCGATTGA 5280

CCACTCCCTT TTTTTATTTA TCTTTTGGAA CATAGAAACT GCGGTTATCT AAACCAAAGA 5340

TTCGTTCCGT ATACTCCCCa GGCTGTGTTT GACGAATTGC CCCTAGCATt CATTTGAATG 5400

GACGCATCCA TGGTATCAAT TTGATGTAAG ATTTCCGCTT CCATAATTCT TGGTCGCACA 5460

GGTGAGCCAT ATTCTAACAA TCCATGGTGG GCTAAGACCA TGTGCCGAAG AACGACAACA 5520

TCTTCTTCTC GATCATCAAT TTTCAAAGCC AAACACGCCT TGGTAATTTC CTCATCCACC 5580

```
AAAACTAAAT GACCAATTAA GTTCCCTGCC ACCGTATATT CTGTCGTAAT AGCACCAGAT     5640
AGTTCTAAAA CTTTTCCTAA GTCATGTAAA ATAATCCCCG CGTATAATAG TGACGTGTTA     5700
ATTTCAGTAT ATTCCTTAGC AATTGCTTTG CCTAGGCGCA ACATTGAGAC AGTGTGGTAA     5760
GCTAAGCCAT TCGCAAAAGC ATGATGGTTC CGTTTGGCTG CAGGAAATTC ATAAAATTCT     5820
TGTTGGTATT TCGTCATGAG ATAGCGCACA ATACGATTCC AATGAGCATT GGTAATTTCA     5880
AAAATCACTT GATTGATTTC CTCGACCATC TCTTCTTTTT TCAAAGGCGC TCGTTCCATA     5940
TATAAACTTG GCTGGCTTGG TTCCTCTGCT GTCGCTAAAC GCATGTGTAA AATTTTTACT     6000
TGGGGTTTAT TTTGATAAAC TTCTCGtTTC CCATTCAGTA AAACAACGTT TCCCGCAGTA     6060
AATCGTGTAA TTTCATCCTC AGAGGCGTCC CAATATTTtC CGTCAATCGT GCCAGATGTG     6120
TCTTGAAAGG TAAAGGCAAT AAACTTTTTC CCATTTTTAG CTACACGTAC ATCGGCATTT     6180
TTAATTAGAA CAAATGATTC AAAAAGTTCT TCTACTGTTA ATTCACGTAA TTTTTTCACT     6240
TGACGCTTCC TCCTTCTAAC GCAATTACTC TTTGTTGTAA GTCCTGATAA TAGCTTACCA     6300
TTTCTTGGTC AGAAGAAAAA CAAATGACTT GCTGTTTTTG ACCAAATTTA GTAAATAATT     6360
CAGCTAATTG GGCTTTTCTT TGATGATCAT AGTGTAACCA GCCATCATCA ATGATAATTG     6420
GACAAATCAG TTGCTCGCCT TGGACTGCTA AAAAGGCAAA ACGAACAGCC ATCATCAATT     6480
GATCTTTAGT GCCAGTTGAT AATTCATGTA AATAAAACAT TTCTTGTTCT TTCGTCACCG     6540
CAACTAGTTG ACCTTCGGCT ACTTGGATTG ATTGATAACG CTGATTTGTT AAAAGTGCAA     6600
AGTAGCTTGA TGCATACTGT AATAAGCTGG GCAATTGTTG TTCCGAAAGC TCTGTTAATA     6660
GGTCCATTAG CAGTTGTCCC GCTAGTTGAT ACCCAGACCA ACGTTGAGCT AGTTCTTTGA     6720
TTTCCGCTTT TAAAATTGCT TGTCGTTGAT ACAACTCATC TAAGGTACCA TCCGTCATTA     6780
ATTGTTGTTT TTCATAAATC AGCGCTTGAT ACTGTTCTTG TATTTTATCT AACTGGCTTT     6840
CTTCCAATGC TTGTTGctGC GCCAGTTGTA AAGAACGGTC TTTTAACGTT GCTTCTGTTA     6900
CCGCTTCTGG AAACAGATTA CCTAATTGCG TCGCTAATTC TTGCCCACGA CCAACTTGGG     6960
CTTGCTGAGT GGTCAGCTGT TGAATCCTAC TTGGAACATC AGCGATTGAT AGAATCCGAT     7020
AACGAATTAG TAAAGGCTGA ATTCGTGTTA ATGCTTCTTG CTGTTTAGTT TTCAACTCTC     7080
GTAACTGCTG TTTGAGATAG CCATCTGCTT GATACTCTTG AGCAAAGCGA ACTTTCTCCA     7140
TTTCTTCAGC AAACCCTTCT AAAAAGTGTA ACTTTTCGGA AAGGCTCTTT CCTTGCAAGG     7200
GTAAATGCGT CATCAATGGG AAGATTTGTT CTTCAAATAC TTGGCGTTTT TGTTGATTTT     7260
CCTTCAATTG ACTCATGAGT TCTTCATTTG TTTGAACAAG CAATAAATAG CGCGTAATCA     7320
ATTCCCGCTG ATTCATTAAA GTATCAATTT GTTGCATTTT TCCTAATTGA TGGACTTGGG     7380
CTTGCTCCGC TACTGCTTGT TCAAGTTGTT CCACTTGCAG TTGTGCCTCA CGTTCTTGAT     7440
TCAGCGCCTG TTGGAACTGT TCATTTAAAT AATCTAGTTG CGAAAGTTTT GTCTGCCACT     7500
CTTCTTTCAC TTCATCAGAG GAAGCTTCTT TTCTAGTGGC GTAgcTAAGC ACTTCTTTCG     7560
```

```
CGATAAAACT AACTCCCACG AGAACTAGGC CCCATTTTAA AGGGCTTGGC AAAAAGAAAC    7620
CGAGTAATAC AAGGATACTT CCTAGCAATA ACCAAAGTGG TTGACTTGTT TTCTGTGCTT    7680
TTTGGGCACG CGTCGCTTCA CCAAAAAATG CGGCATGCGC TTTTTCAAAA TCTGTTAACT    7740
GGTCTTCCAG ATTTTCAATT TGAGAAGCCA AGAATGCTTT CTGTTCTTGT CCTTTTTGAA    7800
GAGCTTGCGT ATACGCAACT TGTTGTTGGC GCCACTTTTG AACGGTTTCG TTGTCATAAA    7860
AAAGCTGTGG TGGTCGTTCA GGCGACCATG ACCATTCGTT CTCTAAAAGG ACCATTTCTT    7920
GACGATTTTG CTCAAAGGTT TGGGTCATCC ATTCCGATTC CGTCATCAAT TGTTGCATAT    7980
CATAGCGTTG ATTCAACAAT TGTTGAATCA ATGCTTCTTG TTCTAGGTAA AAAGCATATC    8040
CTTCTGGAAC ATCGAGCGCT TCTGACTGAA GAGATAACTC TTTGGTCAAA CGTTGTCGTT    8100
CTTTGTCTAG GTACATATAT TCTTGAAAgG TTTCTTCTAA TTGaACTTGG TCTTCTTTTA    8160
CTAAAACAAC GTTCTCTGTT TCTTcAGGTA GCGACTGCCA TTCTTCGTAC aAGGGAAAGT    8220
GaCGTTgcTg TTCCgCAACC TGTAACAAGT cCTGTTGCGT TTCTTTTAAT TGCTGTTGCA    8280
GTGtCACTCG CTGTTGTTCT GTTTCGCGAA TCGTTTCTTC TAATTGTTGA AAAGTTTGTT    8340
GcTGCGCTTC TTTCTCTTGA ATTTGCTCTT TTAGTTCTTG ATAAGCCGCC AACTTCTGAT    8400
TGAGTAATGG CTGCTGTCCT TTACCTTTAA AGATTTTTTG TGCTTCTTTA AAGTACGCAT    8460
TTCGGTAAGT CAAAAGTTGT TGACTACCAG AAACACCAAT GGCTAACAAG GATGTTTGCA    8520
GTTCCTCTTC TGTCAACTTG TCACTCGTAA TTAATTGTTC TTGTTGGAAA GTAAAAACCT    8580
CTTGAAATAA CTTTTTCGTT AGCGGATGTA ACATTTGCTG TAAGGTTTTT TCATCCCCCA    8640
TTTGTTGTTG ATAATAAATG ATTGCTTGCC CTTTATTTTT TTCTTTAAAA CGCTCGACTT    8700
GAACTTCCCC AAACACAGGA TGTGCCAACC ACAAACGACC ACCATACGAG CCACCATTTT    8760
TAGGTGCGTA ATCGCGTTTT CTTTTGCCCT TTGTGGGAAA CCCAAACAGC ATTGCTTGAA    8820
TAAATTGATA CAATGTGGAC TTGCCAACTT CATTTGCGCC GTACAGCAAT TGATTACCAG    8880
ATTCAAATGT AATTTTCTGT TGTTGCCATT TTCCAAAGCC CACAATCTCT ACTGCGAGTA    8940
ATTTCATAGT GGGTCCTCCT GAATAGTAAA GTCTTCATTA ATTTGCTCAT GTGCTAGGCG    9000
CAAGCTCTCT AATCGCCACT GCTCATCCAT TGAAAAAATT CCAGCAAATT CTGGATAACG    9060
TAACAATTCC TTGGTGGTAT CTTCAAAAAT AACTGGATCT AAGTATGTTT TTTCTAATGC    9120
CGTTAATAAG TTGGGACTAG CGGCTAAATA GATCGGTTCT TTCGTTGGTT CCTCTGCTTG    9180
CAGTATTAGG CGATAAACAA AAAAATCACC TTGCGTTTTT TGTAAAAGCT GCTGCTGTAA    9240
GTAGCTTAAT AGTTCACCAT TGACAATGGC TTGCACCACG GTCACATCCA AATGTTGCGT    9300
TTCTTTCAGT GTCAACGCCA TCAATTGCTG TTGCTGATGC GCCTGCCACT TCGTCAGCAA    9360
ACTGGTTTCA AGATAAGACA AAACGCTTTG GGTGTCACGG CATTGCGCAA GAGATACTTC    9420
TTCATTCGTC CAATCTATTT CTGCGACTTT TTCAAATTGA ACAGTCGCTT GATTTTGGGA    9480
AAGTGTGACC ACTGCAACCC CTTGTAAAGA CTGTTCTTTT TTCGTGTGTC CTTGAGGTGT    9540
```

```
TCCAGGATAA ACAATTAGCG GTTGTTCACT AACAATTTGC GGCTGATGAA TGTGGCCCAA     9600
TGCCCAATAA TTATAGCCAG TGGCTTTTAA ATCCTTCCAT GTAAACGGAG CATAATTTTG     9660
TTGCGCTTTC CGACTGGTGT CTCCATGATA AATCCCAATA TGAATATCAG ATTCTGCTTG     9720
TTTGATAGGA AATTCTAGTG CTTTATTTTC ATCAATCCAG GGATGCTCAT AACTAAACCC     9780
TGAAACAGCC ACTTGTTCGC CCGCTTTCGT TTCAAAATAC AGTGTCTCAA CCATTTCTTT     9840
TTCAAATAAA AAGACATTGT CTGGAAAAGA AAACCAATAG CGGTCTTTTT TATAGTAATC     9900
ATGATTCCCA AATGTTAGAA TCACAGGGAT TTCTGCTTGT TTCAACTGTT CTAATGCAGC     9960
CATCAATTGG GCTTGCATTG AAATAGAGGT CTGACTTTGA TGAAACGTAT CTCCTGCAAA    10020
AATAACCAAG TCCACCGCTT CTCGAATGGC AACTGTTACG ATATTCATTA ACAATTGTTG    10080
ATTAGCTTCT CTTAaGCGcT TCGCTAAGGC TGTCGGGATC CCAGTTAACC CTTCAAAAGA    10140
CCGATCCATG TGTAAATCAG CTGTATGCAA TAGTTTCATC GTTTCCCCTC CTTTTCTCCT    10200
TCCTTTATCA TATCATTTTC TGCCAGCTTG GTCAGTAAAA ACATACGAAA ATTTGTTCGT    10260
TTTTTTATTT TCAGTAAAAG AAAAACTCTG ACAAATTGTC TTATAGACAA CTTGTCAGAG    10320
TTCAATTACG TAAATTAATC GTTATACATA TCACGCACAG GAGTCATGAT AATTCGGTTT    10380
AAATCATTGA CAACTAAGCT AAACGCTTGT TCTTTTTGCA TTAATTCGTT GATAACTGCT    10440
GTTTCTTGCA CTTTCGTTGC AATGCTTTGT GCTTTTTCAG CATCTTCATC CGTGAATTCT    10500
TCGCCGCGCA TTTGTTTTTC TTGTAGTCCT TGTTGGAATC CTTGGAATTC TTTAAACAAT    10560
GTGTAAGCTG CTTCGTCTGC TTTTACTTCT TCATACGCTG CTTGTAATGC TTTAAATTCG    10620
TCTAACTCAC GTAGTTCACG TTCAATCTGA TTGGCTGTAT CATAAATGTT ACTCAATGGA    10680
ATCCCTCTTT TCTCAATAGT TTCATCTTTA TTGTACCATT TCTCTTAGGT TTTCCCAACT    10740
AATTTCCTAA TAAGCCTCCG AAAATTCCTT CAATCTTGTC TTTGACTTTT TTCGCGCCGT    10800
CTTTAATTTG TCCCCCAACT TCATTCAATG TATCTCCGAA ATTGCTAGCG CCGTCTTGAA    10860
CAGAATTTCC AAAGTCTTTC AAGTTATCCC GCCAATTAGA ATTGGCACCA CCAGTTAAAC    10920
TATCGTTGCT ATCAACAACT TTCCCACCTG TGGCATAAGC ATCCGCGACA GGAAATTGTG    10980
CTTGTTTGAC TTGTGGCAAG ATGCCACTGG CTTGACTGTT GAACACCTGT GAGGCATACG    11040
TAGCGCTACT GCCTTCCAAA TAATGGGTTT TACTGGTTTC TTGGAAGCCT AACCAAGTAG    11100
CAATGACAAC TTCTGGTGTA TAACCAATCA CCCACTGATC GTTCGTTTTA CTACTATCGA    11160
AGTTGGTTTC TGTCGTACCT GTTTTACCAG CCATTACGTA GCCTGCTGGT TGTGCGTTGA    11220
CCCCGCTCCC AGAACTAAAG ACCCCTAGTA ACATACTTGT AATCCCGTCT GCAACATCTT    11280
TGGTAATGAC TTGTTCTTTC TTGACTTTAG TATTATCAAC TACAATTGCG CCAGTTGAGT    11340
CGACGATTTT TGTAATTAAA TGCGTTTCAG TTTTATAGCC TTCATTGGCA AAGGCACTAT    11400
AAGCTCCCGC CATAGTTAGT GGCGAAACAC CCGTTTGTAA TCCTCCAAGG GCTAAACCAT    11460
AGTACCGATC TTTTTCTGAT AACGGAATGC CAAATTTCTC TGTTTTTTCA TAGCCTTTAT    11520
```

CTAAACCAAT TTGATGTAAT AACCATACCG CTGGTAAATT TAAGCTTTCT CCTAAGGCTT 11580

GATACATAGG AACTTCTCCA CTGTACGTGC GGCTATAGTT TTGCGCTGGA TAATAATCTT 11640

GTGGTTTATC TTCTAAAACA GAGTCTGGTT TATAGCCAGC CTCTAAAGCT GGTGTATAGA 11700

CAGAAATTGG TTTTAAGGAC GAGCCTGGTG AACGTTTTGT TTGCGTTGCA AAGTTAAAGC 11760

CTCGATAAAC GTGTTCCCCT CGACCGCCCA CCAGTGCTTG CACGCCGCCG GTTTTCGGAT 11820

CAATGGCTAC AGAGCCACTT TGAACCATGG CACCGTCTTC TGCATTGGGT GGAAATAGTG 11880

CATCATTTTT ATAAGTTGCA TCCATCGCAT CTTGATACCC CTGATTTAAA GAAGTGTAAA 11940

TCTTGTATCC TTTATTTAAA ATATCTTCTT CTTTAAATTT ATAACGATTG ACCGCTTCAT 12000

CAATTACCGC ATCAAAATAA TATGGATATT TGTAACTATT TTCGTCGCCA ACATAGGTGT 12060

CATTTAATAG ACTCGCTAAG TTAACACTCG CTTCTTGATT GGCTTCTTCT TGAGATAGTT 12120

TTTTATTGTC GACCATTAGT TGTAAAACGG TATTCCGTCG CGCTGTGGCA TTGTCAATGT 12180

AATCAATTGG GTTATAAATT CCTGGCCCTT TTAACATGCC CGCTAGCGTT GCGGCTTCCC 12240

CCACTGTCAC TTGGGACGCA TCCACGCCAA AGTATTTACG AGCTGCATCT TGCACGCCCC 12300

AGACACCATT ACCAAAATAA GAATTATTTA AATACATTGC CAAGATATCT TTTTTCGAAT 12360

ATTTCTTTTC GATTTCAATC GCTAAGAATA ATTCTCTGGC TTTCCTATCT AACGTTTGTT 12420

CTTGCGTTAA GTAAGCATTT TTGGCTAATT GCTGCGTAAT TGTACTTCCG CCACCACCGC 12480

CACCGATTTT CCCGAAGGTC AATTTTCCAA CAACCGCACG GGCAATCCCt TTGATGGAAT 12540

ACCCTTTATG TTGATAAAAC CCGCGATCTT CAGTTGAAAT AACCGCGTCT TGAATATAAG 12600

GAGAAATATT ATCCAGTTCC ACAAAAGTCC CTTTTTGCCC GAATAACTTT CCGACTTCTT 12660

TTCCTGATTC ATCATAAACG CGTGTTGATT CACTTAGCCC TGATTTTAGG GTCTCAACGT 12720

TTGCGGACTT GGCTAAGTAA AATAAATAAA TACTGGCACC TAAAGCGACT ACCAAACCAA 12780

TTAACAAGAA AATTTTATTG ATTTGATACT TTTTCCAAAT CCGCTTTCTG GCTTGATGGA 12840

ATTGTATTAA ATAAGGCTTT 12860

(2) INFORMATION FOR SEQ ID NO: 145:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9086 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

AACGCTAAAA CCTTTAATAA ATAAAGGTTT TAGCnTTTTA ATAAAGAGCC ACCTGCCGGG 60

CTTGAACCGG CGACCCCCAC CTTACCATGG TGGTGCTCTA CCAACTGAGC TAAGGCGGCA 120

CTTTCAAAGG CTAAACTAAC TCGAGGATTG CAACGAATGA AACCCATAGA GTTCATTTTC 180

AAAGGCTACA CTGCGACGAG TAATCGTCAA AGTTCCTTTT GAAAAATCGT ACAAATCTCG 240

TGCACAAGAA ATAGAATAAC GTATAGACGA AAGTTTGTCA ATTGCTCTTT CGTAAAAAAA 300

```
GAAAAATGAT TTTTTTCTAA GAAAAACTGT TGACAATCTC TTTTTTCTCT AGTAAAGTGT     360
AAATCAACAT CACAGATAAC TATTCATTCG TTGAAGAGAA GAGTAAATTT TGATGACCTT     420
CAAAAGAGAG TCCTGATTGG TGAmAAGGGA CAAGGTTCAG AAAAGTTGAm GATGGCCTCT     480
GAGAATAAAT TGCTGATAAG TAAGCAATTT CGGCTAAGCA CCCGTTACCA TGTGCACCAA     540
GTTAATAAAG ACTTGGATAA GAGATGAGAT ATAAAGAATC AGTTATATTT CAAAAAAAGG     600
TGGTACCGCG AATCATTCGC CCTTTACTAG CAAAACGCTA GTGAAGGGCT TTTTGCTGTG     660
AATCACTACG ACTGGCGCGA AGACAGAGTA GATGATGAAC AGTACAAGAT GACCGAAGGC     720
AAGTTGTGTC ATTTGCACGA ATCACTACGA CTGGCGCGAA gACAGAGTAG ATGATGAACA     780
GTACAAAATG ACCGAAGGCA GGTTGTATCA TTTGCACGAA TCACTACGAC TGGCGCGAAG     840
ACAGAGCAGA TGATGTCTGC GATTACTCGA CGCAGGATGG CCTTTGTGAG TGATGAACAG     900
TACAAGGTGA CCGAAGGCAA GTTGTGTCAT TTGCACGAAA GTGTTCCTAC GATTCTCGTC     960
ACAGTCAAAT CCTGTTGAGC ACTACGACTG ATGCAAAGAG AGAGAGTAGC TATTATCTGC    1020
GATTACTCGA TGCTTGAAGG GGTAGTAACT CTAGTCTGTT TGATTTTGAA AATAAACGAA    1080
AATCAGTTTA AAATGAATAG AAATAAAAAA GCAGTGATAT AGTAAGTACA AGACTTGATT    1140
TGTCTAACAG AGAGCCTAGG TAGCTGAGAA TAGGTGACAA AGACGTTTTG GAAAATGGCT    1200
ATTGAGCAAG CGCTCCGAGT GTTAAACACA AGGCGGCGAC GGGGGCTCCC GTTaTCGAGC    1260
TAGGGTATCG AACATATTGT TCCGTACTTG AAAAGGGAAT TTGTGTGAGC AAATTCTGAA    1320
CATGAGGTGG TAACACGCAA AAAAGsGTsC TCAACGATAC AACTCGAGTT GTGTCGTTGA    1380
GGATGCTTTT TTCTTTTTTA TGAAAAAATG CAGGTTGGTA AGAATGAAGG GGGATAAAGA    1440
ATGATTGAAC TAAAAAATAT CAGTGTGACA TTTCAACAAA AAAAACAAGA AATTCAGGCT    1500
GTCCAAGATG TTTCTTTGAC AATTGATAAA GGGGATATCT ACGGAATCGT TGGTTATTCC    1560
GGTGCTGGAA AAAGTACGCT AGTACGGGTA ATTAATTTAT TGCAACGACC AACAGCGGGC    1620
ACCGTTATTA TTAACAAAGA AAATATTTTA ACTTTTTCTA AAAAGGAATT ACGACAACAA    1680
CGAAAAAAGA TTGGAATGAT TTTTCAACAT TTTAATTTGA TGAAAGAGCG GACTATATTT    1740
AGTAATATCG ATTTTTCATT AAAATATTCT GGGCTAAGTA AAAGTGAACG GCGTCAAAAA    1800
ATCAGCCATT TATTAGAGCT AGTTGGACTA AGTGAAAAAC GGGACGCTTA CCCCAGTCAA    1860
CTATCAGGAG GGCAAAAACA GCGGGTTGCG ATTGCCCGCG CTTTAGCAAA TGATCCTGAA    1920
ATTTTACTTT GTGATGAAGC GACGAGTGCG TTAGATCCAA AAACAACGGG ACAGATTTTA    1980
GCCTTGTTAA AAAAGTTAAA TCAAGAACTG AACTTGACGA TTGTCTTAAT CACCCATGAA    2040
ATGCAAGTGG TTAAAGAAAT TTGTAATAAA GTCGCGGTGA TGGAAAATGG CTGCGTTGTC    2100
GAATCAAACG ATATTGTTTC GATTTTTAGT CAACCCCAAC AACCTTTAAC CAAAGATTTT    2160
ATCCGGACAG CAACTCATAT TGACCAAGCA TTAACAACAA TTCTTGAACA TCCTAAATTA    2220
GCTGATTTAG ATAAAAATCA AGAATTAATT GAATTTTCTT ATGTGGGAGA TCAAACGAAT    2280
```

GAGCCATTAA TTGCACAATT GTATAGCCAA TATCAGGTGT ATACGAATAT CCTTTACGGC 2340

AATGTCGAAA TTGTTCAAAA TGTTCCTATT GGCCATTTAA TTGTGGTGCT TTCTGGTGAT 2400

GAGGCCCAGC GTCAACAAGC CCTTACTTAT TTAGCGAAAC AAGGCGTGCG AACCAATGTT 2460

TTGAAAACCT ATCAGCAAAC GAAGCAAAAG CAGAATTTAC AGGTGATTTA AAAAAGGAGT 2520

GAACTGAGAA TGTATCAATT ATTTGAGAAA TATTTTCCGA ATGTTGTCCA ATTAAAACAA 2580

GAGTTTCTTC AAAGTACATG GGAAACATTA TACATGGTTT TTTGGACAGC ATTGATTGCT 2640

GGCGTGTTAG GAGTCTTGTT GGGTGTCGTG CTTGTTAGTA CTGGCCCCAG TGGTGTTTTG 2700

AAAAATCCAC CCCTGTACAG TGTCTTAGAA AAAATTATTA ATGTTTGCCG CTCTATTCCT 2760

TTCATTATTA TGCTCGCACT GATTCAACCA TTAACACGAA TTTTGGCAGG AACGACAATT 2820

GGTACAACCG CAGCGTTAGT CCCATTAGTT ATTGGCGTAA TCCCGTTCTT CGCGCGCCAA 2880

ATTGAAAATG CGTTATTAGA AGTGGATCCT GGCGTTATTG AAGCGGCAGA AGCCATGGGG 2940

ACGAGTCCCT TAGGGATTAT TTTTAGGGTT TATCTAATTG AAGGGTTACC AAGTATTATT 3000

CGTGTTTCAG CGGTGACAAT TATTAATTTG ATTGGATTAA CAGCCATGGC AGGAGCGATT 3060

GGAGCCGGTG GTCTGGGCAA CTTAGCGATT ACTCGAGGAT ACAATCGGTT TCAAACCGAT 3120

GTGACATTTA TGGCCACGTT AATTATTTTA ATTATGGTAT TTATCAGTCA AGCCATTAGT 3180

AATCAATTAA TCAAAAAAAC ATCACATTAG AAAAAAGGAG AAATGACAAA TGAAAAAATT 3240

TAGTAAATTA ATTGGACTTA TTGGGGTATT AGCTTTTACG ATTGCAGGTT GTGCATCGGG 3300

GTCTGTGAAG GATACTAAGA CAGAAACCGT TAAACTAGGG GTTGTAGGAA CAAAAAATGA 3360

TGAATGGGAA TCGGTCAAAG ACCGTTTGAA AAAGAAAAAT ATTGATTTAC AATTGGTAGA 3420

ATTTACAGAC TATACGCAAC CAAACGCAGC ATTAGCAGAA AAAGAAATTG ATTTAAATGC 3480

CTTTCAGCAT CAAATCTTTT TAGACAATTA CAATAAAGAG CATGGAACGA AATTAGTATC 3540

aATTGGCaAT ACAGTCAATG CaCCATTGGG AATTTACGCT AATAAATTGA AAGATATCAC 3600

GAAAATTAAA GACGGCGGAG AAATTGCTAT TCCTAATGAC CCAACGAATG GCGGGCGGGC 3660

GTTAATTTTA TTACAAACTG CAGGACTGAT AAAAGTAGAT CCTGCGAAAC AGCAACTACC 3720

GACTGTCAGT GATATTACTG AAAATAAACG CCAATTGAAA ATAACTGAAT TAGATGcTAC 3780

GCAAACAGCG CGCGCTTTAC AAGATGTCGA TGCTTCAGTG ATTAATAGCG GCATGGCTGT 3840

CGATGCTGGG TATACACCAG ATAAAGATGC TATTTTCTTA GAACCTGTAA ACGAAAAAGC 3900

GAAACCTTAT GTGAACATTG TCGTGGCCCG AGAAGAAGAT CAAGAGAATA AACTTTATCA 3960

AAAAGTTGTA GAAGAATATC AACAAGAAGA AACGAAAAAG GTCATTGCAG AAACATCAAA 4020

AGGCGCCAAT GTTCCAGCCT GGGAAACATT TGGTAAAAAA TAAAGGAGGC ATTTATAATG 4080

AGTACAACAA CGATTCAAAC AATCCAAGAA GCTATTGCTA CAGAAAAGGA ATGGATAATC 4140

CATTTAAGAC GTCATTTTCA TCAATATCCT GAAGCAAGTT TAAAAGAATA TGAAACGATT 4200

AAGCGAATTA AAGAAGAACT ACTAGCCTTA GCTATTCCTT TTGTAGAAGT AGGGGAAACG 4260

GGTGTTTTAG CAACCATTGA AGGAGGTCTT GGCGCTGGCA AAACGATTTT GTTACGTGCA 4320

GATATTGATG CGTTGGAATT GCCAGATGCA ACAGGTGCTG CCTATGCTTC TAAAAATCCA 4380

GGACTCAATC ATGCTTGTGG ACACGATGGT CATGCGGCAG CATTGCTAGG TGCAGCTAAA 4440

GTGCTCAAAA AACATCAGGA TACCTTTTCA GGAACGATTA AACTGGCCTT CCAGCCAGCT 4500

GAAGAAATTG GTGCAGGCGC TCGACAATTT GTGGAAGGAA ATTATTTAGA AGCAATCGAC 4560

CAAGTGTTTG GGaTTCATTT AGATTCCAGT GTGCCGGTCG GAAAATTAGT CGCTACCAAA 4620

GGCGCCACCA ACGCCTCTTG TGATATTTTT AAAAtTGAAG TCAGTGGTCA AAGTAGTCAC 4680

GTCGCCCAAC CACAGAATGG CCGGGATGCy stTTTAGCAG CGGCCaGTAT CGTTGTGGAA 4740

TTACAAAAAA TTGTAGCTCG CGAGATTGAT CCTTTAGATT CTGTCGTAGT AGGAATTGGC 4800

GTTTTACAAG CAGGAACACG CTATAATATT GTAGCAAACC AGGCAACCAT TGAAGGTACT 4860

GTTCGAACAT TTAGTCAGGA AACGCGCCAA TTTGTTTTAC AACGAGTCGA AGAAATTGCC 4920

CATGAAATTG CCCAGTCTAA TCGCACAGAA ATTGCTGATT TTTCTGTTTA TGCGGCAGCG 4980

AATCCATTAA TTAATGAAGA ACAAGCAACC AACCGTGCAC AGCAGGTAGC CAGTGAAATT 5040

GTTGGTTTTG AAAATGTTGT GACCGATCAT CCTAAAAGTT TAGGGGCGGA TGATTTTGCT 5100

GATTACTTAG CGGTAATTCC TGGTATCTAT GGACGGGTCG GTTCACGAAA CCCTGAAAAT 5160

CCTGCTACTC ATTTTGGACA CCATCATGAA CAATTTGATA TTGATGAACG TGCCTTACTT 5220

TTAGCAGCAG AGTACCACGT ACGCTATGCA TTAAATTATT TATCAGAATA ACAGGGATAG 5280

GGAGCCGACA AAATGATGGT TTGTGGGCTC CTTTCGTTAT AAACCAACGT GGCTTGCTTT 5340

CTTGTCATAC TATGGTAGAA TACAAGCATA AGAAAGGTGA TTCACATGTT TTTTGTTTTA 5400

ACTGATAAAT TAACAAGAAA AGTAGCTCCC GATAGTTTAA GAACAGATGA AGCACTGTTG 5460

TGGAGCCGAG CAATTAATTT CTTCATCGCT TGATAGGACA TTTTTTGACT AGAAAAGCTA 5520

GTCTTATTTG AGTGTACGTA CAAGCGGTTG TCGGGATTTC TGTCAGGATG GACGACGTGC 5580

AGACGTGGTT CTTTTTTTGT TGTCCCAAAA TAATCAGTTA ACAAGAAAGA AGTGACGAAA 5640

ATGAGTATTT TAACTATTGA ACATTTAACG CATCGATTTG GCGAAAAGGT CTTGTATGAA 5700

GAGGCTTCAT TGCAAGTGAA TAAAGGGGAT CATTTAGGCT TAACTGGCCA AAATGGGGTC 5760

GGCAAATCCA CCTTAATTAA AATTTTAACG GGGGAAGTAT TGCCaGACGA AGGAACGATT 5820

CAGTGGCAAA AAAATTGCAA GATTGGGTAC TTGGATCAGC ACGTTTCTGT AGAACAATCA 5880

CTAACGATGG TTGATTTTTT GAAACAAGCC TTTCAAGAAC TTTTTGATAA AGAAGCGAAA 5940

CTGACAAAGC TTTATGAAGA ATACAGTCAA ACGGCTTCGG AAAAACTTTT AGAACAAGCA 6000

GGCAAGTTAC AAACAGATTT AGATGAAAGT AATTTTTACC AAATCGACAC GATTATTCAG 6060

GATTTAGCCA ATGGGTTAGG ACTACAAGCA ATTGGTTTGG ATAAAAAGTT AGGGGAGCTA 6120

AGCGGTGGTC AACGTTCAAA AGTGATTTTA GCAAAATTAT TGTTAGAGGC CCCTGATGTG 6180

TTACTTTTAG ATGAACCCAC CAACTATTTA GATGATACAC ATATTCAATG GCTTGTTCGC 6240

```
TATCTGAATA ATTTTGAAGG AAGCTTTCTA TTAGTCTCCC ATGATTATCA ATTTTTAAAT    6300

GAAGTGACGA ATTGCATCGC AGACATTGAG TTTGGCAAGT TAACCAAATA TACTGGAAAT    6360

GTAGAAAAAT CTTTTGCACA AAAAGAGCAG AATAAACAAA CCTATTTGAA ACAGTATCAG    6420

GCCCAACAAG AAAAAATTGA AAAAATGGAA GCCTATATTC GTAAATACAA AGCTGGAAAT    6480

CGAGCAACGA TGGCTAAAAG TCGACAAAAA CAATTGGACC GGTTGGAACG ATTGACTCCG    6540

CCTGGTTCCT TGACTAAGCC AGCGATTGAA TTTCCTTATC AAGGGTTAGT TGCAACGCAA    6600

GCACTAACGA CCCAGAAGTT AGTTGTCGGC TATCGGGAAC CTTTGTTAGA ACCGTTAGAT    6660

TTAATGGTTC ATGTCGGTGA AAAAGTCGCA TTGAAAGGCT TTAATGGGAT TGGCAAATCA    6720

ACATTAATTA AAACGTTGAC GAAAGTGATT CCTTCATTAG ATGGAGAATT TCATTATCCG    6780

CTGAATACAA AAATTGCTTA TTTTACCCAA GACTTAGCGT GGCCTAATGA GCAGTTAACG    6840

CCACTAGACT ATTTATCAGA TCGTTTTCCA GATACAACGA TTAAAGAGCG AAGAAGTCAT    6900

TTGGCTAGAG CAGGTTTACC AGATAAGTTA GCAATGCAGT CGCTAGCCCT GTTAAGTGGT    6960

GGTGAGCAAA CAAAaGTAAA ACTAGCGGAA CTAATGATGC AAACAAGTAA TTTGTTATTT    7020

TTAGATGAGC CAACCAATCA TATTGATGAA GCGGCCAAAA AAAGTTTACA AGAAGCAATT    7080

CACGTTTACC CAGGAACGGT TTTTCTGGTT TCCCATGAAG CAGATTTTTA TGAGGAAATT    7140

GTGGATCGAG TAATTGATAT TGAGGAATTA GTTAAATAAA GAAGAAAGAA CCAAGTTGAG    7200

GGGGAGTCAA CTTGGTTCTT TCTTCTTTAT TATTTATTTA CTCATTAAGC CATCATGGAT    7260

TTTTGTTAAA TTCCAGTTCA TCATGCTGTA GTACGTATCG CCTTCTGTTC CTTCTTTGGC    7320

AAGAGAGTCT GTGAAAAGTG TATCGTAAAT TGGTCGTTTC ACTTCTTTTG AGACCCGTTC    7380

CATACTACGT TTATCGACAC TGGTTTCAAC AAATAACACA GGTGCTTTTG ATTTCTTAAT    7440

GGTATCAATA ATCGTGGTCA TTTGTTCAGG TGTyCCTTGA CTTTCTGTGT TAATTTCCCA    7500

AATATAAGCG GCATTTAAAT CATAAGCTTT GGAGAAATAT TTAAAGGCAC CTTCACTTGT    7560

AACTAATAAT TTTTTATCAT CAGGAATATC AGCAAATTTA GCTTTGGCTT CCTCATGTAG    7620

TTTGCTAAGT TTTTCGGTAT AATTTTTCGC GTTTTCTGTA TAGAAATCTT TATTTTTTGG    7680

ATCTTTTTCT ACTAACACGT CACGAATGTT TTCTACATAT TTAATGCCaT TTTCAATGTC    7740

TAACCAAGCA TGTGGATCTT CTGTTTGTTC tTGACCGGCA CTTGTTAAAT ATTGTGGCGT    7800

AACATTTTTG CTTGTAGAAA AGTAATCTTT ATTCTCAACT TTTTTGGCCG TTTTCATTAA    7860

TTTGTTAAAC CAGCCATTTC CGCCTGTTTC TAAGTTCAAG CCGTTAAAGA ATAAAATGTC    7920

CGCTTCAGAA GCTTTCGCAA TGTCTTCTGG TAACGGTTCA TATTCGTGAG GGTCTGTCCC    7980

AATTGGCACA ATACTATGCA GCTCAATTTT GTCTTGCCCA ACATTTTTCA CTAAATCAGA    8040

TAGGATCGAG TTCGTTGTCA CAATTGCTAA TTTTTCTTTC TTTTCAGCGG CTTGATTCCC    8100

GCAAGCAGCT AACGTTAACC CTGCTAAAAG TGTTAAAAAA AATAAACTAA ATTTTTTCAT    8160

TCGTTGATTC CTCCATTTCT TTCTCACGGT TTACAAATAC TAGGCCTTTC TTTGGTGAGA    8220
```

AAAAGAAAGC CAATAAAAAG AATAAGGCGG CGGTTAAAAC AATCGTCGCA CCAGAAGCTA 8280

AGTTATAAGA ATAACTGAAG AATAGGCCAA TCACTGAACT TAAAATCCCA AaaGTaGAGG 8340

CTAaGCCAaT CATtGTCGGT AAGTGATTTG TTAaTwGATA gCTGTTGCTG CTGGCGTAAT 8400

CAGCtGGCAT GACTAAAATA GTTCCGACGG TTTGTAAmGA AGAAACAGCG ACTAGTGTTA 8460

GTAAAaACAT CAATGCATAG TGGAAAAATT GAaTATTTAA ACCGTAGGCC TGAGCCATTG 8520

TGGGATCAAA AGAAGTGAGT TGCAACTCTT TATAAAACAA TGCCACAAAA ATCAAAACAA 8580

TCACACCGAC GACACAAGTA ATTAAAATAT CGGTATCTGC GACCGCTAAT ACATTTCCAA 8640

ACAGAATATG ATATAAATCG GTGGAACTTT TTGCAAAAGA GATAAAAATA ATCCCTAAAG 8700

CAAAAAAAGA ACTAAAGACA ACACCAATGG CAGTATCATT TTTTAGTGGA CTTTTTTGGG 8760

TAATAAAGCC GATCGATAAT GCAGCTAATA AGCCAAAAAT AGAAGCGCCA AAAATATAAT 8820

TGAAACCAAA CATmTAAGAA ACAGCGACCC CTGGTAAAAC AGCATGTGAA ATCGCATCAC 8880

CCATCAAAGA CATCCCACGT AAAATAATGA ATGAACCAAT TACCCCTGAA ATGAGTCCGA 8940

CAATTATAGA CGTCAAAAGG GCATTTTGTA AAAATTGATA ACGGAATAAA CCATCAATAA 9000

ATGCAGCAAT CATTTTGCAA CCCCCTGTAT TAATAAATCA CCTAACAAAT CACCGTATGC 9060

TTCTTGGAGC GTTTCTGCAG TAAATG 9086

(2) INFORMATION FOR SEQ ID NO: 146:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8546 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

TGCGCTTGCA ATAGTGCAAT GTACTAATCG TTTATTTTCC ATGTTCATCT GATAATCAGC 60

CCCCTAACAT TTTGGACACG CTTTCATTAT TACATTAAAG GGGCGTTCCA AAATAAATCT 120

TTTATTTTTG GAACGtCCAT CTCTAAATCA TTCTCTTAAA AACTGTAATA GCTTACGACG 180

ACAAGTCAGA ATTATTCATC CTCATCATCT ACTGCAACAT TTAGCAGCAA TTCGTTGACT 240

TTCACAATAC TTTCTGCTGA TTCCTTAACA TAGTCACCTT CGATCGCTAA TAAACTTGCA 300

TTTACATAGC TGATAACCAT CGGTAAATTC ATCCCAGTAT ATAAGTCAAA CGAAGCACCT 360

TGCATAATTT TTTTGGATAA AATATTACAA GGTGTCCCCC CTAGTAGATC AGCAAAGACA 420

ACAAAATCTT CTAGCGATTC CGTAATCGCT TCAAATTTTT CTAAAAAGTC AGCTTCACCT 480

tCTTCTGGTA ATAAAGCAAC AGTGGnAATT GTTTCTTGCG GTCCCATAAT CATCTCCGCA 540

CTTTTTTTCA ATTCCTCGCA AAAACGGCCA TGACTTACTA AAACTAAATG TTCACTCATC 600

GATTTTCyCC TGTCTTACTT ATTTTGAAwT AACACCTAAC GCAGACAaTG CGACACAGAC 660

AaTTAATACG ATAAAAATGG CTTTAGTAGA CGTCATATTC TTTTTACCTA GCATCCAATA 720

AACGCCACCG ACAATTCCTG CAGGAAGTAG ACGTGGTAAA ATCATATCCA AGTTATTTTG 780

CATGTTAAGT GTTACATCCC CAATACTTGG TGCCCATGAT AGTTTGACAT TAACCATCGT 840

TGCCACTAAA GCGCCAACCA TGAAGACCCC TAGTAAAGTT GCAGCGTCTG TTAAGGCAGT 900

TAAGCGGTGT TGCATAGTTG TTACCAAAGA AATTCCTTCG CGATACGCAA ATTCCAATTG 960

TTTCCAACGG AAGACCATCA CGGCAATTTG AGCAACAATC CAGATAAAGA TACCAGTTGG 1020

GTTACCATTA ATCGCCATGT TGGCaGCTAG CGCCCCAAAA ATAGTTGGAA TTAAAGCTGC 1080

GAAAATTGAA TCCCCAATTG cTGcAAACGG TCCCATTAAA CCAGTTTTCa AACCTGaAAC 1140

TGTTTGtTTT CCAGCAATGc CTTCTTTTTC TTCaATGGcC AAATCGATCC CTGTAAtAAT 1200

TGTATTAAAG AAGTTTGAAG TATTGAAAAA TTGCGTGTGT GTTTTCATTA CTTCTTTTAA 1260

TTCAGGCGTA TCATCCCCAT AAATTTTACG CAATTGCGGC AAAATCGTAT ACAAGTAACC 1320

TGATCCTTGC ATTCTTTCGT AGTTCCAACC TAATTGGAAA GTAAACAAGC TTCTGCGATT 1380

AATTTGTTTA AAATCTTCTT TCGTCAATTT ATAATTAGAT TTCGTCATCT TCAATCTCTC 1440

CTTCGTCTGA ATCATTCGCT GCATGTTCTT TTTTGGCTGC TGGCCCACTC GGAATCATTT 1500

GACCATTTTT GTAGCTAATT GCGGCTAAAG CGAAACCAAT TAAAGCGACT GCTAACATTG 1560

GTAGTGCGTT AAATACACCA CTGAAGTCTT TCACAACGCT CGCAACAGAT GTTCCTAAAA 1620

GTTGCATGTT TGTAAAGATT GTTCCTAGTA AAGCAGTTAC AGTAAAACCA AGAATTAAAT 1680

AAGGGAAATG TTTTTTCACT GGTAAGTAAC GTAGTAAAAT TGCAAAACCG ACCGCTGGTA 1740

AGACCGCCCC AGCAACAGAC AAGCCATCAC CTAACCATTT CAAATCGCCA TTTAGGACAG 1800

ATACAACTTT TTGTACTAAA CCACCACCAA AAGCAAGTGC TAAGAAGACT GGAATCATTC 1860

GAGAAAGAGA CCAAGGTAAG GCACCCATTA AGAAGTTACG TTCAATTCCT TTGTAGTTCA 1920

TATCTTCGAC CATTTTATCA ATCCGATGTG CAAAGTAGGT ATTCGCAAAA CGAGCCAAAA 1980

TGTCTAATTG AATCATTAAA CTAGCAACTG GTACTGCAAT AGCAGCGATT GCTTGTTCAG 2040

GGTTCATTCC TAAAGACACA GAAAATGCTG TGGCAAGAAT CGTGCCTGAG TTAGCATCAA 2100

TTTTGGAGGC ACCCCCGAAA GTCCCAACTC CTAATACAGT TAATTGCATG CTTCCACCAA 2160

TAATAAGTCC TGCTTTAATA TCCCCCATCA CTAATCCTGC AAAAAGACCT GCAAACACTG 2220

GCGCACTTAA TGAAGAATAA ATTTGTAATT CATCTAAAAT TTGGTATCCT GCGTATAATG 2280

TTAATAATAA GATCTGCCAC CATAAGATAT TCATTTTTTG TTCCCTCCTA TTTTGTTATC 2340

AAACTCATAA AGTCTTCCGC TTTGTCACTT GGCACCATTT GAGCAATCAG ATGGACACCT 2400

TTAGCATTTA ACTCTTTAAA AACCTCAATA TCTTGATCGA CCACATTGAT TGACTTAGTA 2460

ATCGACCGTG TTTCGTTTGT TTGAGACATG TTCCCAACAT TGATTTCTTT AATCGGCACC 2520

CCAAGTTCTA CTAGTTTTAA TAAACGATCT GGTTTCCGTA CCACAATTAA TAAACGTTGT 2580

GAGTCGTATT TTCCTGCTAA AATATTTTGA GCAGCTTTTT CAACTGGTAA CACGCTTAGT 2640

TTCACGCCAG CTGGTGTTGC TAATTTCAAG CCACTTTTTT CAATGGCATT CCCTGCAACT 2700

```
TCATCATCAA CGACCATAAT TCTGGAAATA TTCAATTTTG TCGCCCATAA ATTTGCCACT   2760
TGTCCATGAA TCAAACGTCC ATCAATTCTT ACGCCAATAA TACTCATTGT AATCCCTCTT   2820
TCATTTCTTT ATATAGTGGT TCTTTTACTA ACTGAATTTC TGGTTGGTAC GTTCCTTCCG   2880
TTTCAAAAAT GACGATCTCA TTTTGCCCTT CTTTTAATAA TCCTTTGGGA ATATATAAGG   2940
AAAGCGTCGG ACCAACATTC CAGAAACGAC CAAGATTGGT TTGGTTAACA AAAACAATTC   3000
CTTTGCCAAA TTTAGAAACA TCAATAAATG TATCTTTCAC TTCTGCTAAC TCCACATGAT   3060
ACTGGTAAAA GCTTGGTTGA TCAGGCTGCC ATTCTCTTGA ATAGTCTACC TGTTCACAGC   3120
TTGTCATAGG TAAACAATAT TGTTGCCATT GCGTCATAAA ATGTAAATCT GCCATGACAC   3180
CCGTTCGAAT CCCTTTTTTC TGAGTATCTG CAAATAGTTT ATGCCCGTAA TTGACACGAC   3240
CCATATTTTC CATTAAAATA TCGATTTGAT TATTCTCTTG CGGCAGTGTA ACGTAAATAT   3300
CTTCGCCAAT TTCCGTTTGA TATTGCGTTG CTTGATGTAC CTGATTGACA AATAATTGCA   3360
GACGATCACG ACCGTCAATC ACTCTTAGCT TTTCTTCTGC AGCATCTTTT TCAATGCTTG   3420
tTCGGTAAAG AAGATAGCCg GTATTTTGTC CTAATTGTTC CATGGTTTGC GGATAAACGC   3480
TGACCACTGG CTGACTAATC GTTTCTAAAG TCGCAAACAA ACTCACTTTG TTGGTTAAAG   3540
GAATTGCTGT TTGCGCAAAA GATTCTTTTA CTAATGGTTC AGCTTGTGGT AACGCAGGGT   3600
ACTCTTCGTG AAGCATTTTT TGTAACGCAA AATACTTTTC GGTTGGATTC CCTTGTTCAT   3660
CAAGTGGCGC ATCATAATCA TAGGAAGTGA TTTGTGGTAA ATCAATGGTT CCTCGTGCGG   3720
AGCAGCCATT CATAAACCCA AAATTGGTGC CGCCATGAAA CATATATAAG TTTATACTGC   3780
CTAATGCTAA TGCTTCTCTA ACTGATTCAG CCAATTCTTG CGGATCTCGT TTAATAATTG   3840
GCTCTTTCCA ACGATTAAAC CAGCCATCCC AAAATTCCAT ACACATTAGC GGCCATTTTT   3900
TACCATGTTC TTCAAAAAAG GCTTGCATCA TACCGAAATT TTCTTTGGCT TTTGAGCCAA   3960
AATTTCCGGT TACTAAAATA TCATCTTCAA TCATGCTGCC TGCCCGCAAT GTTGCGCGCC   4020
ATGGCCCGTC AGAAGTAAAA AATGGTGCTG TCACGCCGCG TGCAATCATC AAATCGCGAA   4080
TCGCCCGTAA ATACGCTTTT TCTTCGCCAA ATGAACCATA TTCGTTTTCG ATTTGAATCA   4140
TTAAAATATT GCCACCGTTA GCAAGTTGAT GCGGAACAAT TTTTTCCATT AAAACATCAT   4200
AATATTCCGC AACATGCTTC AGATAGGTTG GATTATTCGA ACGCATCCGT CCTGGTTCAT   4260
TTAATAGCCA GGCTGGAAAA CCGCCAAACT CCCATTCTGC ACAAATATAA GGAGACGGCC   4320
GCACAATCGC ATATAGTCCA AGTTCTTGTG CTAATTTTAA AAAGCGCTCC AAATCTAATA   4380
TTCCCTCAAA ATGAAATGTT CCTTTTTGCG GTTCATGTAA ATTCCATGGT ACATATGTCT   4440
CAACTGTATT GAAGCCTAAA GCTTTCAAGT TATAAAGGGA ATGATGCCAA TCTGAGGGAT   4500
CTACACGAAA ATAATGAATG GCGCCTGATA AAATTTTAAA TGATTGACCA TTTAAAAGAA   4560
ACTCCTCTTT AATTTCAAAT CTGTCCACAT TGTCACCCTT TCTAGTAAAC GCTTTCAATT   4620
GCTGGATAAT TGCTAAATAA AGAAAGCTCT TTATTTAGCA ATTTCGACTA CTTGGATGTT   4680
```

CATAAGATGT TTCTTCACGT TCTCTACGTC TACATGCCCA GTCATTCTTT CTTGGGCATT 4740

AGCCATTCCA GCCGCCATGC CCCATTTTAG CAGTTCGGCA GCTGGGGCAT CCTTGGCTAG 4800

TCCATATGCA AGACCAGCAA TGGTTGCATC TCCTGAACCA ACAGGATTTT TwGcTtGAaT 4860

CGTTGGAATT TTTACCACGA TAGAATTGAT CGTGATGTTT GGCAATTGCA CCATCTTTTC 4920

CTAAAGATAT GACAATCCAT TCAATTCCCG CAAACATTGG TTTTGTTAAA GCTGTTTGTA 4980

CAGCTGCTAA TGGATTTTCT GAAAAATCTT GCCCTAGCAA TCCTTCTAGT TCCTCTAAAT 5040

TAGGTTTAAT TAAATACGGT TTCCATGGTC CTTGAAGAAC TTGCCTTAAA CTATCACCAG 5100

AAGTATCTAG TAGCACTTTA ACCTCTTGTG CGTGAnCTTT TTGAACTAAT TCTTGATAAA 5160

AATCTGATGG TAATCCTTTC GCTAAACTGC CAGAAATTGT GACAATTTCT GCTTGCTTAA 5220

TTAATTGATC GAAATTTTCA AGGAAGTTAC TGATTTCTTC TGGGGAAACA GTTGGTCCAG 5280

CTTCTAAAAT TTCAGTTTGA TTCCCCTCGT GCAAAATCGC AATCGAATCA CGTGTTTCTT 5340

CTTTGATTGA CGTAAATGCT TGGGGGATGT TTGCTTTCTT TAACTCATTT GCAATAAACG 5400

CCCCATGAAA ACCACCAAGA ACGCCGGTAG CGATAACATC TCCACCTAAA TCATGAATAA 5460

CGCGAGTGAC ATTTAGTCCT TTGCCACCAG GGGTTTTGGT CACCTGACTG GTCCGATTAA 5520

CGGTATCAAG CTTTAGATGA TCTAATAGGT AGGAGATGTC AATTGACGGA TTCATTGTTA 5580

CTGTTACTAT CACAGCTATG CCTCCCAATT ACCCTTCGTT TACTTGAACT TTCGCATGCC 5640

AAGAGCTGGC GGTTGCTGCG ATGACTTCAT TTAAACTTTC AATGTTTTCT CGGCCTTCTG 5700

TTTTTAACCA ATCACAAGCC GCAGTTTCTC CTGCTTCAAC GAAAGGTTTC ACGCCATTTT 5760

TCCACGTAGC ACGGCCACAT AAAACACCAT TAAATGTTGA ACCAGCTTCT TTAGCGAAAC 5820

GTAGTGTTTC TTGGAATAGC TCTGTTGAGA CACCGGCACT TAAAAAGATA AATGGTAAAT 5880

CAGTGGCTTG ACTTTGTTCT AAGAAATAAT TCGCCGCTTC CTCTTTTGTA TACGCTGTTT 5940

CAGCCGGCGC AAAACCTTCT ACAAAATTCA TATTAACTGG AACTTCCACT TTCAATACAT 6000

CTACTTTATA TTGGGGTTTA GAAAATTCTT TCATCATTTC ATTCACTTTA TGCGGTTTAA 6060

CTTTTGCATA TTCTAAAGAA GTCGCATCTG CTATTTGTGC ATCATAAGAA ACTAATTCTA 6120

AGTAAAAATGG TAGGTCTTCT TCTGCACATT CACTTCCTAA GCGTTCAATG AATACATGTT 6180

TTTGATGATT GATTTCTGGA TCCTCATCCA CATCATAGTA TAGTAAAAAT TTAATCGCAT 6240

CTGCGCCTTC TTCTTTCAAG CGTAAAACAG ACCAATCTGC TAGTAAATCA GGCAATCGTC 6300

CCGGAGTTGT TGCATCATAG CCCGTTTTTT CATAAGCGAG TAACAATCCA GCCTCGGTAT 6360

CGCGTGCTTT TGCAGCAGGC AATCCATATT CTGGaTCCAA TAAAATGGcA GACGCATAAG 6420

GTGTCAATTC TTTCGAGACT AATTCTTTAA ACGTTTCAAT TTGTGCATCA GTTGGCTCTA 6480

CATCTAGCGC TTTGATCATT TTTTTCAAGG CACCTCGTTG ATCAATGGCT AATGCGCTAA 6540

TGATTCCTTC TTGGGTTGAC AAACGATCCA TTGCCGCTTT TTTTCCAGCT GTTAACGTAA 6600

GCATAGTTCC TCCTCCAAAT GTTGTATTAT TTATAGTCAT AAATCGTTAC GCCTTTAACC 6660

ACACGATTCA CTGTACCTGT CGGTGAAGGG GTATCCGGTT TATTTCCAAC TTTAATAGAT 6720

GTTAAAAGTG CGACTGTTTG GGCTACCATG ATATCTGCnA ATGCnAAATA GCCATCAGGT 6780

AATAGAGCAG TTTCTGCGGa ATATTCAATm GyTGAACCtG AGAAATTACG TTTGCCAGGT 6840

TGAGCAATCG CAAAAACACC CGCCGCAATT TCGTCACCGT GAACTTCTTC TAAGATATCC 6900

AAGTCATAAT CTCTTGTGTA TGGTGTATTA TTaACAAAAC CAAACATTAT TGTTTTTTCA 6960

TTAATAAATG ATTTTGGTCC ATGACGGAAG CCCATAGAAG AGTCAAAAAT TGTGGCAACT 7020

TTGCCGGCTG TTAGTTCTAA AATTTTCAAT TGCGCTTCTC TTGTCAAACC AGCTAAACTT 7080

CCAGAACCAA CATAAACAAT GCGTTCAAAA TCAAGGTTAA CAATTTCATT TAATTCTTCT 7140

TCACGGGATA GCACTTCTTC ACCTAATTTA ATTAGAGCGG TTACATAATC TTGTTTTTGC 7200

TCAAAGGCTG TTTGATCAAA GAGTAGTAAT GTACCCAAAG TCATGCAAGA AAAGCTACCA 7260

GTCATCGCAA AACCACCATC ATTGGAACGT GTTGGCATTA AGAATAAAAA GCTGTTTTCT 7320

TCTTTTTGAC TTATTTGTGC TAGTTGTCCT TCTTCGGCAC AAGTAATCGC TAAATGATGA 7380

ATTGTTTCGA CCACTTGATT AGCTACTTCA ACTGCCGCTA AACTTTCAGG ACTATTTCCA 7440

CTTCTTGCAA ACGAAACTAA CAATGTTGGT TCGTCTTTAA TCAAATAATC TTCTGGTTTG 7500

GCCACAATAT CCGTTGTCCC AATACTTTCG AATGAAAAGG CTTGTGTGTC GCCGTGTGCT 7560

CGGAGATAAG GTACAACCGT ATCACCGACA TATTGCGAAG TCCCTGCTCC TGTAAAAATT 7620

ACTCTTGTGC GTTTTCCTTG CGCTTTGGCC TGGACTTCTT TTAAAAAGTT TTCCAAAGCT 7680

GTTTGATTTT CATGATAAAG TGTCACGGTT TCTTGCCAAA GCTCAGGTTG CTGACGAATT 7740

TCACGTGTAG TGATTTCCGC TCCTAGTTGT TCTAGTTCTT CTTTTTCTGC TGTAAACATA 7800

CTTAACTCTC CTAACTATTC CGAATATGGC GTATCTGATA ATGAAATTGa TCTGCTCGCG 7860

CCACACTCAA CGTAAATTCA ATAATTTCAT TTTTCATGTT ATATGTCTGG CGTGCTAAAT 7920

GAAGGACTGG CGCTCCTTCC GGaATCaTCA ATAATTTGGc ATCTTCTTTT GACGCAATGC 7980

TAGCGTATAA TTCTTCATCA GCTAAACGAA CGGTTTGCTT AAAATCTTCC GAAAATAGAT 8040

CATATAATGG TTTGCTACGT AGTAATTGAT CCGTGAGTGA TAAGAAAAAT TTCACAGGCA 8100

AGTATGTGTC TTCAACCATT AATGGCTCGT TATCAGCAAT CCGCAACCGA CTTAATTTAA 8160

AAACTGCCTC TCCTAACGAA AGATTCAAAT GCTGACAAAT AAATTTATCC GCTTCCAATT 8220

TTTCGAAGGA TAATATACGG GTATGTGGTT TTCGACCCAG CCCTTTCATT TGTTCTGTAA 8280

AACTATATGC GCCAGCTAAA TCAGCGGCCT CTTTTTTTAT ATCTGAAACA AATGTTCCTT 8340

TGCCATGACG ACGATAGATG GAACCTCTAT TTTCTAGTTC TTGTAGCGCT AACCGGACCG 8400

TCGTCCGACT CACTCCATAT TGTGCAGTCA ATTCACGTTC AGAGGGCAAC TTGTCATGAG 8460

GAATCATGCA TGTTTCGATT CGTTCCTGCA GCAAATCAAC AAGTTGATGA TACAAAGCTT 8520

TGTTTTTAAA GGCATTTTCC ATAACT 8546

(2) INFORMATION FOR SEQ ID NO: 147:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3493 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

CGACACATTG TGGGACATTG CTGCACGTTT TGGAACGACA GTTGATAATT TAATGTCACT 60

AAATGGATTA ACATTGGATT CTGTTTTATC GGTGGGTCAA ACCATCCAAA TCGGTTAATT 120

TTGCAATTAT TTAGTkGAAG AAAGACCTwA CAAAAGCAGT AGTTTTTGTA AGkTCTTTTA 180

TCGTGAAAAT AGGTAGGGCT AGCAGCAATT ATTTGATGAG CCTAATCCGT TTTTTTATTT 240

AATTTTTCAC AAACTAAAAT TCGTTTTTCT GGTAGAAATC GTTTTTTTGA GCTATGCTTA 300

GAGAAAGAAT TATTATATTG TTTGTAAAAT TGGAGGATGT AGGATTATGA CTAAAGGTTA 360

TGTAAAATCA GTAACAGAGT TAATTGGTCA AAGCCCAGTT GTGAAATTAA AACGGATGGT 420

ACCAGAAGGA GCTGCAGATG TTTTTGTTAA ACTAGAGTTC TTTAATCCTG GTGGCAGTGT 480

GAAAGATCGA ATTGCTTTAA GCATGATCCA ACAAGCCGAA GCAGATGGTC GCCTAAAACC 540

TGGACAAACG ATTATTGAGC CGACTAGTGG TAATACAGGC ATTGGGTTAG CAATGGTAGG 600

TGCTGCGTTG GGATATCCAG TTAAGATTGT GATGCCAGAT ACTATGAGTA TTGAACGCCG 660

CAAATTAATG CAAGCATATG GGGCTGATCT ATTATTAACC CCTGGTGCCG AAGGAATGAA 720

AGGGGCAATT GCAAAAGCTA CAGCATTAGC AGAAGAACAC GGGTACTTCA TGCCTTTGCA 780

ATTTAATAAT CCAGCTAATC CAATGGTACA TGAACAAAAA ACAGGAAAAG AAATTGTTGA 840

TGTCTTTGGT AAACGTGGCT TAGATGCGTT TGTTTCTGGT GTTGGCACTG GAGGAACCGT 900

TACAGGAGTT GGCCATGAAT TGAAACGGAT TTTTCCAGAT ATTGAAATTG TTGCAGTAGA 960

ACCAACAGAG TCGCCTGTTT TAGAAGGTGG CGAACCAGGT CCACATAAAA TCCAAGGAAT 1020

AGGCGCTGGT TTTGTCCCAG AAGTTTTAGA CACCACCGTT TATCAAAAAG TTGCCGCTGT 1080

TTCTAGTGAA GACGCATTAG AAACTGCTCG TTTAATGGGG CCAAAAGAAG GTATCCTTGT 1140

GGGGATTTCA GCAGGGGCAG CAATTAAAGC TGCCATTGAT TTGGCAGTCG AATTAGGCGC 1200

AGGCAAACGT GTCTTAGCGC TGGTTCCGGA TAACGGTGAA CGTTATCTTT CGACAGCTCT 1260

TTACGAATTT CCAGAATAAA AAACACGCCA AAAAACGAGC ATTTCTCTTA GTTTAGTAAA 1320

ATGCTCGTTT TTTCCACGAG AATCCTACAA TAATTCTTGT CAATCGTAAT AAGAAATAGT 1380

ATTATTTAAG AAAGATAGAC AGGGGAGGGA TACAAAGATG GACAACGTAT TGGTTAAAAA 1440

TGCACTTGCT GAATTAAAAG AAGCCAATAT CCGAATTACT CCGCAACGTT ATGCTATCTT 1500

GGAATATTTA ATCGAAAATC ATACACACCC AACAGCTGAT GAAATTTATC GCGCACTAGA 1560

AGATCATTTT CCAAATATGA GTGTAGCAAC GGTTTACAAC AATCTACGTT TATTTACTGa 1620

AATCGGTTTC GTTCAAGaAA TGAGTTATGG CGATGcATCT AGTCGTTTTG ATTTTAGTTC 1680

GAAGAAACAT TATCACGTGA TTTGTCAAAA ATGTGGTAAA ATCGTTGATT TTCATTATCC 1740

AGGGTTAGAG GACGTTGAAA TGGCCGCTAG TAAATTAACA GGCTTCGAAA TTAATGAACA 1800

TCGTTTAGAA TTATATGGAT TATGTCCAGA TTGCCAACAA GCACAACAGG AGAATGTGTA 1860

AAAAATTACA TACTGTTAAA AATTTATCAA AAGATAAGTC ACTTGTCTAG GCAAATTTCA 1920

CAAGCGTAAG AGTTTTCACT CTTATGCTTG TTTTTTTGTG TGGTTATTGA CTTTTAAATT 1980

CTTAAAAACA AAAAATTAAC TAGCTATTAT TAAGAATTAG CAACAGTTTG TGAATTTTTT 2040

TATATGCGAT TCACAAGCAT AGAACTCGAA TAAGCTTGAT TCCTCTGATA TATTTTAAAC 2100

AGACAGAGGG AGACGTCACC TCTGCAAAAC AAAGATTGTG AAAGAAGGGA TTTACAACAT 2160

GAAAGTCGTA GTCGTAGGAT GTACCCATGC TGGAACTTCA GCAGTTAAAA GTATTTTAGC 2220

AAATCACCCG GAGGCGGAAG TCACTGTATA TGAAAGAAAT GACAATATTT CCTTCCTATC 2280

TTGCGGAATT GCCTTGTACG TTGGAGGCGT AGTTAAAAAT GCTGCTGATT TATTTTATTC 2340

AAATCCAGAA GAACTGGCTT CTTTAGGAGC AACAGTGAAG ATGGAACACA ATGTTGAAGA 2400

AATCAACGTG GACGACAAAA CTGTTACAGC GAAAAACTTA CAAACGGGCG CAACTGAGAC 2460

CGTTTCTTAT GATAAATTAG TAATGACAAC TGGTTCTTGG CCAATTATTC CACCAATTCC 2520

AGGAATTGAT GCTGAAAATA TTTTATTATG TAAAAACTAT TCACAAGCCA ACGTGaTCAT 2580

CGAAAAAGCG AArGATGCCA AACGTGTCGT TGTCGTTGGT GGCGGTTATA TTGGTATTGA 2640

ATTAGTAGAA GCTTTTGTAG AATCAGGTAA ACAAGTTACT TTAGTTGACG GGTTAGACCG 2700

CATTTTAAAT AAATATTTAG ATAAGCCATT CACAGATGTT CTAGAAAAAG AATTAGTTGA 2760

TCGTGGCGTT AATTTAGCTT TAGGTGAAAA TGTCCAACAA TTCGTAGCAG ATGAGCAAGG 2820

AaAAGTAGCA AAAGTCATTA CACCAAGTCA AGAATTCGAA GCAgATATGG TGATTATGTG 2880

TGTTGGTTTC CGTCCAAATA CAGAATTATT AAAAGATAAA GTTGACATGT TACCAAATGG 2940

TGCAATCGAA GTCAATGAGT ATATGCAAAC AAGCAATCCT GATATTTTTG CTGCGGGTGA 3000

TAGTGCAGTT GTTCACTACA ACCCAAGTCA AACAAAAAAT TATATCCCAT TAGCAACGAA 3060

TGCTGTTCGT CAAGGTATGC TAGTAGGTCG TAATTTGACT GAACAAAAAT TGGCTTATCG 3120

TGGCACACAG GGAACTTCTG GTTTATATTT ATTCGGTTGG AAAATTGGTT CAACAGGTGT 3180

GACAAAAGAA AGTGCCAAAT TAAACGGACT AGATGTCGAA GCGACTGTTT TCGAAGATAA 3240

CTATCGTCCA GAGTTTATGC CAACAACTGA AAAAGTTCTG ATGGAATTAG TCTATGAAAA 3300

AGGAACACAA CGCATTGTTG GTGGTCAATT GATGTCTAAA TATGATATCA CACAATCTGC 3360

CAATACTTtA TCATTAGCGG TCCAAAATAA AATGACgTCG AAgATTaGCG ATTTCTGATT 3420

TCTTCTTCCA ACCACATTTT GATCGCCCTg GGAcTACTTA AACTTATagc CCAAGCCGCT 3480

TTAGAAAATA TGT 3493

(2) INFORMATION FOR SEQ ID NO: 148:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1299 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

CTTCTTTGGA GATTAGGACT TTCTCAATAT CTTTTTCTAT CATAGTGCGC TCCTTTCAGC 60

ATTACTCCTT TGTCTTAAAA AGGAGTCTGT AGTGTATTTT ATCAGTTTCT AAAGGAATAC 120

TCAAATAGGA ATTTGCAAAT TTTGGTACCC AAATGATTTC CTCTTGCTCA TCCGTAATGA 180

CCCAAGCACT CTCTCTCGCT TCATTTGGCA CTTTTTGATC AATAAAAACT CTGCTTAATT 240

TTTTCGTAAA CCCTGGTTTT AATGTAATCC GATCACCACT TTGGCGATGA CGAACGGTCA 300

AAGGTGTGGC TGTGGTCAAC GGGATCGCTA ACAACTGACC CGTCCACTGA TTAATCTCTT 360

CAGGAAGAGG AAACGGTTTG TCTGTCGCAA TTAATCCTAG CCACTCGGTT TCCGATAGAT 420

AAATCCCCGC ACCAGGTACT AATAAATGTG TAACTTCTTC TCGCAGTGCA GCCTGTTTTT 480

GTTCTAAGCA TAGCCAATCA TAACTTTTAG CTAGTTGCCA TTGCTGTTCT AACATAACCG 540

AGCCTTGCGG TTGCCGTTCA TCATTCAAAA CTTTTAAAAT TTGTGTCATC TGGCGTTGGT 600

TTAACGAAAC CCCTAATGGC ACTAATGTTT TTTGGaAAAA CGCGGTTAAA AAGAAGTGCT 660

GGTACGCAGG CGTTTCTTTT AATAATCTCT TCAAAGGAAT GCGCCAGCCT TGTTTGGTTG 720

GTTCTACGGC TTCAGATAAT TGAGGCTCAA TTTGTTCTTG AATAAATTCT TGGGCAAAAC 780

GTTGTTCATT ACTAAAAGAA GCAATCTGAT CTAAGAATTG CGGATTTTCC TGCTTTAATA 840

GCGGCAAAAC CTGATTACGG AGCCGATTTC TCAAATATTC ATTGGTTTGG TTCGTTTCGT 900

CCTCAAAATA AACGAATGCT TGGCGCTGTG CAAATTGATA CAATTCTTCT TTTGGATAAA 960

TGAGTAACGG ACGAATCAAT TCCCCTGTCG CAAAAGGACG TCTCTCTTTA ATGCCTGCGG 1020

AATGGGAAAA ATTAGTTCCT CGAATCAGCT TCATTAAAAT CGTTTCCGCT TGATCGTCTA 1080

GATGATGTGC CGTCATTAAG ACCGCCGCAC CTTCAATTTC CATGACTTCT TtAAAAAATt 1140

CATAACGAAA TGTGCGCGCT TTGCTTCTAA ATTtCGTTGC TTTcAgGGTC TcCCAACGTG 1200

TTcATAATAC GTaGcgCTCG CTGCTGACAA TAGGTAGCGA GATATTGGCG CTTCCTGTGC 1260

CGACGCTTCT CGCAATTGAA GGATTGACAT GTGGCAAnA 1299

(2) INFORMATION FOR SEQ ID NO: 149:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4503 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

TTACTAGTCT GTTGGAATAC ATTATGATAG TGTAAGATAA ACGGACTATT GGAAAGTCAC 60

AGCAGAGGAG TTTGAACATG TACACATTTA AAATCGGTAT TCCGGCAGAA GTACACGATA 120

CATTTGTTAA AAATCACCCA TTATGTAATT TACTACAGTC ATCTTCATGG GCAAAAGTCA 180

AGGATAACTG GGGTTCTGAA ATTGTTGGGG TTTATGAAAA GGATACATTA GTTGCATCAA 240

GTTTAGTTTT AATCAAACCA TTACCAGCCG GTTTTACGAT GTTATATACA CCAAGAGGTC 300

CAGTAATGGA CTATACAAAT GAACGTTTGG TCAGCTACTT TATGGCTGAA TTGAAAAAAT 360

TTGGCAAGAA AAAACGAGCT TTATTTATAA AAATGGATCC AGCGGTTCAT TATCAAGATT 420

TTCATTTGGG AGAAGAACAC CAACCGCATG CTGAAGCGAC ATCTATTATT GAGACCTTGA 480

AAGAAGCAGG TGCCAAATAT CAAGGACTGA CAATGGATAT GGGGGCCACC ATCCAGCCTC 540

GCTTCCAAGC GAATATTTAT CGTGAAGATT TTAGTGAGGA ACAATTGTCA AAAAGTACCA 600

AGAAAATGAT TAAAACAGCT GAAAAAAAAG GGGTCGTTGT GCAGCAAGGG CATGTTGACT 660

TTGTTGATGA ATTTGAAAAA GTCATTCAAT CAACCATGGA ACGCCAACAT ATTTCTTTGC 720

GGAATAGTGA TTACTTTAAG AAGCTGTTGA ACATTTACCC AGAAGACTCC TTTATTATGC 780

TGGCGCAAGT CAATTTAAAG GAACGCTTAG ATAGCACGCG ACAACGCTAT GACAAAAATC 840

AAAAAGACTT GAGTAATTTG AAAGAAAATC AAGTGAAAAA GCGCCATAAT TTAGAAGAAT 900

TAGATGCTTC CTTGACTCGG GAACTAAAAG AATTAGAAGA AAATATCGCA GAAGCTGGTG 960

AAATCGTCAC GGTTGCTGGC GCTTTGGCTG TGACATTTGG TCCAACAAGT GAAATCCTTT 1020

ATGCTGGCTT AGATGATCGT TATAAACGCT ACATGCCAAC ATATGTGACT TGGCGAGATG 1080

CCATTCAAGA ATGTTTTAAC CGTGGcTGCG AAAcTGTAAT ATGGGCGGCT TAGAAGGTAG 1140

TTTAAATGAT GGCyTAATTA AGTTTAAAGC GAACTTCAAC CCmACGATTA ATGAATTTAT 1200

TGGGGAATTC GATTTACCGG TTAATAAATT ATTATTTAAA GCCAGTGAAT ATGCCTATAA 1260

ATTACGAAAA CAAAAAAAAT AAATAAAAAA CACAGGGACT AAAAATCAAA CTTGATTTTT 1320

AGTCCCTGTG TTTTTTTCTA TAGTTTTTGC GTTGCATTCC GTTTCATTAA GCGAACAAAA 1380

CGCTAGCAAT TGCTGAAGCA GCTACACCAG CAACAAGAGC TTCTGGCACA CCAGAACTGA 1440

AAATCGTCGC AATAATAGCT AGGTAAACAG CAGTTCCAGC TTTACCAATA ATTTCAGCAT 1500

AAGCTGAATT AAAAAGAAAA TAAATCATGT TCATGACTAA AATGGTATTC GTTGCTGAGC 1560

CTAAAACACC AGCCACAAAT AAGGAGACAG GTTGCGCTTT TTCCTTCGTT AATTTATTTA 1620

GCCATTTGTA AGCGAAATAA GGAACGACCC CAATTAAAAT CCGCGGAATG AGCACAATTA 1680

AAAGTGCCTT CCAACTACCA TGATCGGTAC CGATCACAGG AATAAATGGA GAAAAAACAA 1740

ACGACAAAGG TGTTGGTGCT GTCGTATTTT TAATTAAGCT AATGATCCCA AACGTGCCGC 1800

CCAACATACT ACCGATTTTT GGCCCTAAAA CAATGGAACC AATAATTACA GGAATATGCA 1860

TCGTGGTCGA ATTTAAAGGA CCTAATGTGA TAAATCCTAG TGGGGTTAAA GCCAGTAAAA 1920

GTAAGATAGC TAAGAACATT GCTGTTAACG TAAATTTTTT CGTGTTATTC ATAAATTACT 1980

CCTTGTTATC AGATGCTAAA ATTTTCATCA CTGTTTGAAG AATATCTTCA ACAGTGGCAA 2040

GTGCGCCACG ACCATAATCG CCACAAGCTA ATAGAGCTTC ACGTGGGATT ATTTCTTGAT 2100

AGCCGACTTC TTTTAAAATG CCAATATTTC GTTGGACAAT GGGGTTTTGG TACATATACG 2160

TATTCATCGC GGGCGCAATA ATTTTTGGTG TTTCAGGATA AAGTGCCAAA GCAACTGTTG 2220

ATAATAAGTC GTCGGCAATC CCATGGGCTA ATTTCCCAAT TGTATTGGCC GAAGCAGGTG 2280

CTACTAAAAA TAAATCCGCT TTTTTAGCTA ACTCAATGTG ATTAATTTTA CTGGGATCAA 2340

TTTCCATCAT GACATCCGTA TGCACTGGAT TCTTAGAAAG TGATTGTAAG GTCAGCGGAG 2400

TGATGaACGC TGTGCTACTT TTAGTCATTA AAATTTCCAC ATTGTAGCCA AGCTTAGCTA 2460

GTTGACTTGT AATATCTGCT GCTTTATATG CAGAAATGCT GCCAGAGACG CCTAATAAAA 2520

TAGTTTTCAT TGTTGTTCTC TCCATTTCTT TTCAACATTT TTAACGATCA TTGAACCAAT 2580

CCCTTGTTTT GAATAAGCTT CTTGGACAAT GCCCGTTTCG TTAATTAATA ATCCTTTGTG 2640

TTCTGTTTCA TGAACGTTCA TCAAGTCATT GGCTAAGACA AAGTCACACC GATTTTTAAC 2700

CAAAGCGGCT TGTGCAACTT GGACAAGTTC TTCTTGTGAA ACGTTGACAA GTAATTTAAA 2760

GCCCACTAAA ACAGTTTGGG GTTGCTTGTC ACGAATCATT TGGATGACTT TAGGATTTTT 2820

CTTTAGAAAA ATTAATAATC GGTCCGTATC AGATGAAATT TTTGTTGCTG TTTGAGGAAT 2880

ATCAGCAATC TGATTCAATG TGTTTTGAAC GATAGTTACC AAGGCTTCTT TGGTCTTTGG 2940

CAATGTTTGT TCGCTTAATT GTTGAGCAAA aGAATCAATA AATTGTTCTT CCGTTTGCGC 3000

GGTTTCCGTC GTAAAATCAC TGACTGCCAT TGCGTGAACA ATCGCATCAA ATTGCTTTTC 3060

AGCAAAGAGT TGTTCTAACG TAGTGGCTAA TTCAACCGTT GTTTCTATTT CTTTAATTGA 3120

TAAGTCCAAT TGTTGGGTTG GACGTAATGC ATGTTTGGTT GTTACGTAGG TTACGGTGTG 3180

TCCAGCAGCT AAAAaACTTT CAGCAATTGC TTTTCCTAGT CCGCCAGAAG AATGATTGGT 3240

AATCGAACGA ACATTATCAA TCGGTTCTGA CGTTCCGCCA GCAGTAACTA AAACATCCAT 3300

AATCAAAACG CCTCCTTAAG AAATTCGGAC AATGGTCCGT CCAATATGTT GGCCTGCTTG 3360

TAATGCTTCA AAAACAGGGG GAAGTTCTTC TAAAGTAATC TCTTGATACA CAGCTTTTTC 3420

AGTGATATCT AAATCAGTGG CAAGACGTTG CCAGACTATT TTTCGTTTGT CCATACCAAC 3480

ATTaACCGAA TCAATGCCCA ATwAATGAAC GCCCCGTAAA ATAAAGGGTA AAACGGTTGT 3540

CTCTAATGTA ATTCCAGCAG CGCGGCCACA AATGGCTGAA CGGCCATCAT AACGAAGTTG 3600

TGGTAAAACA GCAGAAGTAA TTTCCCCACC AGTGGTGTCA ATTGCAAAAG CAAAACGCTG 3660

TTTCATCAAT GCTTTAACAG GAGTTGCTAA GAAATCTTCT ACTAATAAGC ATTCACTAAC 3720

ACCTAACTCT TTTAAAAGCT CAAAAGAATC TGGCTTTCTA GTTAATGCAG TGATTGAAGT 3780

GTAGCCTAAT TGATGcAACA TCGCAwTAGA CaATGTCGCA ACaCCGCCTG TCGCACCAGT 3840

TACAATAATG GGGGCTTCTT TGTCGTCGCG TAAGCCaTCG TCTTCTAAAG CTGTGACAGA 3900

TAATCCGGCC GTAAAGCCAG CAGTTCCGAA GATCATTGCT TCTTTGAGAC TCATTTTTTC 3960

AGGAAGTGGA ACTACCCAAT CGCCAGGAAC ACGCACATAT TCAGAATAAC CACCAGTATG 4020

ACTAACCCCT AATTGATAGC CTGTTACGAT CACTTTATCG CCTTCTTTGA AaCGAGGATC 4080

TTCCGATGTG ACAACAGTGC CACTGACATC AATTCCTGGA ATCATTGGAT AATTGCGAAT 4140

CACACCGCCG TCTTTTTTTG AAGCTAAAGA GTCTTTAAAG TTAACTGAGG AGTAGGCTGC 4200

TTTGATTACT ACAGTGCCTT CTGATAAATC GTCTAAAGTA ATGGTTTCTA GTTGACCAAT 4260

AACAGATGTA TCTTGTTCTC TTAACGTAAA TGCTTGAAAT GAATTCATAT AAAAACTCTC 4320

CTTCCAATTA TGTAAAAAAG CAACTCGTCT TATTTTATCA TATTCTGATT TTTTCGTATC 4380

ATTTTTCTTC TTATAAACAG TAAGTTATAC AGAAAAGAAA GATTGTGTGA CAATTAGACT 4440

AGAAAAATAA AGAAAAATGG TTGAAATTTT TCAAAAAGTT CGTATACTAA AAAAGGTAAA 4500

ACT 4503

(2) INFORMATION FOR SEQ ID NO: 150:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11739 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

GCGTTAGGTC AAGTAAGTCT TGGmTCGTTT TTGATGAcGT CTTTTTAtGG TGTTTTTmaG 60

TACTmGTTTA TCGTATTTTC TATTACGCAA AAATCAATCC AATTTATTTT TATTATTGAT 120

GATTGGCATG ATTTTAGGTA CTTTTTTCAG TAGTATTAGT ACGTTTTTAC AAGTGATTAT 180

GGATCCTAAC GAATACGACT TACTCCAAGG AAAGTTGTTT GCCAGTTTTG GTAACGTCAA 240

TAGTCAGTAT CTCATTTTCG TTGGTGTGGC GGTACTCATA ATTTGTGGGA TACTTTGGCG 300

GGAAAGTTAC CGTTTAGACG TTTTACATTT AGGGAACCAT CaAGCACAGA ATTTGGGGAT 360

TGCTGTGAGT CAGTTTCAAT TAGTGTTGTT GTGTTTGATC AGTAGTCTGG TCGGCTTATC 420

AACAGCATTA GTGGGACCGA TTACATTTTT AGGTTTTATC GTGGCCAATA TGAGTTATCA 480

ATTCATGAAA ACCTATCGCC ACCGAGAACT GTTTGTGGCA GGGAGTTTAA TAGCAGTCTT 540

TTTACTGGTG TTTGGTCAAT TGATGGTTGA ACAGGTCTTT CATTTGAATA CCCCGCTCAG 600

TGTGGTGATT CAATTTGTTG GGGGCGTTTA TTTTATATGG AAAATTATTG CAGAAAGGAA 660

GCAACGCACA TGATCCAGAT GAGTGATGTA TCAAAAAAAT ATGGCGATAA ATTTGTCGTT 720

TCAGACATTG ATTTACCTAT TTCTGAAGGC AAATTAACGG CCTTTATTGG ACCAAATGGC 780

GCTGGGAAAA GCACGCTGCT AGCCATGATG AGTCGCTTAA TTCCCAAAGA TACCGGGGAA 840

ATTTATCTGG ACAAGCAAGA AGTAAAAACC TGGAAGCAAT CGGCTTTTTC ACAAAAAATT 900

GCCCTTTTAA AACAAGCCAA CGGTGTACAA CTTAAATTAA CTGTGCGAGA ACTGGTCAAT 960

TTTGGTCGTT TTCCTTATTC CAAAGGTCGC TTAAAAAGCG CTGACCACGA AAAAGTAACA 1020

GAAGCACTGG AACAATTGGG CTTAACTGAA ATGGCGGAAG AATACATTGA TACCCTTTCT 1080

GGGGGCCAAT TACAACGGGC GTATATTGCC ATGATTTTAG CCCAAGATAC CGATTATATT 1140

TTATTGGATG AGCCCTTAAA TAACCTCGAC ATGAATTATG CGGTTCAAAT GATGCAGATT 1200

CTTAAACGCT TAGTGGATGA ACTAGGAAAA ACGATTCTGA TTGTTCTACA TGATATTAAT 1260

TTTGCCGCAA GCTATGCAGA TGAAATTGTC GCCATGAAAG GCGGAAAGCT CTATGCGCAT 1320

GGCGCAACAG AAGAAGTTAT TCAAACAAGT ATTTTAAATG ATCTTTATGA AATGAATATC 1380

CGGATTTGTG AAATTGAAGG AAAACGCTTT TGTCTTTATC ACTAAATCGG GCAATTTAAA 1440

AATAACGTTT TTGGAGGAAA AGAATGAAAA AGAAATTTTT AGCAATGATG GCAGTTTCAA 1500

TGATGGGACT GTTAATGTTA AGTGCTTGTC AAACAAATAA AAAAACAGCA GATTCTGCAA 1560

CAACAGAAAC AACAGCTAAA ACGGAAGTCA CAGTCAAAGA CACCAATGGT CAATTAACCG 1620

TTCCCAAAAA TCCTAAGAAA GTCGTTGTTT TTGATAATGG TTCCTTGGAT ACAATGGATG 1680

CACTAGGTGT CGGTGACCGC GTGGTAGGTG CGCCAACTAA AAATATCCCT GCGTATTTGA 1740

AAAAATACCA AAAAGTTGAA TCAGCAGGCG GCATTAAAGA ACCAGATTTA GAAAAAATCA 1800

ATCAACTAAA ACCAGACTTA ATTATTATTT CTGGTCGTCA ACAAGATTAT CAAGAACAAT 1860

TAAAAGCCAT TGCGCCAACC ATTTACTTAG CTGTAGATGC CAAAAATCCT TGGGCATCAA 1920

CGAAACAAAA TATCGAAACG TTAGGCACTA TTTTTGATAA AGAAGAGGTA GCTAAAGAAA 1980

AAATAACTGG CTTAGAAAAA GAAATTGCTG ACGTGAAAAA ACAAGCAGAA GCTAGCGCGA 2040

ATAATGCGCT TGTTGTGTTA GTTAACGAAG GACAACTTTC CGCTTACGGA AAAGGCTCTC 2100

GTTTCGGTTT AATTCATGAT ACATTTGGCT TCAAAGCAGC AGACGATAAG ATTGAAGCTT 2160

CCACTCATGG GCAAAGTGTT TCTTACGAAT ATGTTTTAGA AAAAAATCCT GGGATTCTCT 2220

TTGTGGTAGA TCGCACCAAA GCAATTGGTG GCGACGATTC AAAAGATAAC GTCGCTGCAA 2280

ACGAATTGAT TCAAAAAACC GATGCTGGTA AAAATGATAA AGTCATTATG CTTCAACCAG 2340

ATGTTTGGTA TCTAAGCGGT GGTGGATTAG AATCAATGCA TTTGATGATA GAAGATGTTA 2400

AAAAAGGATT AGAGTAAAAA AArGCCTAGG ACAAAAATCA CTTTGGATTT TTGTCCTAGG 2460

CTTAAAAGCT GATAAACGGC GGGAACAAAA GCAACTCCTT CGGAAATAAG CTGAAATTCT 2520

CCAAAAATTA AAGAACAATT TTCGGAAATT TCTTCTTATT TCTCGGAGCT AAACACTTCT 2580

GTCCCGACCT TCTTGATGAA AAACAAAAAA GTTTTAAATT AAAAATGTGG TTGACGGCCT 2640

CGAAAATGTT TGCTAAACTA TCCGTAAATC AGTATGATTG ATGAAACAAA TAAAAGAACA 2700

AAATGTGATG AGAAGGAAAG TACTTTTTAA AATAGCTTGC AGAGAGCCTC GTTTGGTGAG 2760

AAGAGGTAGT GACGTTAAAG GGGAAAATGG CCTTTGAACA GGTTGGTTGA AGCCGTTTGC 2820

TTAAACCAAT TCAGGTGCGC CTGTTATAGC GCTACAGTAT GATTGTACTG GAAACGAGCG 2880

GATCTTTTTT AGCGAAGAGA TCAACAAAGG TGGTACCGCG ATTGCATAAG CATTCGTCCT 2940

TTTTAGGACG AATGCTTTTT TTATTTGTTT AAAAAATATG AGGGGGCAAC AACATGAAAT 3000

TGGGGAAAAA AGTAGTAGGT TTGATTGCAA CAGGGTTTCT TTTAGCCGCA TGTGGCGGAA 3060

CCAAAGAAGC GGCAGAGAAA GTAGATTCGG GAAATTTAGC AGCTGAACAA AAAATCAGTA 3120

TTAGTTCACC TGCACCAATC TCAACATTGG ATACAACACA AACAACAGAT AAAAATACCT 3180

TTACAATGGC ACAACATTTA TTTGAAGGCC TTTATCGGTT TGATGATGAT AGTGCCACGG 3240

TGCCAGCTCT AGCTAAAGAT GTCAAGATTA GTGACGATGG GCGCAAGTAC CACTTTACCT 3300

TGCGGGAGGG GATTAAGTGG AGCAACGGCG AGCCAATCAC GGCCCAAGAT TTTGTTTATT 3360

CTTGGAAAAA ACTGGTGACA CCAGCGACGA TTGGACCGAA TGCCTATTTA CTAGACAGTG 3420

TTAAAAATAG TTTTGAAATA CGCAACGGTG AAAAGTCAGT CGATGAATTA GGGATTTCAG 3480

CCCCGAATGA CAAAGAATTC ATTGTTGAAT TAAAACAGGC CCAACCTTCC TTCTTAGCAG 3540

TCGTTTCGAT TGCTTGGTTA GCGCCACAAA ATCAAAAATT TGTCGAAGCG CAAGGCAAAG 3600

ATTACGCCTT GGATAGTGAA CATTTACTTT ATAGCGGGCC ATTTACGCTA GCCAATTGGG 3660

ATGCGACTTC AGATACTTGG ACATTGAAAA AAAATCCAGA ATACTATGAT GCGGATCAAG 3720

TGAAACTGGA AGAAGTTGCG GTTAGCACAA TCAAAGAAGA TAATACTGGG ATTAACTTAT 3780

ATCAAGTGAA TGAACTAGAC TTAGTTCGCA TTAACGGACA ATATGTTCAA CAATATCAAG 3840

ATGATCCAGG CTATGTCAGT CATCCAGATG TGGCCAACTA CTTCTTAGAT TTCAACAAAA 3900

AAGAAGGAAC GCCATTAGCG AATGTTCATT TACGAAAAGC GATTGGCCAA GCAATTGATA 3960

AAGAAGCCTT AACACAAAGT GTCTTAAACG ATGGGTCAAA ACCCCTTAAC GGATTGATTC 4020

CAAGTAAACT TTATGCGAAT CCAGAAACGG ATGAAGATTT CCGAGCTTAC AGTGGCGAAT 4080

ATTTGAAAAA TGACGTCAAA AAAGCTCAAG CTGAATGGAC GAAAGCCCAA GCGGATGTCG 4140

GTAAAAAAGT GAAACTTTCA TTGCTGgCGG CAGACACAGA TCAAGGAAAA CGAATTGCTG 4200

AATATGTTCA AAGTCAGTTG CAAGAAAATC TGCCAGGTTT AGAAATTACC ATTTCATCGC 4260

AACCAAGTAA TAATGTGAAC CAATCGCGAC GTGAAAAAAA TTATGAGTTG TCTCTTTCAG 4320

GATGGATTGC CGGCAGTAGT GAATTAGACT CTTACTTTAA CTTATATGCA GGAGAATCAA 4380

GTTACAATTA CGGCAATTAT CATAATGCCA AATACGACCA ATTGGTAGAA GAGGCACGAA 4440

CGATTAATGC CAATAATCCA GAGAAACAGT TTGCAGAATA CAAAGAAGCG GAAGACATCT 4500

TGTTGAACCA AGATGCTGCC CAAGTACCGC TGTATCAAAG TGCCTCAAAT TATCTAATCA 4560

ATCCTAAATT GAAAGGCaTT AGTTATCACT TGTATGGGGA TTATTTCCAC TTGCGCAATG 4620

CCTATTTAAC AGAATGATTC AGGAAGAAGG AGTAAAAAAT GGAAAAATTA GTCGaACGAT 4680

TTTTAACTTA CGTGAAAGTT AACACCCGTT CAGATGCCAA CAGTCAAACC ACACCGACGA 4740

CGGTCGGTCA AGTGGTTTTA GCCAAAATGA TTGAAACAGA ACTTCATGAA TTAGGCTTGG 4800

CCGATGTGCA TTATAACGAA CAAAACGGCT TTTTAACGGC ACGGTTGCCT GGTAACCAAC 4860

CTGCAGCGAA GAGTATCGGC TTGATTGCAC ATCTGGATAC GGCGGATTTT TCAGCAGAAA 4920

ACATCCGTCC CCAAGTCATC ACTAATTATG ATGGTCAAGA TGTCCTCTTG AATCAAGAAC 4980

AAGGCATCGT TTTATCAGTT GCTGAATTTC CCAATTTAAA AGAGTATCAA GGGGAAACGT 5040

TGATTACTAC AGACGGAACC ACACTTTTAG GAGCAGATGA CAAAGCGGGA ATTGTCGAGA 5100

TATTAGCTGC TGTTGAGTAT TTTTTGGCAC ATCCAGAAGT AGCGCGGGGG GATGTTTGGC 5160

TCGCTTCGGT CCAGATGAAG AGATTGGACG CGGGGCCGAT CAATTTGATG CGCCGAACTT 5220

CCCAGTCGCT TTTGCGTATA CGATTGACAG TGGCCGCGTG GGGCATTTCG AGTATGAAAC 5280

GTTTAACGCT GCCCAAGCTG TGATAACAAT TGAGGGAACC AGTGTCCATC CGGGAACGGC 5340

CTATGGTTCG TTAGTAAATG CCATTAAATT AGGCGAACAA CTTGATCAAT CACTACCGCA 5400

AAAGGAAGTG CCAGAACAAA CACGTGGAAA TGAAGGATTT TATTTACTCA ATAAATTTAC 5460

AGGAAGTATT GAAAAAGCAG AACTTGTCTA TATCATTCGG GATCATGATC AAGAAAACTT 5520

TCAAGCGAGA AAGCAATTTT TGGAAAAACA AGTCCAACGT TTAAATGCGC TGGCGGATAA 5580

ACCCAGACTA ACAATTACTT TCCAAGACCA ATACTATAAT ATGAAGGAAA TCATCGAAAA 5640

AGATTGGACT CCTGTAGAAT TGGCCGTCCA AGCAATGGAA AGTTGTGACA TTGAACCAAT 5700

CATCACACCA TTTCGTGGCG GGACAGATGG TTCTAAGATT TCTTTCATGG GCATTCCCAC 5760

GCCTAATCTA TTTACAGGTG GCGAGAATTT TCATGGGCAA TATGAATTCA TTACCGTTGA 5820

ATCAATGGCC AAAGCTGTTC AGACGATCAT TGCCATCATT CGCTTAAACG CTAATTAAAG 5880

AAACACGCAA GGGAAATGCC GATTTTCCTT GCGTGTTTCG CTGTTTTTCG GTAAAGTATG 5940

TAGAAAAGTA GGGAGTTTTG CTCATGAATT ATCCaTTGAT TTTTGCAGAA AACCAACGAA 6000

TCGAAACCGA ACGCTTGATT TTAAGGCCCG TCACATTAGC CGATGCAGAG GATATGTACG 6060

AATATGCTTC AGATGAGGAA ACGGTCCGCT ATGTGTTTTT GAAAAATCAA ACAATCGCTG 6120

AAACAAGACA AAACATTGCT AAATTTTTTA TGGGGGAACC TTTAGGAAAA TATGGGATTG 6180

AAGTGAAAGA AACGGGCAAA ATGATTGGCA CGATTGATTT GCGGGTCAAT GAAACCAATA 6240

ATATTGGAGA ATTAGGCTAT GTCTTGAACC GTACCTTTTG GGGCAATGGC TACATGCCTG 6300

AAGCCGCGAC GGCGTTAGTT GAACTGGGTT TTGCAAAAAT GAAACTCATG AGAATCTTTG 6360

CTTTACATGA TCAAGACAAT CCAGCTTCTG GTCGTGTCAT GGAGAAAATT GGTTTTACCT 6420

ACGAAGGAAC ACTACCCAAT GCTCGAATTA GTAAAGGGAA AATTGTCACA GATGTTTATC 6480

GTGGGATGAC ACTGGAAACA TGGCAAAATC GACAAAAATA GAAGCGGAAG ACAAGTTAAG 6540

AGAGGAATTA GAGAAATGTT GATGTTTACA GAAAAAGAAT TCGCAGCCTT TGAAGTTGCT 6600

GGATTAGATG AACGAATGGC GGTTATCCGT GCGCAAATTC AACCTATTTT TCAAGAATTA 6660

GACACGTATT TTGCCGAACA ACTAGCGCCT GAACTAGGAA CGGAGTTGTT TGTCCATATT 6720

GCGCAACATC GCCGTCGCAC CGTGTATCCG CCAGAAAATA CTTGGTCAGC ACTTAGTCCG 6780

AATAAGCGCG GGTATAAAAT GCAGCCGCAT TTTCAATTAG GCATTTGGGG CGACTATGTG 6840

TTTATGTGGT TGTCATTCAT TGATAATCCT AAAAATGAAA AGCAGATTGC GCAACCTTTT 6900

TAGAGAATCA GCAATTATTC CAAGCATTAC CAGAAGATAC GTATGTyTCC TTGGATCATA 6960

CGGTTCCTCA AATTACTCCG CTGCCAGAAA CGGATTTAGA GAAAGCCTTA ACTCGTTTTA 7020

GAGATGTAAA AAAAGGTGAG TTTGAAATTG GGCGCATCAT TCCCAAAGAT AGTGACTTGT 7080

GGCAAAACCC TGAAAAAGCA CGAGCTTATA TGTTGGCAAC CTATCAACAA CTCCTGCCAT 7140

TGTATCAATT AGCTGTTGCC CAATAAGAGA AACAGACAGT TATTTTGTGT TTTAGTGTTA 7200

TTTTATCGAA TTATCCAATC GTCAAAAATG AGAaAAGCTC ATAATGTTAG TTTTAATTAC 7260

TAACATTGTG AGCTTTTCTG TAGTTTTTTT GAAAAAAATT AAGGGATGAC TGGCTGATTT 7320

CTATGGAAAA GTACGTTTTG ATAGCCTTTT TTGCTTAAAT AACAAGAGTA GCAACCCTTG 7380

TTTAATAAGG GATTTTCGCG AATAGATGTT GTCAAATGAA AAAAAATGAT GTTTTGAACA 7440

AAGAGTTCTT TGCTATTATA AAGGTATGAA AAAGGTTGTT TACGCATTCT TTGTATAGAA 7500

GAATGTGAAT GAGTTAAAAG GAGGTTTTTA AATGAAAAAA AAGCGTTATT TAATAATTGC 7560

GTGTTTACTA TTTTCCCCTA GTTTTTTTAT AAATGTTGAA GCATCTGAGG GTGGTTCTAG 7620

TTCGGTGGGA ATTGAATTTT ACCAAAATCC GGCAACACCC GCTCCTAAAG ATGCCCCACC 7680

GAAAACAGAT GAGCCAGCTG CGGATCCCAA GGAACCAGCT GGTCCTCTGC AAGGAGATCA 7740

ACGAAgTGGT GGTTCGACAC AGACCACCAC AGCTGGCTCG CAGCTCCCTC GTACAGGAAG 7800

CAAGAGTCAG GCAAACCTGA GCATTCTTGG TCTTGTCTTG ATTGGTCTTG TCGGAATGGT 7860

CCAGAGAAAG AAGGGACGAC ATGAAGCAAA CTAAGTGGCA ACGATTAGCA ACCATTGGCT 7920

TGTGTAGTTC TTTAGTGATT AACGCCTTTT CTGGCGCGAC CGCAGTTGCG GAAACAATCA 7980

CGATTGAAAG TAGTCCGACA GTAGCAAGTA GCGCTAAGGA AGCAACATCA GCAAGTAGCG 8040

CAACGCCAGA AAGTACGGAA AGCAGTCAAG AAACGACTGA AACGAGTCGG GAAGAAGTGA 8100

CACAGGAAAC AGTGGAACAA GAAGATACAA CGGAAATAGC TCAAGAAGAA AATCTGCTGG 8160

AACCATCTGG TATTCAAGCG AAAATTATGC CTAGAGCATT TGTCTGGGCA GCTACAGGAA 8220

CCACGAATAG AACCACGGTT CGTCTTAACG ATGTAGTTGA TTATCGAATG AAACTGGAAA 8280

ATATTACAAC AGATGATACA AATTCAGATA TTGTTGGTCC GATTCGAATT GAATCAAAAT 8340

TGGCAGCTGG AATGACTAGA CCAACGTCAG TTATTATCAA AGTTGGCCCC GAGGAACAAA 8400

ATGTAGTTAT AGGAGAAGGC TCTCATAATG CCAATAGTGG TGGCGAATAT TTTGTTTGGA 8460

CAGAGTCGAC TAGGACGGTC ACTGCATTTA TCAATCGTCT ACACGGGCGA ACTTCTGGAA 8520

ATACTGGGTA TATCAAAACA TTGTATTTTC AGACAAGAAT TACTTCGGGT ACACACAACG 8580

AAAAAAAATA TATCGATTCC ACGATTCGAT TTAATGGATT TGGCGCAGTG AATAGACGAA 8640

ATGAAATGAC GTATGTGGAT AGATTTTCTG GTAACTTAGG TTTGAAAGGG AATTTAAAAT 8700

TTTACGATGA GGATGGAGTG GCTAGTAAGG AACAAATGAG TCTTTCCATA AACCCAGTAT 8760

TATTTAAATC AACCAATTCG AGAGTAGAAG GGACGCTGTT AAGATATCCT TTAACTATGA 8820

CAAATGCTGG TAGTGGGCTT TATAGGTTTG ATACAGGACT AAAAAATATA GGTACTAGTA 8880

ACGGGGGTTC GTGGTTTTTC TTTAGTAACG ATATTGCGAT TAAwTTTGCT ACAGCTACCA 8940

CCAATAGAAT TGATAAAATT GTGTTAGCAC CACCTAAAGC AACACTTCCA GACAGCCGAA 9000

CAGGTGTGAT GACTTCCAGA GCGTACACAT CAACAACTTA TACGAGAAGT AACTATGATA 9060

TTTCaCaTAT AaCTTTACTA AAGCGACTGG TAAGACATGG CGTATAACGA ATCGTAGTTC 9120

TAATAGCGGA GTAAGCACCA ATATTATGTT TAGTAACAAT CTTGTAGGAA CAACAGGACA 9180

GTTCAATAAA GGGAATTCTG GTGCTACTAT TAATTATTAT ATTTATTACA AAAAGCTTTA 9240

TGAGAATTTT GTTAATAGTA ACGGACAAAA AATTACACCA CCATCAGGAT TCACACAAGG 9300

AAAGCGAACG GTTATTAATA GTGAAGCGTA CACTTTTAAA CAAAGTGGCA CCTTGCCAGA 9360

TACCTATCAA GCAGACGGCA AAACCTATAA GTTCAAAGGT TGGTACAAAG GCAAAACCAA 9420

GCCAAACACC TTGACCACCA CCAAAGCACC AAGCTATGCG GTGACTTATA ACAATGATGA 9480

TGATTTGACG GTGGTGTATG AGGAGATTGA GTTTTTTGAC TTCCCAGCGC TGACCTATCA 9540

ATTCGGGTTC GTGGACGAGT CTGGCAAACG AGTAGATGCT TCAACAATTG GGTTGACTTA 9600

TGACAAATGG CGTGGGGAGT TGTTAACAAA TAATGTCAAT GACTGGAAAA CAGTCAGTCT 9660

TGAAAAAGGT CAAGTTGCAC CAACCAAAAA CAATTTGAAG GAAATTGCTT ATCCTGCGCA 9720

ATCTCTTGAA ATACTGAGTG ATCGAAGTAC ACAATATAGT GCGGCTAATT TAACCTTTAC 9780

ACTGCCCAAA TACTATGGGA AAATCAGTGT TTATAACAAG AGTGGTACTT TTGATGCAGG 9840

CTACCTGTTT CCAAATATCC AAAATAATAA TAGTTCAGAA CCATCTGTTG AAGGCTTACA 9900

CAAAAATCTT TCTAGCTGGT TTGAACTAAA GAAAAGTGGG AACCAATCGT TTGTTTTTAA 9960

CAAACCGATT CCAGCAGCAC CAATCGATGT TCAGGTGCCT GCTTACTTGC GAGAGATTGT 10020

CTATAATCCC GATCTGATAT CAAATACAGC CGTCTATTAT GTAATAGATA AACCTGTCTA 10080

CTACTATTTA ACCAACCGCA AAGTCACCGA AAACTTCGTA GACACAAGCG GTGCCAAAAT 10140

TACACCACCA ACAGGTTTCA CCCAAGGAAA ACAAACAGCG ATTACCAGTG ACCCATACAC 10200

CTTCAAACAA GCAGGTACCT TACCTGATAC CTACACAACA GGCGGTAAGA CCTACAAGTT 10260

CAAAGGCTGG TACAAAGGGA AAACCAAAAC CGAGCCTTTA ACCACGACGA AAGCACCAAG 10320

TTATGGGGTA ACCTATAATG ATAATGATGA TTTGAATGTG GTGTATGAAG AAGTAACAGC 10380

TATTTATCCA GGGTTCTATG CGCAATTTGT TGATGAACAA GGCAAAGCTT TTACTAATTC 10440

GCTAACTCTT TCTGAAAATT ATACTGAGTT TCTCAGAAAG ACAGGCACTA CTGTTTTTGA 10500

AACGAAGGGC GCCTTTTACC CAATGGTAGG TAAGCGCGAT GCAACAGTGA CAAATAAGTA 10560

TAAGATTGAA ACAACGAGTA GTGTTGCAGT ACCAAATAAT TATTGGACAG CCATGCCTGC 10620

TGACATTGCT AATAGAGGGT ACGCATTAAA TAACTTGGAT ATTACCAATC AAATGAAATA 10680

TGTTGACAAG ATAGAAACTA CTGACACGCA CACTACAATG TACGATTATA CTGGAGTTTC 10740

GGACATTGGG GGCAAACCAA ACAATGCTTC TCAGACAGAA TTAAATCCTA TTGATAAAAC 10800

ATTTGTTAAG CAAACAGATA ACACTTTTGC GATAAAATCA ATTTGGGGAA ATGGaACGTT 10860

TTCGCATATA GTTCCCAAAA TGTTATTGTT AGCAGAAAGT GGCAGTTTTT CAAGATTATT 10920

CGGGTTTGAT GGTACTCCAG ACTACAAGCA AACCGTAAAC TACAAAGTCA CCCGTAAACA 10980

AGTCACCGAA AACTTTGTTG ATGCCAACGG TGCCAAAATC ACCGCACCAA CAGGCTTCAC 11040

CCAAGGCAAC AAAATACCCA TGACTAGCAA CACCTTCAAG TACACAGCCG CAAAAGCCTT 11100

GCCAGCAACT TATACCGCAG GTGGCAAGAC CTATACTTTC CAAGGCTGGT ACAAAGGCAA 11160

AACGAAGCCT AATACTTTAA CAACCAGCAC TACGCCAACA TATAACACAA CTTTTGATGG 11220

CAATGACGAC ATGACGGCGA TTTATAAGGA AGAAGTGCCC AAAGCAAGTG TCGCTTTAAC 11280

GCGCACTACC GCTGAAACCG TGACTAGCGG CGGCAATGTC ACCTGGCGCG CCACAATTAC 11340

CAATACAAGC CAAGCACCAC TAACCACGGC GACCATCAAG AAATCAACCG CTTGGACAAC 11400

TGGCTTAGCA GCACCAACCG CCATGATTGT CACGCCAGCT GGCGGAACCG CAAAAACAGT 11460

TCCCGTCACA GCGACTACGT GGACCAACGG TGTTAGCTTA GGCACTGACA TTCCCGTAGG 11520

AAAATCTGCC ACGGTTCAGT TCACTACGAA AGCGACGGGG ACAGCGGGGC AAGTCTTGCG 11580

CGCTGGGATT ACCACCAGTG GGAATTACAG TGGTGTTTCT ACTTCGGCCA CCGTACGTGT 11640

GAAAGACAAT GATCAAGCAA TTGTAACACC AACGGCTGAA GGCTTCATCA GTGTTCCAAC 11700

CTTCAATTTT GGTCAAGTTG GGGTAGCGGG AAGCACGCA 11739

(2) INFORMATION FOR SEQ ID NO: 151:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4045 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

GAATTTCTTC AGTATCCTAA AACAAGAAGT GTATTATGGA AAAATCTATC AAaGTCAAaT 60

GAATTGrTaG AAGCAaTTGA GAACTACATT TATTATTACA ATCATCACCG AATCAAGGAA 120

AAACTTAACT GGAAAAGCCC AGTAGAATTT CGACAATTCA ATCAAAAAAC TGCATAAAAA 180

TAGAGTGGAA AAATCCACTC TATAAAAAGT CTAACTTTTT GGGGTCGCTA CAACAATACG 240

TTTGTTTTTT AATCCTTCAG TTTATAACAT TATAAGAGAA AAGCGAGAAG AATCATCCTC 300

CTTCTCGCTT TCTTGTTACC CTAAACCAAC ACGTTCAAAA ATCGTGTCAA CATTTTTTAA 360

ATGATAATGG TAATCAAAGG CATCATCTAG CTCTTCTTTT GATAAGACAG ATGTAATTTT 420

TTCATCCGCA TCTAGCAACG GACGGAACGC CGTTTGATGA TCCCATGCGT AAGCTGTTTT 480

TGGtTGGACC AGATCATACG CTTCTTCACG TGTCATGCCA TGATCAATCA ATTTTAACAA 540

GACACGTTGG CTGTAAATTA GACCATAGGT CGCATCCATG TTACGTTTCA TATTTTCTGG 600

GAAGACCGTT AAGTTTTTGA CAATATTACC AAAACGATTC AACATGTAAT TTAGTAAAAT 660

CGTCGTATCT GGAATAATAA TCCGTTCAGC AGACGAGTGA GAAATATCTC GTTCATGCCA 720

TAAACTCACA TTTTCAAAGG CAGTAATGAC ATGACCACGA ATCACACGTG CCAAACCAGT 780

CATATTTTCA GAACCAATTG GATTCCGTTT ATGCGGCATT GCCGAAGATC CTTTTTGTCC 840

TTTTGCAAAA AATTCTTCTA CTTCACGTGT TTCTGATTTT TGTAGACCAC GAATTTCTGT 900

TGCAAATTTT TCAATACTTG TCGCAATTAA TGCCATGGcA GAAACATATT CTGCATGTAA 960

ATCACGTGGT AAAACTTGTG TCGAAATCTC TTGAGCACGG ATCCCCAAAT GTTCGCAAAC 1020

ATATTCTTCA ACAAATGGTG rAATATTGGc AAAGGTACCT ACGGCTCCAC TGATTTTACC 1080

CGCTTCGACA CCTTTGGCTG CATGTTCAAA ACGTTCAATG TTGCGTTTCA TTTCAGAATA 1140

CCACAACGCT AATTTCAAAC CAAATGTAGT CGGTTCTGCA TGAACGCCAT GCGTTCTTCC 1200

CATCATGAmC GTATGTTTAT ACTCTTTGGC TTTGGCTCCG ACAATCTCCG TAAATGTTTG 1260

TAAATCTTTC CGTAAAATAT CATTCGCCTG TTTCAATAGA TAACCATAGG CAGTATCCAC 1320

AACATCTGTG CTGGTTAACC CATAGTGTAC CCATTTTCGC TCTTCACCAA GCGTTTCTGA 1380

AACCGCACGA GTAAAAGCGA CAACATCATG ACGTGTTTCT GCTTCAATTT CTAAAATACG 1440

TGCAACATCA AATGATGCAT TTTCTCGGAT TTTTTGTACA TCTTCTTTAG GAATTTCGCC 1500

AAGTTCTGCC CAAGCTTCAT CCGCCAAAAT TTCAACTTCT AGCCATGCCT TGTAGCGGTT 1560

CTTATCTGTC CAAATTGCGC CCATTTCGGG TCTTGTATAA CGATCAATCA TTGTGCCTTA 1620

TCCCCTCTTT ACTCCTCAAT TTTTGAAAAA TAATTACGAC CAAATCTTTG TTTCTTTAAT 1680

TTCTTCCAGT GTTTCATCGA TTTCTTTGGt TAATATAGTA ATATGcCCCA TCGCTCGTTG 1740

CGCTTTTGCT TCwCTGTGGC CATAATAGTG GAAATGCCAA TCTGGATGTT CATTAATAAT 1800

GTCCATCGTT TCATAGGCTT CATTTCCAAG CACATTGACC GTGACTGCTT CAGACAACAG 1860

TTGAATCGTT GGCATTGGCC AATTACATAA ACCACGAATA TGCGTATCAT ATTGGCTAAA 1920

AGAACAAGCA TCTATTGAAA AATTCCCAGC ATCTTCTGGA CGAGGTAAAA TTTTATTTAC 1980

ATAAATCCCG CCTGTTTCTG TATAAAACAT TTCAATTGAA AGGGTGCCGA CTAAACCAAC 2040

TGCTTTAGCA ATCACCTTAG CAATTCGTTG CAACTCACCA CTGAcTtCGT CATCAATTCG 2100

GGCTGGTGCA ATCGTTTCAT GTAGcTTACG CTCTTTTTTA ATTGTTTCCA CAACTGGAAA 2160

TGTCGCAAAT TCGCCACGAC CATTACCAGA AATCAGAACA GATAATTCTT TTTCGTAAGG 2220

AATCCAAGCT TCCAAGACAC AGGTTCCTTC TCTCAGAAGA TTCATGGATG GCGCTAAGTC 2280

TGAAGTACTA TGTAGTACAT AACGACCTTC CCCTCGCGTA TTCTTCAAGA TACACGGATA 2340

ACCAATGCCA TCAATTGCTT CTTGAATATC AGTCGGACTA ACAATCGTTG CATAAGGTGC 2400

AATCACAATA TTATTATCTT CCAAGAAACT TTTTTCCATT AAGCGATCTT GCGTAATCGC 2460

TAAAATGTCT GTTCCTTGCG GTATATTCGC CAAATGTTGA ATCGCTGTTA ACGCATCAAC 2520

ATCTGCATGT TCAAATTCAT ACGTAATCAC TTGGGCACGT TTTGCCAATT TCTCTAACGC 2580

TTCAGGATCA TTATATGCTC CTAATAATTG CCAATCGGCA ACTTGTGCAG CAGGACAATC 2640

AATTTGTGGA TCTAGCACAC CCACTCGAAA CCCCATTTCT CTTGCACTAA GGGCCATCAT 2700

TTTTGCTAAA TGTCCGCCAC CAATGATACC AATCACTTGA CCAGGTAATA GTGGTTTATC 2760

CAAGTTCATC ACAACTTTCC ATCACAGTTT TATCCTTAAA ATCCTTCTGC GGTTTCTCGT 2820

CGCAGTCTAG CCTTTAAGAG TGGCCCTnTT TACATGaTAC mAGGrAAAAG GcTCAACTTC 2880

AAGTGCCAGC AGTGTTTCCA TTAGGTTTTC TCATCAGAAT TTATTAACAT ACTTTAGGaT 2940

AAAACAGCGG TAACTTACTA CTCGTCTCTT TGTAACCAAG CTTTTTTCTC TTTTTCTCTT 3000

CAAATTGATG AAAACATTAA ATTTTTCTAA ACGATGATGA GGTCAAAAAA GTATGAAAGA 3060

AAAGCCACCT GACCTTATCT ATCATACTTT TTTGAAACTA TTCGTTACTT TCTGAGGTTT 3120

CTTTTAAACC ATTGAATAAA AGATTTAAAA TAATTGCGGT AATGCTGCTC ATGACAATTC 3180

CATTACCAGT AAACATGCTA ACAGTTTCTG GTACTTTATT AAAGAGCGTC GGCATAATAT 3240

TAAAGCCTAA ACCAAAGCCA ATCGAGATGG CAATAATTAA TAAATTTTTG TCATTTGAGT 3300

AGTCAACTTT TGACAACATG CGCATTCCTT GGACAGCCAC CATTCCAAAC ATAACCAGCA 3360

TTCCGCCGCC TAAAACCGGC TCTGGGATAA TTTGGGCCAT CGCACCAATT TTTGGTAATA 3420

AGCCTAAAAT AATTAAGAAA AAAGCAGAAA AATAAATTGG TCGACGTGTT TTGATCCCAG 3480

ATAATTGAAC TAACCCGACA TTTTGAGAGA AACCTGTATA AGGAAACGTA TTAAAAATCC 3540

CCCCTAAAAT GACGGCTAAA CCTTCTGCAT GATACCCACG ACGTAAATCA TCTTCGCCAA 3600

TTTTCTTCCC TGTAATATCT CCTAAGGCAA AATAAACACC TGTTGATTCA ACCaTACTAA 3660

CAATtGaAAT AATAATCATC AACACAATAG AAGAAAGATC AAACGTTGGT TTGCCAAAAT 3720

AAAATGGCTG TgGAAAaTGa aACCACGTAG CTTGACCAAC AGGAGACAAa TCCACTAATC 3780

CCAGAAACGC AGCTAAAATT GTGCCACCGA CTAGGCCGAT TAACACGGCA ATTGACTTGA 3840

TAAAGCCACG TCCCCAAACT TGCACAACAA TGATTAAGGC AATTGTTACA AATGCGAGGA 3900

GCAAGTTTTT AGGATCACCA AAACTTTTGT CCGTTGCTAA GCCGCCACCC ATTTTTTCAA 3960

CAGCAACAGG AATTAAGGTC AAACCAATAA CTGTAATGAC TGTGCCGGTG ACTAAAGGCG 4020

GAAATAGACG TTTAATTTTT GAGAA 4045

(2) INFORMATION FOR SEQ ID NO: 152:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7574 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

nAnnAATATA CTATAAAGTC CAAAGAAAGC GCTTTAATTC GATGGAGAAG GTGGTTTAGT 60

GAAAAGATTT TCATTTTTTT TACTAATTTT ACTTGCTTTA ACAGGTTGTA AATCCGGTGA 120

AAAAGAATTT GATGAAGAAT CTCTTCAAAA TCTAAAGGAA ACGkCACAGT CTTAyTCAGA 180

AACAGAATTA CAAAATGGTG ACGTTCGTTT AAATGAATAT ATTTCTTTGA AAGGGGAGAT 240

TGTTGAGAGT GACAGTCGTT CCAGTTTAAT AAAAAAAGGT GATCGTTTTA TTTTGAAAAG 300

TGGTTCTAGT AAATATCAAG TTTyTAATGA GCAAAAGAAA AAATTGAAGA TTGGTGACGA 360

AGTGACAGTT TACGGAGAAT ATTACGGCTT TTTGAAAGGG ACATTAATTG AAAGTGAGGA 420

GAATCATGAT TCAGCCACGA ATTAGACAGA ATCCTTTTGA AACATACGAA GAAGTTGCAG 480

AAGGATTTGA AGATTTATTA GCGCCACTGG AATTGTTTTT TGATCGAGAG TATCAAGGTC 540

ATCTGGATTT AGGGACACAC GGAACGGTTT ATTCTAAAGG GACACGTGAT GCGGAGGCAT 600
TTTTACGACC ATTATGGGGA CTAGGTCCTT ATGTAACGCA AAATGAAAGT GAGTATTTGA 660
ATGACTTTCT AACGGGAATT ATCGAAGGAA CCGATCCTGA GAGTTCATCT TATTGGGGaA 720
AAACAAAAGA TTATGATCAA TTGATTGTTG AGATGGcAGC ATTAAGTACA TTTCTTTTGt 780
TAAACAAAGA AAAAACGTGG GACCAATTAa CAAAAGAACA GCAAAACAAT TTACACAGTT 840
GGCTGATACA AGCCAATGAA AACATTATTC CGCCAAATAA TTGGCATTTC TTTAGAGTGT 900
TGCTGAtTac GtGAAcATgT GAAgCCGTAT cnAAAGAaCA GAtCGCTGTT GATTTGGCTG 960
TAATnGATCG TTTTTAtGtC GGTAATGGTt GGTATTATGA CGGAGTCGAG ACACAAGTTG 1020
ATTATTATGT TTCTTtCGCT ATTCATTATT ATAGTTTATT GTATTGTCGT TTTGCCCCGG 1080
AAGATACCGC AAGAGTGGCT ATTATGAAGG AACGAGCAAC TTTGTTCGCG CAAGAGTTTA 1140
AATATTGGTT TACTCAGCCA GGAGAAGCGA TTCCCTTTGG ACGTAGCTTA ACGTATCGTT 1200
TTGCTCAAGT GTCATTTTTT AGTGCTTTAG TTTTTGCAGA TGTTGAAGCA TTGCCGTGGG 1260
GTGAGATTAA AGGTCTAATC AGCCGACATC TTCATCAATG GATGAACAAA GATATTTTTA 1320
CAACAGATGG ATTATTGTCG GTAGGTTATG ATTATCAAAA TATGGTCTTT GCCGAAGnTA 1380
TaATGGTCCT GGTTCACCTT ATTGGGCCTT tAAAACATTC ATTTtGTTAG CGGTTCCAAA 1440
AGATCATCCT TATTGGCAgC AGAAaCACaG CCtATTAGCT TtCCaGAAAA ACACTTACCT 1500
cACCgAAAGT CGGAACTACT ATCAaGTAAA TGATGCAGGG ACACATGGTT TAATGTTTCC 1560
AGCGGGACAA TTTATTAATT ATCAAGCTCa TGCGCATGAT AAATATAGTA AATTTGTTTA 1620
TTCTTCCCAT TTTGGTTTCA GTACTATAAA GAGTGATTAT TGGTATTACG AAGGTGCTTA 1680
TGATAATTGC TTGGCTTTAG CTGAAGACGA CCACTATTTC CGCACGAAAG GATTAGATGA 1740
TCAGTACGAA ATTCTTGATG ATCGTATCAT TCATCAATGG CATCCTTGGT CCGATGTAGC 1800
AATCAAAACG ACAATTGTAC CATTAGAAGG CCAACATTTA CGGATCCACG AAATTGAGAC 1860
ACAACGTGCA TTAGTGGCTT ATGAAGGTGG CTTCTCAATC CCGCTCTTCG ATGAAAAGGT 1920
CACTTGTGTT TCTGATCAAA TGGCTGAGGT AAAAAATGCA AAAGGTGTGT CAAAAGTAGA 1980
AAATATCAaC GGCTTTTCTG AAGCyGCGAT TATTCGAACA GAACCGAATA CAAATTTACT 2040
TTATCCGTTA ACGGAGTTGC CATATTTGAA AGCAAACTTG TCGAAAGGGA AACATCTTTT 2100
AATTTCATTG GTGACAGGCG TACTACCGAA TGAACAAATT GAGCCAGTTA AAGTTAGGTT 2160
AAAAGAAAAT CAGCTTTTAG TTGAAGAAAA AGTTGTAATT TTGGGGAATT AGGGGGAGAA 2220
ATACACAATG ACAGAAAATG GACAGATTGA GTGGTTGCCG AAGCAAATCA GTTTTGCTTT 2280
AAACAAAGTT AGAGGAAATT TAGAGACATT TAAAGAATTA GTTCCACCCG CGGCAAGTTT 2340
AAATCAAATT TATTATCCTG AAGAAAaTAT TGACTGGACG GCCAGCTTTT GGCCAGGAAT 2400
GTTATTTTTA GCAAAGGAAT TAACAAATTC AACAGAGTTC GATGAGGTGA TTGCCACACA 2460
GATGGCTTCT TTTCAACATC GCTTGGATGA ACAAATTGAG CTGGAAACAC ATGATATAGG 2520

CTTTTTGTAT ATTTTAACAG CCATTGCTGA TTATCAAGTC AATGGTCATG AGGCCTCAAA 2580

AGAAATGGCC TTGCAAGCTG CGGACTTATT AATGAAACGT TATTCACCAA AAGCCAAGAT 2640

TATTCAAGCT TGGGGGAATT TAGACGATCC AGAGCAACGA GGTAGAATGA TTATTGATTG 2700

TTTAATGAAC TTACCATTAT TATACTTTGC AGCAAAAATG ACTGGCAAGC AAGAGTATTA 2760

TGAAGCTGCC TACAATCATG CGAAACAAAC ACAAAAATAT ATTGTGCGAG AAAATCACAC 2820

TACGTTCCAT ACGTATTAATT TTGATACGGA AACAGGCGAA CTTGTACGGA AAGACACAGC 2880

AAGGATACTC AGATGATTCT TGTTGGGCTC GCGGCCAAGC CTGGGGAATT TACGGCTTTA 2940

CGTTAAGTTA TTTATATACA GGAGATAGTA GTTTTCTGGA AACAGCAAAA AATGTTGCTG 3000

ATTACTTTAT TCmAGAATTA CCAGAGGrCa AAATCTGTTA CTGGGATTTA ATTTTCAATG 3060

AAGGGAGTGA AGAAGAACGT GATAGCTCAT CTGCTGCAAT TGCCGCATGT GGCTTACTAG 3120

AATTATCTCG TCAGCTTCCA TTAAATGATG AAAAACATGG CTATTATGAA AAGGTTGCGC 3180

TAGAACTATT ACAAGCATTA GCGGAAAAGT ATACAACGGT TTTACAACCA GAGTCGAACG 3240

GCCTATTGCT TCATGGTGTC TATGATAAAA AAACAAATAC AGGAGTCGAT GAATGCATGA 3300

TTTGGGGCGA CTACTTTTAT TTGGAAGCGT TAACTCGTTT AGCGAAAAGC TGGTATTCTT 3360

TTTGGTAATT CATGGAGGAA AACTCCTAAA AAATAATAAA TAGCGGAGGT AGTACAAATG 3420

AAATTTTGGA AAAAAGGCTT AACAGCGGCA GCGCTGTTAG CAGTGGCGGC AGTAACTTTA 3480

ACAGCATGTG GTGGTTCAAG TGAAAAGAAA GCAACTGAAA AGAGTGAAGA TGGCAAAACA 3540

AAATTAACAG TAACTACTTG GAATTATGAC ACGACCCCAG AATTTGAGAA ATTATTCAGA 3600

GCTTTTGAAG CGGAAAATCC TGATATCACT ATTGAACCGG TGGACATTGC TTCAGATGAT 3660

TATGACACAA AAGTAACAAC GATGCTTTCA TCAGGAGATA CGACGGATAT TTTAACCATG 3720

AAAAACTTAC TTTCATATTC TAATTAcgCG CTACGCAATC AATTGGTGGA TTTAACCGAT 3780

CACGTTAAAG ATTTAGATAT CGAACCTGCC AAAGCAAGTT ACGAGATGTA TGAAATCGAT 3840

GGTAAAACCT ATGCTCAGCC TTACCGTACA GATTTCTGGG TATTGTATTA CAATAAAAAA 3900

ATGTTTGATG AAGCCGGAAT TGCCTATCCC GATAACTTAA CTTGGGATGA ATATGAAGCG 3960

TTAgCGAAAA AATTATCTAA ACCAGAAGAA CAAGTATATG GTGCCTATCA ACATACTTGG 4020

CGCTCAACCG TTCAAGCGAT TGCTGCTGCT CAAAACAATG CCAATTTGAT TGAACCAAAA 4080

TACAATTATA TGGAAACTTA TTATGATCGC GCATTGAGAA TGCAAAAAGA TCAATCACAA 4140

ATGGATTTTG GAACAGCAAA ATCAACAAAA GTAACGTATC AATCACAATT TGAAAATTCA 4200

AAAGCGGCGA TGATGTACAT GGGTAGCTGG TACATGGGGA CTTTATTAAC AAACATTGAT 4260

GATGGCAAAA CAAATGTCGA ATGGGGGATT GCCGAAATAC CACAACAAGA AAAAGGCAAA 4320

GCAACTACCT TTGGCTCACC GACAAGTTTT GCAATTAATA AAAACAGTAA AAAACAAAAA 4380

GCTGCTCAAA AATTCTTAGA CTTTGCTTCA GGTAAAGAAG GTGCAAAACT TTTAGCAGAA 4440

GTAGGGGTGG TTCCTTCTTA TAAAACAGAT GAAATTGATA AAATCTACTT TGCAAGAAAA 4500

GGAaTGCCTT CAGACGAGTC TCACAAAAAG CCTTTAACCC AGATACAATT AATTTAGAAT 4560

TCCCATCTGA TAAAAACGGT GCCGCAATTG ATAAAGTATT ACAAGAmGmA CATGATTTAA 4620

TTATGGTCGG CGmCGAAAAA CCAAAAGATG GTATTGCTAA CATGGAAAAA CGTGTCAAAG 4680

CAGAGATAGA CTAATTATTA TTTTCTGCAA AAACACAAAA GATTGATGAA AAAAATAAGC 4740

TGAAATGAGT AACCTACAAT GAATTTCAGC TTATTTTTTT AAATAAAGAA TGTAGCTTGA 4800

GTTTGGAAGG AGGGCCTGTT AAGTTCACTT AACAGAATAC AATTATGGAA ATTACTTCAA 4860

AAACAAAAAC AGCAAATAAA TTGAGACGAA AAAATACGTG GACTGCGTTA TCTTTTATCG 4920

CACCAAACTT TATTGGCTTT TTCCTCTTCA CTTTAATCCC GGTTATCTGT TCACTGATTT 4980

TAGCTTTTAT GAGTTGGGAT TCTTTTTCAA CCCCTGAATT TGTcGGAATG AAAAATTTCA 5040

CCAAAATGAT TCATGATGAT ACGTTTTGGA TTTCGTTAAA ACAAACGTTT ATCTATACGA 5100

TTGGTGTTGT TCCGTTAACA TTGATTTGCT CATTAGGACT AGCGATTTTG TTGAATCGTA 5160

AAATTAGAGG GATGAAATTT TTTAGAACGG CTTTTTTCTT TCCTTATGTA ACGTCATTAG 5220

TTGCGATTGC GGTCGTTTGG AATATGCTAT TTCATCCAAC GATGGGACCA ATTAATCAAT 5280

TTTTGAAATT ATTTATTGAA AATCCTCCTG GCTGGACATC TAGTTCTACT TGGGCTTTAC 5340

CAGCAATTAT CATCGTTAGT GTGTGGCGCT TTATGGGGTA TTATATGATT TTGTATTTAG 5400

CTGGTTTACA AAGTGTGCCA AGAGAATTAT ATGAAGCAGC GTCAATGGAT GGCGCGGGTA 5460

AATGGAATCA ATTTTTAAAT GTTACCTTAC CTTCACTACG ACCAACGACC TTTTTCGTGA 5520

CTATCATGCT TGTAATTAAC TGTTTCAAAG TGTTTGACTT AGTTCAAGTT ATGACTGGTG 5580

GTGGACCAGG ACGTGCAACG AATGTGTTAG TTTACGAAAT TTACAATGAA GCCTTTGTCA 5640

AATTTAATTT TGGTTATGCA TCAGCCATTG CGATGGTCTT ATTCATCATC GTTCTAGCGA 5700

TTACTGTGGC CCAATTTAAA TGGAATCAAC ATCAAGAGAA AGTTTAGGAG GAGAAAAGAT 5760

GGAAGCAACT TTTTCAAAAA AAGAAAATAC TCGAAAAGCA AAAACGaAAA AAGCAGAAAA 5820

AGTCACACCT GGTGGGATCC TTGTAATCAT TCTCTTAACT TTACTTGCGT TAATTACTTt 5880

GGTGCCTTTC ATTTGGATGG TTTCAGCATC ATTTAAAnCn AACAATGAAG TTTTCACCAT 5940

TCCAATTCAA TGGATTCCTA AAAGTTGGCA TCCAGAAAAC TATTCCGTTA TTTGGGAACG 6000

CATTCCGTTG TTAACTTTCT TTAAAAATAC ATTATTTTTA AGTATTGTGA TTACGATTAT 6060

TCAACTATTT ACTTCAAGTT TTGCAGCATA TGGCTTTTCA AAAATGAATT TTAGAGGAAG 6120

AGATACACTT TTTATTACGT ACATCGCAAC AATTGCTATT CCTTGGCAAT CGTATATGAT 6180

TCCACAATTT ATTATGATGC GTCAATTAGG TTTGACTGaC ACGTTATGGT CACTTGTTTT 6240

aTTACAAGCA TTTAATGCAT TTGGGGTTTT TTTACTGAAA CAATATTACA GCAGTATTCC 6300

AGATTCACTA TGTGAAAGCG CGCGGATTGA TGGGTTAAGT GAGTGGGGAA TTTATTGGAA 6360

AATTATTTTA CCATTAACTA AACCAGCTTT AGCTaGTCTC ACGATTATTA CTTTkGTGAA 6420

TACATGGAAT GACtATATGG GaCCTTTCAT TTATTTATCT TCTACTGAGA ATAAAACGAT 6480

TCAGTTAGGT TTGAAAATGT TCGTCGGTTT ATTTGATGCT GAATATGCCT TAATCATGGC 6540

CGCATCAGTG GTTTCTATTT TACCAGTAGG GATTGTCTTT TTAGCAATGC AAAAATACTT 6600

TGTTGAAGGT ATTGCTACAT CTGGAATGAA GAATTAGGAG GAAGTTTGAA TGTTTACTAA 6660

GCAAACCTTT GATAATAATA TCTATATGAA AATTTTTCGC TGGCTTTACA TTTTATTTAT 6720

CGGTAATCTA GGGCTTCTCT TAGTGAACAT TCCTTTTTTC ATAGCGGTGA TTAGCTTAGA 6780

CATTGATCCA CGAAACCTCC CACTTTTTGT AGTAACGTTG TTGCCGATGG GGACAGGGAT 6840

GATTGCACTT TTAGGTTTAA TCGATACTTT TAAAGAAGAA AAGGAGCTGG ATCCATTTAA 6900

AACTTTTTTC CAAAAGTTTC GTCAGTTTGG GTTAAGAGGC TTTCTTATCA GCTTGCTTGG 6960

ATTAGGAAGT AGCATCATCA GTGTGACAGA TATATTCTTT TTTGCCAAAA CAACGATTGG 7020

TAAATGGTTT ATTCCATTGA TGGTCCTTCT GCTAATTTTT GGGTTAGCTA TTATGCTGAA 7080

TGCTTGGTAT TTTCAAGTAC GAAACCCGCA AGCTTCTTTT AAAGATGTGT TTCGGATTTC 7140

TGTTTACTAT GGACTGCGGA AATGGTATGT CAGCTGTCTA AATGTTTTTT TACTATTCTC 7200

AATGTTTGCA ATGATGTTTT TGAAGCCACA GTTTGGGTTT GTTGTAACCC CTGTCCTGTT 7260

CCTAGGGATT ATTTATTTAA ATACTGGGAA ATTGCATGAA AAGCAAAAAA AGAATAAATA 7320

ATTTTTACAT AATAAGTAAT CTCGATTTTC AGGAAAATGA ATGAAGGTGG GGTTGCTTAT 7380

TTTGTTCTTG CCTTGTTTAT TAAAGTGGAG GGAGCGAACG AATGAAGAAA AAAATAATTA 7440

TTTTAAGTAG CTTTTTGATA GGTGCTATTT TTGTGCTCGT TATTTTTACT ACCAGCCAGC 7500

AAAGTATGAT TAATAACgAa GCmATCCAAG TCgArGAATT AATTGTTwAT GGCGATAAAA 7560

ATATGAATCG TATA 7574

(2) INFORMATION FOR SEQ ID NO: 153:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1229 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

TGTnTAATGT TCGTGTTTTA TnTTTTAATA ACCTTTTTCG AAAAAATATA CGAACGTTCG 60

GGTTTAAAAA CGAACAAAAA GATGAAAATG TGTTGAAATA TACGTTTTCC TTGCGTATAC 120

TGATGAAGTG AAAAAACGAA CGAGGAGTGT ACGAATGAAC AAGAAAGTTC CTTATTTAAT 180

TGTGGCACCG GGGCTAGTAC TATTATTATT TTTCTTAATG ATTCCATTAA TTACTAGCAT 240

TTTGCCAACC ATTTTTACTG ATCATGGCTT GACACTAAAT CAGTATGTTA CTTTTTTTAA 300

AGATGACTAT AATGTTTCAA TTTTCTGGCG AACGATAAGA GTTTCGTTAA TTGTGACGGG 360

TATTTCAATC GTGTTGGGCA TTCCAACGGC ATACTTTATT GCTGGAGTTT CCAAAAAATG 420

GCGTGGCTTT TTAATGGCTA TGACCTTGTT TCCATTATTA ACCAATTCAG TTATTCGAAG 480

TTTTGCTTGG ATTAACATTT TAGGAAAAAA CGGTGTAGTG AATACGTTGT TATTAAAAAC 540

AGGCTTAATT GAGCAACCGC TGAATTTATT GTATACAGAA TTTGCAATTA TTATCGGTTC 600

AGTGTATTTA TTTTTACCAA CAATGATTAT GACCTTAGTT GGGGTTATGG AAAACATTGA 660

AGGAGAAATG CTAGAAGCGG CTGAGACGCT AGGCGCCAAT CCAATGACGG CTTTTCGGAA 720

AATTGTTTTA CCTTTATCAA TTCCAGGCAC CATTGTAGGA AGTATATTAG TGTTTACAGG 780

AACTTTAACG GCTTATACGA CGCCACAGCT GTTAGGTGGT AACCAAAAAA TGATGATGTC 840

GACCTTCTTA TATCAAAAGG CTAATACTTT AGGTGATTGG CAATCGGCGA GTGTTCTTGC 900

ATTTATTATG ATTTTAACGA CATTAATTGT GATGAAAGGC CTAGACATGG TCGCTAAGAA 960

AGTAGATCGG AGAGAAGCAA ATsATGCGTA AsAAAAAGGA CTATCTGTTA TTGCAATTCT 1020

AGTTTTTsTG TTTTtATTTT TACCGTTATG TTTGATTGTG GTGcnTCTTT tGGAaCAGCC 1080

GCAGCGATTC AATTTCCGAT TAAAGGATTA CTTTAGATTG GTACGCCAAA GCATTACAAT 1140

CGGAAACGTT TATGGACAGT TTAAGTTAAG CTTGTTGAAT GGGGTATAGC GACTGTTTTA 1200

GCCATTAATA GTAGGGATCC GGCCTCCnA 1229

(2) INFORMATION FOR SEQ ID NO: 154:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9749 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

CTTGGTTATC TGGTATGGCC ATCCTCTGTG TGAGCAAAAA TTATGGAGAT CTATTnGTAA 60

TATTATTTGT GTGACTCCGT ACCCTCTCCT TGCTGTGGCT GATAAGGGCA TCTATGTGGT 120

TGAAGCACGA AAGrAGGgtT mCGTGTATTg rTaATCCTGT AGcgTATTaA CaATwACCCa 180

cATTTGaTTa CCaATTATTT AATTTAAACA ATTATCGTGT ACTTTCAACT AGAACTCCaA 240

AAAATAATTT TTCAGATCAA aTAGAGTTAG ATATTTATAG TAGAGGAATG GGTGGTATTG 300

GGTTGCCAGG AGATTTATCA TCAGTATCTA GATTTGTAAA AGCAACTTTT ACTAAGTTAA 360

ATTCTGTATC AAGAAGTTCA GAATATGAAA GTATTAGCCA ATTTTTTCAT ATTTTAAGTT 420

CTGTCGAACA ACAAAAAGGA TTGTGTGATG TTGGTGATGA AAAATATGAG TATACGATTT 480

ATTCTTCATG TTGTAACCTG GAAAAGGGAA TTTATTACTA TCGTACGTAT GACAATAGTC 540

AAATTACTGC TGTGGATATG AATAAGGAAA ATTTAGAGAA GGATAGCTTA ATTGTTTATC 600

CAATGGTGGA AACACAACAA ATTAACTATG CTAATTAATT TGTTGTAATT CTTGTTTTTA 660

TAAGTATTTT CGAAATGAGT TTAAATAAAC GGGaATAATA GAaGAGTAGA TTTaAGTTAT 720

TGATAACTAC TAGGTAATTG ATGATAAATT AAmCAAATTA GATGCTTTAA GGAAAGATGA 780

CAAAAAAGTA CAGTATTACA TGAAAAAGTG GATTTTTGAG AAAAAAAGGG TTTAACAATT 840

```
AATATCAAAA ACAACTATTC TTATAAAACG AATAGTTGTT TTTTTGTATT CATATTTTCT     900
ATAAAAAAAT GAATAGTAAA CACTTGTAAT TTTTTAATAA GAAgAATAGC TTGTTATAAG     960
AGTAACTACG ATTGTCCGCA TAATTTTTTt GAAGgTGgTA AAAAAAAtAT AGrGkTATAA    1020
ATGtGaTAGG kGtATTAGtA tTTATTTAaa AGrCtAAAAA AtGtGaAGGA GkTTGaATCg    1080
AAaTGaAAAA AGTATGtGCT ATTTTAATGA TTTCTGCGTC AGTTTTAGGA ATGGTGCCCA    1140
GTAATGTACC AATTTTTACA GCAGAAATAA ATTATGCGCA TGAAAATGAA TTACAAATGA    1200
ATCCGTTTGT TATAGCGTTT GTCGATAAAG TGAAGGGTTC TGTTGCAAAG TTTGAAAATC    1260
AAACGCGCCC TCTCTGAACT CCAAATATCT TAGGCTGGTA TTCCCATTAA CCTTGATTTC    1320
AGTAGACACC GAAAATCCGA AGAGCGTTCC ATTTCTTCGG TTCTTTTTAT ATATTCCTCG    1380
AATGGTCTCC ATGCCCTTAA TCGTGGAAGA GGCTGTACGG AGACTTTGAT AAAATTTATT    1440
CCGTCGTTTA ATAGGTCGAT GGTCTTGTTC TATTAAATTG TTAAGATACT TCACAGTTCG    1500
GTGCTCTGTC TTAGTATATA AACCCACACT CTGTAACTTT CTAAAGGCGG AGCAAGaGAA    1560
GGTGCTTTAT CGGTCACAAT TGCTTGCGTA TCCCGTTTCT TTCGTAACCA GATATCTAAG    1620
GTTAAGCCGT CCGCATCAAT TGCACGATAA AGATAATGCC AACGTCCCTT AATTTTGATA    1680
TAGGTTTCGT CCATTTTCCA TGAATAGAAG GATTGTCTAT TTTTCTTCTT CCAAAGATCA    1740
TAGAGGACTT TGCTGTACTC TTGCACCCAA CGATAAATCG TAGTATGACA AACATTTATT    1800
CCACGATCAT ATAACAATTC CTGaACTTCA CGATAGCTTA GATTGTAACG CaGGTAGTAA    1860
CCAACAGCGA CAATAATGAC GTCTTTTTTG AATTGTTTGC CTTTAAAATG ATTCATTACT    1920
CTGTCCTCTC TGTCTTTTTT CTCAATTTTA CACTAAAATA GATTTTTTGG AAAACTTTGC    1980
AACAGAACCT TTTTTTGCAA TAGAATTACT AATGGGAAAA CAAATTAATG CGATAACCAG    2040
ACCGATAGCA ATTTCGTGAA AAATAATTGT TATTAAGAAT AAAACTATAC TAGATATGAA    2100
TAAACTTAAA ACATGTATTT TATCTTCTTT CATAGGATCC TCCCGTTATT GTCATACAAT    2160
AGTTGTTTGA TTTATATTTT AAATTATCAA TAAAAGAAAA GCTAGGAAAA AAGCATTCAA    2220
ACGAAAACCC AGTTAATACA GATACAGGCG TTGTGATTGT CGCAGTTGAA GATAGTAAAC    2280
CGATTATTCA ATTAGCAGAC GGAACAACAA AGAAAGTTGA AGCTAAAGAA ATTGGAGCAG    2340
ACGTTCAAAA AGACGGAACA GTAACGGTTA AAGGTTCAGA CGGAAGAATG AAAGTTTTAC    2400
CAAAAACAGG AGAAACAGAA AATATTGCAT TGTCTGTCCT AGGTTCATTA ATGGTATTAG    2460
GATCAGCATT CATTTTTAAA AAGArAaTCT AATTAAGTAT TTATAAAAGC ATGGTCGCAA    2520
GTTTTACACT CTCCTACTTG AAGTtTCCGA ATGAAAACTC TATAATGAGA AATGTGAAAA    2580
TGTTCACTTA TTATAAATTA AGTtTGGAGG CTTTTATTTA TGAAAATTGC AATTGCAGGG    2640
GCCGGCGCGA TGGGTTCCCG GTTCGGACTG ATGTTGAAAC AAGGTGGTAA CGATGTTTTG    2700
CTAATCGACG GCTGGCAGGA GCATATCAAT GCCATCAAAG AAAACGGATT GAAGGCTAAT    2760
TATAATGGCG AAGAAATTAC TGTTAAAGTC TCAATTGTTA ATCAGAATGA GGTGCCCACT    2820
```

GGTGAGCAAT TTGATTTGAT TATCTTATTC ACGAAGGCAA TGCAGTTGGA AAAAATGCTG 2880

CAGGATGTTA AACCATTAAT TGCTGATCAC ACAGAAGTTT TGTGTCTTTT AAACGGTATC 2940

GGTCATGAAG ACGTAATCGA AAAATTCGTA CCGATGGAAA AAATCTTTAT CGGCAACACC 3000

ATGTGGACTG CTGGTCTAGA AGGTCCCGGC AAGGCTAAAT TATTTGGCAG CGGGTCTGTT 3060

GAATTGCAAA ATCTAGGTAT TGGCCAAGAA GAATCAGCGA AAAAATTGGC GGAAACTTTG 3120

TCCGCGTCTG GTTTAAATGC CAAATATTCC GACAACATTC ATTATTCAAT TTATCGTAAA 3180

GCTTGTGTCA ATGGTACGAT GAATGGCTTA TGCACGATTT TAGACGTCAA TATGGCGGGG 3240

CTTGGTGCAA CCAAACCTGC TCATGACATG GTAGTGACTA TTGTTAACGA ATTTGCAGCT 3300

GTTGCTGCCA AAGAAAACGT CAACTTAGAC ATTCCCGAAG TAATCGAACA TGTAGAAACT 3360

TGTTTTGACC CGACTACGAT TGGTATGCAT TTCCCATCGA TGCACCAAGA CTTGATAAAA 3420

AATAACCGTT TGACCGAGAT TGACTACATC AACGGAGCCA TTTCTCGTAA AGGTAAAAAA 3480

TATGGAGTAG TAACTCCTTA CTGTGATTTC TTAACGCAAC TGGTTCATAG TAAAGAAGAA 3540

ATTCTAGGAG CGAAATAGTA ATCATTAGAA AGAGGATGAG AAAATGGAAG AAAAATTAAC 3600

TCCAAAAATA TTTTTAAATA AAGTATTAGC GGGTACCGCA ACGGGGATCA TCGTAGGGTT 3660

GATTCCTAAC GCGGTATTGG CAGCAATTTT GAAATTATTT GGTGAAAATA CCTTAGCAGT 3720

AACGATCACA CAAATGGCAG TTATTTTCCA ACTAGCTACT CCACTATTGA TTGGCGCTTT 3780

GATTGCGGTG CAGTTTGGCT TCAAACCAAT GCAAATGTTA GTAGTTGCCG GTGCAGCTTT 3840

CGTTGGCTCC GGTGTTGTGA AATTCAATCC TGACACAGCT ACTTATATCG GTGCTGGTAC 3900

CGGAGACATC ATCAATACCA TGATTACTGC ATCAATCGCT GTCGGCATGA TCTTGTTGAT 3960

CGGCGAAAAA TTTGGGTCCG TGGCAATTGT TGCAACACCA ATCGTTGTTG GTATCGGCGC 4020

TGGTCTAATC GGCTATTATT TATATCCTTA CGTAACCAAG ATTACTGCCG CTATCGGTGA 4080

CTTAATTAAT ACCTTTACTA CTTTGCAACC AATTTTGATG TCAATCTTGA TTGCATGTTC 4140

CTTTGCTTTC TTGATTATCT CGCCGATCTC TACAGTTGCC ATTGGTATGG CAATTCAATT 4200

GAACGGTGTT TCTGCTGGTG CAGCCGCTAT GGGTGTTGCC GCAACAACAG TTGTCTTGGT 4260

AGTTAACTCT TGGAAAGTCA ACAAACCAGG GGTTACTTTG GCCATCGCAT TGGGTGCTAT 4320

GAAAATGATG ATGCCAAACT TGTTTAGGAA ACCAATTATC TTGGTGCCAT GTCTATTCAC 4380

CGCTATCATC TCAGCAATTC CAGTAGCGTT ATTCTCAGTC TCAGGGACTC CTGCTTCTGC 4440

TGGTTTCGGT TTGGTAGGGT TGGTAGGTCC GTTGGCTTCA TTGGACGCCG GTTTGAGTAT 4500

TATTCTACTA TTAATCAGCT GGTTTGTTGT ACCAATCGTA GCTGCTTTCG TAGGTCAAAT 4560

CTTATTTGAA AAAATCTTGA AATTGTATGA TCGTAAAGAT GTTTTCGAAT TTTTAGGATA 4620

GTCAACTAGA AAAAGTGTGA GCCTTGTGCT CACGCTTTTT TGTGCAATTG AAGTTGGTGA 4680

ATCTCAGCGA AAAGAGAAAG GTACCTTAAA TAGTGTGTTT CTGGTCAATA GCACAATGAA 4740

ATATCTAAAT ACATTATGAA CTGCTTTTCG TTTTATGAAG AGCAGTTTGT GCTATTTGAG 4800

AAATTGCTCA ATTTAATCAT TAAGTAGTTT CATCAATGGT GACTATGCTA CTGCAATTCT 4860

TGTTGTACCA ACCGGGAGCG TCAATAATTT TGTGTAAATA ACTCGTCCTT CTGCAAAATA 4920

ATGGGTTACT CAGTAAACAT TGAAGCTAAT GTATCCGTCA CTTGTTGGAA GCCTTTATGG 4980

CTTCTATTTA AGAATTTTTG ATTGTATGTA TCAAAGATGG ATACTAGGAA ACGCTCTAGT 5040

GATTCTTCAT TTTGAAACTG CTCTTTTCTG CGGCTGTATT TCTTGATTTG TTTATTGAAA 5100

GATTCGATTA AATTAGTTGA GTAAATGCTT CGGCGGATAC CAGGCGGAAA CTCATAAAAA 5160

GTCAATAAAT CTTGATTTTT TAtyCAnkGC TGCGTCACTT TAGGATAAGT TTTCTGCCAC 5220

TTCTCAATCA TACTCCCTAA AAAGGTATTT GCCTCTTCCT TTGAGCTAGC TTGATAAACA 5280

GCCTTAAAAT CATCACAGAT TTCTTTTCGA TCTTTGaCAC GTACTTTATG AGCAATATTA 5340

CGGTAAACGT GAAcACAGCA ATGTTGGTAT TTAGCtTTAG GATAAATCTG GTGAATCGTA 5400

TCTTTCATGC CTTTTAAGCC ATCGGTAATA AACAGTAAGA CATCATGAAC CCCTCTGGAG 5460

TAAATGTCCT GTAACAACTC ATTCCAAACG TAGGTCGATT CAGTCGGAGC AATCGCATAA 5520

CTTAGTACTT CTTTAGTGCC GTCTTCTCGT ATACCAATGG CAATATAGAT TGCTTCTTTG 5580

GAGACGGTTT GCCGTTTTAA AGAAATATAA GTAGCATCCA TGAAAATAGC GACATACTTA 5640

TCATTTAGAG TTCTAACTTT AAATGCATTT ACTTCTTCAG TCTGCGATTT CCGACATGGT 5700

AATACCTTTT CGGAAGAGGT GAATGACGGT CTCCTCTAAG GTATCATTCG TCCGTCTATA 5760

AGCAGGAACA GTCTGTTGCT TGAACTCGCC GTTGCGGTCG CGCGGAATTC TGTGAATGTG 5820

TTGTGGTAAC GGTTCTGTTG CAAAGTTTTA AATCTACTAT CAAATAAGGT AGAATAATAG 5880

AAAAAGATAG CAGGAGGAAT GACGGTGAAT CATTTTAAAG GAAAGCAATT TCAGTAGGAT 5940

GTGATTATTG TAGCCGTGGG CTACTATCTT CGTTATAACC TTAGCTATCG TGAAGTTCAA 6000

GAAATCTTAT ATGATCGTGG CATTAACGTT TCTCATACGA CGATTTATCG TTGGGTGCAA 6060

GAATATGGCA AACTACTCTA TCAAATTTGG AAAAAGAAAA ATAAAmAATC CTTTTATTCA 6120

TGGAAAATGG ATGAAACGTA CATCAAAATT AAAGGaAAAT GGCATTATTT GTATCGAGCC 6180

ATCGATGCAG ATGGTTTaAC CTTGGaTATT TGGTTACGTA AAAAACGGGA CACACAAGCA 6240

GCCTATGCTT TTCTTAAGCG GTTAGTGAAG CAGTTTGATG AACCGAAGtT GtAGTCACAG 6300

ATAAAGCCCC CTCTATTACA AGTGCCTTTA AGAAACTAAA AGaATACGGC TTTTATCAAG 6360

GGACAGAACA TCGTACCATT AAATACCTGA ATAATTTGAT TGAACAAGAC CATCGTCCAG 6420

TAAAGAGACG CAATAAATTC TATCGAAGTT TACGCACTGC CTCTACCACG ATTAAAGGCA 6480

TGGAAGCCAT TCGAGGATTA TATAAGAAAA CCCGAAAAGA AGGCACTCTC TTCGGGTTTT 6540

CGGTCTGTAC TGAAATCAAG GTATTATTGG GAATCCCAGC TTAAATCATA GATACCGTAA 6600

GGgATTTTAT TCTTTATTTA AAACTTTGCA ACAGAACCAA TAATTTGATC TACTCATTTC 6660

TAATTCTTGG GAGAATTTTT CTTTTGATTT GGTACCTCTC ATCTTTTTTA AGCTTTCTCT 6720

AAATGCTAGT TCCAAGTTTC TCCCTCCTGT TTATGTAAAT ATTTTGCACG TTTTCCATTT 6780

AATTATATCG TTACACCTTG TAAATCACAA GTGATTAATC ACAAATCACT TGTGATTAGT 6840

GATTGTTATT GGTGTTGCCT AGAACGATTC TAAGCACATG TTTAATTTTA AAGTGGGCAA 6900

TTATCTTTAA AATGTCCTAC AATGATTCAA TAAAAGCATG GCTGCTAGTG AAACTTGCGA 6960

CCATGCTTTT AAAAATACTT AATGGGGTCC CGAGCGCTTT AGCTCCTTGG AAGCTGTCAG 7020

TAGTATACCy AATAATTTAT CTACATtCCC yTTAGTAACG TGkAACTTTC CAAATTTACa 7080

AAAGCGACTC AtAGmmTtAT TTCCTCCCGT TAAATAATAG ATAACTATTA AAAATAGACA 7140

ATACTTGCTC ATAAGTAACG GTACTTAAAT TGTyyACTTT GGCGTGTTTC ATTGCTTGAT 7200

GAAACTGATT TTTAGTAAAC AGTTGACGAT ATTCTCGATT GACCCATTTT GAAACAAAGT 7260

ACGTATATAG CTTCCAATAT TTATCTGGAA CATCTGTGGT ATGGCGGGTA AGTTTTATTA 7320

AGACACTGTT TACTTTTGGT TTAGGATGAA AGCATTCCGC TGGCAGCTTA AGCAATTGCT 7380

GAATCGAGAC TTGAGTGTGC AAGAGCAACC CTAGTGTTCG GTGAATATCC AAGGTACGCT 7440

TGTAGAATCC TTCTTCAACA ATCAGATAGA TGTCAGACGC ATGGCTTTCA AAAACCACTT 7500

TTTTAATAAT TTGTGTGCTT AAATGGTAAG GAATATTCCC AACAATTTTA TACCTCTGTT 7560

TGTTAGGGAA TTGAAACTGT AGAATATCTT GGTGAATTAA AGTGACACGA ATGTTCAGTT 7620

TTAATTTTTC TGACGATAAG TTGAATAGAT GACTGTCTAA tTCAATAGaC GTTACCTGTT 7680

TACTTATTTT AGCCAGTTTC GTCGTTAAAT GCCCTTTACC TGTTCCAATT TCGTAAACGG 7740

TAwCGGTTTC TTTTAAATTC AATTGTTTTA TTATTTGGTT GAGTACTTTT TCACTCGTTA 7800

AAAAGTTTTG AGAATATTTT ATATTTTTGT TCcATGTAAT CACTCCTGAA GTGATTACAT 7860

CTATAAATAA ATACAGAAGT TAAACGATTT GTTTGTAATT TTAGTTATCT GTTTAAAAAG 7920

TCATAAGATT AGTCACTGGT AGGAATTAAT CTAACGTATT TATTTATCTG CGTAATCACT 7980

GTTTTTAGGG TTCTGTTGCA AAGTTTTAAA TAAAGAATAA AATCCCTTAC GGTATCTATG 8040

ATTTAAGCTG GGATTCCCAA TAATACCTTG ATTTCAGTAC AGACCGAAAA CCCGAAGAGA 8100

GTGCCTTCTT TTCGGGTTTT CTTATATAAT CCTCGAATGG CTTCCATGCC TTTAATCGTG 8160

GTAGAGGCAG TGCGTAAACT TCGATAGAAT TTATTGCGTC TCTTTACTGG ACGATGGTCT 8220

TGTTCAATCA AATTATTCAG GTATTTAATG GTACGATGTT CTGTCCCTTG ATAAAAGCCG 8280

TATTCTTTTA GTTTCTTAAA GGCACTTGTA ATAGAGGGGG CTTTATCTGT GACTACAACn 8340

TTCGGTTCAT CAAACTGCTT CACTAACCGC TTAAGAAAAG CATAGGCTGC TTGTGTGTCC 8400

CGTTTTTTAC GTAACCAAAT ATCCAAGGTT AAACCATCTG CATCGATGGC TCGATACAAA 8460

TAATGCCATT TTCCTTTAAT TTTGATGTAC GTTTCATCCA TTTTCCATGA ATAAAAGGAT 8520

TkTTTATTTT TCTTTTTCCA AATTTGATAG AGTAGTTTGC CATATTCTTG CACCCAACGA 8580

TAAATCGTCG TATGAGAAAC GTTAATGCCA CGATCATATA AGATTTCTTG AACTTCACGA 8640

TAGCTAAGGT TATAACGAAG ATAGTAGCCC ACGGCTACAA TAATCACATC CTGCTGAAAT 8700

TGCTTTCCTT TAAAATGATT CATCGTCATT CCTCCTGCTA TCTTTTTCTA TTATTCTACC 8760

TTATTTGATA GTAGATTTAA AACTTTGCaA CAGAACCGTT CATGGGTATA GCATGAATTG 8820

TTTAACTGTC CTCTAGGATT CCCGTCCTAT GGCAGTCGGT TCTTATTTTG TTAATTGTCT 8880

TGTAAGTTGC CTTTATATCG CACTAAGTTT AAAAAAGTAA AAAGATTTAT CTCTCAGATA 8940

ATTGCTATAT CAAGGTAACT ACGAGATTCT AAACTAATTA TCCTACAAGT GTGACAACAA 9000

ATAAAGTTAT CAATGGAATA ATTTTTTTCA TAGCTCGTTT GCTCATTCTA ATAGCTCTCC 9060

TCTTATTTAA ACTTAACTTA CAAATTAATA ATAGCATAGA TTAAAATAAA AATAAACAAA 9120

AATGCAAAAA TTTGCGCTTT TTCATCAAAA AAATTTATTA CATAATAAAA AAACCAAAAA 9180

ATCATTCCAC GAATGCACTT TTTTGCGCCT CGGAAGGAGT ATTTCAAAAT GTTTCTTTAC 9240

GAACTATTCA AAGAACAGCG AAAACTCTAT AATTTTACTC AAGAAGAATT TTATGAAGGA 9300

ATATTCAAAA AAAGAGCTGC CTCTTCTTTT GAGGTACACA ACACGCATGA TCTCAAAGTT 9360

AAAGATTTAC CTGTTTTATC AGATAGAAGC ATGATGAGTA TCTTAGAAAT AATTCATTAT 9420

GCTAAAGAAG AATTTATTTC TCCtTACGAC GAAGACTTAA ATTCATTGTT TGATATATTC 9480

CAGAATAATA CCGAAAAAGA AAATAAAGAC TACATTTACA AATTATACAA AAAGTCTATA 9540

GAGCTAAAAG AACATTCTAT AATATATTGG AATCTTTACT TAATAATAAA AATCCAATGT 9600

TCTGAATACG ATTCAAGAAT TGTCCCTACA GACTCACAAG ATTTATCTGA GTTAAAAAAA 9660

ATGaTTTTAT CAAaACaAAA GTTTACTCTT TACGACTATA AAATTGTTAC TAATCTATCT 9720

CTTGTATTTT CTTACAAAGA ATTGCAACC 9749

(2) INFORMATION FOR SEQ ID NO: 155:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9021 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

AAAATCAGAT ATCTTATAAn nGnAATCAAG CACGTATTCG CGATAAGGGT AAAAAATTAC 60

GTAAGAATCG GCACAGCGCT TCTTTAAAAA GTAAATTAAA CTTTGACAAC TTTTTCGAAT 120

AATGGTACTG TTAAGAAGTA AATTTGTGAA TAGAGGAGGT CGAAGAACAT GGCAACAAAG 180

aAAGCAGCTC TTGCTTGTTC CGTTTGTGGC TCTCGCAACT ATTCTAAATC CGTTAGTGAA 240

GGTAAACGTG GCGAACGCTT AGAAATCAAT AAATTCTGTA AATATTGTAA TCAATACACG 300

TTACATAAAG AAACGAAATA GATTTAGAGG AGGGTATCAA CATGAAATTT TTCCGCAGCG 360

TTGCTGATGA GATGAAGCAA GTAACTTGGC CAACAAAAAA ACAATTGCGT AAAGATACAT 420

TAGTTGTAAT TGAAACATCT ATTTTATTTG CAGCATTGTT CTTTATCATG GaTACTGTTA 480

TCCAAACGGC ATTTGGcTGG ATTTTAAAAT AATTAATTCT AGTATTTCCC GACGAGGGGT 540

GCTATAATAG TATTCGAGAA AAAGCTTCGG GAAACTGAGC TTTTTTTATT TGCTCGAAAA 600

AAATAAACTA AAACATTTAA AGGAGCCGAA TCAGTAATGG AAACTTTTGA AAGAAATTGG 660

TAcGTATTGC ACACCTATTC AGGCTATGAA AATAAAGTAA AwGCAAACAT TGAATyACGT 720
GCACAAAGCA TGGGGATGGG CGATTATATT TTTCGTGTCG TTGTACCAGA AGAAACAGAA 780
AAAGAAGTCA AAAATGGTAA AGAAAAAGAA ATCGTTCATA AAACTTTCCC tGGTTATGTG 840
TTAGTGGAAA TGATTATGAC CGATGATTCT TGGTATATTG TTCGTAACAC GCCAGGCGTA 900
ACTGGTTTTG TAGGCTCACA TGGTGCTGGT AGTAAACCCG CTCCTTTATT ACAAGAAGAA 960
ATTAACCACA TTTTACGTTC AATTGGCATG AGCACCCGTC AATCAGACTT AGAAGTAGCG 1020
CTAGGCGATA CAGTGAAAAT TATCGAAGGT GCCTTTTCTG GTCTTGAAGG TGTTGTGACA 1080
GAAATCGATG AAGAAAAACA AAAATTAAAA GTAAACATCG ATATGTTTGG TCGTGAAACA 1140
AGTACAGAAT TAGACTTTGA ACAAGTCGAT AACATTGATT AATAAACAAG CTGACAAACT 1200
ATTGTTCAGC TGATTGATAC AGATTGTAGG GTCTGGAAAA GCAGTCAACT GACTTCTTTT 1260
CCAAGCCCTT TTTTAGTTCA TCGCAGAAAG GAtGrAAAAA AATGaACATG CCCAAAAATA 1320
TCnGTtATTT TTCTTTGCTA ATGGGTCTtG tTCTATTATT AAGTGCTTGc CAAATTGGGG 1380
CaACTACGAA GGATGACAAC CAAGCCGCCa CAAAAGAAGC AACTGTTGAG TTAAACCGCA 1440
CAACAACACC AACGCTTTTT TTTCaTGGTT ACGCAGGAAC TAAAAATTCG TTTGGCTCGT 1500
TACTGCATCG CTTGGAGAAA CAAGGTGCCA CAACTCAAGA ATTAGTGCTA CTCGTTAAAC 1560
CTGATGGGAC CGTGGTTAAA GAGCGAGGAG CTTTAAGTGG CAAAGCGACG AATCCCAGTG 1620
TTCAAGTTCT ATTTGAAGAT AATAAAAACA ATGAATGGAA TCAAACAGAA TGGATAAAAA 1680
ACACATTACT CTATTTACAA AAAAATTATC AAGTGAACAA AGCCAATATT GTCGGGCACT 1740
CTATGGGTGG TGTTAGTGGT TTACGTTATT TAGGAACCTA TGGGCAAGAT ACATCGTTAC 1800
CTAAAATTGA AAAATTCGTC AGCATTGGAG CACCTTTCAA TGATTTTATT GATACGAGTC 1860
AACAGCAAAC CATCGAAACG GAACTAGAAA ACGGCCCCAC AGAAAAAAGT AGCCGCTATT 1920
TGGATTATCA AGAGATGATT AATGTTGTTC CAGAAAAACT GCCCATTTTA TTAATTGGTG 1980
GTCAATTAAG TCCAACAGAT TTAAGTGATG GAACGGTGCC GTTATCTAGT GCCTTAGCAG 2040
TCAACGCCTT GCTAAGACAG CGAGGAACTC AAGTCACTAG CCAGATTATT AAAGGAGAAA 2100
ATGCACAACA TAGTCAATTA CATGAAAATC CTGAAGTAGA TCAATTGCTA ATCGAATTTC 2160
TATGGCCGAG TAAAAAaTAG CAaAAAGGCA CATGTTCCAT AAAAAATGGA ACATGTGCCy 2220
TTTTCTATAC TTTGGCTAGC GAAATAGCTT TTTTCAGCTA TAATAAATAA AATTATTTAT 2280
TTTTATGAGA TATTGGAATG GAATCCTCAT GATTTAAATA AGCACAAATC AAGGCTTCAA 2340
AGGTGATTAA TGTTTGGCCA AAAAACATTT GTGGCTGAGC GACACTTGTT GAAAACGGGC 2400
TGTAAGAGTC AGTTGAAAAA ACGAGATCTG CCAGTCGACC AAGAGTTGAA TTTTTGGCGC 2460
CAACAAAAGA AATGATAGGA TGTTTTTTCT CTTTCGCAAG CAGAGCTGTT TCTTTTAAGT 2520
ATTTTTCTTC ACCAGAATGT GAAAGGATGA TAAAAAGAAA ATTCTGATTA TTTTGACTAT 2580
TTATCGAATG AGTATAGGCT GTTGTGATGC TGGGGATGCC ATGAAAATTG AAAACTTCAC 2640

TAATGTAAGC CGCCAAATGT CGTGAAAAGT CATGACCTAA AATAACAAAT AAATGACTCT 2700

TGTGTTTGGC AAGCAGTCGA CAAAACTCGT TGGTTGTTTC AAATGAAGGT GCTGTTTCAA 2760

AAGGGATTGG CTGAGAAAAA TGGGCTTCTT TTAATTTGTA AATAAGTTCA TTATAACCTG 2820

ATAAATTTAA TTTTTTGGCT AAACGAATGA TGGAACTAGG CGAAGAAAAG CAGTGTTTTG 2880

CAACTTCCCG AATACTCAAA TCGGCCGCAT TCGTACTTTG TTGATCCAAA TAATACAAAA 2940

TCTGaCTTTC TGTTTTCGTT AAATGaTACG TTTCTTTTAA GTAATCTAAG TTCATGAGCA 3000

AACCTCCTTT ATCAGATTAT AAATAATTTT GTGGGAAAAA GGTATCAAAA AATAAAAAAA 3060

TGAGGTGTTT TTATGAAAAG GGCAACAAAG CAAAGGCTGT CTTTGGCAGC AATCATGGTT 3120

CTACTTCTCT CGGGCTGTGG AAGTGTTGGG AAAGAAACCA AAAAGCAAGA ACAACAGGTA 3180

TTACGGGTCG GGATTGATTC GGAATTATCA ACGGCAGACG TGTCGTTGGC AATGGATAAT 3240

ACCGCAGCAG ATGTAATGAG CCAAGTAGGG GAGGGACTTT TCTCCTTTGA CGAAAAAGGA 3300

GAAGCGAAAC CAGCATTGGC AACTGAAAAA GTACAGCCCT CCAATGATGG TTTAAGCTAT 3360

ACTTTTACGA TTCGAAAAGA TGCAAAATGG AGTAACGGCG AGCCAATCAC AGCAAATGAT 3420

TTTGAATACT CTTGGAAGCG CACAGTGGAC CCAAAAACAG CTTCCCCGCA AGCGTATTAC 3480

TTTGAAGGGT TAAAAAATTA TCGTGCTATT GTTGACGGTA GCAAATCTAA AGAAGAGTTA 3540

GGGGTAACAG CCATTGATGA CCATACCTTG GAAGTAGAGC TAAGCTATCC TATGAGTTAT 3600

TTTCAACAAT TATTGGCGGT ACCAGCTTTT TATCCTTTAA ATGAAGCATT TGTCGAAAAA 3660

ACGGGCAAAA ACTATGGTAC ATCAGCTGAG TCAACACTTT ACAATGGCGC CTTCACATTA 3720

GAAGGTTGGG ATGGCACGAA TAATACTTGG TCCTATGTGA AGAATAAAAA TTATTGGGAT 3780

CAAGCGAATG TTTCGCTAGA TAAGGTGGAT GTCCAAGTAG TTAAAGAAGT CAATACTGGG 3840

AAAAATCTTT TCGAAGGGAA AGAATTAGAT GTTGTAAAAA TTTCTGGAGA AATTGTTGCA 3900

CAAGAACAAG GCAATGCAGC TTTGAAAATT CGTGAAATTC CTGGAACGTA TTATATCCAA 3960

TTAAATACGC AAAAAGATCT TTTGGCAAAT AAGAATGCAC GTCGAGCAAT AGCATTATCA 4020

TTGAATTCTG AGCGTTTAGC TAAAAATGTT TTAAATGATG GCTCAAAAAA AGCACTTGGC 4080

TTCGTGCCAA CAGGTTTCAC TAATCAAGAA ACGCAAAAAG ATTTTGCAGA GGAATTAGGA 4140

GATTTAAATC CTAGTGAACC AGAAAAAGCG AAAGAGTTAT GGCAAACGGC TAAAAAAGAA 4200

TTAGGAATTG AAAAAGCGGA GcTAACGATT TTAAGTTCGG ATACAGAAAA TGCTAAAAAA 4260

ATCAGTGAGT ATGTTCAaGG AGCTTTAGCA GATAATTTAG AAAATTTAAC AGTCAATGTT 4320

TCACCAGTTC CTTTTAATAA TCGTTTAGAA AAAAGTCGCA GCGGAGATTT CGACATTGTG 4380

GTTGGTGGCT GGACGCCaGT ATATGCTGAT CCAATCGATT TCTTAAACTT ACTGCAATCA 4440

AAAAATTCCA ATAATTTTGG TAAATGGTCT AATAAGACCT TTGATCAGTT GCTTCAAGAA 4500

GCAAACGTAA CTTATGCAAA TAAATATGAA GAACGTTGGA AAACATTACA AAAAGCGGAT 4560

CAATTGGTTG CGGAAGAAGC CCCCCTAGTT CCTCTTTATC AATTAACAGA AGCACGCTTA 4620

GTGGCCGATT CTGTCCAAAA TTTAGTCTAT GGTCCATTAG GTTCAGGCTA TTACAAATCA 4680

GTCTCTATCG GCGACAAGTA AAGGTTGGGA AAAAATGAAA AAAAGTCAAG CTGTTCAATT 4740

AATTCAGGAA CTTTCAAATG CTAATGGTGT TTCAGGCTTT GAAACAGAAG TCGTTAGAAT 4800

CCTTCAACAC GCAACTGCGG ATTTTACTAT TCAACGGCTA GATTCCATTA AAAATTTATA 4860

TCTTGAGAAG AAAAATAATC TAAGTGAAGG TCCAGTAGTG CTTTTTGATG CTCATAGTGA 4920

CGAAGTGGGT TTTATGATTC AAGCAATTAA AGAAAACGGC TTGTTGCGTT TTTTTCCGTT 4980

AGGTGGCTGG GTCCCTAATA CGATTTCTGC ACAAAAAGTT CGGATTCGTA ACCGGGAAGG 5040

TACCTATCTG CCAGGCGTAG TTACTAGTCG GCCCCCTCAT TTTATGACAC CAGAGGAACG 5100

TCAGCGACCG TTAACCATTG CTGATTTGAC GATCGACATT GGTGCTACCT CAAAAGAAGA 5160

AGTAATTGAA ACTTACAAAA TCGATCTTGG TGCACCGGTC ATTCCAGATG TGACGTGCTG 5220

TTACAACGAA CAGACAGATC TTTTTTTAGG AAAAGCTTTT GATTGTCGCA TTGGCTGTGC 5280

TTGCTTAGTG GATGTCATGG AGGAGTTGAA AGAAGAGACG CTTCCATTCA AGTTAGTTGC 5340

GACAGTAACT GCACAAGAAG AAGTTGGCGA AAGAGGAGCA CTTATTGCCG CAAAACAAGT 5400

CAATCCTGAT TTAGCTATTG TTTTCGAAGG CTGCCCGGCT GATGATACGG CAGAAACGCC 5460

AGAGATGATT CAATCAGCAA TGGGAAAAGG ACCGATGTTA CGTTATTTTG ATGTTTCTAT 5520

GATTACGAAT CCGGAATTTC AAGAGTATGC ACTAGAGATT GCTAAAATAC ACAAAATTCC 5580

TGTTCAAGTT TCTGTTCGAA GTGGTGGTGG AACCAATGGC ATGGCGATTA CACAAGTTCA 5640

AGGAGCCCCG ACGATTGTTG TAGGAATTCC TGTCCGTTAC GCTCACACAC CCCATTGCTA 5700

TGTAGATTTT CAAGATTACC AAGCGGCGAA AGAATTAGTC ATTCAATTAA TCAAAAATTT 5760

AGATGCTGAC AAAATTCAAG CACTGGTTCA GCCATTGTCA AAGGAGTGGA ACAAATGAAG 5820

ATTTATGTAT CATGCGATAT TGAAGGATTA GCAGGTATTG CGACATTTGA TATGGAAAAA 5880

GAAGACACCG TTTTATTCCG AGAATTATAT CACCAACATG TGGCTTGGTT AATTGAAGGG 5940

ATTCAAAAGA GTGCCAAAAA TGAACAAATT ACAGAAATTA CTATTGCCGA TTCGCATAGC 6000

CGTGGATTGA ATTTAGCTTA TGCGCGCTTG GCAGAAATGG ATGAACGAAT TTCCTTAGTG 6060

AGCGGCTTTC CTCGCATGGA TTATATGATG AGTGGCTTGG ATAGTAGCTA TGATGTGGTC 6120

TTTTTCTTAG GTTATCATGC CGGTATTGGT AAACAAAAAG GGAATATGGA TCATGGGTAT 6180

AGCGCTAGTG TTGCTTATGA TTTAAAAATA AACGACTTAG CGATGAATGA AACGACAATT 6240

AATGCAGcTT ACGCTAGCGA ATTAGGTGTC CCTGTTGGTT TAATTATCGG TGAATCAGGC 6300

TTAGAGGAAC AACTTTTTCA AGAAAAAATG ATGCCCGAAG TTCCCTTTGT TTCCACCAAA 6360

GAATCTCTGG GACGTTATGC AATCAAAAAT AGACCCATGC AACAAGTTCG GGAAGCTATA 6420

GTAGCAACTA CAAGTCAGGT GTTAACTTCC TTTGCGTTGT CTGAATTACC CCGATATGCC 6480

TTGCAAACAC CTGCAACTGt AAAATTGCAA TGTGTGACGA CAGCACAAGC AGATCGTATT 6540

GAAATGTTAC CAATGATTAA ACGCATTGAT GGTAGAACTG TTTCATTTGT AGGTGAAACG 6600

ATGAAAGATG TCATGAATGG AATTGTCGCT GTGGTTGGTT TAGGCGGAAC AAGTTATTAA 6660

AGTAAGAAGC TTACGGAAAA AACGAATGGA CATTTTTTTC TGTAAGCTTC TTGCTATAGT 6720

TATTTCTATT TTTGTTTCTC TAGAATTTCT TTGCTTAAAC GTTGTGCAGT TGGTGTCACC 6780

GCAAGACcGC CTTCAGCGGT TTCCTTGAAA ATTTGTGGCA TTTGGCGTCC CACTTGGTAC 6840

ATTGCGTGGA TGACTTCATC TGGCGGAATG ACACTTCGAA TCCCCGCTAA AGCCATGTCA 6900

GCTGAGATGA AGGCTTGAGA AGAACCTAAT GCGTTTCGCT TCACGCAAGG GACTTCAACA 6960

AGTCCCGCAA CAGGATCACA AATGAGGCCC ATCATATTTT TTATCGTAAT GGCCACAGCT 7020

TGAGCAGATT GATCAGgCGT ACCGCCAGAA GCCGCAACTA AAGCGGCACT TGCCATTGCA 7080

CTAGCAGAAC CGACTTCAGC TTGACAACCA CTTCGGCACC ACTGATTGAG GCATTATTGG 7140

CAATCACTAA GCCAAAAGCC CCAGCTGTAA ATAGAAAATC CAATTGTTGT TGATGCGTTA 7200

ACTTTAAGCG ATCAATAACT GCCAACAAGA CGCCAGAAAC TACTCCGGCA CTGCCTGCGG 7260

TTGGTGTAGC ACAAATCAAG CCCATTTTTG CATTAACCTC ATTAACAGCA ATCGCATTTC 7320

GCACGGCTTG TAAAATGGTC TCACCGCTTA AAAAGTTGCC TGAATGTATG TAATCATTTA 7380

AGCGGGTTGC ATCGCCACCA GTTAGCCCTG TTACTGAGGT AACACCAGCG ATTCCTTCAG 7440

CGACGGATTG TTTCATAACG GCTAGATTTT TTTCCATTGT TTCAATAATT CGTTCACGAG 7500

AATACCCGCC ATGAGCCATT TCTGTTGCAA TCATTAATTC GCTCACAGAA GGGTAATCTT 7560

TCGCTTGAGC TACCAAATCT TCAATTGAAT AAAACATCGT ATTACTCCTT TACGTTCATT 7620

TAGTCAAAAT AGTTAACGCT ATATATATGA GGGATTTCAG CTAATTGCAT GACAATGTCA 7680

CCGACTTCGG CTTGATCAAC TTCAATAATC ATAATCGCTT TTTCCCCTTT AGATTCTCGG 7740

GTAACGGTCA TTGTACTAAT ATTGATATCG CTAGCTGAAA GGATGTTGGT CACCTTAGCA 7800

ATCATGCCAG GAACATCTTG ATGAACAACG ATAAATGTTG GCGTGCCCAT ACTTAAAGAC 7860

AATTTGAAGC CATTTAATTC AGAAATTTGA ATGTTTCCAC CACCAATTGA AATTCCTGTC 7920

ACAGATAGTT TTCGGTCGCC AGAAGAAACC ACAATTTTAA CGGAATTAGG GTGGTCTGCT 7980

TTTTCACTTT TAGGAATAAA ACAAACTTCC ATTCCTTGTT CATAGGCAAT TTCCAAAGAA 8040

TTGGATAGTT GCTCATCATC TGGTTCCATG CCTAAAAGGC CGCCAACTAA GGCGATATCT 8100

GTTCCGTGAC CACGATAAGT TTTGGCAAAA GACTCATATA AATAAATGTC AACGCTATCA 8160

GGTTGCTGTC CAAAAATACT GCGAACCACT TTTCCAATTC TTGCCGCTCC AGCTGTATGA 8220

GAACTGCTTG GACCAATCAT AACAGGTCCG ATAATATCAA AAACACTATT AAAGCGTAAC 8280

TCTTCCATTT CAGCACCTTC TTCATTAGAA TATGTTTTCA TTGGAAGTCT AACACAAAAA 8340

ATGGCGGGGA ACTAGGCAAT CGCTTGTGTT TTAGCGAATA ATTTAAATTT TTATCAGAAT 8400

GAAAGCAAGG GGCCTTCGTT GACAAATCAG TGAAGAGATG CTAAATTCTA GATAATAAAA 8460

GACGAAAAGA GGGATTCGTG CGTTGAAAAT TTAAGTTAAT TTTTTAGAGA AACTGTATAA 8520

CTTTGCCTGA ATATAACGGG TGGAGTGTGC GGATTTTCGG AAATTAGCTT AAAGAATTCA 8580

ACCACGGGTC TTTTTGTATA CCTTGGATTG ATGCCGCTGA TTTCCGTAAA GCTGCATCAA 8640

TTTTGCTTTA TGGAGGTATA TTTATGTCGA AAATTGAACT AAAACAACTA TCTTTTGCCT 8700

ATGaTAATCA AGAAGCGTTG CTTTTTGaTC AGGCAAATAT CACGaTGGAT ACCAATTGGA 8760

AATTAGGaTk GATTGGCCGC AATGGCcGTG GGAAAACAAC CTTATTAAGA TgTTACAAAA 8820

GCAGTTGGaT TACCAAGGAG AGATTCyTCa TCaAGTCGAT TTCGTCTATT TTCCACAAAC 8880

AGTTGCAGAA GAACAACAGC TCACTTATTA TGTCTTACAA GAGGTGACTT CTTTTGAACA 8940

GTGGnAATTA GAACGAGAAT TAACGCTTTT AAACGTTGAT CCTGAnGTTT ATGGCGTCCC 9000

TTTCTTCTTn ATCAGGCGGC G 9021

(2) INFORMATION FOR SEQ ID NO: 156:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6491 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

ACAATAAACG GGAGGGATCG TTTATGAATA TTGAAAAAAT GACAACCACG CTACAAsAGG 60

CGATTGCTGA AGCTCAAAAA GTTGCGGTGA CACGTCAACA TCAAGAAATT GATATCGCAC 120

ATTTATGGAA AATTTTTCTT CAACCAAACC ATTTTGGACG TAATTTTTAC ACAGATGCTG 180

GTTTAGATGT TGACGCATTT GAACGAGAAG TAGATAACGC GTTAGATGAA TATCCAAGTG 240

TGGCAGGCGG TAATGTGCAA TATGGTCAAA ATTTAAGTCA AAATTTGTTC CATTTATTGC 300

AAGAAGCTGa TTCACTTCGA GAAGAtTCCA AGATGaATTT TTATCaaCCG rAATTGtACT 360

TTTAGCTTTA ATGrAATTGa AAAATTATCG TTTaACmAAA TATTTAATGc mACmAGGcAT 420

TACGGaAAAA GAGTTwAGrA AAAATATTGA AGaGATGAGA GGAGGAGATC GTGTGACTTC 480

TCAAAACCAG GAAGAACAAT ATAAAGCACT AGAAAAATAC GGTGTTGACT TAGTACAACA 540

AGTAAAAGCA GGAAAACAAG ATCCAATCAT TGGTCGTGAC GAAGAAATTC GTGACGTTAT 600

TCGGATTTTA TCAAGAAAAA CTAAAAATAA TCCAGTCTTA ATTGGTGAAC CAGGTGTTGG 660

TAAAACAGCG ATTGTTGAAG GATTAGCACA ACGAATCGTT CGTAAAGATG TTCCCGAAAA 720

CTTAAAAGAT AAAACCATTT TTTCTTTAGA TATGGGCGCC TTAATTGCGG GAGCAAAATT 780

CCGTGGCGAA TTTGaAGAAC GGTTAAAAGC TGTTTTAAAA GAAGTGAAAA AAAGTGATGG 840

CAAAATCATT TTATTCATTG ATGAAATTCA TAATATCGTC GGAGCTGGTA AAACAGAAGG 900

CAGTATGGAC GCCGGAAATT TATTAAAACC GATGCTAGCA CGTGGTGAAT TGCATTTAAT 960

CGGTGCCACA ACGCTTGATG AATATCGCCA ATATATGGAA AAAGATAAAG CCTTAGAACG 1020

TCGTTTCCAA AAAGTATTAG TCAAAGAACC AACAGTAGAA GATACAATTT CTATTTTGCG 1080

TGGCTTAAAA GAACGATTTG AAATTCACCA CGGCGTGAAT ATTCATGACA ATGCTTTAGT 1140

```
TGCTGCCGCT ACTTTATCCG ATCGCTATAT TACTGATCGT TTCTTACCAG ATAAAGCGAT   1200
TGACTTAGTC GATGAAGCAA GCGCAACGAT TCGAGTGGAA ATGAATTCGA TGCCAACAGA   1260
GCTCGATCAA GTAACACGTC GGTTAATGCm ACyAGaAAtT GAAGAAGCGG CCTTGAAAAA   1320
AGAATCAGAT GATGCAAGTA AAAAACGGTT AGCCAACTTG CAAGAAGAAT TAGCAGATTT   1380
GCGTGAAGAA GCCAACTCAA TGAAAATGCA ATGGGAAACG GAAAAAGAAG AAGTTAATGC   1440
GGTTTCCAAT AAACGTGCAG AGATTGATAA AGCAAAACAC GAATTAGAAG ATGCTGAAAA   1500
TAATTATGAT TTAGAACGAG CTGCTGTTTT GCGTCATGGA ACAATTCCGC AATTAGAACA   1560
TGAATTGAAA GAATTGGAAG AAAAGAACGC TAAAGATAAC GTCAAAATGG TACAAGAATC   1620
GGTTACTGAA AACGAAATTG CGCAAGTGGT CGGTCGTTTA ACTGGCATTC CTGTGACAAA   1680
ATTAGTTGAA GGCGAACGAG AAAAATTAAT GAAACTGAAT GAAACATTAC ACAAACGTGT   1740
GATTGGTCAA GATGAAGCCG TCGATGCTGT TAGTGATGCG GTGATTCGTT CAAGAGCAGG   1800
CTTACAAGAT CCAAATCGcC CACTCGGTTC GTTCCTTTTC CTAGGACCAA CTGGTGTTGG   1860
TAAAACAGaA CTTGCTAAAG CTTTAGCTGA AGATTTGTTT GATTCTGAAG ATCATATGGT   1920
ACGGATTGAC ATGAGTGAAT ACATGGAAAA ACATGCCGTG TCTCGTTTGG TTGGTGCCCC   1980
TCCaGGCTAT GkTGGTTwTG AAGAAGGTGG CCaGTTAACG GAAGCTGtTC gTCGrAACCC   2040
tTATACAATT GTCtTATTAG ACGAAATTGa AAAAGCGCAC CCAGATGTCT TTAATATCtT   2100
ATTACAAGTT TTAGATGATG GTCGTTTGAC AGATTCTAAA GGTCGGGTCG TTGATTTTAA   2160
AAATACAGTT TTAATTATGA CAAGTAACAT CGGCTCTCaG CTTTTATTGG AAGGTGTGAC   2220
ACCTGAAGGA ACAATTCCAG AAGAAGTTGA AAATCAAGTC ATGAATATTC TAAAAGGACA   2280
CTTTAAACCA GAGTTCTTAA ACAGAATTGA TGATACCATT TTATTTACAC CATTAAGCTT   2340
GGaTAATGTG AAAGGAATCa TTGGTAAAAT GACCGCACAA CTTGCTCATC GTTTAGAGCA   2400
ACAAGAAATT GTGTTGGAAA TCACTGATGA AGCTAAAACA TGGATTGCTG AAAATGGGTA   2460
TGAGCCAGCT TATGGAGCTC GTCCGTTAAA ACGTTTCATT ACTCGCGAAG TCGAAACACC   2520
ATTAGCAAAA GAAATTGTTT CTGGACGAGT AATGCCAAAA ACAAAAGTCA CGATTAGTCT   2580
ATTAGACAAT CAATTAGTTT TTGAAAATGA ACCGATAGAA GAAGTGTAAA AAAGAACGAG   2640
CGAATTCGCT CGTTCTTTTT TTATTATGAC AATCGTTGCT TAATTTCTTC CAACAAATCC   2700
TCTTCTGAAG AAACAACCCG ACCATTCAAT TTAATCAAAC CAACTGTGTA AAGGTTTAAA   2760
TAATGGAATT GATTTTCAGC GATTTCGTTC AAAGCATCAA GCTTTTGGTG ATTATCAGCA   2820
CCTTGTTGGC GGCTGTCTGT GTAAAGGGCA ACCACCGGGA TTCCTTTGGC ATAAGCAACA   2880
CCAATTTCTG AAGcTACGCC CGCATCAATG GTTGGTCCGT CTAATAAAGC AACCAATAAG   2940
TCGCTAGCAA GGACGTTTTC TGTATCTGCT AAGGCAATCA TTTTGCTATC GGCATAAGCA   3000
GATTTGTCGT TAATCGCCGC ATTTTCTTGT GGTAAATAAA GATCGATTGT TTTATCCAAT   3060
TGTCGAATTT GTTCGACTAA ATAGGCATTA TAACGTAAGT CAGCTTGCGA AAATAAAGGT   3120
```

CCTGCAAAAT AAATTTTTGT CATTTGTCTG TCTCCAATCC AAGGGTTTAG AAATAATCAC 3180

TGAAATAAGC TTCTTGGTCA GGAAAAATCC GTTCCAAAGA ATTTAAAGTT TCTTTATCTG 3240

TATCAATACG TTCAATTCGA TGTAAGTACG AAGGGCGTCG GTAATTCATG ACTAAACAAG 3300

TTAGTGTTTG AATATCTAAG TGAACGGCGG TGCCTAGAGG TTCATCCGTA ATGGTCACTT 3360

GATCATTTTC ATCCCAAATT AAGCCAAAAA TACCATTATT CCACTCAGCG ACAGGATCTT 3420

TAACGACAAA ATGGAAGGGC TTAGCTGTAC TTTCAAAAGG AAAGTTTTCT AAAAAGGCTT 3480

TGACATCAAC AATTCGCGCC ATATAATAGG GCTCAATGCT TTCTTTAATT TGGCTGTCTT 3540

CTAAAAGAAA GGCTAATGGT TCGTTTTTAT AGATATCGCC CTTGACCCAA TAAACCATCG 3600

AAAAATGCGC AGTAATAAAA TTCCaCAAGC CATTTCGGGC TTCTTGGTTC AAGTAAAACA 3660

TTTCCTTAAT ATGAAATACC TCATCGGCcA CCCAATAGAA AAGTACTCCT AAAGGTTCTT 3720

GATTGGCACC ATAATAGACA GCAGCAGTCC GTTCTTCCTC GTTTTCGAAA CGCCAATATT 3780

CTTCCCAGTT GAAGGCACTA CGAATGAGTG CGCCGTGATT TTGACGAGCG AATCGCGCAT 3840

AGACATCAAA GACATCTGGG TGATCGACAG CTAACCGTTC AATCATTCCT GGAACAGGAA 3900

CGGTTTTTGG TAATTGCGTA TCCCGAATTT TAAAAGATAA CTTGTCGGAC ATAATTTCCC 3960

AACCTTTACG TCGGTAATAA GGGATGTTGT AAGGGAAAAG ATAAGAAATC CACTGTTTAT 4020

CTTGGCGCAT TTCTTCTAAT GCCGTTTGAA TCAAGTCTTT CATTAAACCG TGGTTGGCAT 4080

ATTCAGGATA GGTACCGACA CCAGTAACGC CGCCCATTTT ATATAAAGCG CCATGGATAT 4140

TGACTTCGCA TGGATAGATA GCAATTTGTG AAATTAATTG ATTTTCGTGA AACCAGCCAA 4200

ATACTTTTGA AAGTTCTAAA ATCGGTTGTT TTGATTTGAT AAATGCGCGT TTGTTTTCAA 4260

AACCACTTTC TTCAATATCT GCTTCAGTCA CTTGAAAAAC ATAGGAAAGA AGTTCATTAA 4320

ATTGATCAAT GTGTTCyTCC yCaACAGGCT TTAAnGGTCa GTTGCTTtcC rAAATtCyTG 4380

gTTCcATCCc ATTTTTcCAT CTCCAATAAA TTTAcTGTTT TTAgTATACC AAATTTTTCG 4440

AAACTCTGTT AGCCTTTCGA AACCAAAAGA ATAAAAGGGA AGTGCTTTAA ATCTTCTTCA 4500

AAAATTGGTA TAATGAAAAG TGCTGATTAT CCTAGACCTT TAAACATGAT AATTTGATAA 4560

AAAAGGCAAC TAAAATATAA CAATTGATTA ATGAAGCTAG CAGGGAGCTA GCTTTTCTTA 4620

AAAGAAATGA GGCAAAAAAT GAACAATAAA GAAATGAAAG CAAGACAAGA GAAAATTCGT 4680

AATTTCTCGA TCATTGCCCA CATTGACCAT GGGAAGTCAA CTTTAGCCGA CCGGATTTTG 4740

GAAAAAACAA ATACAGTTAG CAGTCGAGAA ATGCAAGATC AATTACTTGA TTCAATGGAT 4800

TTAGAGAGAG AACGCGGCAT TACTATCAAA TTAAACGCAA TTGAATTAAA CTATACAGCC 4860

AAAGATGGTG AAATCTATAC TTTCCATTTG ATTGACACAC CAGGGCACGT CGATTTCACC 4920

TACGAAaGTTT CTCGTAGCTT GGCAGCTTGT GAAGGGGCTG TTCTAGTTGT TGaTGCGGCG 4980

CAAGGaATTG AAGCGCAAAC GCTAGCaAAT GTCTATTTGG CATTGGATAA TGACTTAGAA 5040

ATTTTACCTG TTATTAATAA AATTGATTTA CCCGCCGCTG ATCCAGAGCG TGTTCGGACA 5100

GAGATTGAAG ACGTAATTGG AATTGATGCA TCGGAAGCTG TTTTAGCAAG TGCAAAAGCA 5160

GGGATTGGGA TTGAAGATAT TTTAGAACAA GTGGTGGAGT ATGTACCAGC TCCATCAGGC 5220

GATATTGAGG CTCCTTTAAA GGCTTTGATT TTTGACTCTA TTTACGATAG TTATCGGGGG 5280

GTCGTTTTAA ACATCCGTGT AATTGACGGT GTCGTTCGTC CTGGGGATAA AATCCAAATG 5340

ATGAGTAACG GTAAAACGTT TGATGTAACA GAAGTCGGCG TTTTTTCACC GAAACCGATT 5400

GCTCGTGATT ATTTAATGGT TGGTGATGTG GGCTATATCA CCGCTAGCAT TAAAACGGTT 5460

CAAGATACAC GGGTCGGGGA TACAGTGACT TTGGCCGACA ATCCAGCAGC AGAAGCACTA 5520

CCAGGCTACC GCAAAATGAA TCCAATGGTT TATTGTGGTT TATATCCAAT TGATACGTCG 5580

CGCTACAACG ATTTACGGGA AGCATTAGAA AAATTACAAT TAAATGATGC GGCGTTACAA 5640

TTTGAACCGG AAACATCGCA AGCTTTAGGG TTTGGTTTCC GTTGTGGTTT CTTAGGTTTG 5700

CTGCACATGG ATGTTGTTCA GGAACGTTTG GAACGAGAAT TTAATTTAGA GTTAATTACA 5760

ACAGCACCGT CTGTAATCTA TCACGTTAAT AAAACTGACG GAACAACCGT TGTTGTTGAC 5820

AACCCAGCTG AATTTCCAGA ACCAGTAACG ATTGAATCTG TGGAAGAACC TTATGTTAAA 5880

GCGCAAATCA TGGTGCCAAA CGATTATGTA GGAGCAGTAA TGGAATTATC ACAACGTAAA 5940

CGTGGCGAAT TCATTACAAT GGATTACTTA GACGATTATC GTGTAAACGT AGTTTATGAA 6000

ATTCCGTTAT CTGAAATCGT GTTTGACTTT TTCGATAAAT TGAAATCAAG TACAAAAGGC 6060

TATGCATCCT TAGATTACGA AATGGCTGGC TATCGTACCA GCCGCCTAGT GAAAATGGAT 6120

ATTCTATTAA ATGCTGAAAA AGTGGATGCT TTAAGCTTTA TTGTTCACCG AGATTTCGCA 6180

TTTGAGCGTG GTAAAGCGAT TGTTGAGAAA CTGAAAAAAC TAATTCCACG TCAACAGTTT 6240

GAAGTCCCAG TTCAAGCGGC GATTGGTCAA AAAATTGTGG CTCGTTCAGA TATTAAAGCC 6300

TTACGCAAAA ACGTACTGGC TAAATGCTAT GGTGGCGATG TTTCTCGTAA ACGTAAATTG 6360

TTAGAGAAAC AAAAAGAAGG GAAGAAACGG ATGAAAcAAA TTGGATCCGT GGAAGTTCCT 6420

CAAGAAGCCT TTATGGCGGT TCTGAAAATG GACGACCAAG ATAACGCGAA ATAGCTTGTT 6480

TCCAACGGAA A 6491

(2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1014 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

CCTACGAAAA ATATGGCACA AAGAAAGTTG GAAAAATGAC AAACTTGTCG ATTTTCCAAC 60

TTTCTTCATC TATTCGTGGT TGGCTTGCTG AACTGCATCC ATCACGGCAT AACGGAAGCC 120

ATTTTTTTCT AGAGCTACGA CACCGCGAAT CGTACTTCCT CCTGGAGAAG TCACTTCATC 180

TTTCAATGCT CCTGGGTGTT TTTTCGTGTC TAAGGCTAAG GTCGCCGTTC CTTTTATCAT 240

TTGAGCGACA ATTTCATAAG ATAAGTCTCT TGGTAACCCT TCTAAGACTG CTGCATCTCC 300

CAATGCTTCC ATAAAGATAT CAACAAAAGC TGGACTGCAA CCAGCCACAG TACCAAAAGT 360

TTCCAGTTGC GCTTCACGAA TTTCTTTGAC TAACCCTAAA GTAGCAAGAA GTTCCATTGC 420

CAGGTCCTTC GTTTGGCCTT CAATTGTTGG TGCCAAGGCT ACGCCAATCA CGCCTTGATT 480

CACGCTAACT GGTGTATTAG GAATGGCATG AACAACTTTG GCTTCTTTAT TGCCAAGAGC 540

CATTTGTGCT TCTTGTATGG TATGACCTGT TGCAACAGAA ATCAATAGAC TTTCGTTGGA 600

AAAATGAGGG GCTAATTCAT GAAGAACAGG TAACACGACC TTTGCTCCTG TCGCAATAAA 660

AATAACCTTA CATTCTTGGA AAAGTGCATA ATCTGTGACA AGTTGAAAGC CCCACTCTTT 720

TTGAAGTGCT TCTGCGGTTC CACTTGATCC GCCTTTCACA TATAAATGTT CCTTCTTGAC 780

TGCCTCAGCT TTCAGTAACC CTTTAATCAT GGCACCGCCC ATACTGCCAG CACCGATAAA 840

GCCAATTTTC ATAGCATTCC CTCCTTAGTT AAATAArGTG TTtGTGkCAA AAGTCAATTA 900

GCTTTtGACA CnAACACTAA CAGTCGkACt TTTTcCGtAC aGCtGkTTCG TTAaTTaATT 960

gGTAAtcCAT wACGaCkGCT CaAAcGCgCA ATTTTTCaGA AGtCyGaACC AACT 1014

(2) INFORMATION FOR SEQ ID NO: 158:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16484 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

TTATAAAAGA CCCTTCCTTA AAAAGCAATG TTTCTTCCAA TCAAAAAAAC GAATTATCTA 60

AAAAATGCAA CTTATTTGCT CAAAAAAGCA AGCGATGATT GACAAACGGT CGTTTTTAAC 120

TGATAATGAG AATCGTTATC AACTAAAAAT AAAGGAGTTA CACAATGAAA CTTTTAAAAA 180

AGACGGTCCT AATTGGTACA ACCCTTCTTC TTGGTTCATT CTTACTCGCA GCTTGTGGTA 240

ATACGAATAA AGAAGCCAAC AACGCTGACA AAACACATGA AGTAACAGAT ACCTTAGGCA 300

ATAAAGTAAC CGTCCCCGCG AAACCCAAAC GGATTATTGC GAGTTATTTA GAAGATTATC 360

TAGTTGCATT AGGAGAAAAA CCAGTGGCAC AATGGACAGT TGGACAAGGC AGCATTCAAG 420

ATTATTTAGC GAAAGAATTG AAAGATGTCC CCACTATTTC CTATGACTTG CCATATGAAG 480

CGGTTCTAAA ATTTGAACCT GACTTATTAT TAATCAGTTC ATCTGCTCTA GTTGAAGGCG 540

GTAAATACAA AGAATACAGT AAAATTGCGC CAACTTATGT AGTCAAAAAC GGCGAAAATG 600

TCACCTGGCG TGATCAATTG GaAGATATTG CCACTGTTTT AGATAAAAAA GAACAAGCGA 660

AAAAAGTGTT AGAAGATTAT GATACCTTAA CCAAAGGCGT CCAAGAATAT CTTGGCAAAA 720

AAGATGCTGG CAAATCTGCG GCAGTCTTAT GGGTAACCAA CAACCAAGTC TTTATGGTTA 780

GCGATAATCG CTCAAGCGGA ACCGTGCTCT ATCAGGACTT AGGCCTCCAA GTTCCAAAAT 840

TAGTGGAAGA AATTTCTAAA AACGCTACTG CGGATTGGAA TCAAGTTTCT TTAGAAAAAT 900

TAGCTGAGCT TGACGCAGAC CACATTTTCC TTGTAAACAG CGATGAATCA GCACCTCTTT 960

TCCAAGAAGC AATTTGGAAG AACTTACCTG CTGTGAAAAA TAACCAAGTT CATACCTATG 1020

ATAAAAAAAG TAGTTGGTTA TACAACGGAC CTATTGCGAA TACTCAAATT GTTGAAGATG 1080

TAAAAAAaGC GCTCTTAAAT TAAAAAAATG CtTGGATGAA AATAATCATC CAAGCATTTT 1140

TTTTAATTAG CAATTGCCGT TTCTAAACCA ACTTCAATCA TTTCATTAAA CGTCGTTTGG 1200

CGTTCTTCTG CCGTTGTTTC TTCGCCTGTC ACCAAGCTAT CACTCACCGT CATAATCGCC 1260

AATGCTTCTA CATCAAATTT TGCTGCTAAA TAATATAACA TTGCCGCTTC CATTTCAATC 1320

GCTAATACCC CGTATCGACC TAAACGAAAG ACATCATCTA AACTATCTTT GTAAAAAACA 1380

TCATCTGACA ATACGTTACC AACATGGGTA GTAAAGCCTT TTGCTTTAGC AATTTCGTAA 1440

GATTTGAGTA AAAGATCAAA GCTCGCAATT TGTGGAAAAT CATATTTAGG AAAATCATTG 1500

CGAATCGCGG AAGAAGGCGT CGCTGCTGCT TGTGCGATAA CTAATTCACG AACGTTGACT 1560

TTTTCTGAAA TAGAACCACA CGTACCAACA CGAATTAACT TCTTCACATC GTAAGAGTTA 1620

ATTAACTCAT GGGCATAAAT ACCTGCTGAA GGCATCCCCA TCCCTGTTCC TTGGACTGAA 1680

ACACGCTGTC CCTTATACGT TCCTGTGTAA CCTAACATCC CACGAACTTG ATTATAGCAG 1740

ACAGGATTCT CTAAAAAAGT TTCCGCAATA TATTTTGCTC GGAGGGGATC CCCTGGTAGT 1800

AAAATTTTAT CGGCAATTTC ACCTGGTTTT GCTTCGATAT GCACACTCAT TCTGGTTTAC 1860

TCCTTTTTTA TAAAGCTGCT AGTGTTGCTT TCACTAACTC TTTGAATTGT CCTTTAACGC 1920

GTTCAGTTGT TTCAACTACT TCTGCATGAT TTAAGCTACT TTGCATACCA GCTGCTAAAT 1980

TCGTAATACA AGAGATGCCT AGTACTTTCA TGCCACTATG CACAGCTACA ATGACCTCCG 2040

AAACAGTTGA CATGCCAACT GCATCCGCAC CCATCGTCCG TGACATTCGA ATTTCTGCAG 2100

GCGTTTCATA TGTCGGGCCA GAAAAGCCCA TATACACGCC TTCTTTTAAA TCAATATTTT 2160

GTTCCGCAGC CACTTTTTTT GCCACTTCCC GATATTCTTG CGTGTAAGCA TGACTCATAT 2220

CTGGAAAACG CGGCCCGATT TCTTCATCAT TTTCACCAAT TAACGGATTA TCGCCTGTAA 2280

AATTAATGTG GTCATTGATT AACATCAAAT TTCCCGGTGT ATATGTTTCA TTTACGCCAC 2340

CAGCTGCGTT GGTTACAATC ATTGAATGAA TGCCTAAAGC AGCCATCACA CGAACGGGAT 2400

AAGTGACCGT TTGCATCGAA TGTCCTTCAT AATAATGAAA ACGCCCTTGC ATCGCCAAAA 2460

CTTTTTTTCC AGAAAGCGTT CCGTAAACTA ATTGTCCAGC ATGTCCAACG ACTGTTGAAA 2520

CAGAAAAATG GGGAATTTCA GAGAAAGGAA TAGCAATCGC ATCGGTTATT TCATTAGCTA 2580

ATTCACCTAA CCCAGAACCT AAGATTAGAC CGAAGTCTGC TTGTTGCACC CCTTTTTCTT 2640

TCAAGAaCTC AGTTGTTTCG TTTAAATGAT TCGTTAACGT TGTCATGTGT TGCCTCCTCG 2700

AATTATTGTA ATTGATTTAA AAAGCTTTGA CCGTTTTCCG TTGCAGGAAC TTCAAAGTTT 2760

TCAGCAATCG TTGCTGAAAT ATCTGAATAA AATCCTTGTG GCAAGCTGCC TTGTTTTGTC 2820

```
ATTTTTTTAC TATAAGCCAA TAAAGGCACA TATTCACGTG TGTGGTCTGT TCCTGGGAAA    2880
GTTGGATCAT TCCCGTGATC TGCTGTAATT AAAAGCAAAT CATCTTCGGC CATAGCTGCC    2940
ATAATCTCTG GCAAACGACC ATCAAAATCT TCGATGGCTT TTGCATAGCC CTTTACATCA    3000
CGACGGTGAC CATACAGCGC ATCGAAATCC ACTAAGTTAG TAAAACTTAA ACCTGTGAAC    3060
TCTTTGTTCA TGACCGTTAA TAATTGATCC ACCCCATCCA TATTGCTCTT CGTGCGGATA    3120
GCTTCAGTAA TTCCTTGTCC ATTAAAGATG TCGTTAATTT TCCCAACAGC GATCACATCA    3180
TTGCCGTTTT CTTTCAACGA ATCTAAAACA GTATGGCCAA AAGGATCTAG AGCATAGTCA    3240
TGACGGTTAC TGGTTCGAGT GAAATTACCT GGTTCACCAA CATACGGGCG AGCGATAATC    3300
CGACCAATCA TATAAGGTTC GTCTTTCGTA ATGTCACGAA CATATTGACA AATTTTATAC    3360
AATTCTTCTA ATGGAATAAT GTCTTCATGC GCCGCAATTT GTAAAACAGG GTCTGCTGAA    3420
GTATAAACGA TTAGATCCCC AGTTTTCATT TGATGTTCGC CATAATCATC AATGACTGCT    3480
GTCCCGCTAT ATGGTTTGTT ACAAACAACT TTTCGCCCAG AAAAATCTTC AATTTGTTTC    3540
AGTAATTCTT CAGGAAAGCC GTTTGGGAAT ACCCGAAAAG GTTTTTTGAT GTTCAAGCCC    3600
ATAATTTCCC AGTGACCGGT CATAGTATCT TTTCCAACAG AAATTTCTTC TAATTTGGTA    3660
GCATAGCCAT CGTGGTCTGC CACTGCTTTG ACACCTGTTA AAGGAGCAAT AGTTCCCAAT    3720
CCAAGTTTTT CTAAATGAGG AATGGTTAGT CCCGCTTCTT TGGCAATATG ACCCaGTGTA    3780
TCACTGCCGA CATCTCCAAA TTTTTCAGCA TCTGGTGCTT CACCAATACC GACTGAATCt    3840
AGTACGATTA AATGAACACG TTTAAACACA CTTGCTTCCT CCTCTAAGGT CGTAAATACG    3900
TCAAGAATGA CAATTATTTT ACGCTAGCCT TTTAAATAGT CTCGAGAATC AAACAGAATA    3960
ATCTATTTAT TCTCAGCTCT TCAATTATAC CATGAATCCG CTATTATTGC CTGAAAGAGC    4020
ATTAAAATTT TCACTGATTC CTTTGATTTT AGCAGGGCTT AAAAATTGAA AGAAAAAAAG    4080
TCTCCCACCG TAATTAGGAA ACTTTTTTAT TGTTTTTTTC GAATAGCCTG CTCCTTAAAA    4140
AGAACACGAC TCTTTATCCA ATCGAAACTT TTAACACTCC AAGCGGTCTC CCTCTTTTGG    4200
CACATTTAAA TACCAGATAA CCGCTTCTGA TAAAATCACA CCAATGCCAA TGGGCGGATA    4260
AAAATAAATA CCGATATATC CGATTGCGAA AAGACAAAGA TGAATGATTT TTGTTGCCTT    4320
ACTATAAGAT TTCGATTTTC TTAACTTAGC TGCACCTTGT GGATCTTTCA CTACATTGTC    4380
TTCATAAATG ATCCTTGTTA GCAAGAAGTA GGCAAAATCC CaAACCGTAT ACACAACAAG    4440
ATAAAAATAA ACAGGTGCCC GCATCCACCA AAATTCACTC ACCCAAGCAG TTGCTACCGG    4500
AATGAAGGCC ATGGAGAATA ACCAAAAATT ATTTGCCCAA AACGCTTTTT TAGAAAACCA    4560
ATGAGTTACG GACATTATGT AATGATGATT GTACCATGCG ATACAGATAA AAATAAAACT    4620
GACAATAAAC GATAAAAAAT ATGGGATATT TTCTATCAAA CCTGCCAAAA CACCCACTTG    4680
AGGAGTTTTA ATTTCTAAAA TCATGATGGT CATGATGATG GCAATGACTG CATCACTAAA    4740
TGCTTCTACG CGTGTTTTAT TCATTGAACG TCCTCCACTT TTCCTATGCT TACTACAAAC    4800
```

AAAAGTATAG CAGAAAAACC AACGATTTTT GAAAAAAAGC TGCTAGACCT GTTTGGTTTT 4860

TTTCAAAATT TCAACACTTC CTTCGAATTG TAAAGAACAG CTAAAAGAAA AmAGAAAAGT 4920

tACCTTTATC AACTCGCTTT TTTACAGTGA GGCAGTGCCT AATGCCGCTT GTGTTCAACT 4980

GATTATCCTA AGAAAAGAAA TGGCCCCCCG TCGCTTTGAC GGAGGGCCTA TTTCTTTTCt 5040

TATTTCGTTT TGACGTAATT TTTCCCGTTC GCTTTTGGTC CCGATGATTT TCCGAAGAAC 5100

ACAACTAGGA CTAAAATGGT TAATACGTAT GGTGCTGCTT GTAGATAAAC TTTTGGAATT 5160

TCTGAAATAA TTGGTAGGTT TTCCCCAGCA ATACTTAAGT TTTGTGCAAA GCCAAAGAAT 5220

AAGGCTGCGC CCATTGCCCC AAGTGGATTC CATTTACCAA AGATCATTGC GGCCATTGAA 5280

ATAAACCCTT GTCCTGCAAT TGTGGTAATC CCAAAACGAC CCGCAATACT TTGAGCAAAG 5340

ACAGcTCCGC CCATACCGCC TAAGAAGCCG GATAGTAAGA CACCAGAATA ACGCATTAAG 5400

TAGACATTAA TTCCCAATGT ATCTGCCGCT TGTGGATGTT CCCCCACAGA ACGTAACCGT 5460

AAGCCAAAAC GTGTTTTATA AATGACAAAC CAAGCTAAAA TCGCAATCAA AATTGCTACG 5520

AAAGCTGGCA AAGAGGTTTT TGCAAAGAAT AGTTCACCAA TCACTGGAAT TTTTGACAAT 5580

CCTGGGAAAG AAAAATAACC AAATGATTGT GTAATCATAT CTGTTTGTCC TTTTTGATAA 5640

ATAACTTTTA CTAGGAAAAC CCCTAGCGCC GGTGCCATCA AGTTAATGAC AGTACCACTG 5700

ATAATATGAT CTGCTCGTAA GTTAATAGTT GCAACGGCAT GTAAGAGAGA AAAGACAATC 5760

CCGACAACGC CTCCTACTAA ACAAGCAATC CATGGTGTCG CCGCGCCAAA TGTATCAGCA 5820

AACGTCAAGT TAAAAACGAC CGAACTGAAA GCACCCATTA CCATAATACC TTCCAAACCA 5880

ACGTTAACCA CACCACTTCG TTCTGAGAAC GTTCCGCCTA ATGCTGTCAA GATGAGCGGA 5940

GTTGCATACA CTAATGTTTG CGTGATGATT GAGGAAATTG TTGCTAAATC CATTAGATGT 6000

CTCCCCCCTC TTGATTGTTT TTGTCTGGCG CCGCTTCAAT TGTTTCTACA ATCGCTACTT 6060

CTTTTTTACT ACCAGACGCT TTTGCTAACA AGAAACGAAT TAAATAACTA ATTCCCACAA 6120

AGAAAATAAT CGCTGCAATA ACAACATCAA CTAATTCGGT TGGTACACCA GCTATCAACG 6180

GCATCCCTTG TCCACCAAGT TTTAACATCC CGAACAGTAA TGCGGAAAGT AAAATCCCGA 6240

TTGAGCTACC TGCTCCTAAT AGAGACACGG CCATACCATC GAACCCAATG CTCAACGATG 6300

TGTTTTGAAT AAAGAAGTTT TGGAAAGTTC CCAATCCTTG GACCGCACCA CCTAAGCCAG 6360

CTAGCGTTCC TGAAATCATC ATGGATAAAA TAATCGTTCG CTTGCTGCTC ATTCCAGCAT 6420

ATTCAGAAGC AAATGGGTTT AACCCCACCG ACCGAATTTC AAAACCAAGC GTTGTTTTTT 6480

TCATCAAGAA CCACACAAGA ATAAGGAAAA TGAAGGCCAA GAAAATGCCG ATGTTCAAAC 6540

GAGAGCCATC ACTGATACTT GTTAAAAACG GTGAGCGTAA TGATGCATTC GCACCAATCA 6600

TTTTGGTGAT CCCTTTGTTC GCCATAATAC TTGGTTTCAT GACATTATTA GCAATATGAT 6660

TTCCTGTGTA TAATAAAATA TAATTCAACA TAATCGTAAC AATAACTTCA CTCGTCCCAA 6720

AATAGGCACG CAAGAAACCT GGAATGGCTG CTGCTAAAGC TCCTGCAACT GCCCCAACAA 6780

TAATCGCCAC CGGCAACGCA ATCATCCGGG GTGCATCAGG CATGGATAAA GCGACCCAAA 6840

TACTAGCCAC CCAACCACAT AAGGCTTGTC CAGAAAGCCC GATATTAAAG AATCCAGCCG 6900

AATTAGCCAC CGAGAATCCT AAAGCGGTAA AAATCAACGG TGCTGCGGTG ACAAAAATCT 6960

CACCAATACT TTTCTTTCCT TGGAAAGCTG TTTCAAACAT CGCTTGATAC CCAGCCACTG 7020

GGTCATAGCC AAAGATAACC ATAATAATTG CCCCTAAGAG GAAACCCATA ATAACAGATA 7080

ACACAGGGAC TAAAATATTT CTTATTTTTT CAGAACGATT ATTCAACAGT TTCACCTGCC 7140

TCTACTTTTG ATAACTCAGC ACGGGCTTCT TCCAATGAGT AGCCAGCCAT TAACAAGCCC 7200

AATTCATTTT CTGTCGTTTC TTTCGGATCA ACAATTCCGA CAATTTCCCC CGCATGAATA 7260

ACAGCGATAC GATCAGAAAC GTTTAAAATT TCATCTAATT CAAAACTTAC TAATAAAACA 7320

GCTTTATTTT TATCTCTTTG TTCAATTAGT CGACGATGAA TGTATTCAAT AGCTCCGACA 7380

TCTAAACCGC GTGTTGGTTG AGAAACAATC AACAAATCTG GATCGCGGTC CACTTCTCGG 7440

GCAATAATCG CTTTTTGCTG ATTCCCACCT GAAAGCGCCC GAGCTGGTAC TAATTCATTA 7500

GGAGTCCGTA CATCATATTC TTCAATTAAT TCACGGGCAT GTTTGTTAAT CATTCGATAG 7560

TTTAAAATAC CTGTTGAACT GTATGGTTTT TGGTAATACG TTTGTAGGCC AATATTTTCA 7620

GCCAATGTCA TCTCTAATAC CAAGCCATAT TTATGTCGGT CTTCAGGAAC GTGGGCGACA 7680

CCCATTTCCG TAATTTCCCG AGGTTTGCGA TTAGTAATCT CTTTTTCATT CAAATGAATG 7740

GACCCGTTTT CGACTTTGGT TAATCCTGCT AAGGCTTGGA TTAATTCTGT TTGTCCATTG 7800

CCATCAATCC CCGCAATTCC AACAACTTCC CCTGCTCGGA CATCTAAGCT AAGATTTTTC 7860

ACTGCATCTA AGCCGCGGCT TTCTTTCACT ACTAAATCTT TAATTGATAA AATTGTTTCT 7920

TTCGGTTGCG CTGGTTTTTT CTCTGTTTTG AAGGAAACAG AACGACCCAC CATCATATCG 7980

GCTAACTGTT GCGAAGAAAC GTCTTTCACA TTGACGGTGC CAATACTTTG CCCCCGGCGA 8040

ATAACCGTAC AACGGTCCGC CACTTTTTTA ATTTCATCTA ACTTATGTGT AATTAAAATG 8100

ATCGATTTGC CTTCATCGGT TAGTTCTTTC AAAATAATCA TTAATTCATC AATTTCTTGT 8160

GGCGTTAATA CTGCGGTTGG TTCGTCAAAA ATTAAAACGT TCGCACCACG ATATAAGGTC 8220

TTTAAAATTT CAACCCGTTG TTGCATACCC ACAGAAATAT CGCGAATATA GGCATTGGGA 8280

TTTACATCCA ACCCGTATTG CTTAGACACT TCTTCAATCT CTTTTTTCGC TTTTTTCCGG 8340

TCTAAAATAC CGCCATGGGT AGGTTCACTA CCTAAGATAA TATTTTCAGT CACAGTAAAG 8400

GCTTCAACTA ACATAAAATG CTGATGGACC ATTCCGATTC CTAAACGATT GGCTGCAGTA 8460

GGACTTGTAA TGTTGACTTT TTCGCCATTC ATATAAATCT CACCTGAGGT CGGTTCCAAT 8520

AGCCCTGATA AGATATTCAT CAAGGTCGAC TTGCCTGcAC CATTTTCACC TAGCAGTGCG 8580

TGGaTTTCAC CAGGTTTTAC TTGTAAATCA ATTTGGTCGT TGGCTTTGAA CGTCCCAAAT 8640

TCTTTGGTGA TTTTACGCAT CTCAATTACG TTTTCTTCTT GAGACAAGGT CTATTCACTT 8700

CCTAACAAAA TTATATTGGC AATACTTTAG AGGGTAATTT TGACCCTTAA CTAAGTAGTT 8760

GAAGAACTTC AACTATTTAG TTAAAAATCC AAACATCATA AAGGGTTCCC TTGGCAGCCG 8820

TCGCTGCCAA GgACCACTTT ATTTTTAGAA GGAACAATAA TGAGCCCTTC TTTTTTCATT 8880

TAGGAGCTTA AATTATTCTG GTTTTTCTGG GACTTTTACG TCACCTGAGA TTACTTTATC 8940

TTTTGCTGTT TTAACAGCTT CTTTTGTTTT GTCGTTTAAA TAGCCGTCTG TTAAGTCAAC 9000

GCCACCATCT TTTAATCCAT AAACTAAATG TTCGCCACCA GGGAATTTGT CTTCTAACGC 9060

ACGGTTGGCA ATATCTTGAA CCGCTGTGCC GACACCTTTA AGCGTTGAAG TTAACGTGAA 9120

GTTGTCTTCT TTGCCGTCTT TTGTTTTGTA CTTGCCATCA GCATCTTGAT CGCGGTCAAC 9180

GCCGATTACC CAAACTTTGT CGCCAGAACC TGATTCATTC AAGTCTTTTG CTTCTTGGAA 9240

GACCCCTTGT CCAGTCGCAC CAGAAGCATG GAAGATGATA TCAACGCCGT TTTGGTACAT 9300

TGCAGCAGCT AAAGCTTTCC CTTTGGCAGG ATCAGCAAAT GAAGCCGCAT ATTTCGTATC 9360

AACAGTAATT TCTTTACCTA ATTCTTTCGC AGCATCAGCC ACACCTTTTT CAAAACCAGC 9420

TTGGAAACGG TCAATTACGA CCCCTTCTTC ACCACCAACA AAACCGACTT TGTTCGTTTT 9480

TGTTTCATTT GCAGCAGCAA CACCGGCTAA GTAAGCTGCT TCATTATCTC TAAATGTTGC 9540

AGAAACGACA TTCTTTTTGC CATCGATTTG ATCATCGATT AAAACAAAGT TTGTATCAGG 9600

GTTGGCATCT GCTGCAGAAG AAATTGCATC TTTTAGCAAG TAGCCAATAC CAAAGATTGT 9660

GTTGAATTTA CTTGATACCG CTTGGTCAAT ATTGGTTGTA TAGTCAGCTG CATCATTCGA 9720

TTGAATATAT GCATACCCTT TTGAACCTTC TGGTAAATCA TGTTCTTTAC CCCAAGCTTG 9780

CAATCCTTCC CAAGAAGATT GGTTGAACGA CTTGTCATCC ACGCCGCCTG TATCTGTAAT 9840

GATTACAGCA CTATGCGCTG CATCCCCTTT GCCACCGCCG CTTTCAGCGG TTTTGCCTTT 9900

GCCACCACCA CATGCTGCAA GTGAAACTGA TAAACCTAAT GCAATCAAAC TAAAACCGAA 9960

TAATTTTGCT TTTTTCATTT CTGTAAAGCC TCCTAAAAAT AGTTTAATGT ATTTAAACAA 10020

CTAAGTTGTT TCTAAGGTTG GTCTGGGACT TTTACGTCAC CAGAGATGAT TTGTTCTTTG 10080

GCTTCTTTAA CTGCCGTTTT GGCTTGATCG CTTAGGGCTT CTGTTGTTAA ATCAACACCG 10140

CCATCTTTCA ATCCATAAGT TAAATGTTCG CCGCCAGGGA ATTTATCTTC TAATGCAAGA 10200

TCAGAGATTT TTTTCACAGC AATATTTACA CCTTTAATGG TTGAAGCTAA CGTTAAGTTA 10260

TCATCCTTGC CATCTTTGGT TGTATATTTA CCGTCTTCGT TTTGGTCACG GTCAACACCA 10320

ATCACCCAAA CTTTATCTTT TGAACCTGTT TCATTTAGCG CTTTGGCTTC TTGGAAGATT 10380

CCTTGTCCTG TAGTCGCCGC TGCGTGATAA ATAATGTCGG CACCTGCTTG ATACATAGAA 10440

GAAGCTAAGG CCCGTCCTTT TGAAGCGTCT GCAAATGTCC CTGCATAAGT AGACGTAATT 10500

TGAATATCTT TGCCTAATTT TTTGCCAGCG TCTGCTACAC CTTTTTCAAA ACCAGCTTGG 10560

AAACGGCCAA TAACAGGTCC TTCAACCCCG CCGATAAATC CTACTTTGTT GGTTTTCGTA 10620

GTGTTAGCTG CTGCCACACC TGCTAAATAC GCAGATTCAT TATCTCTAAA TGTTGCAGAA 10680

GCGACGTTGT TTTTCCCATT AACTGTATCG TCGATTAGAA CGAAATTCGT TTCTGGGTTG 10740

GCGTCCGCTG CATCTACTAC TGCATTTTTT AGTAAATAGC CAATGCCAAA AATCGTTTTA 10800

AATTGACTTG AAATGGCTTG GTCAATGTTT GATGTATAAT CAGATGCTTC ATTTGATTGA 10860

ATATACGCGT ACCCTTGAGG TCCTTCTGGA AGTTTGTGCT CTTTGCCCCA CTCTTGCATT 10920

CCTTCCCACG CAGATTGGTT AAAAGATTTG TCATCAATGC CGGCTGTATC TGTGACCATT 10980

ACGACACTAT GTTGTGGATC ACCTTTTGTG TTGGCGTTCT TTTTGCTATT GCCGCCACAT 11040

GCAGCTAACC CAATGGTTAA GGCAACTGCT GCCAATCCTA GACCTAAAAA TTTTGCTTGT 11100

TTCACTTTCC GAAAAGCCCC CTACTATTTT ATCTATTATT TCATTCGCAT TTATGCTATT 11160

GAAATCTACT TTGAGCCCAA CGAAACCTTA TGTCCTTTTT TTAAAAATCT TTATCTGTAA 11220

ATGAATAGGG CAATAATGCC TCAACTGTCA TTGTTTTAGT CGCGCCATCT TTACCAACTA 11280

GCGTGACCGG CATTTGTGGA TCAAAGAATT CAGCCATTAC TTGACGACAA GCGCCACATG 11340

GAGAAATGGG TGCCTCTGTA TTACCTGCTA CAACTAAATG TTGGAAGGTC CGTTCATCTT 11400

CTGAAATGGC TTTAAAAATC GCCGTTCTTT CTGCACAATT TGTTAAACCA AAGGAGGCAT 11460

TTTCAATGTT GACTCCTTGA TACATTTTCC CGTCTGCAGT CACTAAGCAA GCGCCCACAG 11520

GAAAATTAGA ATACGGTACA TAGGCGTGTG CCAATGCTTC TACGGCACAA TCAAGCCATT 11580

CTTGTTTTAC TGTCATCTTT TTCGCCACAC CTTCGTTTAA TTTGAAATAC GAATTAATAC 11640

CCTTCGCCAG TTGCGCCAGA AACAATTGCT ACACCCGTAC TTGCACCGAT GCGTGTTGCC 11700

CCCGCTTCAA TCATTGcTAA CGCTTCTTCT GCATTATGCA CGCCACCAGA AGCCTTAACG 11760

CCCATGTCAG GACCAACTGT CTCACGCATT AAACGAATGT CTGCTACTTT AGCGCCGCCT 11820

GTTGAGAAGC CAGTAGATGT TTTAACGAAA TCGGCACCTG CTTCTTTGGC TAATTCACAA 11880

GCTTTAACTT TTTCTTCGTC TGTTAATAAT GCTGTTTCAA TAATCACTTT GACTAACGCT 11940

TTTCCTTTGG CTGCATCTAC TACGCCTTGA ATGTCTTGAC GTACGTAATC GTATTGtTGT 12000

GATTTTAATG CACCAATATT AATCACCATA TCCACTTCAT TCGCACCATT TTTAATCGCA 12060

TCAGCTGCTT CGTATGCTTT AACTTCTGGT GTATTTGCCC CCAAAGGAAA GCCAATAACG 12120

GTACACACAG CGACGTCGGT ATCTGCTAAT TGCTCGCTAG CAAAAGCGAC CCAACAAGGA 12180

TTGATACATA CAGAGAAAAA GTTGTACTCC TTAGCTTCGT CAATAATTTT TTGGACAGCG 12240

GCTTCTGTCG CTTCTGGTTT CAAAATAGTG TGGTCAATCA TGCGGTTTAA TTCCATGATG 12300

AGAATTCCTC TTTTCTTTTT GTTTTATTCC GTAATAATAT CGTGAATCAA AATAGGCTCT 12360

TCGCCGGTTG GTCCAATTTC AATGTTTTTA TAAAGCAATT GTTCGACTTC TGTCACGTCT 12420

TCACGATTGG CATAAATGGT TAGTAATGAT TCGCCTTTTT CAACTGGGCT CCCTACTTTT 12480

TTGTGAAGTT TCAAGCCAAC TGCAAAATCA ATCGTATCTT CTTTCGTAGC ACGACCAGCG 12540

CCCAGCAACA TTGCGGCAAC GCCAATTTCA TTGGCAACGA TTTTTTGTAC CACGCCTGTT 12600

TCTTTGGCAG GTAAATCAAA TTGATAAGTT GCGGTTAATA ATTTTTCTGG CTGGTCAACA 12660

ACGGAACTGT CGCCGCCTTG ATTTTCAATC ATTTCTTTAA ATTTTGCGAG CGCTCGACCG 12720

CTTGTTAATG CTTCTTCTAA TAAAGCCCGT GCTTCGTCTA ACGTCTCGGC TTTTTTAGCT 12780

AAAACGACCA TCTGACTTCC TAAGACATAA CACATCTCCA CTAAATCTTC TGGCCCATTT 12840

CCTTGCAAGG TCTCGATTGC TTCAACGATT TCTAAACTAT TTCCGATGGC CTCCCCTAAA 12900

GGTTGTGACA TGTCTGAAAT TACTGCCATC GTTTGGCGAT TTGCCAGTTT TCCAATGCGA 12960

ACCATTGTTT CCGCTAATCG ACGAGCACTT GCTAGGTCTT TCATAAAGGC CCCTTCACCA 13020

GTAGTCACAT CTAAGACAAT CGCATCTGCC CnTGCGGCAA TTTTTTTACT CATAATGGAA 13080

CTTGCAATTA ATGGAATTGA GTCAACAGTG GCGGTCACAT CGCGTAACGC ATAAAGTTTT 13140

TTGTCTGCTG GTGCTAGATC ACCTGACTGA CCAATTACTG CCACTTTTGT ATCGTTCACC 13200

AATTGAATAA AACGTTCTTC TGTCAATTCA ACTTTAAAAC CAGGAATCGC TTCTAATTTA 13260

TCCAAcGTAC CGCCCGTGTA GCCTAAGCCA CGACCAGACA TTTTTGCTAC ATTGACCCCG 13320

ACACTAGCAA CTAACGGCGC TAAAATAAGC GTCGTAGTGT CGCCTACGCC ACCTGTAGAA 13380

TGCTTGTCGA CTTTGATTCC TTCTATAGAA GAAAGATCCA CAATTTCGCC TGAATGGACC 13440

ATTGCTAACG TCATTTGTGT AATTTCTTCG TCTGTCATAT CTTGGTAAAA AACAGCCATT 13500

AAAAAGGcAC TTACCTGaTA ATCTGGaATT gAACCmTCTG tGtATCCGTT GATAATAAAT 13560

TGAATTTCTT CGCTTGTTAG TTGATGCCCG TCaCGTTTCT TTTCAATAAT GTCTACCATT 13620

CTCATGAATT CTTTCCCTCC GATTTCCCGa TAATGTATTA TACAACAAAA TCGTTCAGCT 13680

TTAAACCATT TAGTAAACCA GTTTACTATT ATTTTTTTCT TGTGAAAACG CAAACAACAC 13740

GAACATTCTA TTTCTAAGCT GTTTTTTAAA TCGGTTTTGT GCTTTTTCTA GCAATAAACG 13800

TCGCATCGAA AACTTTATTT GGAACTTTCC CTGTTGGATC TGCAATAGCT TCCACCAAAA 13860

ATTTTGCTGC ATAATAGCCA ATATCAAAAC TCGGCTGATA GACAGTTGTT AAACTCGGCA 13920

CCAAATATTC GGAAATTTCC AAGCCATCAA AACCAACCAC CGAAATATCG TCAGGAATTT 13980

TCTTGCCCGC TTCTTCAATG GCTTGATAGG CCCCCATCGC CATTTCATCA TTGCCACAAA 14040

AAATTGCTGT GATTTCTTCA TTTTTCAAAA CTTCTTTCGC CGCAGCATAG CCGCCATGCA 14100

AGTTCAATTC CCCAGCACAA ACATAATGCG GACGAAAAGG TAAATTATTG TCCTGTAACG 14160

CATGCTGATA CGCCGTCACT CGTTCCGTCA AACGATAATA GCCGGCACTT TCACGCAACA 14220

TGCCAATATG CTGATGCCCT TGTTGAATCA AATGATGGAC CGCTTGATAA GCACCTTCGT 14280

ACTCTTTAAC GATTAATCGT CCACTATCTC GTGGATTAAT CCCACGATCG ACAAGAATGA 14340

CAGGAATCGG ATGTTCTCGT CGGTCTTTTA AAATATGATT TTCTGGCAAA ACATTCGGCG 14400

TGGCAAAAAT AATGCCATCC ACCGACCGAT GAATCAACTC GCTAACATAT TGCATCGCTT 14460

GCTCGGAATC ATGCTTTGAA TTGCATAACA AAATCATGTA GCCTAACGAA TTTAAATATG 14520

TTTCTACACC TTCAATCACT TTTGAGAAAA AGAAATCCGT CACATCTGGC ACAATCATAC 14580

CAATGGTTTT GGAATGACGG TTGATTACAT TCGAAGCAAA GTAATCTGGT TTGTAATGAT 14640

TTTCAGCGAC TACCGCCAAC ACCTTGTTGC GGGTTTTTTC ACTGAAACGA CTCCCTTTAT 14700

TATTCAGAAT TTGTGAGACG GTTGTGACAG AAACACCAGC CATATCAGCA ATTTCTTTAA 14760

TCGTTAGTTT CATCGCGCCC CTCCCAACGT CTCCTTTTGC ATTGACAGAC ATTTGTAGAC 14820

GTGTTACAGT CTTCAAAATA CTTATCGAAT CTTTTACATT TTAACAAAAA AAAGACAAAA 14880

GAAATAGACT AAAAAGATTA AAAAACAAAA GAAAATGGAG GCTATTTTTT TCATGGAAGA 14940

ATCCCGTGTC CGTCAATTAC CAAAAATCGA GCTGCACTGT CACTTAGATG GGTCGATTCG 15000

ACCAACCACT TTACGCACCA TTGCTGAAAA ACAAAACATT CCATTACCCC AAGACGAACA 15060

AGCCTTGAAA GAGTTAGTGG TTGCCCCTGA AAAATGCACC GATTTGAATG ATTATTTAAC 15120

TCGGTTCGAC TTCGTGCTCA CCTGTCTACA AACAGCTGAG GCTTTGCAAG CAGCGGCGTA 15180

TGACGTGATT AGCCAAGCTG CCGAAGATGG CGTGGCTTAC ATAGAGGTGC GTTTTGCGCC 15240

TTCCCAACAT ACAGAAAAAG GCTTACGGTT ACCTGAAATC GTGACTGCTG TTTTAACTGG 15300

CTTAAAACAA GGCGAAGAAG ATTTTGGTGT AAAAAGCAAT GCACTGTTAT GCGGCATGCG 15360

GCATGACCAA CAACAAGCCA TAGAAAAAAT CGTTCACTTA GCGCATGATT TCAGAGAGAC 15420

AGGCGTTGTC GGCTTTGATT TAGCTGGCAA TGAAGTCGAT TTCCCTCCAT ACACATTTGA 15480

AGACGTATTA GCTTTAGCCA ACCAATTGTC AATTCCATTA ACACTCCACG CTGGGGAATG 15540

CGGCTGTGGC AAAAATGTCG CTGATGCTGT CACTTTAGGT GCTACTCGGA TTGGTCACGG 15600

TATTGCTTTG AAAGATACAC CCGAATATTT GGCGCTCTTA AAAGAAAAAA AGGTTCTTTT 15660

AGAAATGTGT CCAACTAGCA ACTTCCAAAC AGGAACCGTC AAAACATTGG CTGAATACCC 15720

TTTTCAACAA TTTATAGAGG CTGGCTTAGC CGTCTGCATC AATACCGATA ATCGAACTGT 15780

TTCAGATACT ACGTTAACGA AAGAATTCAT GAAATTGGCC ACTTGGTATC AATTAAGTTA 15840

TGACGAAATG AAGCAACTCA CTAAAAATGC GCTTGCAGGT GCTTTTTTAT CGCCAGACGA 15900

AAAAAAACTG TTGAACCAAA AAATTGATCA AGCGTATCTA TTCTAAAAAA GAACGCCACT 15960

AAAGTGGCGG TTCTTTTTTA CTCTGCTTCT TTAACAtTCG TAAAATATAA TAGCCTTTGT 16020

CTTTTGTCAC AATTTCAACA TTGCCAAATA CTTCAGCCAT TTTCTTTTCT GCACTAGGTG 16080

CACCTTGTTT CTTTTGAATA ACCACCGTTA GTGTGCCACC AACTTTCAAC AACGGAAGAG 16140

CCCCTGTTAA AATGCCATGG ACAACTTTTT TGCCTGCACG AATGGGCGGG TTGCTAACAA 16200

TCGCCGCGTA GGTCGTTTCG TTTAAGGTTT CATACACATT GGAAGAATGG ATATCCACAG 16260

TAGTAATTTG GTTACGTTGG GCGTTCATTT GAGCTAGCCC AACTGCCCGT TGGTTGACAT 16320

CCACCATTTC CACCAGTCGA CCAGTCGCCG CTGCTAAGGA CAACCCAATT GGTCCATAGC 16380

CACAACCAAC ATCTAACAAA CGGCCTGCTG GTAAATTTTC CCATTCAAAG GCATCAATCA 16440

ACACACGAGA ACCATAATCC ACTGTTCCTC TTGAGAAGAC CCnG 16484

(2) INFORMATION FOR SEQ ID NO: 159:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 8160 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

CCGAGTAGTA CCAGAGAGGm AnTCCATGTT wAAAAcATTA CTTCTTTmCA ATCTAGTTAC 60

AGTAATATGa AATTTTGTTG AAGTGTTTTT AATTCGCCAG AAAAAAAGGT ATACTTGTTA 120

TAATrACAAA ACTAAAAAAT TACATTTACA TATAAGTTAA AAAAGGAAAG TTGGGGACGT 180

ATCAATGAAG AAAGAATCAA TGTCACGTAT CGAAAGAAGG AAAGCACAAC AAAGAAAGAA 240

AACGCCAGTA CAATGGAAGA AGAGCACTAC TTTATTCAGC TCGGCGTTAA TTGTTTCATC 300

TGTAGGAACG CCCGTTGCGT TACTACCAGT GACTGCTGAG GCAACAGAAG AGCAGCCAAC 360

AAATGCGGAA GTTGCCCAAG CACCTACTAC GGAAACTGGC TTAGTAGAGA CACCAACAAC 420

AGAAAcTACG CCAGGAACTA CGGAACAACC GACAACGGAT TCGTCAACAA CGACTGAATC 480

GACAACTGAA TCATCAAAAG AAACACCAAC AACACCAAGT ACCGAGCAAC CAACAGCTGA 540

TTCAACTACA CCTGTGGaAT CAGGaACGAC TGATTCTTCA GTAGCAGAAA TTACGCCAgT 600

AGCTCCTTCA GCAACCGAGT CTGAAGCAGC GCCTGCGGTT ACACCCGATG ATGAAGTAAA 660

AGTACCAGaA GCTAGAGTAG CTTCTGCGCA AACTTTTTCA GCGTTATCAC CGACGCAAAG 720

TCCTTCAGAA TTTATTGCCG AGTTAGCTCG TTGTGCACAA CCTATTGCGC AagCCAATGA 780

TTTATATGCA TCAGTGATGA TGGCTCAAGC AATCGTTGAA AGTGGTTGGG GAGCAAGTAC 840

GCTATCTAAG GCACCAAACT ATAACTTATT TGGGATTAAA GGCAGCTACA ATGGACAATC 900

TGTCTATATG GATACATGGG AATATTTAAA CGGCAAATGG TTAGTGAAAA AAGAACCTTT 960

CCGTAAATAT CCTTCTTACA TGGAATCATT CCAAGATAAT GCGCACGTGC TAAAAACAAC 1020

TTCTTTCCAA GCGGGCGTTT ACTATTATGC TGGGGCTTGG AAAAGCAATA CAAGCTCGTA 1080

CCGCGATGCA ACTGCTTGGT TAACAGGTCG TTATGCGACA GATCCTAGCT ACAATGCTAA 1140

ATTAAATAAT GTCATTACCG CATATAACTT AACTCAATAT GATACACCAT CTTCTGGTGG 1200

AAATACTGGG GGCGGAACAG TTAATCCAGG AACAGGCGGC TCGAACAATc AATCAGGAAC 1260

GAACACGTAC TATACTGTAA AATCAGGaGA TACCTTGAAT AAAATTGCCG CsCAATATGG 1320

yGTGAGCGTT GCTAATTTAC GCTCATGGAA yGGCATCTCT GGCGATTTAA TTTTCGTTGG 1380

TCAAAAACTy ATCGTGAAAA AAGGTGCTTC AGGTAACACT GGTGGCTCAG GCAGCGGTGG 1440

tTCTAACAAT AATCAATCAG GAACGAAyAC GTACTAyACt tAAAtCAGGs GATACCTTGA 1500

AyAAAATTGC CGCCCAATAT GGCGTGAGyG TTGCTAATTT ACGCTCATGG AATGGCATyT 1560

CTGGCGATTT AATyTTCGyT GGTCAAAAAa TtATyGTGAA AAAAGGTACT TCAGGTAACA 1620

CCGGTGGCTC AAGCAATGGT GGTTCTAACA ATAATCAATC AGGAACGAAT ACGTACTACA 1680

CGATTAAATC GGGCGATACC TTGAACAAAA TTTCTGCACA ATTCGGTGTT AGTGTGGCTA 1740

ACTTACAAGC CTGGAATAAC ATCAGCGGTA GTTTGATTTT TGCTGGTCAA AAGATTATCG 1800

TGAAAAAAGG CGCCAACTCA GGTTCAACGA ATACGAACAA GCCTACGAAT AATGGTGGCG 1860

GTGCGACAAC ATCCTACACG ATTAAATCAG GTGATACGCT GAATAAAATT TCTGCACAGT 1920

TTGGCGTGAG TGTTGCTAAT CTACGTTCAT GGAACGGGAT CAAAGGCGAT TTAATTTTTG 1980

CTGGTCAAAC AATCATCGTG AAAAAAGGCG CTTCTGCAGG TGGCAATGCT TCTTCAACAA 2040

ATAGTGCATC AGGCAAACGC CATACAGTTA AAAGCGGTGA TTCACTTTGG GGCTTATCAA 2100

TGCAATACGG AATCAGCATC CAAAAAATCA AACAATTAAA TGGCTTAAGC GGGGATACAA 2160

TTTATATTGG TCAAACTTTA AAAGTTGGTT AATTTAAGAT TGAAAAAAAT AGCTATCTTT 2220

GGTAACATGA GTTTTATAAT AGAAAAAAAC AATGAGAAGG AATAGTAGAA AGAACCATGT 2280

GATAGAGAGC GTATGGCTGG TGGAAATACG TACAGAAGCT TTTGAACTCG CCTTTAAGTT 2340

ACTTTTTTGA ACAAACCAAG TAGGAAAAGT CGGTGACGAT CGTTAAGACG TTTGAGGTTA 2400

AGGTGACGGT GACTGACCGT TTCTTTGACG AATTTAGGTG GTACCACGTT GCATTTGTAT 2460

GTTCACGTCC TATAGGAATT TTTCCTATAG GACTTTTTTT TATCACTAAG TGGCTAGTTT 2520

AAGTCAGAAT AGTGACAATG GCTGCAGGTG GCAAATTTTA AGAATGAAAA ATTTTATATT 2580

ACTAGGAGGA ATAACATGAG CTACAATCAC AAAGAGATTG AGAAAAAATG GCAAAAATAT 2640

TGGGCTAAGA ACAATTGTTT CAATACATTG GACGACCCAA ATAAAGAAAA ATTTTATGCA 2700

CTAGATATGT TTCCCTATCC ATCTGGACAA GGCTTACACG TAGGTCACCC GGAAGGCTAT 2760

ACAGCAACCG ATATTCTTTC ACGTATGAAA CGTGCGCAAG GCTATAATGT GTTGCATCCA 2820

ATGGGCTGGG ATGCGTTTGG TTTGCCAGCA GAGCAATATG CGTTAGATAC AGGAAATGAC 2880

CCAGCTGAAT TTACTAAGAA AAATATCGAA ACATTCCGTC GCCAAATTAA TTCACTAGGA 2940

TTCAGCTATG ATTGGAATCG TGAAATTAAT ACCACTGATC CTGAATATTA CAAATGGACA 3000

CAATGGATAT TTACAAAATT ATATGAAAAA GGGTTAGCTT ATGAAGCAGA AGTTGCGGTT 3060

AACTGGGTCC CTGAATTAGG AACTGTTATT TCAAATGAAG AAGTCATTGA TGGAAAAAGT 3120

GAACGTGGCG GTTATGATGT GGTTCGCCGA CCAATGCGTC AATGGATGCT GAAAATTACT 3180

GCTTATGCAG ATCGCTTATT AGAAGATTTA GAGCTTGTTG ATTGGCCAGA GAGTATTAAA 3240

GATATGCAAC GAAATTGGAT TGGACGTTCT GAAGGAGCCA ATGTGACCTT TAAAGTCGCT 3300

GGCACAGAAG AAAGTTTCAC GGTGTTTACA ACCCGTCCTG ATACCTTGTT TGGTGCAACC 3360

TATACTGTTC TAGCTCCTGA ACTAGAACTA GTGAAAAAAA TTACGACACC TGAACAAACA 3420

GCAGCTGTAG AAGCATATAT TGAAGAAACC TCAAAAAAAT CTGATTTAAA TAGAACGGAT 3480

TTAGCAAAAG AAAAAACAGG TGTTTTCACA GGTGCGTATG CTATAAATCC AGTCAATGGC 3540

CAAGAAATTC CAATTTGGAT TGGCGATTAT GTTTTAGCAA GCTATGGCAC AGGCGCAATC 3600

ATGGCGGTCC CAGCACATGA TGAACGGGAT TACGAATTTG CGAAAACATT TGGCATTGAT 3660

ATCCTACCAG TAATCGCAGG TGGCGACATT ACAACAGAAG CCTATACAGG GGATGGACCG 3720

CATATCAATT CTGATTTCTT AAATGGATTA AACAAAGCAG AAGCCATCGC TAAAATGAAT 3780

GAGTGGCTAG AAGAAAATCA CGTAGGGAAA AAAGAAGTAT CTTATCGTTT ACGTGACTGG 3840

```
TTATTCTCTC GTCAACGCTA CTGGGGTGAA CCAATTCCTG TGATCCATTG GGAAGATGGA    3900
ACAACCACAA CGGTTCCTGA ATCTGAGTTA CCTCTACGTT TACCAGTAAC ATCGGATATT    3960
CGCCCAAGTG GAACTGGGGA ATCGCCATTA GCAAACATTG ATGAATGGGT CAATGTCGTC    4020
GACCCTGAAA CTGGCATGAA GGGAAAACGT GAAACGAATA CTATGCCACA ATGGGCTGGA    4080
AGCTCTTGGT ATTACTTACG ATTCATTGAT CCTCATAATA AAAATGAAAT TGCTGATTTT    4140
GAAAAATTAA AACGTTGGTT ACCAGTTGAT ATCTATATTG GTGGTGCCGA ACATGCGGTG    4200
CTGCATTTAC TTTATGCTCG TTTTTGGCAT AAATTCTTAT ATGATATTGG TGTGGTTCCT    4260
ACCAAAGAAC CTTTCCAAAA ATTATACAAC CAAGGTATGA TTTTAGGAGA AAACAACGAA    4320
AAAATGTCTA AATCACGTGG CAATGTTGTA AATCCCGATG ATGTGGTGGC TAAATATGGT    4380
GCGGATACGT TACGTCTTTA TGAAATGTTC ATGGGCCCAT TAGATGCTTC CATTGCTTGG    4440
AATGAAAATG GATTAGAAGG AAGTCGTAAA TTCTTAGATC GCGTTTGGCG TCTGATTGTT    4500
GATGAAGAAG GCAAAATGCG TGACCGAATT ACCACAATTA ATGATGGCCG TTTAACGAAA    4560
GTTTATCATC AAACGGTTAA AAAAGTGACA GAAGATATGG CAAACTTGCA CTTTAATACA    4620
GCGATTTCTC AATTAATGGT TTTTGTGAAT GAAGCCAATA AAGTGGATGC CTTACCTTAT    4680
GAATATGTGG AAGGATTTGT CCAATTACTT GCGCCAATCG CGCCACATAT TGGTGAAGAA    4740
CTATGGCAAA TTTTAGGTAA CGAGGAAAGT TTAACTTATG TCCCTTGGCC AACCTATGAT    4800
GAAACGGCCT TAGTAGAAGA TGAAGTGGAA GTAGTTTTCC AAGTGAACGG AAAATTACGT    4860
GGCAAACAAA ATGTCGCTCG TGGGTTAAGC AAAGAAGAAT TAGAACAAAT TGCAATGAAC    4920
CATGAAGCTG TTAAAGAATT TATTGAAGGA AAAACAGTGC GCAAAGTGAT TGCTGTTCCA    4980
GATAAATTAG TAAATATTGT TGCAAATTAA GTTTACATGT TTTTATTAAA AAAAGAGCCG    5040
GACGTATCAA AACGTCAGGC TCTTTTTAAG GTTGTTTTGT TTTACCTTTT TGATTCATAT    5100
GTTCTTTACA TGTGATTGTT ATGACTGAAA GCAAGGCGAT GAAGAGAAAA AAGTAACCAA    5160
GCGGCGTACA GAAAAAGGCG GGTTCTTTTA AAAGTCCATC GCTAGATAAT GTTGATCCAA    5220
CTAAAATAGA GATCATAAAT AATAAAAGAC TAATCAAGCC AAGAAATAAG CTAATACGAT    5280
ATTTTTTCAA CTTCAGTTCC TCCTTTAATG AACTATCACT AAGATAAATC AAAAAGGTTA    5340
ACTTTTTTTT AGGCAACTGT TAAAACTTAG ATAAGAAATC TATTTTCTAA GTTTGTCGTT    5400
GTTGGTAATT GGTTAAGATC ATACGTGAAA AGGGACCGAC AATCAGCAAT TGATAAGGAA    5460
GGGCAACAAT AAAGTTCAGG CCCCAAGCTT GGAAATACGC ACTTAGTGAC AACGTCTGGC    5520
CTTCCATTAA TAAGCCAAAT AAAGACATAC AAGTCACCAT GCTGATAATC ATGAGACTCG    5580
AAATGGTCAG AATAAGTAAT AGCCGTTTTT CTTTATTAAA GGGTAAAGAA AAAGCGATTT    5640
TCTTAGCTAA GACACCAACG ATGAAAACAT CTAAAATAAA AGCAACGATA AATCCAGGAA    5700
GCAAACCTTT TGATAAGGCG TACAATGAAA ATTCTCCATG CAATAATAAA TTATAGGTGC    5760
TCATTCCGAA AACCATCAGA AAGCACATTA ACGTAgTAAA AAGCATTCCT TCTTTTTTAT    5820
```

TAGTAGGCAT TTTTTTCTCC TTTCATTTAT TGAGAGTAAT GATTGTTCTA GGCAGAAAaG 5880

GACAAATTCT CTCGACTTTT ACTAGTGTAG CAAAAATAAA AGTTCTTCAT AAGTTAGAAT 5940

TTGCTAAGAA TTTGATTCAA ATAAAAAGGT GTACTAAGTC AGCAGTACAC CTTTTTATTC 6000

GTTATTTAGT TTATTTTTGT GTTATTTGTA ACGTCCTCTC CAAACCATTT TTGAGAAATT 6060

TTACTAAGGG TGCCATCTTT TCGTAACGTT TCAAAGGCAG TATTGATTTT TTGGACTAAT 6120

TGATTGTCTG ATTTGCGGAC GCCCACAGCA AAATCTTCAT TGTCATAGCC TACATGAGAA 6180

ATAGTATAGT TTTTTAAATT ATCTTCGTGG GAAAGATAGT AGTTGGCGTA AACGCGATCG 6240

ATTAGGAGTC CGTCAATTCG ACCAGATTTT AAATCTAAGA AAGCTTCATT AAAGCCGTCA 6300

TATAAAATAG GTGTTTGGTC TTTAACAAAT TTTTTCAAAA CGTCAGGCTG ACTTTCGAAG 6360

CCATCATAGC CAGAAGAGCC GTTTTGAACC CCTAAAATTT TGCCTTGCAT GTCGCTCGCT 6420

GTTGCAATGT TTTTTTCTTT TAAAGAAACA AGTACTTGGT CGTTCGTCAT GTAAGGTTGT 6480

GTGAATTGAA CTTTTTCGGC CCGCTCGCTC GTTTTAGTGT AGCCGTTCCA AATAAGATCA 6540

ATGGTTTGAT TTTGTAATTC TGTTTCTTTC ATAGACCAAT CAATCGGTTG GAAGTCAACG 6600

GAAATGCCAT AAAGTTTAAA AACCGCTTTG GCTAAGTCGA CATCAAAGCC GACAATTTTG 6660

CCTGATTTAT CTTGAAAACC CATGGGCACA AAGGAGTCAT CTAAGCCAAT AATAATCCGT 6720

TTTTCTTCGT TAATCCGTGT CCATTGATCT TCGTTGCTTT TTCTTTTACC ACAACCTGCA 6780

AATAGGAGTA AACTACAACA GATAACCAGC AAGGCTAAAG AATATTTCTT TTTCATCATA 6840

AACTCCTTTC TATTGGACGG GGGCAACCGT CAAAAGATTG TCTGCAATTT TTTCAGCGAA 6900

AGCCATATCG TGGGTCACGA CAATCTGTGT CATGCCTTGT TTCCGTAAAT CTAAAATAAC 6960

TTCTTCCACT TGTTGGCGTA AAGCTGGATC TAGCGCGCTC GTTGGTTCAT CGTACCCAAG 7020

AACTTTTGGC TTCATTGCTA AGGCACGAGC AAGTGCAACT CGTTGTTTTT GTCCACCGGA 7080

AAGTTGGTAC GGATAGAGTT TTTCTTTACC AGCCAAGCCC AATTTTTCCA GTAATTCTAA 7140

AGCTTCTTGC TGACTTTTTG CTTTTTCTTC TTTTAGTACT AATGTAGGTG CTAAGGTAAT 7200

GTTTTCTAAA ACTGATAAGT GAGGGAAAAG TTGGAAATCC TGGAAAACGA CCCCAACCAC 7260

TTGATCGGCA TTGTCCATTT CaGCAGGATT GAAAGGAACA CCATCCAACA AAAGCTCACC 7320

AGAATCAATG GTTTCTAAAC CAGCCAAACA ACGAAGTAAA GTGGTTTTTC CGCCGCCTGA 7380

AGGACCAACA ATCGTTAAAA TTTTACCATC TGGAATTTCT AAATTTAAGT GATCAATAAT 7440

GGTCCGGTTG TCAAAATGTT TGGTTAAATT TTTAATAGTT AACATGGTAG AGTCCTCCTA 7500

TTTATAATAT TGATAGAATT TTTCAATTTT CTTAGAAGCA ACAGTCAAAA TTGCCGTAAA 7560

GAGTAAATAA ATCACTCCCA CTAAAGCTAA TGGCACTAAT GTAACATCGC GGCTCATCGC 7620

AATTTTACCA GCACGTAAAA GATCGCCTAA GCCAAGAACA TAAATAAGAG AAGTATCCTT 7680

GACCAAATTG ATAATTTCAT TTCnTACGGA AGGCAGCACG ATTTTAACAA CTTGTGGCAA 7740

AATAATCCGC GTGACTGTTT GAAAACGTGT TAAACGTAAA ACCTTAGCAG CTTCATATTG 7800

ACCTTGGGGA ATTGCTTGAA TCCCCCCGCG GAAAATTTCA GCAAAGTACG CAGCATAATT 7860

TAAAATAAAT GCGAATAAGG CCGCATCGTA ACGTTCAAAG ACAATACCAA TTAATGGTAA 7920

GCCGTAAAAA ACAAAAATTA ATTGTAACAA TAGCGGAGTG CCGCGCATTA ACCAAATGTA 7980

GAGGTTAATT AAGTAACGTA AAGGTTTAAA CGAACTTTGT AAAATAAAGG CAATGATGAT 8040

CCCCAGTGGA ATCGAGCCAA GGAGCGTAAA GAAAAAGAmT TTCAGTGTCA TGCCAGTCCG 8100

CTAACTAAAG CAGGTAAAAT TTCTAAAATA TAATCCATCA TTTTTTCCTC CAACAATTTT 8160

(2) INFORMATION FOR SEQ ID NO: 160:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6253 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

CAnCCCAAAT TTnTAGCTTG ATTCTGCAGC TTTTGATAAC aTTTGGTCAA TTTTTGGTAA 60

TCCATTAAAT CAACTAACTG ACCTTGTTTT AACATGkkrs tAAAGTCTAA ATCCCAAAGC 120

ATTTCTTTAG TCGGCTGTTG CAATAAATAA TACATAGAAT GACTACTAAT AATTTCAGTG 180

GTCGGTTCTT CAATTGTTTC ATAAAAAGTC ACAAAATATT CTCGCTCAAT CACAGTAATA 240

ATTTCTTCTT CTGAAAGCTC ACAATCTTCT AAATTTTGAT AGAATATCAA TAATTTAATC 300

CGTTCATCTA TTTTATTCAG CGAancGACt AATTCAGGAA AAAGAACCGA CTGACAATTT 360

TCTACTAATT CTCTTGCCTT ATTTTGTTCA TTTTTATAAA AGTTGTATAA CTTTTTCCGC 420

TGTTTTAACG AGTTTAGTAA GTTATCGGGT ATATTCCAGT AATATTGTTT CTTCATAGCA 480

TTCCCACTCT TCTTTTTTAT GTACAATTTT ACAAAATTGG TCTTAGCCAA TTTCGTCTCT 540

AGTTATTCTA TCAAAAAAGT GATGCTTTTT TACTGACaAT AATTACTCGT TTTACGCCCA 600

AAAAGTGAAA ATCTAGCTGT TAAGTAATTA TTGGGGCTTG TTyCCCTAGG gCCGAaAAAA 660

CTTTTTCCTT TTAAAAGTCC AGAAAAAGAC CTACACTAGA GATAGACAAA CGTTGGAGGG 720

ATTTAAATGG ATTTACACTT AACGAATAAA TTAGCATTAA TTACTGGTTC AACCAAAGGA 780

ATTGGCAAAG CAATTGCGAT TGAAATGGCT CGTGAAGGGA CCGATGTCAT TATCAATGGG 840

CGCAATGAAG CCGAAGTAAC TAAAGTTGTT AAAGAAATAC AAACAATGTT TCCAGACACT 900

CATCCTCAAG CAGGGACTGC CGATATTTCC ATTGAAAGTC AACGAGCTAC TTTGCTTGAA 960

AAATTCCCTA AAGTCGACAT TTTAGTGAAC AATATGGGGA TTTTTGAGCC AATGGAATAC 1020

TGGGACATCG ATGACGCCAC TTGGGAAAAA TTTTTTACTG TGAACGTGTT GTCAGGCAAT 1080

GCATTAGCAA AAGCTTATCT ACCTAAAATG CTTGCACAAG ATTTTGGTCG CATTATTTTC 1140

ATCGCTAGCG AAGAAGCGGT GATGCCTTCT GGCGaAATGC CCCAATATAG CATGACAAAA 1200

ACGATGAATC TTTCCTTAGC TAAAAGTTTA TCCAACTTAA CTGTTGGCAC ACATGTCACC 1260

GTTAACACGG TTATGCCTGG CTCAACCCTT ACCGAAGtGT AGAAAAAATG TTGGAAGATA 1320

TGTACGCTGA TTCAGACATT CCCAAAGAGG ATTGGGAAAA AGATTTCATG AAAAATCATC 1380

GTTCTCGTTC ACAAATCCAA CGGCTCATTC GACCAGAAGA AATTGGTCGT TTTGTTACCT 1440

TTGTGGcCAG CCCAGATTCT TCTTCTTTCT CAGGAGAAGC CTTAAGAATC GATGGCGGCT 1500

TAGTTCCAAC AATCTTCTAA AATAAACAAA AAAGACAGCC AGCAAAAATT CACATTTTTn 1560

GCTGGCTGTC nTnTTArGCA TATTAATTGA TTACTTmCCA AAAGTAAGTT ATATTTATCT 1620

CGAATTCGAA wTAAAAAGAG GTGAAmCAAT GAACCAACAG CAAGAAGCTT TAAAAGCCTA 1680

TATCGGTTTA TTAAGAACCA GCCATCGACT AGAGCAACTT GCCAAGCAAG ATGTCACTTG 1740

TTATGACTTA AACATTACAG AATTTTCAGT ATTAGAGCTG TTACTCCATA AAGGTCCTCA 1800

GACCATCCAA AAAATTAAGG AGAAAATTTT AATCGCTAGC AGTAGCACCA CTTATGTTAT 1860

TGACCAATTA CATAAAAAAG GCTATGTAAC GCGCACCCCC AGTGAAAAAG ACCGACGCAT 1920

TACTTACGTT GAATTAACAG AAGCTGGAAA AACATTAATT GAAGAAATTT TCCCGACGCA 1980

TGCAAAGCGA ATTGCAGAAG CATTTGAACA ACTCTCTTCC GAAGAATTAA CACTTCTTCA 2040

AAAAACTTTA CGAAAAATAA CAAATGAaAC GAAATGAGGA AATAATGATG AAAAAAGAAG 2100

ATCAATTATT AGGAATCCAC CACGTTACAG CTATGACAAG TGATGCAGAA AAAAACTATC 2160

ACTTCTTTAC AGATGTTTTA GGGATGCGTT TAGTCAAAAA AACAGTTAAT CAAGATGATA 2220

TCTATACTTA CCATACCTAT TTTGCTGATG aTTTGGGTAC ACCAGGTACA ACCaTGACCT 2280

TTTTCGATTT TCCCAATAAC CCTAAAGGAT TAAAAGGAAC CAATACAATT TCAAGAACAG 2340

GGTTCCGGGT TCCTTCAGAT GCAGCTTTGA CTTATTATGA AAATCGCTTC AaTGAATTTG 2400

CTGTCAAACA CACAGGCATT TCTGAAGAAT TCGGGAAAAA AGTCCTTCGC TTTTGGGATT 2460

TTGATGATCA AGCGTATCAA TTAATCTCTG ATGAATTAAA TCAGGGCGTT GCAGCGGGCA 2520

CCCCTTGGAA AAAAGGACCT GTTCCAACAG AATTTGCGAT TTATGGCTTA GGACCTGTCG 2580

AAATAGCTAT TTCCTATTTT CATGAATTCA AAGAAGTCTT TGAAGAAATT CTAGGCTTTC 2640

ACGTAGTGGC ACAAGAAGGC AATCGCTATT TACTAGAAGT TGGACAAGGT GGCAATGGTG 2700

CCCAAGTCGT TTTGGTAGAC GATGATACTA GCTCACAAGC GCAACAAGGA TATGGTGAAG 2760

TACATCACGT TGCATTCCGC CTAGCGGATC GTAAATCACT TGGGACTTGG CAAGCGCTCT 2820

TTGATCATTT AGGCTTACAA AACTCTGGCT ATGTCGATCG TTATTATTTT GAATCATTGT 2880

ATGTTCGCAT TGGGCATATT TTAGTCGAAT TAGCCACCGA TGAACCAGGA TTTATGGGGG 2940

ATGAACCTTA CGAAACATTA GGAGAAAAGT TATCTCTTGC GCCATTTTTA GAAAACCGTC 3000

GTGAATATAT TGAGAGTGTG ATCAAGCCTT TCAATACAAA ACGAGCCTAA GGAGGAAAAA 3060

ACATGCATTC AaTTTTTAAA AAAGGACATC CTGAAGCACC TGTCTTTGTG CTACTTCACG 3120

GTACAGGTGG TGATGAAACA TCTCTCCTAC CAATTGCCCA AGAACTAAAT AAACAAGCTA 3180

CTGTGCTAAG TATTCGTGGT GATGTTTCAG AAAATGGAAT GAATCGTTAT TTTAAGCGCC 3240

TAGCGGAAGG TCATTATGAC TTAGAAGATC TAGAAAAACG CGGCGAGGCG CTTCATAAGT 3300
TTATTCAACA AGCCGCTAAC GAGCATCAAT TTTCATTGGA TAAAATTATT TTTATTGGCT 3360
ATTCAAATGG GGCCAATATC GCTATTCAAT TATTGCTTAC TCATCCCGAC AGCTACCATC 3420
AAGCTGTCCT CTATCATCCC ATGTTTCCTG TTGAATTGAC CAATCAACCA GACTTGACCG 3480
ACACTTCTGT TTTATTATCT CTAGGAGAGC ATGACCCGAT TGTCCCACTT CCTGAGAGTA 3540
TGCGTGTGAT TCAACTATTT CAGAATCATG GAGCAACCGT ACAAGAGGTT TGGACACAAA 3600
GTCATCAATT AACTTATCAA GAAATTAAGG AAACACAAAC TTGGTTGGCA CATCTGTCCT 3660
CTTAACAAAA AAGTGGGTCC TAAGTCATTT CGACTTAGGA CCCACTTTTT CTATTTTAAG 3720
ATAATTCACT GTCTGAGCTT CTTTTTATTC TGCCATTTGA ACAACTATTT TACCAACTGC 3780
ATGGTGCGTT TCACTGAGTG CATGTGCATC ATAAATCCCT TGTCTAGAAA AAGGGAAAAC 3840
TTCACCAATA ATTGATTTAA CTTTTCCAGC TGCCATCAAA TCAGCAATTT TTTGTAATTG 3900
TTCTCCATTT GGTTGAAGCC AAATACTTTC AGCAGAAATA TTTTTTTCGG CTGCCAATTG 3960
TTTATCTTCA ATTCCCACAA TTGAAACAAG ACGACCGGTG TTTGGTTTTA AGACGGCAAA 4020
ACTATTTTTT TGAACCTCAC CACCCATTGT ATCAAAGACT AAATCAACAT CAGTTAATAC 4080
CTCTGCAAAA TTCGTTGTAT GATAATCAAT CACTTCATCT GCACCAATTT TTTTCAGTAA 4140
GGTATGATTT TTTGCGCTAC GGTcGTGATG ACATGTGCGC CCGCTTCTTT TGCTAGTTGA 4200
ATCGCATAAG TACCGACCCC ACCTGCACCC GCATGAATTA AAACAATTTC TCCTTCTTTA 4260
AGGTGACCAT GATCAAACAA TGCTTGCCAA GCGGTTAAAC CAGCCAACGG AACGGCAGCC 4320
GCTTCTTCAA AGCTAATTGT TTCAGGGATT TTtGCTAATA AATGATCATC CACAATCGTT 4380
ACTTCTGCAT AGGTACCAAA ACGAGTAGTT TCAGGACGAG AAAAAACTTT ATCGcCAACT 4440
TGCCAATCCG TTACTTGACT CCCcACCTCT GTAATCACAC CAGCGACATC CCAACCAAGA 4500
ATAATCGGAA AGGACCAATC AAACATCTGT TTtAAATATC CTTCACGCAA TTTCCAATCA 4560
ATCGGATTAA TTGATGTCGC GTATTCTTTT ACCAACACTT GATGTTCCGA TAATTCTGGC 4620
AGTGTTACTT CTGCTTCTTC AAGTACCTCT TTACTTCCGT ACTGATTAAT CACAACAGCT 4680
TTCATAAACC AACTTCCTCT CATTTGTTTC ATTTTTCACA CTTTTAGTAT ACGnCCTATT 4740
GCCTTCTTTT TATATAAGAA TATGCTTTAG TTAAATTGAG AGAAAGTAGC TGTTTTTAGT 4800
CTATTATTCA TAaCGTAAAG ATTCAATTGG ATCTAATTTT GgCTGtCTTC GAGCTGGTAA 4860
AGTTCCTGCT AAAAAAGCaA TAAACATAAT AACTAAGATA ATCGTTAGTG AAGATGGCAA 4920
TGAAAATTGA ATTAATTTAA ATCCTGTTAA AGCCTTTAAA AAAGAATCCG TTGCCAAACG 4980
ATTAACTAAA TTTCCAACCC CAACAGCTCC TAAAATGCCT AATATTGAGC CAAAAAAGCC 5040
AATTAAAGCC GCCTCaACAC TAAAAATTGT AAAgACtTTC CCATTGCTAA GACCCATGGC 5100
TTTCATTAAA CCAATTTCTC GCGTTCTTTC TTGGACAGAC ATATACAGCG TGTTAATAAT 5160
GCCAAAGcTT GCAGCTAACA AGGCAATTGC ACCAAACATG GTTAAGACAC CAGTGATTGC 5220

ATTGATTATA TTACGAATCA TGCCAATTTC ATCTTCGACA GTAGTCGCTA AATAGCCAGC 5280

TTTGTCTAAA TCTTTTTTGA TGTCCTTAAT CTGCTCTGGC GTGCTGTCTT TTTTAACTTC 5340

AGCAATAATC ATCGCGTATT GATTTTTTAA ATGTTCTGGA AGATCTGCTT GATTGATTGA 5400

GACAACTTTA TCAATCAATG CTTTATTCAT CAGCGAGAGC CCATTTTGAA TGACACTGGC 5460

ATTTCTAACG CCAACAATTT TTGCTTCAAT AACTTGCTCT TGTCCTTTCA AGGAACTTGA 5520

GATACCTAAT TGAACTGTTT CCCCCACGGC AGCTTTACTA GATGTATAGC CGAGCGCTTT 5580

TACATATTCT GGTGACAAAT TAATTTCGAA GTCCTGGCTC GTTTGCGAAA CTTTACGGCC 5640

AGCCGCTAAA TCAATGGTCA TCTCATCTAG CGCAGAAGTT GCTGAAAATA CATATTTATG 5700

CTTATCTGCT CCTTTTATAT AATCGATAGC AACGGATTTC ATTGGTTCCA CTGATGTAAC 5760

ATCCGAGATT TTTTTAATCT TCTCAATATC TTTTTCAGCA AGCATACTTT GTTGTTGAAT 5820

CGTACTGGTT TTTTTCTCTG GATTATATTT ACTTGGTTCT GTTCCGTTAC CAACATTCAT 5880

CTCCATTTTT GGCTGAATAA ATAACTGATT TGCACCACCG ACACTGCCGA CTTGTTTGTC 5940

AATATAATCA TTTACACCAA TATTGACTCC TGTAGTTAAA GCTATGGTAA ATGCTCCAAT 6000

AAAGATTGCA ATAATGGTCA ATACGGTTCG ACCTTTATTA CGCATTAAAT TGGTACTAGC 6060

TGATTTTAAA ATATCTCTAA ATTTCATTTA CTCATCTCCT CCTACAATCA AGcCGTCTCG 6120

AACATGAACT TGCCGATCAC AACGTGCTGC TAAGTCTGGA TCATGTGTCA CAATAATtAa 6180

GGTAATcCTT tATTTTTATT TAAAyCAAaT AACAATTCTT CAATTTTCTT CCCAGTGGCA 6240

GAATCTAAAT TCC 6253

(2) INFORMATION FOR SEQ ID NO: 161:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3312 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

TGTGGCAACT TAAAAAAGAA CAAAATGTTC CCATTTATCG CAGTATTATG GAGATGATTG 60

TTCAAAAGAT TCAAACAGGC GAATTATTAC CAGGCGAAAA AATCCCTTCG GAACGTAAAC 120

TGGCAGAGTA CTTGTCTGnC AATCGTTCAA CCGTTGTTCA TGCATTAGAC GAACTGGTGG 180

ACTTAGGCTG GATTGTGCGA AAAAGAGGGA GTGGAnGTnT GTCAACGAAG GAAAATGGGG 240

GGTAATGATG ACGCCACGAA CGGATTGGCA TCGTTATTTA GAACAAAATG TGTTCAAACA 300

AGTTCATCCC TTGGTGACAG AGATTGAGTT ACGTGCTAAA CAAAATTTGC CGACTGACTT 360

AGATATGTAT ACAGGCGAGT TGCCGTTAGA CTTAATTCCC AGTTTCGATT TTCCAGCATT 420

AAATTGGAAA CAATTTTTAA AAGAGGAAGC ACAAGATGAA TTAGGTTATT TGCCGTTAAG 480

AAAAGAAATC CAAGCCATTA CAGCAACGGA ATATCAGCTT CATTTACCAA GTGAAAGTTT 540

ATTGCTTACT TCAGGTGCGC AACAAGCACT TTTTTTAGTT TTACAAGTTT TATTGAGACA 600

```
AGGAGATAGT GTTGCCGTAG AAGATCCATC TTTCTTCTAT GCCTTACCAA TTTTCCAAGC     660
TGCTGGTGTG CGCTTATTTG GTGTGCCCAT GACAGAGGAA GGGATTGATC TTGAAGCGCT     720
AGAACAAACA ATTCGACAAC ATCGCATCAA AATGGTTATG GTTAATCCTT CCTTTCAAAA     780
TCCGACTGGG ACAGTGATGC CTTTGAGAAA AAGAGAACAG CTTGTGAAAG TCTGTCAAAC     840
GTATCAAGTT CCTATTTTAG AAGATGATGT TTTTGGGCAG TTAAGTTTTA TCCCGAAAAC     900
AGAAATTCCT CCTTTAAAAA AATTAGATCC AGATAATGTC TTGTATATCG GTTCGTTGTC     960
AAAAATTTTA GGTTCGACCA CGAAAATTGG TTGGTTAAGT GCGCCCGCCT CCGTGACGAA    1020
ACAAATTGCA GAAGCCAGAA AAATGATGGA TGTTTCGTTG AGTATTTTTC CACAGATGCT    1080
CGCTAAAATG GCGATTGAAG ATCCTTCTTT CTCTGAAAAA ATTACGTTAT TAAATAAACA    1140
AGTGGAACAA CGAGCAACAG CTGTATATCA AGTATTTAAA TCGTTGTCAG AATGGGAAGT    1200
TTCGCCAGTA AAAGGCGGCT TTTATTTATG GGCACATTGG CGTCAAGGGG CACTTAAGCC    1260
AGAAGATTGG CAAGTGTTTT TACGAGAAGG TGTTTTGGTA GCACCAAGTG TAGCTTTTAG    1320
TGAAAAACGT GGAAGTATTC GCTTGAACTG TTCCCGAATT TCACCAGAAG AAATGCCTCT    1380
TTTTTGTGAG CGAATGGTAC GGATTACACG ACAGCTTTCT GAAAATAGAC AAACAAAAAT    1440
AGAACAGTAA GTAAGATTTT TCTTATAAAA ATCATTTGCT GAGAGAAAAC GCTACAACTT    1500
CCCTTTTCTA AACGAATATC CAATGCTATA CTTAAAAAGT AAAAACAAAA AAACAAATGA    1560
GGTGTTTTCC ATGTCTTGGG AACAAGTTTA CCAACAATGG TTAGATGAAG AAAATATTCC    1620
AGAAAATTTA AAAAATGAAT TAAAGGACTT AAATACAGAC CCTGAAAAAT GTGAAGATGC    1680
ATTTTACGCA CCATTAGAAT TTGGGACTGC CGGGATGCGT GGGATTTTAG GCGCCGGTAT    1740
TAATCGGATG AATATCTTTA CGGTTCGTCA AGCAACAGAA GGACTTGCAC GTTTTATGGA    1800
TACGCAAGAT CCTGAAACAA AACGTCGCGG TGTAGCTATT GCCTATGATT CACGTCATAT    1860
GTCTCCAGAA TTTGCAATGG AAGCAGCTAA AACATTAGCT AAACATGATA TTCCTTCTTT    1920
TGTGTTTGAA AGCTTACGAC CAACACCAGA ATTATCTTTT GCGGTCCGCT ACTTTAAAGC    1980
TTTTGCGGGA ATCATGATTA CGGCTTCTCA CAATCCAGCT GCCTATAATG GCTACAAAGT    2040
CTATGGTGAA GACGGTGGTC AAATGCCGCC AGCTGATGCA GATGCATTAA CTAAATATGT    2100
TCGTAGTATC GAAAATCCTT TGAAAATTGA TGTCTTATCA GATGAAGAAG TTGCCCACAG    2160
CGGTTTAATT AACATTGTTG GTGAAGAAGT CGATAACGCG TATTTGAAAG AAATTAAAAC    2220
AGTTACGATT AATCAAGAAT TAATTAACGA AATGGGTAAA GAATTAAAAT TAGTTTACAC    2280
TCCCTTACAT GGAACTGGGA AAATGTTAGG AGAAAAGGCA TTAAAACAAG CCGGTTTTGA    2340
AAAATTTGTT TTAGTACCTG AGCAAGCCGT TGCTGATCCT GACTTTACTA CAGTTAAGTC    2400
ACCAAACCCT GAAGAACATT CAGCTTTTGA ATACGCGATT CGTTTAGGGG AAAAAGAAGG    2460
CGCTGATTTA TTGATTGCAA CAGATCCAGA TGCTGATCGT TTAGGTGCTG CTGTGCGGAT    2520
GCCAAACGGA GACTACCAAG TATTAACAGG AAATCAATTA GGTTCAATTA TGATTCATTA    2580
```

TATTTTAGAA GCCCATCAAC AAGCAGGCAC GTTACCTCAA AATGCTGCGG TTCTAAAATC 2640

AATCGTTTCT AGTGAACTTG CTACAGCGAT TGCTGAGAAA TGTAATACGA AGATGTTTAA 2700

TGTTTTAACA GGTTTCAAAT TTATTGCTGA AAAAATTCAA CAATACGAAG AAGACCACAG 2760

TCAAACGTTT ATGTTTGGTT TTGAAGAAAG TTATGGTTAC TTAGTGAAAC CTTTTGTTCG 2820

TGATAAAGAT GCAATTCAAG CGTTAGTTTT ATTAGCAGAA GTAGCGGCTT TCTATAAAAA 2880

ACAAGGAAAA ACATTATACG ATGGTTTACA AGATATCTTT GAAGAATTTG GTTACTTTGA 2940

AGAAAAAACT ATTTCCGTAA CAATGAGCGG AATTGAAGGC AGCGGTAAAA TCAAAGCTTT 3000

AATGGCAAAA TGTCGTGAAC AAGCGCCAAC TGAGTTTGCC GGGATTCAAG TTGCTCAAAC 3060

GGAAGATTTT AAAGAATTAA CACGCACTTT TGCAGATGGT CAAACAGrAC AATTACAAAC 3120

GCCTCCTTCT GATGTGTTGA AATATCATTT AGAAGATGGC AGTTGGATTG CTATCCGACC 3180

ATCAGGAACA GAACCAAAAA TTAAATTCTA TCTAGCGACA AAAGCAACAA GTAGTTCAGA 3240

AGCTTCTGAA AAAATTGCTG CGTTTGAAGC nGTCGTTAAT GAATTAACAA AATAAAATAA 3300

ACGAAACAAG CT 3312

(2) INFORMATION FOR SEQ ID NO: 162:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14286 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

TATCTTCGTG AAATTCAAGA AGACTTAGCA ACCCTTCCTG AAAATACCTT AGAGATTCAG 60

ATCCATCACG GAAAAATTTC ATTACATACG ACTTCCAATG ATTTTCTTTC CCACTATTTT 120

ktAtTATTTA AtTAtTaCTG kGtCCaTyCT ACTGACTTCa TyCTTTTTCm AGcAATTATC 180

AAAAAAGrAA ACmATTCTGT CTGGAAAGTG AGTCmAGrAA CTAATTTTAG CTCATCTTAC 240

ATTTACAAcG GTTAAaTAAC ATTAATAAAC TATTTGGTTT GTATGGAATT TCCGTCCATT 300

TTTCCCCCTC CGCGAGTAAA GTGATTACAG GTTCTGAATT TCAAGTCCTA TATGCTTTTC 360

TTGATGTCTA CTGGACCATC TTTTTAAATA CCTTAGCCTT ACCACGTTAT ACCTCTACTT 420

ATTTTGAAGA TATACTGACG CAGCATATTA AACCAGAAAT TCTTGAACGC TTGGAACACG 480

CTTCTCTACT AAAATTAAAT TTACTTTTAC ACATCGCAAA ATATCGCTTT CCATATACTT 540

GTGAAGATGA TCTCCAAAAA GAATTTGAAG AACATCCTTA TTTAAGGCTT TTCTCAGACT 600

CAAAACAAAC TATTTTTTGT TCCCAATCAA ATTTTTCAGA AGATCAACAG CGGGTTGTAG 660

ATATCATTGC ACCCATTGCC ATTGATAATT TTGATTCAGC AAAACGAGCC AAAGATTTAC 720

TGGCTGCTCT AAATGaAATT TCTTGCCCAG AATATTTCTA CGTTGATTCT TTATTAACCA 780

ATTTTTGCGA aGTTTTTCAA ATTCCAAAAA TGCGTGAATT TGATCGTCAT TACTTCTTAC 840

TTATTTTATT TAAAAATATT ATCTATAACC AATTATTTAT TTCACAGCAA CCAAATACCG 900

CAATTCTAGA TTTTTTAGTC AATAAAAATA GTACTTATCA AAAAGATTTA CAGAAAAAAA 960

TCTATGCTTT TTATCAAAAC TTTTCAGCGC ATTCCCCTTA TCAGTTAAGA TTTAAGGATC 1020

AGGAGAACCC AATTTGGACA TGTGAATTGT TGTTACGAAT TTATAAGCAC TATAGTCCCC 1080

CCAAAAAAAT TAAAATTGGT GTTGCATATT CTCGCGATTT TTACATTGCT AGCTTTATTA 1140

TGATGAAAAT CAAGCAACCA TTTAGTGAGG ATATTATTTT TGAACGAACC GATTTTTCTG 1200

GTTGTGATAT TGTGATTACT GACTATCCCt TATTTGATTT ACCAAAAGAA ATAGACCGAA 1260

TTTACATTTT AGAAGAAGAA TTAACCCGAG AAGATTGGGA ACTAATTTTC AGTAAAATAA 1320

GTATCGCCAT TTTTGATTTG CAATATTAAA ACAAACATCA AGCGCCAACC TGTCTCACCA 1380

GCCATAATGA AAACAAGCGA CCTACAAACC ACTTCTGCCA GTTTTAGTCT GACAGAAAAT 1440

AGTTTGTAGG TCGCTTGTAT ATTACTGTTC TTTTCTATGT TACCTTCAAA GACAATCTAG 1500

TAACGTACTT CAGTCCTTAA AGTGCTACTG TTCCACCGCT GATCAGTTTT TAAGCCTGAG 1560

ACGArGATCC AAAGTGTTTT TTGTCTCAGG CTCTTACATT TTTTCCAACA AAGGTATTGc 1620

TaGGCGCTAC TTGAAGCTGA AAACTAGTTT CTCTATCGAG ATTCCTCATC AGTAGACTCA 1680

AATTTAAAAC TTTTTTCTAT AAAAGTGaCT TCTTTTGCGT ATGTTTTTAC AAAAGAAGAA 1740

TAGTCTTTGG GATCTTTGAC CTTTTCCAAT TGGTCAACAG AGACTTCTAA TTCAAATCCT 1800

GTCCCATCTT TAAAGAGATA ATATTTTGAG ATGTCTGTAT ATCCTTGATT CCGTTTTAAA 1860

ATATAGCCGG TTTGTACATC AGCAATTTTA TCAATTGGGT ATGCTTTGAT AAACTCTTTA 1920

CTTGTATCCA AATAATCAGA AAGACTGTTG ACAAGATTGC CAGAAAGGTA CATACCTGCC 1980

GATGTATAAT AGTCTCTTTT TTCATAAAAA CTGATATTGC CTAGATATTT TGTATCTGGA 2040

TACTCCTCTG TAGAAGGTAA AATTAAAAAG TTCTTTCCGC CACTGGTTTC AATATCGGCA 2100

TCGCCGATAA TTTCTATTTG GTAGCCTTGA TTAAATAAAA GTTTATTTTC TTCTATTGTG 2160

ACTTTTAGTG GTTTTTCTAA ATCAACAGGT CCAGTTGTGG TAAGTCCAGC TTCTTTTAAA 2220

GGATAAACAT CTTCTGATGA AGCGGAGTAA TCCCAATAAT CATCTTTCtC CGCAAGATCA 2280

ATATATACTA TAAAGCCAAT TGCTACAGCA ATGCCTATCC AAATAATTTT TTTGTTCTTT 2340

TTTTGTTTTT TCTTTTTTTT CGTCTGTGGT TTTACCACTT TTTTATTGGG GTTAGGAAGT 2400

TGCTTTGCTG TATCTGTTCT TGTGACAGGA AAAAACGACC ATTTTGGCTT TTGAAAAAGC 2460

GCATCATATG kATCATAAAA ATTCTTTTGT TTAGCCATTT TTTCTCTCCA CTAATTTTTT 2520

TGATAAGGCA ATTTCCTCAG ATGCTTCGGA GACTAATCTT TCATAATCAT CTGTGATACT 2580

TTCCGACAAC ACTCGTATGG TACTTAAGAT ACCATTGGTT GCAAGTGTTA TTTCwTCTGT 2640

TGACAATTTA TCTTCTTTAA TGCTTAAAAA CCGTTcTGAT GCTCGTAAAA TCTCTGGTAC 2700

CTTATGATAT AAAAAATCAC TAAAAtGAGT AATTGcTTCA GGATACTTCA CTAAGTGTTT 2760

AAAGATGCCT TtAGCGGAAG CAACACCCAG TTTTACTTGT TCAATTTTTT TTAAATTTCG 2820

TGACTTATTc GTTAGTTCGT TTAAATAAAT AATTTGTTTT TTTAATTCGT CCATAATTCT 2880

TTGGTACATC TCTAGGTCAC TGTCCGTTAG ATGGTGCgTC GTTTGATATT TTTTTAGATT 2940

TTTTTTATAa CGGAAATACG TACGACCATT TCTATAAGTA ACCCAACCAA TGTACCCAGT 3000

CCCCACTAAC AAAACCATTA TCCCATTCAA ATCATGTACT CCTTTTTAAA TTAGTATATA 3060

TTACTGAATA ACACACTGTC CAGCTTAAAA TTGACAAATA AGTATGAAAT GTTACTCCCC 3120

CAATTACTTG TGCACTTTTA TTGTATCATA TAGGAAAATT GTTCTGTTCC GCATTTGGAC 3180

AAATTGTCTA AATGTCTAAG GGAATCTTTT TTTGTTTAAT ATGTAGCAAA AAACACTTGC 3240

AAGAATTGAT TACTTCTCAC AAGCGTTTTT TATTGTCTAA AACAACTATT TACATTTCAT 3300

TTTATTAAAT AGTGAATAAA TAAATTACTG CAGCTAATGA AGCaCCTAAG ATAGGTCCCA 3360

CAATTGGcAC CCAAGAATAA GCCCAATCGG AATCACCTTT TGTTTTCACA GGAATAATTT 3420

GGTGCGCTAA ACGAGGTCCT AAATCACGTG CAGGGTTAAT TGCATAACCT GTAGGTCCGC 3480

CTAATGATAG ACCTAGTCCT AAGACCAAAA TCCCAACAAG CATTGGGTTA ATGCCAGCCG 3540

CTAAGTCGTT TTGAGAAAAG GCTAGTAAAC CAAAAACCAA GACAAATGTG CCGATAATTT 3600

CTGTTAAAAC ATTGGCTGGA TAGTTACGGA CAGCTGGACC AGTGGCGAAG GTTCCTAAAA 3660

TAGCACCTTT GTCTTCTGTA ATGTTCCAAT GGGGTAAATA AGCTAACCAG ACAACTAAAC 3720

CACCAATAAA ACCGCCTAGG ACTTGAGCTA CGATATAAGG TAAAACCATG CCCCATTCAA 3780

AGTTTCCAGT AATTGCCATC GCAACTGTTA CAGCAGGATT TAAATGCGCT GGACTCATAT 3840

ACCCAGACAT GTAAACAGCT AATGTAACAG CAGCGCCCCA ACCTAAAGCA ATAACGACCC 3900

AACCAGAAGC AAAGGCTTTG CTTTTCTTCA AGTTAACTGC GGCACAGACG CCATCCCCTA 3960

GTAAAACTAA AATCATCGTT CCGAAAAATT CACCGAATAA TTGTGTCATC ATCGAAGTTC 4020

CCATTTTTTC TTCACCCTAC TTTCAAATTT TTTAAATCTG ATTCTTCAAT TACTTTTTCT 4080

AATGTTTCAA TGTGTCGTGC TCTTTCTTCC GCTGTCCACT GATAATGCTG TGCCATTTCT 4140

TCAATGACTC CCGCTTTCAC TTGGTCCAAA CGATCACGCA TAAATAATAA GTGGTTGGTC 4200

CGTCGCAATA AAAAGTCGAC GGGTGTTAAA GCCATTTCTT CTTCCATCGC ATAGTTTAAA 4260

GAGACAGTTT CTGTTAAGGT TAAACCAGGA ATTACTTTGG CATCATCAAT CATTTCAAAA 4320

ACAGTTGGTA CGTTGGAGCC ATATAAATGA GCTAAGTATA GTGCATCCTT TTCACTTAAC 4380

CCTTTAGACA CGCCATGTTT TGCCAACGCT TCTAATTCTT CGTCTACCGT TGCAGGGTTC 4440

AATTTTCCGC CTGAAACTGG ATAGTTTTTA GAATCAATTA AGGTGAATTC TTTCTGGTAT 4500

TCTTCTGCTA AGATGGTTTG AATCATTTTC ATCGCACCTT CTGCCATTTT ACGATAATCT 4560

GTTAATTTCC CACCAGCCAA AGTGAGCAAG CCATCCGCTG ATCTTTCCAA TGAGCTGCCT 4620

CGAGAGACAG CTGAAGGATT GACTTTATTT TCCACCAAAC TGTCTTCCAG ATGGTTTAGG 4680

ACTTCTTCAA TTTCTCGTTT ATCTGTTTGA TTTTTTTGGT ATCTGTCAAC GACATCAATG 4740

ACTTCATCAA TACTTTTATC TGAAAGTTTG CCATTGTTTC CCCCATTATA ATCTGAGCCA 4800

CCGTTAGCAG AAATTAAAGG TCGTAAGCCA GCCCAGCTGG CTTCAATATC ATCAATTGTC 4860

AATTGTACCT CTGGATAACG ATTATTGACT ACTTCCAATA GGTAATCCAC ATCACTTTGT 4920

TCTACTGTTG GATGTTGAAA GTCTCCATGG TAATCTGTAT CCGTCGTCCC AAAATATGTT 4980

TTTTCTTCTC TTGGAACAAC GAAGACCATG CGGCCATCTT GTTTACCTGT ATCAAAATAC 5040

GTTGGTTGAG GAACATTTAA ACGGCTTTTG TCGACAACTA AATGAACCCC TTTGGTTGGA 5100

CGCATTTGTG GTGTAAATGA ATCGTTTTTG TCTAAGCCTC TTAATTTATC TGACCAAGGA 5160

CCAGTAGTAT TAATCACAAC ATCTGCGTGA ATCTCAAACT GTTCGTCAGT TAGTAGGTCC 5220

TTTGCTTGAA TGCCAGTAAT TTTCCCTTCA TCATTGTATA AAAAGCCCAC CGCTTGAACC 5280

CGACTCACAG CATGCGCCCC GTCAGCTACT GCTTGTTTAA TATTTTCAAT AACTAAACGA 5340

GCATCATTAT TTCGGTAATC TAAATAAACA CCGCCACCTT GTAAACCTTC ACTTTTTAGT 5400

TGTGGCTCAC GTGCCAAAAC CTCGTCTTTT GTTAACAGAT AGTTGGCGTA CTTGGTACCT 5460

GTTACATCAG CTAACTGATC ATATAAATCC ATTGCTATTT TGACAGAAAA TAGATTAAAT 5520

GTTGCATTTG GCTCATCATA AATCGGCAAT AACATGGGAT CCGCTTTAGG GATGTGTGGC 5580

GCAATGGATT GAACGACGGC TCGTTCTTTC ACTGTATCTG CTACCACCTc CACGTCAAAC 5640

GTTTTCAAAT AGCGGATACC ACCATGAACT AATTTGGTTG AtCGTGAAgA GGTGCCCTCT 5700

GCGAAATCCT GCATTTCAAT TAACCCTGTC TTCATTTTAG CGGCACTGGC TTGCAGGGCA 5760

ACCCCTGCGC CAGTAATGCC ACCACCAATA ATTAAGACAT CTAAGTGTTG TGTTTTTAAA 5820

GCGTCGATTG ATTGTCGGCG TGTTTCTATT GAGAAAGTCA TAGAAAATTC CCCTTTTCCA 5880

TTTACTTATT TTTACGTTTG AATTGTTGGG TAGCTGCTAC TGCTTGTTGC CAACCTTCAT 5940

ATAAATCTTC TCGTTCTTCT TCCGCCATAA TTGGTTCAAA TTGTTGTCCT TCTTCTTGGA 6000

ATGCTTTGAT TTCTTCTAAG TCTTTCCAGA AGCCAACCGC TAAGCCCGCC AAGAAAGcTG 6060

CGCCTAAGGC CGTTGTTTCT AAGTTATGGG CTCTTTGAAC AGCAGTGTTT AAAATATCTG 6120

CTTGGAATTG CATCAAGAAA TCATTGTTTG CTGCTCCACC GTCTACTTTT AAGACGGGAA 6180

TATCTATGCC TGTATCTTCT TTCATCGTGT CAATAATGTC ACGAACTTGA TACGCAACTG 6240

CTTGAAGTGT CGCTTTGACA AAATCTTCTC TCGTGGTACC GCGTGTTAAT CCAAAGACAG 6300

CCCCACGCGC TTGTGAATCC CAATATGGTG CGCCCAAACC AGTGAATGCC GGAACCACAT 6360

AAACTTCATT GTGCCCTGTT GACGCTTTGG CTACTGCTTC AGATTCAGCA GCCGTTTGCA 6420

ACATTTTTAA GCCATCTCGT AACCATTGAA TCGCTGAACC AGCCACGAAG ATACTTCCTT 6480

CCAAGGCATA ATATACTTTG CCGTTAATGC CATAGCCAAT AGTGGTAAGC AGATTATTTT 6540

TAGAAAGTTG AGGTTCTTCG CCAGTGTTCA TGACAATAAA TGAGCCTGTT CCATAGGTAT 6600

TTTTAACCAT CCCAGGTTCA AAGGCCATTT GACCAAATAA GGCGGCTTGT TGGTCACCAG 6660

CCATCCCAGC GATAGGAACT TCACTTCCGT AAAAATGGTA ATTTTTTGTT AATCCATACA 6720

CTTCAGAGTT CGAAACGACT TTCGGTAACA TCACTCGTGG GATATTTAAA AGATCCAAAA 6780

```
TTTCTTGATC CCAATCTAAA TCATGGATAT TAAACAACAT CGTCCGACTT GCATTTGAAT     6840
AATCAGTGAC ATGAGTATCT CCCGTTAATT TCCAAACGAG CCAAGTGTCA ATTGTCCCAA     6900
ACATCAATTC TCCATTTTCT GCACGTTCTT GCGCACCTTC TACGTGATCT AAGATCCAAC     6960
GAACCTTCGT TGCTGAGAAA TAAGCATCAA TGATTAACCC AGTTTTTTCA TGAATCATTT     7020
CGCTGTAGCC ATCTTCTTTT AGTTGATCAG CAATAGGTGT CGTTTGGCGT GATTGCCAAA     7080
CAATTGCGTT ATAAATGGGT AACCCAGTTG CTTTGTCCCA AACAACGGTT GTTTCCCGTT     7140
GGTTTGTAAT CCCAATTCCG GCAATATCTG TAGGTTTCAC TCCTGACTCA ATTAAAGAAC     7200
CTGCAATAAC GGATTGAACA GAATTCCAAA TTTCATTGGC ATTATGTTCC ACCCAGCCTG     7260
CATTAGGAAA ATATTGAGTA AATTCCTTTT GTGAACTACC AATTTTGTTT CCTTTTTTAT     7320
CAAAAATAAT TGCACGAGAA CTAGTCGTTC CTTGATCAAT CGCCATAATG TATTTTTCTT     7380
CTGCCATGAT AATTCCTCCT AAACGTTTTT GAAATCGTTT TCTTTATGTC CCTATCTTAC     7440
AAATAAAAGA ATAACTAAAC ATGTCAAAAA CTTTTTTTAT TTTAAAAAAA TTGAAAGGAT     7500
TTCTACTAAA ACCAACAAAG ACAACGCTTT CAATAATACT TTTATTATCT CTAATTTTTC     7560
AATAAAAAGA AACATTTTTG TTTAAAAAAG CCTAATTTTC GACTTTTAAT TGTAAAAAGG     7620
CTGGTGCGAG GATCAGACAA TTTCCGTCAC ACCAACCTTT TGTTTATTTC TCCCGCAAGG     7680
TGCGTATTTT CTTTTTTAAT TGTTGCACAT CATAGGTCGA ACTTAATTCT GACAAAACAT     7740
AGACTTCCGG TGTTCCTTGA TAGCTCGTTT GCTGATAATT GGTCACAATT AAATCATATG     7800
TTTGTTCAGG CTGATACCTC TCCACCAGCG TGCCATTCAT ATTTTTAAGA CTAAGCATTA     7860
ATACTTGTGT TAACGCCGCT TGATAAAGAG CATCCATTGC TAAATGAATT CCAATTTTTA     7920
TTTCACCAAT CATTTTAAAT TCAATTAATG CCAACACGCT TAAATATTTG ACCAGAGAAA     7980
GTTCATGCAA ACTATCTCCT GGTTGATACG TTTCATCAAA ACAAGCTAAA GCTTTTTCTA     8040
ATAACGTTGT TGAAAAAGAA GCTAATTCCC GACTAGATAA TTGCTGTTCT TTTTCCCAGA     8100
TATGGGCCCG ATCAAAGACT TCCAATTCTC CTTTAAAAAA ATACAGTTTT CCATGAATTT     8160
GTGAAAAATA GTAAGCCATG GTTTTCTCAA GAGTAGGTTC AAATTTATGG GGCCGATAAT     8220
ACAAGACCAC ATTTTCCAAG ATAAACGTAT CAGCTAAAGA AACAGCGGAA CGACGTCCCC     8280
GAATTAAGGA ATATTCATCA AAAGCTGATT CTGGTAAGAC CGACATTGTG ATTAGAAAAA     8340
CAAAATGAAT CATGCTTTCT CCTTCGTCTA ATTCTAATGA ATAGCGACTA GAAAAACGGA     8400
AAAACAATTG TCGGACTTGC TTAAATAACG GATCTTCTTT AAAATCAATC ATTTTCTTTC     8460
TCAATTGTTG ATATTTTTTT GAAGTGAcCG CTAGGCGCTT TTTCGTAATA CTCAACCAAA     8520
GGCTAACTTT TAACGCATTA TGTTCATTAA AAGTAAAACC TAAGCCAGTC TCAATCCCTT     8580
CAATTATCCG TTTATTTAAC GGCGACATCG TTTTTTTCTG GTGTATTTCA TAGGGAGTAA     8640
TATACCAATA CAACTGAAAG TAAAAGTAAC GAATCTGCAA TTCTTCTCCC TGTAGTTGTC     8700
CATTTTTTAT TTGCAAATCA AAGACGGCTA ACAATTGATT GAGTTCCTTA ATCTTTCGAA     8760
```

ATAAAGAAGA CTCACTAAtC aTAAATTTAG TCGtTAATTG AACAATAGAA AACTCTTTAT 8820

TACGATATAA ATAATCtAAA AGTTGAAATT TAAGAGACTC TTTTACAAAT GCTTCAATTA 8880

AATGATTCAA CGAAAAAAAA TCCGATAAGG TGATTGATAG CTGTTGGCCA TCATAAGTAA 8940

GCTGACAATC CTGTCCAAAA GGTTTCAAAA AATAATGTAA ATCCTCTAAG TCTTTTTCAA 9000

AGGAGGcTTT TGTGACCGCT AAAAATGCTG tTAAATCAGA GACTGAGCAG GTGCCGCCTG 9060

CTAAGATGGT TTGTTGTACC GTTAATAATT GTCGCATCGC TTTCTTTTCt AATAATGTCT 9120

CAACTTTCAT CGTTCTCTCT CCAATAAGCC GAAGCACCCA TCAAATTTGT AATATCATAA 9180

ATATTGGTTA AGGAAGGTTG ATAAAAATAA TCTTTTTGTA ATAGGTCTGT CAGTGTCTGA 9240

CCTGTTTGGA TACTTAATGC TAAAGTATTT ATTTTTTCTA AACAATTGTT TTTTGAACAA 9300

AGCTGTGCAC CAAGGACCCG TTGCGTCACT TTATCATAGA TTAATTTGCC TAAAATTTCT 9360

GTCCCATGTT GTAGAGGCGG CGCTGGCTGT CTAACAATTA TGGAAGCAAG TGTTTGCGGA 9420

AAAAATAAAC CTTCCGTTTC GGTTAGACCT GTACTAGCAA GGTAATAGTC CCCCACTTTT 9480

GTTCCCATTG TTCGCAAAGA GCCAATGAAT CGGTGTGTTT TTTCTTCTAA ATTGTTGGCG 9540

ACTACTAAGC CAGTTCGTAC CGCATTGTTC ACTAACGGAG CATAAAAAGT TTCCGCGACT 9600

GGTTCATTCA TAACTGAAAT GCAATCACCG ATGGCAAAAA CATTCGGCAC TGAAGTTTGT 9660

AAATAAGCAT CCACCGCGAT TGTCTGATCC AGATTCCGTT GAATTTTTTT ATCCAAATAG 9720

GCTAATTGAG GATGTAAGTT TAAAGCAAAA ATCCCACTGT CACAAGAAAT TTCCTGCTCG 9780

CTTGTTTCAA GAACAATACC ATTTGCTGTT TCTTCAATTC CTAGGACCGT TTCTTCAAAA 9840

TGGAAAATTA CTGCTTGTTT TTCCAAAGAT TTTTGAACTT CTGCGACCAT TTCTTTATCA 9900

AAATATTTAG GCAATAGATT TTCCAAGCTT TCAAAGACAT GGACCGTTTT CTTCATTTTC 9960

ACAAGAAAAT CAATAGCTTC CATTCCAATG GGTCCTGCAC CAATGACGGC TACTGTCTGG 10020

CTATTTTCCA GTAATGGAAC AGCCGCTAAA GCACCTGATA AAAATTTATA TTTGAGTAGT 10080

TTTTCTGTTT GACTACCACG AATTTGCGTG GAAAACTGGC TCGCGCCTGT CGCTAAAATC 10140

AATTTATCAT AAGAATACCA CTGTTGCTCT TCCTTTCGTG TCCAAGCAAT CAATTGGTTC 10200

TCAACATCCA TTGCCACAAC TTCTCTATTT AAAAGCAACT GAATCTTTTG ACGACGTAAT 10260

TCCTCTTCCG TTATGTAACG TGCTTCGTGT AATTCATTGA TGGTATGATT AAAATACGCA 10320

CTTAAGCCAC CAGATAAATA CCCCACTGTT GCTTGTTTAT CAATTAAAGA AATTTCTGCT 10380

TGGGGATATT TTTTCCTTGA AGCAATTGCT GCGGAAATAC CTGCGAATGA TGCACCAATA 10440

ATCACAATTT TCATTTGTgT CCTCCTCTGT CAGGACCTAc TATACTTCTC TCAGTTTATC 10500

ATGATTAAAA AGAaTGCTAA AgCAACAACt ATAGtTTTCC CTTAAAATAA TTATtTTTGC 10560

TGAGCGCATT GCTTTTCAAG CACTTTCATT TTCATTTATA ATAAATGTTA CTACTGTTTG 10620

AGGAGGAGTT TTCATGTCTG CAACAAAAGA GCAATTGGaA CTATTTTTAT TTTATTTATC 10680

TGAAACCCAT ACAAAAAGTT TATCTTTACA TGACTTAGTT ACAAGTCGAC CACGCCCAGA 10740

AGAAGCGCGT ATTTTATTAA ATATTAACGA GGTTTTCACT TACTATCATA GCGCTCGTGT 10800

TCTCTACACA AGTGTTCCTG CTCTTGAAAA CAACAAGTCT GAGCCTTTTT TTCAAGCCTT 10860

CGAAAATTTT TATTTTGAAT TAAAACAACA TTTTTTCAAT GAAGAAGATG AAACCAATCA 10920

ATTGAATGAA CGTCTAGAGG AAATGAAAAT TGCCTTTGAA CAATTGACGG ATGACTACAA 10980

TGTTCTATAA TAGAAGAGAA CAATTCGTTG AGAAGGGGTC CTTATGAAGC AAGAAGAAGT 11040

GATTGAACGT TTAAAAGAAG AACTAAATTT GCCATTTTTT AATGGTATGC TAGAAGAAAA 11100

GAACTATTCT GAAGCAGATT ATCAACAAAT CAAGAAAGAG TTAATTCAAT ACTTTGATGA 11160

TTATGTGCGA AATGTTGAAA ATTAACAAGC TAAAACAACC TTTGATTTTA ATAAAATCAA 11220

AGGTTGTTTT ATTATTAAAT AAAAAAAATA ACAAAAAAAC CACCGGCAGT TTTGCCGATG 11280

GAAAAGGAGT ATTTACTCAA TAAGAGTAAA ATGAAAAATA AAAAGGTTGT TGTTGGTATG 11340

AATTAATAAT ACAATACACT TTCTCTTTTG TCCATAAAAA AGACATGAAA AAAATATTAA 11400

ATTCGTTAGT TATTTTTATT AGTTTTTTAT TTTTTCAAAA GAATGTTGCT CATCCGCGTA 11460

ATGGCAAAGG TTAATTCATC CGTTGCTTGA TCCTGCTTTT CCTCTGAAGC ATTAAATAAA 11520

TCTGCATCAA TACCTGTTTC TGGATCTGTT GCTAAAATCA TTTCATTGCA TCCTGCTACA 11580

TATTCTGTAT AAGCATGTTG AAATTTTTTG TGAATGCCAA GTACTTGTGG CGTCGGTTTT 11640

AATGTTCCGA CTTTTTTGAG CATTAATTCA TATTTTTCAG TACCTTCTTT AAATTTTGCt 11700

TGAATTTCTG CAATTCGTtC GGCTGTTAAA TCATTCACTT tATTCTTATC GATCGCTTGA 11760

CGAACTTCTT CGTAATATGG ATyCATGACT TCGCCAATGy CTtCTGTkGT TTTAACGATA 11820

TCATtAATTG TTkGAACGTA AAATCTTAAA TTTGGTCTCA TAACCATCCC TCTTTTCCCT 11880

TTTCTCTtTC ACTCCATTTT ACACTTTTTC TATTCCATCT GACAAGTGGC ACTTTTTCAG 11940

TCTTCATTCT TTTCATTCTT AACAGAGAAA ACATAGTGAG GCATCGTTTT GCCACGGTAT 12000

GTTTTAATGT ATCGATCCCT AACAAGCATC TGGTTTCGAA TAGCGACATT CATAGAGGCT 12060

AAGTTCGTAT CACGAATCGT TGAAATAACT TCTGGAAAAT TTAATACATC GAATGCGTAC 12120

TGTTTTGTCG CTGCGGCAGC TTCAATCGCA TACCCTTTGT GCCATTCTTT CACACATAAA 12180

TGATAACCAA TTTCTGGATA GCTCTTACCT TTTACTAATT GATTGGTGAG ACCACACTCG 12240

CCGATGACTT CGCCTGTTCG CTTCTTCTCT AACGCCCATA AGCCGTAGCC GTATTCTTGA 12300

TAGGACTGTA AGTTCCACTT TAGCCAGTTT GCGACCTCTT CATCAGAAAA AGCGTGTTCA 12360

TAGGCATACA TCACATTTTC ATCTTGCAAA AATTTTTTCA ATTCCACTTC ATCTTCCGAC 12420

TGCCATTCTC TGACAATAAG TCGTTGTGTT TCGAAAATAA TTTTCCGTTT CAAGCCCTTT 12480

CACTCCTTAC ATAAAGAAAA ATTGACTGGT TTAAGTACCA GCCAATTTTC TTTTCTTATT 12540

TATTAAGCAC CTTTACCTGC TTTACGTTGG AAGAACTTCA CAATTTCCAC AATTGGGATG 12600

ATAGCAAAAG AAGTACCTGC TACAATTGCC CATTGATATG CATCTAAATG TGTGACGCTG 12660

AATAAATCAT TAAATCCTGG AATAACAATC GTTGCAGCTA ATAGCAAGAA CGAAACAAGA 12720

ATCCCATAGT TAAATGATTT GTTGCGGAAT AAGCCAACTT TGAAAATTGA TTGATAAATC 12780

GACTTCACAT TAAAGGCATG GAACAATTGA ATTAATCCCA AAGTAGCGAA GGCCATGGTT 12840

AAAGCATCGC CGTGTTGTAA ATCATACAAC GcTTGCGCCG ATAAGTTTGT GTTAGCTGCG 12900

GTATGTGCTG GGAACATAAT TGACATTTTG TACACAATCA AAGTCAATGC CCCTTGCGTA 12960

ATCCCTTGAT AAACAACACT GCTTAAAACG CCGCCAGAGA AGAAGTTCGA TTTTTTACCA 13020

CGAGGTTCAT GGCTCATCAC ATCACGTTCT GCTGGTTCCA CACCTAACGC AATCGCTGGG 13080

AAAGTATCCG TTACTAAGTT AATCCACAAT AAGTGAACAG GTAACAAGGT ATCCCAATTG 13140

AGCATTGTTG CAATAAACAA GGTTAATACT TCCCCTAAGT TAGCTGAAAG TAAATATTGA 13200

ATCGTTTTTT GGATATTTGA GAAGACTTTA CGTCCTTCTT CTACTGCCAC AATAATTGTT 13260

GAAAAGTTAT CATCGGCAAG GACCATATCT GAAGCACCTT TTGATACTTC TGTTCCGGTA 13320

ATTCCCATAC CAATCCCGAT GTCAGCAGCT TTTAAGGCTG GGGCATCATT TACGCCATCA 13380

CCAGTCATCG CAACCACTTT TCCTTCTTGT TGCCACGCTT TAACAATCCG AACCTTGTGT 13440

TCTGGTGAAA CACGAGCATA GACAGAGTAG TGTCCCACGA CTTGTGCGAA TTTTTCATCT 13500

GACAATTCAT TTAATTCAGC GCCTGTAATG ACTGCATCGT CGTCGCCTTC TTTGATGATG 13560

CCTAAACGTG CCGCAATTGC TTCTGCTGTA TCACGGTGGT CACCCGTAAT CATAATTGGA 13620

CGAATACCCG CTTCTTTCGC GACTTTTACA GCATCTGCAG CTTCTTTTCT TTCTGGGTCA 13680

ATCATACCGA CTAAACCAGC AAATGTTAAA TCTTTTTCCA CAAGTTCAGA ACTCATTTCG 13740

GCTGGAATCG TCTCAACATA TTTATAAGCC ATTCCTAAAA CACGTAAGGC TTGTTTCGCT 13800

AATGAGGTGT TGGTTTTCAA AATTTCTTGA CGTTTTGTTT CATCTAATGG ACTTGTTTCT 13860

CCATTTGATA AGATTTCCGT ACAACGTTTC AATAATTCAT CTGGTGCACC TTTGACAGAA 13920

ACTAGGAAGC CACCTGTTTT TAATTCATGA ACAGTAGTCA TTAATTTACG ATCAGAGTCA 13980

AAAGGAATCT CTGCCACACG AGGTTCAGCT GCTACTTTTT CGGTTACGTT AAACGCATGA 14040

TCTAAACCAA ATTGTACTAA GGCAGTTTCT GTTGGGTCAC CAATTAATGA ACCATCTTGA 14100

GCGATTTTTG TATCATTCGT AAAGTTCATG ATTTTTAAAG CCATATTATC GGCAGGGATT 14160

TCTGTGGAAG CAGATAAAAC TTGACCATCT GTATACAACG CTTCAACGGT CATTTGATTT 14220

AAGGTTAACG TACCTGTTTT ATCTGAACAG ATAATGTCTG TACTTCCTAA TGTTTCAACG 14280

GCTGGT 14286

(2) INFORMATION FOR SEQ ID NO: 163:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6723 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

CAATGCTCAT TTCCACTCnT CCTAAATCTG TTATTGCCAG TAGTTTTGCT CTAACTTTTC 60

```
TAAATaATGk GaATAGTAtC CATGCTGTCG aACCTCAAAT TCTTTTAAGT AATCtAATGG    120
AATACTCyTA CGtTCTAAGT TTTTTTGCAT ATTTAAATAA TCATCAACAT CCaCTAAAAA    180
TACTTGATGA CTTTTGTTGA ATTGAATTAA TAAAAATGCT GTTCCACCGC TTAATTTAAA    240
AnCCTTTAAA AAATTTTTTT GGTGATCTTT TAACATTGGT TTATTTCCAA CATAAAATGG    300
AAATGCAGTT TTATTTTCAG TAGATTTGCA ATCGAAGGCG ATTGGCCGTC CCTTCAAATG    360
TCCGATAAAA TCACATCCTG TCTTGTTAGT TGGAATCACT ACGGGTTTTC CACCAACCCT    420
TATAGTTTTA GTTCCATTAG GTATTTTTGC TACTGTGCCT TTTCTGTTAC GACAATACCA    480
CTCGTTGGTT tGTTCAATCA TTTTTTCAAA CTGACTCCAC GCTTTCATCA CAGCACCTCT    540
CTCTTAATTT CCCTCTGGAT TTAACCAGTT ATGCATTGTA ACTTTTGGTA TTGAATCGAC    600
AGAAATCTCC TCTTGATCAT TTCGCATATT GTATTCACTA ATTTTTTTTA TAGCTTGATT    660
TTTTGTCTTT ATTCCCTTAC GTTCCCAGCT TAACAATATG CGGTCCATAT ACTTTAGACT    720
GAATGCCTGA TTAAGAACTG CCTCTTTAAG GGCTAATTGA ATAAGGTCAT CTGGATAACC    780
ATCTTCTGTT TGCCACTGTT TAATCATTTC CATTTCGaTA GGCGAAAGTG GTCtCCCaAA    840
ATTTTGTTCg ACGGcAGCGT ATATATTTAA tTTaTTTTTA TCTGTTTCTG GTGAAAAAGA    900
ATTATTTGTG TTTATGTTTT GTTTATGTTT AATTAATGTG CCACTGTTCG TTGACTGCGT    960
TGTACTCTTT AGTGTCACTT TATCGTTACT CTGTTGTGTA CTCTCTTGCG CACTATTTAA   1020
CGTATAAAGT ACGCATACCT TGTAAGAAGT AGCTTTTCGA CCATTACTTT TAAAATCAAT   1080
CAAACCTAGT TGCTTTAACG CGnTTCTATT TTTATTAATT CCTGAGCGTG ACAAACCTGA   1140
CAAAGATTCT AATGTTGCAT TAGCTGCTGT AAACCATGTA GCCCATCCTG CTTTGTTGTT   1200
TATGGACATT AATGCACGCC ATAAAGCAAT CTGACCTGAT GAAAGCTTTT GTTTATAAAG   1260
CAAATAATCG TCAAACGCAA GAATCTGTTG TAAATAATTC AATGTCACAC CTCCACTCTC   1320
TATGCAATAA TTATtTTtTT ATTGGTTATC TtTTCTATtT CTtCTTTAAA TTTTTCTGGa   1380
TCACTATTCT TGTCGCTTAA ATGGATGAGA TATATTTCTT CTGTTTTTGT TAAATCAGTT   1440
GACTGAAAAA ACTTTTTGCA GGCATCTATG CTCATATGCG TTCTTAAAAT TCGATCTTGA   1500
ACACTTTTTG GCAGTTTGCT CTGTCTTACT AACTTGATAT CATGATTACA CTCAACGAGC   1560
CAATGAGTGA CGTCTTTAAA CGTTTTAGGT AAATAGTTAG TATCTGTAGC AAATACTATT   1620
TTCTTTCCAC TAGGTGAAAG AATAAGAAAG CCCAGTGGTT CTCTCGCCCG TGCCTTTTTG   1680
TCATCGTGAA TAGTGGCAAA AGGTTTAACT AACCAATCAC CAATTTTCTG TTGCTGGTTA   1740
GCTTTTAAAA TATGTGACCG TCTATTTATA CCTAATCCTT CTAAAGTTCC TCGTGAAGCC   1800
CACACATCAA ATCTTCCAGC GAGTAAAATA TCATTGATAT ACTTTGAGTG GTCACCATGT   1860
TCATGTGTTA CAAGTAAACC TTGAATATTT GAAAAATTAA TACCTTGTTT CATAATATCT   1920
TTAGGCTTTA ATCCTGCTTC TAACATTAAG GAAGAGTTTC CATCTGCAAG TAAATAATTA   1980
TTACCTGCAG ATGAAGAACC TTGTATATTA ATTTCAATCA TTAAAAGCCA CGTCCATCTG   2040
```

TTTCGCTAGC AATATTTGTG GTTAATTCAT CAAAAAACGC TGTCTGGGTT GGTTCGTTTT 2100

CTATTTCTTG ATTCGTTGTT GCTTTTTGAA TCGGCTCTTC TGGTTCGATG ATAGTCTCAT 2160

CTGCTGCAGT TTCTTTATCG AGTTTTACTG CTTCTTGTTT GACTTCTGTT GGTTCTTCAT 2220

CAAAATCGAA CACTTCTTGT GCTGTTTGCT CAGTGACATC TTTTCTAACA GATTTAACCA 2280

CTTCATCATC TTGCGACTGA TACTGCATTG CGACATAGGC GTTTTCAAAA TTTTTAGGAA 2340

TTTTCTTAAC AATGkTGTTT CTCATTTTAC GAACAATCAT GGATTCTCTA CTTTGTGGAG 2400

ATTTCCATGC GGGACTAATA TAAGTCTGGT ACTCTTCGCT ATCCAAGATG TCATCTAAAG 2460

ACATTGTTTT CAAATTGTTC ATGATTTCTT GCTTTTTATT TTCGATTTCT GTTCTCTGAC 2520

TCTGAGTTGC TTTGAATTTA TTTTCAGCTA GTCCAAACGT CTCATTCATC AAATTTTGGT 2580

TGATGTGGGC AATTAAGTTT TTAACAACGT CTGCACGCTC TGAAATATGG AACTCAACAG 2640

ATCCGTCTGT CATTTCAACT GGATAAACAA CTCTAACCAC TTTTCCTTTT CCAGTGGGTC 2700

CCCACTCTGG ATCAGTCACA GACAGCCCTT TGTATCCAGG GTACGAAAAA TGGTCCTCTT 2760

CCCGAACTTC CCAATGTCGA TGAACATGCG CCACATTTCG TCCGAACTTT GAAAGAATGG 2820

CGTCATTACC ATCTCCTTCA ATTCCCATTT CAATCTGTTT AATCCAAACA TCTTTCGGCT 2880

TGCCCGTGTT TGGATCATTC ACTTTTCTTT TAACATTTCG GGTTTGAAAA TATaCTTCGC 2940

GAGGAATCGC CGATGCATTC ACTTGTAATG CTGCGATGGT CATTAGCGTT TCTGTTATAT 3000

TCGTTGAATC TACATCATTA ATTGACAAAC TGGCATTTGT aAGCATTGTA TTAAtTGCTT 3060

GAAwTGAGCT GATTACGCAT TGTTTTTGAT ATTCATTCAT ATTAATTCCG TTGCTTAcAA 3120

GCTGCGAttC AACTTGTGGT AAAAACGTAT CGTTGATTTT AGTTAATCTA TTCTCGAAAA 3180

TTTTATTTTT TGCAATTTCA TTGGTCATCA ATATCTATTC CTTTCTCTGT CAAAGCAGCC 3240

ATCCATACTT TTTTAATTTC GTCGGGACAA TGACTCATAG CATCTTTCCA AGTTGGCCAT 3300

CTTCCGTTTT TGTCATAGAA TTTATATTGA TATGTTAAAC TTTGCTGATT ATGTGGTTGC 3360

TCTTCATCAT GCtTAACTGC ACAAATTTCG CAGGTTCCTT CCGGAACGTT TCCGATCATC 3420

ATAAACCCAG TTTCGTCTTT TAGAAATCCC ATCTATTTTG CACCTGCCTT TTCAATTGaA 3480

TGCTGTTCAA TTCTAATTTT TTTAyCTTTT TCGCTTACAT ACATCGCAAT GACTTGTGTA 3540

TCTACTTGAA CAGAGTCTCT ATTGTGATTA GTCAAACCTT CAGCGTTATC GATAAAAATT 3600

GGAACGATAT AGCCTTCCTG TTTCATTAGA GTGTTGGAGA CATCTAATCC AGCTTGCATA 3660

CGGCTACCGT TATTTAGAGA ACTGAACGGT ACTCCGTCAA TCATCGGTTC ACATACTGCT 3720

TCATTGAGTC CTCCATCTTC AAAGAAATCG AACAATTTCC ATTTAACAAC TGAGAAATGA 3780

CTATTTATAA TTTCTTGCAG CATGTTCTGT TTTGTTATAA AGAATTCTTC AAATAGTACT 3840

AATTTTTGTA ATACTTCACC TTTCTTATGA GATAATTGGC GCTCTTGTTC ATTGAAATCT 3900

TCAATAATCG ATAGCTGTCT TTCATACTCT TTTAGCAAGG CTAATTTTTC ATCAATCATT 3960

GCGATTTCTT TATCTATTTC TGTAATCTCA CTTGTTTTAG CGGAAGTTTG TTCTAGAATA 4020

GCTTCGTTAC TTTGAGTAAT GTACTCTTGC AATTTTTTTA TTTCTTTAGT GATAGTTGAA 4080

TATTTTTCAG TTGCTTCAAA TGGTATTTTG TCTAATTTTA ATGAGGAAAT CTGCTGCTCT 4140

ACATCAGCTA AGTTTTCCTT CACATCTTCC AAGTGTTTTT TAGCAATGTT ATATGCTTCT 4200

GTTTTAATCA ATAGTTGTTC TTTTAACTCG CCGATCCCTT TTTTTAGGgC TTCGCGATCT 4260

TTGTTATTTT GAATGCCTTT TTCTCGaATT TCAGAGAGCT TGATTTGCTT ATCCGCTTCA 4320

AACTGGGCTT TTATTTCTTT ATTTGTTAAC TCAATTTCTT CTGCACGTTT TTGCTCTTCT 4380

TCCTCATGAT GTCGTTTCAT TTCATCTTGa TCCTTGACAT CATATGGACG ATTACAGTGC 4440

TGACAAACTA AAAGATTTTC ATTGAAAGAT AATTTGGTAT AAACTAGACC GCCTGTAAAT 4500

TCTTCAGCTT CTACTTCATC ATATTTGTCG TACAATTCTT CATGCTTTTT ATTTAATGCA 4560

ATCAGTTCAT TATCTTTTAT AGAAaCCAAA CGTTCTGTTA CATTAAGACT TGATTCTTCA 4620

TCTGCATACG TCTTCTGGGC TTTATTGAGA TCAGCAAAGA GTTTTGACTT ACCTTGCTCA 4680

ATGCCATTAA TGCGTGCATT CTGTGCATTA TCATGCTTCA ATTTAGCAGC AGTTAATTCT 4740

TCTTGTTTTG TATTAAGACT TGCAATTAAT TCTGAAATAT TACCACCATT TCTAATAGTA 4800

ACGAGCTGGT TTTCGATGTC ATTTTTCTTT AATGTCAACT CGTTGCGAGT GGTTAATAGT 4860

TGCTCTTTGT TGATATTTTC AATATCTGGc AATGCTGCTT GAATACCTTC AATTTTTACA 4920

GGAATATTTT TCAGCGTTTC GTTGATcGCT TTTGTTCTTG AAGCACACGC TCACGAGCTG 4980

TTTTAATGTC ATCATTACCA ATAATTtCTT TTAATTGGTG AATTGATGGC GATTCGTTGA 5040

TAATCTCTkC ATCTGTCTTG CTGCCAAAGT AtTCAAAAAG CTTTTGACGT CGTtCATCTG 5100

CCACTAACTG CTCACAAAAA TAAGTCACAC TAGTCAGATT CTTAAATGTG TCTTGATCTA 5160

ATACTTTTTC TACTTCATCG TCAAATGCTT TCTTAGTTGT GGTCTCTAAT CCGTCTACGA 5220

GGTATTTTGT AAACATCTCA TAAGACTTGT GTTCCGAATT TCGTTTGATA ACCTCTTTGT 5280

CTCCTCGGAC CTTTTCAAAC TCCTTTGCCT GACCATTGAT AGTCAAAACG ACAGTTACTG 5340

ATGTCTGTTT ACCGCGAATT GGCTCGCTAT TTTCATCAAG TGGACGCCAC TGAATCTTAG 5400

TACGTTCTTT TGAATCTTTG TTGAACAAGC ACCAGAGAAA aGCATCaTAG ATAGTTGTTT 5460

kGCCTGCATC GTTGTCTCCA AAAATATCAA TACTTTTACC ATTTGGTTCA ATCATCAGGT 5520

CCGATATACC TTTAAAATTA TGAATACGGA TTGACTCCAA GCTAATGTTT TTCATACAAT 5580

TAATCTCCTT CAATTTCGAT ATTGTCTTCT AAGATTGCTC GTATAACTGT TTCTTTGCGT 5640

TTGTTACAAA TTTCAGTAGC GACATCATAC CAAGCACTAC TATGATTTTG ATTCAGCTCT 5700

GTTGTTAAAT CATATCTATA ACTTTTACTA ATTATTGGAT CAACTAACCA ATCCATACGT 5760

GCAATTAAGG CTGTAGCGTA TAATTCTCGT TCTGCGCAAT TTTCAGTCGC GTTTTTAAAC 5820

CATTCTTCAA ACCGTTTTGA CATGGTTAAT TTATTTGACA CATTATCCAT TTCATTTCCT 5880

CCAGATTTTG TGTTATAATT CTCATGTATA ATTTTTGTAT GGGACTTAAT CGTTTGTGGA 5940

CGAGTGAGTC TCTTTTTTTG TGTAAGCTAA ATATTTTGCA TCGTCGTATT TCATGAACCA 6000

AACAACTGCA ATTGGTCCAA TAATCAAAAG TAGATAGcTC GCTGGAACAC TATTTTTAAT 6060

TaCTAACCCT AATATAAAGA CAAGAATAAA TGATCCTAAA ATTCTAGCTt CGTATAAGCT 6120

TTnTGGTTTT CTTTCcTTCA TTTATTTTCC ACTCCTATTT ACCCATAGCA CTCTGATAAA 6180

CATTAAAAAA AGTGCTAATA TAATCGCTGT TAAAAATTGT GATTTAATCA AACATAAAAT 6240

AAACACAAAG AATAAAATTC CTACGGAAAG TGTCATCATT GTTTTAATTG CAAGTTCCTT 6300

ATTCTTCACT AGAGACCCTC CTATATAAAG TTTCTTTCAT ACCATTCTTC TAAATCATCG 6360

ACTTCTATTC TTACCAATCC ACCATTTCTT TTAGCAGGTA ATGGATCAAT TTCTCGTTTC 6420

ATCCATCTTG TAACAGTAGC AGCTGAAACA TTTTTTCTTT TTGgCTACTT CTATTGCTTT 6480

CAAAGGTGGn AATTTcTTTG aATTAACTTT TCGaTCkTTT TTgaTTTAAA ATCAATAACT 6540

TCTAGTGcCA TTkGwATyCT CCTTCCTCAT GtATCCTAGT kGTTCCCAAT ATGGAAAACG 6600

TTGCTCACTT AAATAACGAA TATCGATGAA AGCAAGATCA CACAAGCTAC TAAGTAGTGT 6660

GATTCAACAA TTACTCAnCA AATAnCCATT GCGnAGCACT ATTTAGAAAT ACTGTCAGAA 6720

AAA 6723

(2) INFORMATION FOR SEQ ID NO: 164:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27360 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

TAAGCGAATT GGACATCATA TGTCAATAAC AAATGTTGAT AAAAAAGGTC AATTAAACGC 60

TATTTTGAAA AATAAAATCA GCTTTTAGAT AATTTTGCCA AAAATTTCGC TTAATGAGTA 120

TTAAAAATAT TTTTTTATTT TATGTGCTCA AATATTTTTC AAAAAGAAAA AATTGTATTA 180

AGGTAATAGT AGAAAGGGAG GAGATTTAAT GACAGAACCA AACTTAAAAG ACCCAAGAAA 240

GTTATTTTAT AGTGATGGAT TTCCGCAACA AGATCAAGAA ACTCCCGCAT TACAGGACAA 300

AATGGTCCCT AAACCAGATT GTGGGGAAGA CAGCTATGTC GGAAATCACA AATTAGAAAA 360

TCGTCGGGTA TTAATTACAG GCGGTGATTC TGGGATTGGA CGAGCTGCGG CGATTGCTTT 420

TGCTCGTGAA GGCGCAGATA TCGCCTTACA TTTTTTCCCA GGTGAAGAAA AAGACGCCGA 480

AGAAGTGGCG CACTATATAC GAGAAGCTGG CCGAAAAGTC GTTCTTTTGC CAGCAGATTT 540

AAGAGATAAA CAGGCACCTG AAGAATTGGT TGCACAAGCT CATGAAGCAT TAGGCGGCTT 600

GGATACCCTC GTTTTAAATG CAGCCCAACA AATTTCCTGT GCCGCTATTG AAGAATTACC 660

AATGGAACAA GTGATAGACA CTTTCCATGT CAATATTATT GCGATGTTTG GCATTGTCAA 720

AGCAGCAGTG CCTCACTTGC CTGCCGGAAG TAGTATTGTG ACAACGACTT CCGTCCAAGC 780

CTTTAATCCC AGTGAGCATT TGTTGGATTA TGCAGCAACG AAGGCATCAA TTGCCAACTT 840

```
TACCGTTGGT TTAGCAAAAC AACTGGCACC AAAAGGCATT CGGGTAAATG GCGTGGCTCC    900
AGGACCAATC TGGACACCGT TACAGTTAGA TCATGGACAA CCAATTGAAG AGTTGCCAGA    960
ATTTGGGCAA CATTCTTTAC TGGAACGTGC GGGTCAGCCT GCTGAATTAG CCCCTGTTTA   1020
TGTTTTCTTG GCGTCCAATG ATGCAAGTTA TGTTACGGCT CAAGTTTATG GCGTAACTGG   1080
TGGAGAAGCA ATTAATCTTT AGGAGGGATG TTCCGTGAAT AAAGTTGTTA AAATACTTTT   1140
GCTTCTAGGT TCACTGATTT TAATGGCATT ATTTGGCGCG GTGGCCATCA CCACTTGGTC   1200
AGCCGATACA GCTTGGCAAC CACCAGATAA TGTCCAATGG TTCTTGCAAG ACAACGAGTA   1260
CGCACGTCAA GCAATCTTTT GGGTAGCGCT GGTTTTACTA GGAATTATGC TCTTTCTCTT   1320
TTTATTCATT CTTTTATTCC CAAGAAGAAA AGGCACGTTT GAATTAAAAG AAGAGCAAGG   1380
AAAGTTAGTG TTACACCGTA AAGCCATCGA AGGATTTGTA CGCTCTAGCT TGCAAGACAG   1440
TGACTTTCTG GAAGCACCAA CCATTCGTGT AAAGGCGACA AAACGGCGCA TTAAGGTCAA   1500
TGTTAAAGGC ACACTGAAAC GAACGGCAGA TTTAATTGGG CAAACAGAAG CTTGGTCAAA   1560
AACAATGGAA GAACGGCTTG CTCGTTTGGT GGGAACTGGT CACAAGATTA CGATTGATGT   1620
TCATTTTGAA GATATTCAAT CAATGAAACG ACCACAACAA CAGCAAACAC AAGCACGTGT   1680
GGAGTAGGAG GGATGTACAA TGGTTGAATG GATAAAACAA TATGCCGTTC CGTTAATTGG   1740
TGGCCTAGTC GGCTTGGTTT TTGCAATCTT GTTTCTGTCA ATCGGCTTCT TTAAAACATT   1800
GTTAGTGGTA CTCCTGACAG GACTAGGTAT TTGGCTAGGT TTGTATTTGG AACAGACAGG   1860
CATTGTCAAA AATTATTTTC ATCATAAATA GGAGGTTTTG GCTATGGAAA ATGCAAATGT   1920
ATCGATTGTT AAAGGTGAGT TGACGTTTGA GGATAAAGTA ATCCAAAAAA TTATTGGCAT   1980
TGCTTTGGAA GAAGTCGACG GTTTGTTAAC TGTGGaCGGT GGTTTTTTCT CGAATTTACG   2040
AGATAAAATG ATTAACAGTG AAGATGTTAC AACAGGCATT CACACCGAGG TCGGCAAAAA   2100
ACAAGTCGCA GTGGATATGG ACATTGTAGC CGAATACGGG AAAGATATCG AGACCATTTA   2160
TGATCAAATG AAAGAGGTTA TTACTCGCGA AGTCCAACAA ATGACTCACT TGGAAGTAAT   2220
CGAAGTCAAT GTCAACGTAG TAGATATCAA AACACAAGCA GAGTACCAAG AAGAAAGTGA   2280
AACGrTCCAA GATAAACTTG GCAACGCCGC TGAAGCGACA GGTAGTTTTA TGTCAAAACA   2340
AACCAACAAA GCCAAACATG CGGTAACTAA AGGCAAGACA AAAGTCAAAG AAAACACGGC   2400
ACCTCGTGTT CAATAATTGG TTATTATTTT ATTAAAAATA AAGCAATAAT AGAAGAAACT   2460
AGGAGGAATG AACCATGAGT AACGAAAAAT TCAATAATGT ACCAAAACCA GCAGGAAACG   2520
GACCACACAC ACCAGAAGAT CATGCAATCA AAGGTGAATT AACATTTGAA GATAAAGTCG   2580
TTCAAAAAAT TATCGGTTTA GCCTTAGAAA ATGTCGATGG TCTATTAACA GTCGATGGCG   2640
GCTTTTTCTC AAATATTGCT GAAAAATTAG TAAACACAGA TAACGTCACA GCGGGTATTG   2700
ATACCGAAGT TGGTAAAAAA CAAGTCGCTG TTGATATGGA TATCGTGGTA GAATACGGCA   2760
AAGATATTCA AGACATTTAC GAAAAAATGA AAGAATTAAT CAGTCGTGAA GTGAAAAAAA   2820
```

TGACCCACTT AGATGTGATT GAAGTGAATG TTAACGTAGT TGATATCAAA TCAAAAGAAG 2880

AATACGAAGA AGATAGCGAA ACGGTTCAAG ATAAAGTCAC AGGCGCTGCC AAAAGTACCG 2940

GTGAATTTGC TTCTGAACAA ACAGAAAAAG CAAAAAAAGC TGTGAATAAA GGAACAGAAC 3000

AAGTCAAAGA AAATATGGAA CCACGTGTAG AATAAACAGT TAAAGGGAGG AAACAATCAT 3060

GGGCTTTATT TGGGCATTAA TTGTCGGCGG GGTCATTGGG GCAATCGCTG GAGCAATTAC 3120

TAAAAAAGGA TCATCGATGG GCATTATTGC CAATATCATT GCAGGGTTAG TTGGTTCAAC 3180

AATTGGTCAA GCCATTTTAG GCACATGGGG ACCAAGCTTA GCTGGGaTGG CTATTGTGCC 3240

ATCGATTATC GGGGCAGTGA TTTTAGTTGC CATCGTTTCA TGGATTTTAG GACGAAAATA 3300

GAGAAAGTTT AAAAAAATAG GGCATAAGTC GTTTCGGCTT ATGTCCTATT TTTCTTTTTC 3360

TATGTATGAG GAAGTTTCGT ATGAAAAGTA AATTTTGAGC AATTAGCCAA GAGCTTATTT 3420

TAATTAAAAT CATTAGTCAT TTTGGCTTTT CAACAAAAAT CAACAACATT TGTTGAAGTT 3480

ATTTGTAGAT AACGTTTATC ATTCACAAAT TTACATTTTT TTGTAACGGA TATCTCGTAT 3540

AGTAATAACT GCAAATGAAT TTCAGAAAAT TTTTTGATAA GTCATCAAAG TGAGGGGATA 3600

GCGGTCTTTT ATTGAAAAAT AGCAACACAA AATAAACAGA ACAAAAGAGA GAGGAATTCA 3660

AAGAATTATG ACAGAACAAA CACCAACAAA GCTTTTGAAT AAAGGCTTTA TCAGTATTAC 3720

AGTGATTAAT TTCATTGTCT ATCTTGTTTA TTATTTATTA ATGGTTATTA TTGCGGTGAT 3780

TGCGCAAGAT AGCTTACACG CTACTTTAGG TCAAGCGGGT TTGGCCTCAG GAATTTACAT 3840

CATCGGGACA TTGCTTGCAC GACTTTTCAT GGGGAAACTC CTTGAACTGA TTGGGCGCAA 3900

ACAAGTCTTA CGTTATGGGG CTTTGTTTTA TCTTTTAACG ACTGTTGCGT ATATGTATAT 3960

GCCAAGTATG GGCATTCTTT ATCTGGTCCG TTTTTTAAAT GGGTTTGGTT ACGGAACCGT 4020

TTCGACAGCG ACCAACGCAA TTGTTACCGC CTATATCCCT AAAAACAAAA AAGGGGAAGG 4080

GATCAATTAT TATGGTTTGA GTACTAGTTT AGCAGCCGGC ATTGGGCCAT TTATTGGGAT 4140

GTTATTGTTA AATGTTTCCA ACTTCCATGT CATTATTAAC TTTTCAATTA TTTTAATTTT 4200

GTTAACAACC ATTGCTTGCT TCATTTTCCC AGTGAAAAAT ATTGAGTTAA CACCTGAACA 4260

TCGTGAAGCT TTATCTAAAT GGAATTTTGA TAGTTTTGTT GAAAAGAAAG TTTTATTTAT 4320

TACATTTATC GCCTTTTTAA TGGGTCTTGC TTATTCAAGT GTCTTATCTT TCTTATCTTC 4380

TTATGTCAAA GTAATCGATT TAGTTGACGT CAGCACCTTC TTCTTTATCG TCTATGCGGT 4440

CGTGATTACC TTAACACGTC CATCGACAGG TCGGATTTTC GATGTCAAAG GTGAACGCTA 4500

CGTCATGTAT CCGAGCTACA TTTTCTTAAC ACTAGGCTTA TTCTTATTAA GTATGACTAC 4560

GTCTGGTTGG ATGTTATTAG TTTCAGGTGG GTTAATTGGC TTAGGTTATG GTACTTTCAT 4620

GTCAAATGGC CAAGCGGTTT GTCTTCAAGA AAGCCCAAGT CCTCACCGTA TCGGGATTGC 4680

TTTATCGACT TATTTCATCG GTTTAGACCT TGGATTAGGG GTTGGTCCGT ACGTTTTAGG 4740

TGAATTACGT AACTTTATGT CATTCCAACA AATGTACTTC TTAGCAGGCT GTATCCCAAT 4800

CGTTTGTACC ATTTTATATA TGGTCTTCCA GAAAGCTAAA AATGATGCCA AAGATTTGTC 4860

TCTTGAAACG ATTGAAGAGA TTGAACACGG CTCAGAACTA TAAATGGACA AGGAGTGATT 4920

TCAATGCCTA ACGATGAACG GATTTTAAAC GTTTTACAAC AAACACTAAC GGAAATTCAA 4980

CGCATGAACA CGCGTTTAGC TAAAATTGAA ACAGAGTTAG CCGATGTCGA TCAACATCAC 5040

GCGACTTTGG ATCATTACTT GCAAGAGCTA GAATTGGTGA CAGAAAATAA CAACGTCATG 5100

TAAAGTGAGT AGTTAGTAAA TGAATGTTCT ACGTTTCGGT ATCACTAGAG TGGAACATTG 5160

AAAATTTTCG TAAAAAACGT CAAAAGGTAG CAGTTTTTCT GCTACCTTTT GGCGTTTTTT 5220

GTAATTTTTA TCACTTAAAA TGTTTCACGT GAAAACTTTT AAAAGTGGCG TGGACTTTGT 5280

TTGGTGACAA TTTGTGTATT TAAATTTTCT AAGCTTTGCG CTAGTTCACC AATGCTAGCA 5340

GAAAAATCAA CATAATAAGT GGGAACCGTT GATTTGACAC TAACTTCTGA AAAGATATCT 5400

GTAATAATTA CATCGAAAGG CTTTTCTGGT GAAGCGACAA AAATTAATTT TTTATCCACG 5460

GCTTTACTTA GCTTTTCATA AAGGGTATCT AAAATTAAAT TACTTTGATT GGTCCTTACT 5520

AAGACACGAA TCGGTGGTAA AGATTTCGTT TTCAATGTAC AAAAAAGAAT AAACAAGCAG 5580

TCAACTAACA ATTTTTCATT GCCAATCAAT GTCGGAAGCA TTTCAATTGT AAGGCGCTTC 5640

ACATCTTTGT AGAGACTTGG ATTTTTTGAA GTGCTTCTTT TAAAAAGCTT TCTTGCACAA 5700

ATGGATTGGC AATCGTGACA AAGGTCTTAA AAAAGACACT TAAATTTTGG ATAGAAGTAT 5760

ATAGATAAAG ATACTCGTGA AATGTCAATT TAATTTGATA AAAAGACGTT AGTTGATAAA 5820

TGATCTCTTT AATGGGTGAG CCTAATAGCT GGACCATATG CATCCGTGCT GCAGGAATTT 5880

CCAAGACATC TTGTGTGGCG TAgcTGaCAC TAATCGTCAA TAAGAAAAAG GCAAAAAGGC 5940

CTTCTTTTTG TATAAATTTT GTTGGAACTT GGTGTAAGGC TTCTTCTTTA AAAAACAATT 6000

CAAATTCACT CAGATTAACA AAGGGATTAA CAATCGCTTC TTCCTCCACG AAAGTTTCTT 6060

TGATAAAAAA GCCCTCTTTC CAACGCGCTA AAATCACCGC AGAAAAAATC AAGTCTTTTG 6120

TAAAATCTTG TGTTGTATTG AAAGAAACAG AGTGTTTTAA ACGTCGAAAG AGACTATGGA 6180

CTTCTTTTTT AGTAATCGTG TGCAACGGCC AATCCCCTTC AGAGTAATTG TTCCAAAAAA 6240

GCATCATAAA GAAAATCCGG ATATCTTTTT CTTTGCCCAC CAACACATTG GTACGTTTTA 6300

AATTGATTTC AATCTGATAT TTTTGGAGCC ATTGGTTTAA TTGTCCAATT TTTCGTGTAA 6360

AGGTTGATGA AGAAATCTCT AGGACTTTTT CAAAATGTTT GCTATCAATT GCATTATGGT 6420

TAAAGGCATA GGTCAGTACT TGATACGCGA AACTTTTTTT GACAAAGTAG GGCATCAAAA 6480

ACGACAGTTG GAAATAGGGT GCAAACCAAA ATGTGATTTC TTTGGCATTT GCTGATAAAG 6540

AAAATAAGTT CTCCCAGCCG TTTTTTTCAA TTTCTGTTTG GAGCAAAGCT ACAGATTGTA 6600

GAATGGTATT TTGTTGCATT TTTAATTGAG TCGCTAAAGT AGAAATCGCC AATGTTTGTT 6660

CTTCTGCTTC ACGACGTTCG AGTAACTCTT CTAAAAGAGC AAAACTTTTA CGATAAGGTA 6720

CATCAAGTAA CCCTGTTAAC ATCCGATTCC CTCCAAACTG TTTTCCCCAC GCAATTTAGC 6780

AAAATAAATT TTGCGAACTA CTTGTTTTAG GTTATAGTAT TCGTCTAAAT AAGGAACGGG 6840

ATGCCAATAA AAAGTTTTAC TGGCGTGAAA AAGCGGATTA TTTTTGTCAG TAATGACTAA 6900

GTCAGGCACT TCTTGAATAT GATTAACAAA TTTCAAAGGA ACATCGCAAT AATTATTGAT 6960

TTCCTTCGCT AACCAATTTT CTTGTATCTG ATTAATAGAT GAATGAAGTA AAATTTTAAC 7020

AGGAATTTGT TTATTTAATA AAACACTCCG AATAATTAAA AAAATGGTTT CTTTAAATAA 7080

CGCATCTAGT GGATGTTGTT TGTCAAAAGG CAAACGCTTT AAGAAATAGT TGGTTGTTCG 7140

TGAAACATTT GGGTATTGTT TTTCATACAA AGAGACATTA GCAACCTCAT AGCTCAATTC 7200

CTCTTCCACA GGGAAATGAA TCAGTCGTTC ACTGGCATTG GATAAATTAA CCAACAAAAA 7260

GACCACTTCT ATTGCGCTTA AATCCAACTG TAATTGCGTA CATTTTTCAA AAATCCAAGT 7320

AGCGTGTGGA CTGTTATCCA TTCCCATGAT CGCTAAGTTG GCAAAGTTGA TGGTTTCCAT 7380

TTGGACCAGA GAGTACATCG TTAATGTGCT GAAGAAATCA TAATAAAAAT GAATCTCAGC 7440

TTGCCAAGTG TCAGGCTCTT TTTGTTCATC AATAAAACAA TCATTCACTA ATAGTTGGAA 7500

TTCTGAATAA GGAAAGGCCA AATTAATAGC AGCAACTTCT GGATAATTGG TGATTGTAAA 7560

ACCGGATTGA TAACGGAGTA AACCTATGGA CATAATTAAT TTTAATTTGC GTAAATCAGT 7620

ATATGTAAAG TAAGGGTAAC GGTTTTTGAT GATTTCAATG AAGTTCTGCA ATTGTGCTTG 7680

ATTAGTCTTA GATAGAAATG GTACTAGAGA AGTTCCTGAA ATGAAGAATA AGGAATAAAA 7740

GAAACTGCGA AAATCTTTCT CGTTACCACA AATTTTTTCT CGTCTTTTTA AGTTTAAATA 7800

ACAATTAAAT CTAGCTAGAT GATTATTTAA ACGGTCACTG ACTTGTTTAA AAATAGAAAA 7860

GTTAATAAAG TGCTCTTCAC AAAAATCTTT GGCTGAAGTG TAGCGTTTTA AAAGTAAATC 7920

CAGGGCTAAC ATATAGGTTA GTGACCCGCG GACAACAGCT TCTTTCATTA ACTCAATCGG 7980

GATATTTTCA TCGGCGACTA ATTCTATGGA CTGTTGATTA TTAATAATAC GAATATGTAC 8040

ATTTTCATGT GTATTGACTC GAGAAATTAA CAAAGTTAAT TGATTCAATA AATTGTTGAT 8100

ATTGTAATGG GAAAATTCAT ATTTTCGTCG GAGTACTTGC ATGGGGACTT TTTCTTTACT 8160

GTCATACAAG TATTCAATTA GGGAAAGTAA CTGAATTTCC TGTTTGGTTA ATAAGTTATA 8220

CATAATAACC ACCTTTCTAA TTGGTTATGA TTGTTTGATT TTTCAGATAA TACAGATAGG 8280

GCATTCTTTT CTATATGACA TTGACGATTG CCAGTGAAAG AACTAGTTGT TCTTATGCTA 8340

CTGTATCTAT TCTTAAAAAT CAAGAAGAAC TGAATGTGTC AAATGGAAGA AAACATGGAC 8400

AAAAAATGAT AGAGCAAAAA ATTATTAAGA AAAACAACCA GTTTATATAA AACATTTAAA 8460

TAACAAAAAT GATAATTTGA GTCTATTGAT TTTTTAGCTT AAATAGCTTA TACAAATAGG 8520

TTTTTTTGTC TGGTTAGTTC TTATGAAAAA AATTAAAAAA AACGGTATAG TGGAAGAAGG 8580

AATAGTAAAA ATAAGAAAGG ATAAGGACGT GTTTTCATAA AGATGAAAAG GGTTTCTGGG 8640

ATTTCATCAA CTTTTTCTAT TTGGAGGAGA AGTTGTTGAT AATTTTTGCT GAAAAATCTC 8700

GTTATACTCT TCTTATTCTT ATAGAAGAAT CTATAGGAAA AAGCAGAAAA TGAATCCGCA 8760

TTCGAAGTAA GAGTGCGTTT TGAGGAGATA AAAAGAAAAG GTGTGTTGGC TGTCTGTTTT 8820

AGCAAGGAAG AAAGACAACA GAAGAGCACT CACTTATTTG GTTTAAAATA AAAAATTGGT 8880

ACGAAGTGAA CGTTCTCTTC TATGTGTCGT TAGTAGAGGA AGGATGAAAG AAATGAGAAA 8940

GAATGGTCCA ATGGTAAACC GTTGGCTCTA CGGGTTGATG TGTTTGTTAC TTGTTCTAAA 9000

TTATGGCACA CCACTCATGG CTTTGGCGGA AGAGGTTAAC AGCGATGGCC AGTTAACGTT 9060

AGGAGAAGTG AAGCAAACCA GCCAGCAAGA AATGACCTTA GCGCTTCAAG GAAAAGCACA 9120

ACCAGTAACA CAAGAGGTTG TAGTGCATTA TAGTGCCAAT GTGTCAATCA AAGCTGCACA 9180

TTGGGCAGCG CCCAATAATA CGCGCAAGAT TCAAGTGGAT GACCAGAAGA AACAGATTCA 9240

AATTGAATTG AATCAGCAAG CGTTAgcAGA TACGTTAGTC TTAACGTTGA ACCCTACAGC 9300

TACAGAAGAT GTGACGTTTT CTTATGGACA ACAGCAACGA GCGTTGACGT TAAAGACTGG 9360

TACTGATCCG ACAGAATCAA CGGCAATCAC GAGTTCGCCA GCCGCATCAG CGAATGAAGG 9420

TTCAACAGAA GAAGCATCTA CAAACTCCTC TGTTCCTCGT TCGTCCGAAG AAACTGTCGC 9480

CAGCACGACA AAAGCGATAG AAAGTAAAAC AACTGAATCG ACGACTGTCA AACCGCGCGT 9540

AGCAGGACCA ACAGATATCA GTGATTATTT TACAGGTGAT GAAACAACGA TTATCGATAA 9600

TTTTGAAGAT CCGATTTATT TAAATCCTGA TGGAACACCA GCAACACCGC CGTATAAAGA 9660

AGATGTGACC ATTCATTGGA ACTTTAACTG GTCGATTCCA GAAGATGTGC GAGAACAAAT 9720

GAAAGCAGGC GATTACTTCG AGTTTCAATT ACCTGGCAAT TTGAAACCTA ATAAACCAGG 9780

TTCAGGTGAT TTAGTTGATG CAGAAGGCAA TGTCTATGGA ACCTACACAA TTAGTGAAGA 9840

TGGTACGGTT CGTTTTACCT TTAATGAGCG AATCACGTCT GAAAGTGACA TTCACGGGGA 9900

CTTTTCTTTA GATACTCATT TGAATGATTC AGATGGGCGG GGCCCAGGAG ATTGGGTGAT 9960

TGATATTCCT ACACAAGAAG ATTTGCCGCC TGTAGTGATT CCAATTGTCC CAGATACCGA 10020

ACAACAAATT GATAAACAAG GCCATTTTGA TCGAACGCCC AATCCTAGTG CGATTACTTG 10080

GACGGTAGAT ATCAATCAAG CGATGAAAGA TCAAACAAAT CCAACTGTGA CGGAAACATG 10140

GCCAACAGGG AATACCTTTA AGTCCGTGAA AGTCTATGAG TTAGTGATGA ATCTTGATGG 10200

AACAATTAAA GAAGTGGGTC GCGAACTTAG TCCAGATGAA TATACCGTTG ATAAAAATGG 10260

CAATGTGACG ATTAAAGGTG ACACCAACAA AGCGTATCGT CTTGAGTACC AAACGACGAT 10320

TGACGAGGCG GTTATTCCAG ATGGCGGCGG CGATGTGCCT TTTAAAAATC ACGCGACGTT 10380

AACAAGTGAT AATAATCCAA ATGGGTTAGA TGCTGAAGCA ACTGTTACCG CCACATATGG 10440

CAAAATGTTA GACAAGCGCA ATATAGATTA CGACGAAGCC AATCAAGAAT TCACTTGGGA 10500

AATTAACTAC AACTATGGTG AACAAACCAT TCCAAAAGAC CAAGCAGTCA TTACAGACAC 10560

AATGGGGGAT AATTTAACGT TTGAACCAGA TTCTTTACAT TTATATTCAG TGACATTTGA 10620

TGACAAAGGA AATGAAGTCG TTGGAGCAGA ACTTGTGGAA GGAAAAGATT ACAAAGTGGT 10680

AATCAACGGA GACGGTTCCT TTGCAATTGA CTTTTTACAT GATGTGACTG GCGCAGTCAA 10740

GATTGATTAT AAAACCAAAG TTGATGGAAT TGTCGAAGGC GATGTTGCCG TGAATAATCG 10800

TGTGGATGTT GGCACTGGTC AGCATTCAGA AGATGATGGC ACAGCCAGTC AACAAAATAT 10860

TATTAAAAAC ACTGGTGCAG TTGATTATCA AAATTCAACG ATTGGTTGGA CGTTAGCTGT 10920

GAATCAAAAT AATTATTTGA TGGAAAATGC CGTGATTACG GATACGTACG AACCAGTTCC 10980

TGGCTTAACT ATGGTACCCA ATTCGTTGGT TGTCAAAGAT ACAACCACTG GTGCTCAGTT 11040

GACGTTAGGC AAGGATTTCA TGGTAGAAAT AACTCGTAAT GCAGATGGTG AAACAGGCTT 11100

TAAGGTAAGT TTTATAGGGG CGTATGCCAA AACAAGTGAT GCCTTCCACA TAACTTATAC 11160

TACCTTTTTC GATGTTACCG AGTTAGACGC TAACAATCCT GCGTTGGACC ATTATCGAAA 11220

TACCGCTGCC ATTGATTGGA CGGATGAAGC AGGAAACAAT CATCATTCAG AAGATAGTAA 11280

ACCGTTTAAA CCTTTACCTG CTTTTGATTT AAATGCGCAA AAAAGCGGTG TTTACAATGC 11340

CGTCACCAAA GAAATCACTT GGACGATTGC GGTTAATTTA AGTAATAATC GTTTAGTCGA 11400

CGCCTTTTTG ACGGATCCAA TTTTAACCAA TCAAACCTAT TTGGCTGGGA GCTTGAAAGT 11460

CTATGAAGGC AATACAAAGC CAGATGGTTC GGTTGAAAAA GTGAAACCAA CGCAACCGTT 11520

GACGGATATC ACAATGGAAG AACCAAGCGA GAAAAACCAA AATACTTGGC GTGTTGATTT 11580

TCCTAATGAT AGTCGTACGT ATGTGATTGA ATTTAAGACG TCTGTTGATG AAAAAGTTAT 11640

CGAAGGTTCG GCTAGTTATG ACAATACCGC ATCTTATACA AACCAAGGTT CTTCACGTGA 11700

TGTGACAGGA AAAGTTTCTA TTCAACATGG TGGCGAATCA GTGAAAAAAG GTGGCGAATA 11760

CCACAAAGAT GATCCAGATC ATGTGTACTG GCATGTAATG ATCAATGGCG CCCAATCGGT 11820

TTTAGACGAT GTGGTTATTA CTGATACACC CTCACCAAAC CAAGTGCTAG ATCCCGAGTC 11880

ATTGGTGATT TACGGTACCA ACGTAACAGA AGACGGAACT ATTACGCCAG ATAAATCTGT 11940

TATTTTAGAA GAAGGAAAAG ATTACACACT GGaAGTTACC ACCGATAATG aAACAGGACA 12000

ACAAAAAATT GTCGTTAAAA TGGCCCATAT TGAAGCACCT TATTATATGG AATATCGTAG 12060

TTTAGTGACT TCTTCAGCGG CGGGGAGTAC AGACACGGTA TCCAACCAAG TGTCAATTAC 12120

TGGAAATGGT TCAGAAGTCG TTCATGGGGA TGACAATGGC GATGTGGTCG TTGACATTGA 12180

TCACAGTGGC GGGCATGCCA CAGGGACTAA AGGCAAAATT CAGCTGAAGA AAACAGCCAT 12240

GGATGAGACG ACTATTTTAG CAGGCGCCCA TTTCCAAATT TGGGACCAAG CTAAAACACA 12300

AGTCCTACGT GAAGGTACAG TAGATGCCAC CGGGGTTATC ACATTTGGTG GGTTGCCACA 12360

AGGGCAATAC ATTTTGGTGG AGACAAAAGC ACCAGAAGGC TATACAGTTT CGGACGAATT 12420

AGCTAAAGGC CGAGTCATTA CTATTGATGA AGAAACTTCA GCCGAAGGAG CACAACCAAC 12480

CATTATTAAA AACGATGTCA ATAAAGTATT TTTAGAAAAA ATGGATGAGA AGGGTAAAAA 12540

GTTAGTCAAT GCTCGCTTTA AATTAGAGCA TGCCGTAACC ACGCCGTTTA CTCATTGGGA 12600

AGAAGTTCCC CTTGCGCCGG ATCGAACCAA CGCGAATGGC CAGTTAGAGG TGGATAGTTT 12660

AAAACCAGGG CTTTATCAGT TCACAGAAAT CGAAGCACCG ACAGGCTATC TTTTAGACAC 12720

GACCCCCAAA CGATTCATCG TGACACAAAA TACGAGCGGA CAAATTCGTG ATGTTCATGT 12780

CAAAATGCTT AATTACCAAG GTTCTGCTGA ACTAATTAAA AAAGACCAAG CAGGCAATCC 12840

ATTAGCAGGT GCTGAATTTT CAGTCCTTGA CACCACAGGA CAAGCAGTTC GAGAACACTT 12900

AGTTTCGGAT GCAAACGGAA AAGTCACAGT GACGGATTTA GCCCCAGGAA AATATCAATT 12960

TGTGGAAACC AAAGCGCCAG CAGGGTACCT TTTAAACACT GAACCAAGTG CTTTCACGAT 13020

TGCAGCAAGC GATCGGGGCA AACCAGCAAC AGTTATAGCA ACGGCTAACT TTGTTAACTA 13080

TCAAGGCACG GCTAAATTAA TCAAAAAAGA TGTGAATGGA CACTTATTAA GTGGTGCGAC 13140

ATTTAAAGTG CTTGATGCGA AGGGAGAAAC GATTCAAACA GGCTTGACGA CAAATAATCA 13200

AGGGGAAATT GTTGCAGAGC ACTTAGCCCC AGGAAAATAT CGCTTTGTAG AAACCAAAGC 13260

GCCAACAGGC TATTTATTAA ATACCACGCC AGTCCCATTT GAAATTGCTG AGAAAAATGC 13320

TGGTAAACCA GCGGTCGTGG TTGCTAGTGA CAACTTTGTG AGTTACAAAG GGGCTTTCCA 13380

AATCGTGAAA ACGAATAGCG CAGACCAACC ATTAGCAGGT GCTGTTTTTG AATTATATGA 13440

TCACAATAAA CAATCATTAG GGATTACAGC AACGAGTGGC AAAGATGGcA AAATTATCTT 13500

TAGAGACTTG GCGCCAGGTA CCTATTATTA CAAAGAAATC AAAGCACCAA AATTACCAGA 13560

TGGCGCAGAT TATATTATTT ATCCTGAATT AGTAAAAGTA GAAATTCGTG GTGATTTCAA 13620

AGGTGATCCG GAGATTTTCC AATTAGGGGC CTTCGCCAAT TTCAAAGGAC GCGCCGTCTT 13680

TAAGAAAATT GATGCCAATG CGAACCCACT TCCAGGAACG ATTTTTAAAT TGTATCGAAT 13740

CGAAAACGGG GAAAAAATCT TTGAAAGAGA AGTAACTGCT GAAAAAGATG GTTCATTGGC 13800

TATGGAGGAT TTAGGTGCTG GTAGCTATGA ATTAGATGAA CTGGATGCAA CGGATGGCTA 13860

TATCGTCAAT AAACAACCCA TTTATTTTGT AGTGAAGAAG AATTCAAATG ATAAACAACC 13920

ACTAGATGAG TTAGAGTTTG TAAATTATCA AGCAGAAGTA ATGGGACGTA AAGTCAACGA 13980

GCAAGGTCAA ACCTTAGCGG GTGCAGTTTT TGCAATTTAC AATGCCGATG AGCAGAATCA 14040

GCCCCAAGGT TCACCGATAA CATTCTTGAA TCGTGCAGGA GAAAAAGTTT CTGAAATAAC 14100

AACGGATAAG ACTGGCGAAA TTTACGCTAA AGGGCTAAAT GAAGGGCATT ACGTTTTAGT 14160

GGAAACGAAA GCACCAACAG GCTATCTGTT AGACACAACG CTACATCCAT TTGATGTAAC 14220

CGCCCAATTA GGAAAAGAGC AGCCAATTGC TTTAGGCGAT CTTATCAATT ATCAAGGAAC 14280

TGCTCAATTA ACCAAAGAAA ACGAAACAGG TGAAGCATTG GCAGGTGCGG TGTTTAAGGT 14340

CATTGATGAA ACAGGGCAAA CCGTAGATGG ACAAACCAAT CTGATGTCTG ACAAGCAAGG 14400

CAAAGTCATT GCGAAAAaCT TAGCACCGGG AACGTATCGT TTTGTGGAGA CaCAAGCGCC 14460

AACTAGCTAT CTTCTTAATG AAACGCCAAG CGCAAGCTTT ACGATTGCCA AAGACAACCA 14520

AGGCAAACCA GCCACTGTGG TACTTAAAGC ACCTTTTATT AATTACCAAG GTGCTGCCAA 14580

GCTGGTGAAA ATTGATCAGC AAAAGAATGC CTTAGCAGGT GCTGAATTTA AAGTGACAGA 14640

TGCAGAGACA GGGCAAACTG TCGCTCGTTC ATTACGTTCT GACAACCAAG GGTTAGTTCA 14700

AGTGAACCAC TTACAACCAG GAAAATATAC CTTTGTGGAA AGAAAAGCAC CGGATGGTTA 14760

CCAACTGTCT AAGCAAGCTG TCGCATTCAC TATTGCGGCA ACAGCGAAAG ACAAACCTGA 14820

ACTCGTGAAT GCGGGCACGT TTGTTAACGA GAAACAACCT GTATCCAAAA AAACAAAACC 14880

AAATCAGCCA ACAACGAAAC AAGCAGCTAG AGAGACAGGT TGGCTTGGTT TACCGAAAAC 14940

CAACACACAA GTCAATTACT TCTTTGTCTT TATCGGCCTC ATGTTGGTCG GTTTGGCAAG 15000

TTGGCTCTTC TATAAAAAGA GCAAGAAATA AGCAAAAGCT ATAAAAAAAC AGAAAGrGCC 15060

TGGAACAAAA ATCACTTTGG ATTTTTGTTC CAGGCyCAAA ACCTTATAAA CGGCGGTAAC 15120

AGAACCAACT CCTTCGGAAA TAAGCCGAAA TTCTCCAAAA ATTAAAGAGC AATTTTCGGA 15180

AATTTCTTCT TATTTCTCGG AGCTAAACGG TTCTGTACCG ACCTCTTTTT TCATCTTAAG 15240

CTAATCAACT GAACAAAAGC AGCTCCTTCT TATTGACATT CTCAGGTAAG CCTTCATCTC 15300

AGTTTACCCT TTGAACAGAA CTTACTGGGT TCTTGGAGAT AACCGCTTTT GCCTCAGCCT 15360

TTGTTTTTCG ATGCTATCTT TAGGTATTTT GGCTAAAATA TTGTACACTT TACATGTGCA 15420

AGAAATTTAG TTAGAGATAT GAGGATAGAA TGATGAAAAA GATGATGGTT TTATACAACG 15480

AAACTTCAGG AAGTAGTGAA AGTAAAGAGA TTGCAGAACG CTTTAAAAAA GCTGCAGAAG 15540

CGAGAGGGGA AGCGGTTATT TTGCAACCTT CCAATCCAGA CATTGATCCT GAAGAAATGC 15600

GGAAAAATGC CAAAGAAAAT CAAGTGGGTG TCTTGGTCGT CATTGGTGGT GACGGGACCA 15660

TTCATCATGC CGTTCAAAAT TTTAAAGACA CCATTCGTGA CTATCAAATA GGGATTATTC 15720

CAGGTGGAAC AGTCAATAAT TTTGCACGTG TGTTAAGTAT TCCTTTGAAA GAAGAAGACG 15780

CCTTTGAGAC CATTTTGGCA GGACAGACCA CACCAGTTGA CTTTGGGATG GTAAATCAAG 15840

ATGTGATGAT TTCGACCTTA ACGATTGGAC TATTAGCAGA TACAGCAGCA AATGTGACCC 15900

AACAGGAGAA ACAAAAATAT GGTCCTTTAG CCTTTACAAA GCAATTTTTC CGCTTATTAA 15960

TGAAAAAGAA AAAATACAAA CTAAAAATTG ATGGCGATGA AAGACGTTGG CATGGCAAAG 16020

CCCAATTATT AACGATGACC ATGACGAATT CTGCTGGTGG CTTTACCAAT TTCGATGCCA 16080

ATGctACGCC AGATGACGGC GAAGTCCACA TTATCATTTT ACCTAAGTTA GTTTTTTACA 16140

AGTTTGTCTA TTATTTACCT AAAATTATCA GAGGACAATT AAATGAAATT CCTGGTGTCG 16200

TGTATTTCTC TGGGAGTCAA TTTAAAATTA GTGGCGAAAA AGACAAGAAA GTCCAAACAC 16260

GAACAGATGG CGACCCTACT GATGATTTAC CGATTAAAGT AACGGTGGAA GCTGGGGGAC 16320

TCAACGTGTT TGTTCCTGAA ACGTCCACAT CTACGAAAGA ATAAAAGAAA CAACTGTTAT 16380

TTTTTCGTGG AATAGATTTC TCATTTTGTG TACAATAGAA CAGAAGAGAG GCAGTTGAGT 16440

GAACAGTGAA ATGGGATGAA TATGTTGATA TATTCGATAC AGATGTCGTT TCTAAAATTA 16500

TTATTTTAAG AAACTTGCCA ATTAATGAAG GCTGGATTGA CATCAATGAG TTAGCAGAAG 16560

AATTACAATT AAACAAAAAA TCAGTCAAAA AATACGTAGA ACTTCTAAAA GATGACATTA 16620

CGTATTATGT TGTGGATAAT GATGCTTCTT TGGCTTTTGA AAAGGGGCAT GGCTATCGCT 16680

TGAACTTGGC GTCTAGCCAA GCGTTCCAGA AACTTTATAC AGCAATTTTT CGGGATAGTT 16740

TAACTTATCA AGTAATGGAA GATTTATTCT GGGGTCGATT TAATGGAATT GTGCCAGCGA 16800

CAGCAAAGTA TCATACTAGT GATGCCACGA TTCGGCGCTT GCTCAAAGGG TTTACAGAAA 16860

TTTTAGCACC ATTGGGTTTA GTAATTAATA AAAGTACTTG GAAAATTGAA GGCAATGAAG 16920

CGTATTTGCG CCATTTTGCT TACACTATTT TTATGGACTT AATCCGTGAT ACTTGGCCTT 16980

TTGACTATAT ACAAGAAACG GAAGTCGCAC AAACCGTCCA AAAAATGATG GCTTTTTTCA 17040

ACGTAGAAGT TGCGGAAACG GTGACACGCC GAGTTAAATA TATGGTCGCT ATCTCAAAAG 17100

TCCGTAGTGC GCAAGGCGCC GAATATCGAC CTTCAGCCGA GCAAGAAACT TATTTAGAAG 17160

GTAATAAACA TTACCAAGCC TTTATAGAAG AGATGACAGC TACAACCCTT TTCCTTAACA 17220

AAAACGATTT AACTTTTGTG TACTTTTTCT TGTTAGCCAA TGACCAATTT TATCAAGATG 17280

AAGCCACTGC ACAAGCAATT TTAGCGCATT TTGATGAGAC TGAAGCACCT GTTTACATGT 17340

TGACTCATTT AGTCCAACAT TTCTTATTAG AAGAAATACA CATGACGCCG CAATTGGAGA 17400

AAATTCGTCC GCAAATTTTT TATTATTTGT TTGCTACACA TTTATTTGCC GAACTTTACT 17460

ATGTCCAAAA TGTGAAAATG TACAATGTCT TTTGGAATAA AATCAAAGAA AAATATCCGG 17520

TTCTTTTAGC AGAGTTAATT CAAGGTTTAA ATCAACTTTA CGAACAAACA GGCAATGAAT 17580

TATTTACAAA TAAAGAATTT TTGGCTTTGC GTTATTCTAG TGTTTTGGCG CTTTCCAATG 17640

ATTTAATTCA TCGCGAACAA AAAATTGTTT TAGGGTTAAA ATCAGGCTTG CCAGTACTGG 17700

AAGAAGAACG GGTGAAACTT TCGATTGAAA ATAATTTTCG TCGGTTTTAT GCATTGGAAG 17760

TTTTACGATA CAGTGAAGCA AGCAAAGAAG AATGGCAAAA TATTGATTTA CTATTAGTGA 17820

CTTCTTTGGC AGATGTGGAA ACGGGCCATC CAGAGATCTT TTTATTTTCA AGTGAGTTAA 17880

CGTTTGAGCA ATATAATCGT TTGAACCAAA AAATAGTAAA AATTAACGAG AAAAGAACCA 17940

CGTCACTTTA ATAAAAAATA GGCTATTTTG ATGCACATAT CTGATAAATT GTAAAAAAAT 18000

TTTTCTAAAA AGTAAAAAAA GACAAGAAGA AAACGTCAAT AAGTAGGTGT TGATTACCTA 18060

CTTATTGACG TTTTCTCGTT GTTTTTTAAT AAATACTGCG TCAATTCTTC TGAAAATTAT 18120

GTCAAAAAGG TACTTTTTTT GCTTTTTTTA TCGTAAATAA TAGATGATTA GCTATTGATA 18180

GCAAAGGAAC TACTACTTTT TTTAACAGTC AGTTAAAAAA TATTGAGGTT TCTCACAAGA 18240

ATTTTTCTAT ATAATAGAAA GAGCAAGGGA AGGAAGGCGG GATAAGATGA AAAAACAGAT 18300

GGAACAAATT GAAGAAATCT TAAGTTGTGA AGAAAATAGT GCAGGCATCC GATTAAAGGA 18360

ATTGGTTGAA GCTTTAGAAT TGGAAGTCAC AAATCAAAAC TTGTTAAAGG TAACAAGTAT 18420

CTTACATATG AATCCAAAAT TCAAAAAAAT TTATGCATAC GAAGACAGTC GTGTTATTAC 18480

GCTGTATCAA CTTTTACAAA ATAAGCCGTT GGAAGTAACG GAATGAAAAG AGGGAAACAC 18540

ATGAAGAAAA CAATGCTCAT TCTTCTATTT GGTTTATTGA CAGTGGTTGG TCTGCCAATG 18600

GTGACCGAAG CGGTTGAACC AGTTAACGCA ACAGATACAA CGATTTTTGG TGCGCAAGCG 18660

ATGGTGCCTA ATAGCACAGA AGATCAAACA GATAAGTTAC AGACGTTGCT GAGTCAGACA 18720

GCAAAGGAAG GACGGGCATT ATTTTTACCA CAGGGAAGTT ATGCTCTGTC AAAAGATATT 18780

GCGATTTCTT CAAATTATCA ATTAATTGGC GATACAACAG GAGCCACTAT TTTACATAAT 18840

GCTACAGGTG CACCGATTCA GCTAACAGAC ACAACGTACG GGACGAAAAC CAATGTACGA 18900

CTTCAAAATA TTGCTTTTGA CGGCATCAAT GTGACCTTAA AATTGACGAA TCAATTAACG 18960

CTAGCAAACA ATATCTTTTA TAATCCATTA AAAGGATTCG TGGTCAATTT AAATGCAGAC 19020

ATTGGCGTTA AGATTTCAGG CAATATTTTC ATGCGAGACA CGGCACATAT GCAATCGGGA 19080

GGTGATTTTA ACCGAGCGAT TTACATTGGT GGTTATTCAA CGCCAAGTCG CTTCCAGTAT 19140

ATGTCAGATG TAGACATCGT GGACAATCTA TTTGGTTTAA AAGTGACAGA ATTAGATGCG 19200

ATTAAAAGTA CCAGTCGTTC AGACTTAGCA GCGACTATTA CACGTTTGCA AACAGCCATT 19260

GAAGCCGGTG CTATTTCAGT ACCTAATGAA CAAAATTATC TATCCACCGG GGTTAATTCT 19320

TTTAATATGC TGAAGGATGT AACTGTTCAA CATAATTTTT TCTATTCACC CTATGATAAT 19380

GAAAACTTGA ATGGTCTAGT GGGAGATCAT GCAATTTATT TCCGTGGCGC CCAAAACATC 19440

ACGGTAGTAG GGAATCATTT ACGCGGACTA CAAAATGGTC CAGCTGGCGG ATTTAAGTTT 19500

AAATCAGGCC GTAATATTAC GATTATGAAT AATTATTTGC GAAATACAGG TCTCATTATG 19560

TATGGAACAC CAGAGATCGG TTTGGCAGAA ACACAAGCAG AGGGGGCTAT TTCAGAACTC 19620

TCTAACTGGT TAGTGGCTAA TAACATTTTT GATTGGAAGT ATTGGGACAA TCAATATGCA 19680

ATTGGGATGG AATACAATCG ACATACAGGA AATAACAATG TGTTCAATGG CGTATTTATC 19740

AACAATCAGT TTGTTAACTA CCATAATATT CCGCAAAACC GACGTCGGGA ATTGTTAATC 19800

GCTAGTGGCG GTGGGTTCCG ACCAGAAACG AGTTTTGTGA AAGACAATAC TCGAGATGAT 19860

GGCCTTAAAA ACGGCCAATT ATTGGTAGAA AATTGGACAG AGGCGGACTA TCGCTTAATG 19920

CCAGCAACAT GGGAATCATT AATTTCACCA ACACTTTATG AGCAATATAA AAATACGCCA 19980

ATTCCGGTTA GAAATACTTT AGCAACCCCT GTGGCTACAA CGATTGTGCA AGGACAATCA 20040

ATCGATCCGC AGCAACTGGT TGCTAATACA AATGACGCGG ATGAAGCAGT CCCAGCCGCT 20100

AAAATCGTCA ATCCAGAGGT CTTAAATGAA ATCGGCCAAC AAAAAGTGAC TGTTCAATTA 20160

ACTTATGAAA CAGGGTCATT GGTAACAGTG AACGTACCGG TGACGGTAGA GGCCCCTGCA 20220

AAAAAGCTTG ATTTGTCCCA GCTACAAACC GTTTATGCTT CGATTGGTGA AGCAAACCAA 20280

TATACAGTTT ACTCTTGGCA GCTATTCACT GCCATTGGGC CTAAAACAAT TGTTCCTTCC 20340

TATTATCAGC AGGCGACACA ATTGTTAGCA GAAGGTCAGG AGAGTCAGGA CAAAACCCAA 20400

GAGCAAGTGG ATCAGTTAAC AAGTAATTTG CAATCTGCAA TGAAAGTATT AGTCAAAAAA 20460

GCCGATATCA CGCTAGAGCG TACAGAAGCA GAAAATGAAC TGGCAAGCGT TCATAAATTA 20520

GACGAATCTG TATATACCAA AGACTCTTGG CAAGCCATGC AAGAAGCCTT AATAGATACA 20580

ACAACTGGCG AAGGAAGCTC TAAGCAGTTA CAACAACTGC TTGCTTGGTC AGATGAAGAA 20640

TTATTAGAGC CAACACTTGG TGGCTTTAAA ACGCCTGCAG ATGCACAAAa GCGTATTAAT 20700

CAATTAACAC AAACCATAAA AACGGCGTTA CTTCTGTTAG TTGAAAAATC AACAGAGACA 20760

ACTAGTAACA CAAGCGAAAG CAGCACCTCA AGCACAACCA GTGAAACCAG TAATACAAGC 20820

GAAAGCAGCA CCTCAAGTAC AACAAGTGAA ACAAGCAATA CAAATGAAAG CAATACACCA 20880

AGTACAACCA GTAAAACCAG TAATACAAGC GAAAGCAGCA CCTCAAGCAC GACAAGTGAA 20940

ACAAGCAATA CAAATGAAAG CAACACACCA AGTACAACTA GCGAAACCAG TAATACAAGC 21000

GAAAGTAGTA CCTCAAGCAC GACAaGTGAA AGCAGTAGTA CCAGCGAAAG CAGCACGCCA 21060

aGCACGACAA GTGAGAGCAG TAGTACCAGC GAAAGCAGCA CyTCAAGCAC GACAAGTGAA 21120

ACCAGTAATA CAAGCGAAAG CAGCACGCCA AGCACGACAA GTGAGAGCAG TAGTACGAGT 21180

GAAAGCAACA CGCCAAGCAC AACCAGCGAA ACCAGTAGTA CCAGCGAAAG TAGTACCTCA 21240

AGTACGACAA GTGAAACAAG CAATACAAAT GAAAGCAACA CACCAAGTAC AACCAGTAAA 21300

ACAAGTAGTA CCAGCGAAAG TAGTGCTTCA AGTACAACCA GCGCAACCAA CAATACAAGT 21360

GAGAGTAGCA CCCCAAGCAC TGTAAACGAA AGTAGCCAAA GCAAAGGCCA GAATTCTGTC 21420

ATTTACGCGG TAGAGTCTAA TCAAGATCCT AATGACGCTC AAAGTAATTC AAAGCCCTCT 21480

GCCAAGGCTA GTCAAACAAA GGAATCAGTA GCAGAAAACC AAGCAACTAA ACAGATACAG 21540

ACAAATCAAG AATCGTCAGG AACAGTCAAA AAAGCGGATA ATACGACCAA AATTGCCAAG 21600

AAAAAATTCC CCAAAACAGG GGAACAATCT TCAGCAGTTG GGAGCTTCCT AGGCCTGAGC 21660

TTTTTATCGT TAGCCATAGC AACCTATTGT TTTAAAGTGA AAAGATAACT CTTTAAAAAC 21720

GGATAAGTAA ACATCCTGAA ACAAGAAATA GGACTTGTTT CAGGATGTTT TCATATAAAA 21780

TTAAAAGAAC TACAGTTTTT AGTGAAAAAA CACACAAAAG TACTTGAAAT TCTGTGATAT 21840

AAGGGTAAAG TAAAGGAGGA CAAAAAACAA AGGAAGTGAA ATGTTTGTAT ACACCAGATG 21900

AAATTTTAGG TATTAGTATT GAAAATGGTC AGAAAAAAAT TCAAAAGCCT TTGGTAGCAA 21960

AACTCATTTT AGGATTTATC GGCGGTGCCA TGATTTCCCT CGGCTATTTG GCTTATGTGC 22020

GTGTTTCCGC CTCTATTCCA GCAGATTTAG CCAGTGTTCA AGCTTTAGTG GGAGCTGCGG 22080

TTTTCCCAAT TGGTCTAATT GTTATTTTAA TGGCTGGTGG TGAGTTAATC ACAGGAAATA 22140

TGATGGCGGT GAGTACGGCT TGGTTTGCTA AAAAGGTCTC TTTCAGAGAA TTACTCGTGA 22200

ACTGGGTCAC AATTACATTG GCAAATATGG TTGGCGCCCT ATTTGTTGCC TATTTTTTTG 22260

GCCATTATTT AGGTTTAACC AGTAGCGGCG CCTATTTAAA TCAGGTTCTT TCTCTTGCAG 22320

AACATAAAGT TGCAGCAACT TGGCTTCAAG CAATTATTTC AGGTATTGGC TGTAACTGGT 22380

TCGTAGGTCT GGCCTTATGG CTTTGTTATG GTGCGAAAGA TGCAGCGGGG AAACTATTAG 22440

GCATCTGGTT TCCAGTAATG ATTTTTGTTG CAATTGGCTT TCAGCATAGT GTGGCGAATG 22500

CATTCATCAT TCCAGCCGCA ATTTTTGAAG GCGGCGTTCA GTGGGGAGAC TTTTTCCGTA 22560

ATTTTATTTT TGTTTATTTA GGTAATATCA TTGGTGGTGC CATTTTTGTT TCTGGCTTTT 22620

```
ACTATTTAAG CTATCGCAAA CATTAAATCC ATTTAGCAGA TAAAAACCTA CACAAAGGCT    22680
AAATGCCTTT GTGCAGGTTT TTTCTTTTTA AACAAAAAGG AAAACGGTTG AAAAATTATT    22740
GCTAGATTTT TAAAACAATG TTACATTAAT GTTACAAAAA GATTTCTGGA GGTAACGAAA    22800
AATGAAAAAA ATTGCTTTGT TCAGTATGTT AACGTTCAGT GTATTGTCTT TAAGTCTAGC    22860
AGGATGTGGA AACAAAAAAA CAGCAaGCAC AAATGATTCT AAGCCAAAGC AAGAAACAAA    22920
GAAAGCCACG CAGAAATCCT CTAGCCAACA AGAAATGAAA AGTAGTCATT CGTCTGTCAC    22980
GGGTCAAAAT TCTAATGTGA CAGGGGAAAA TCCGTCAGAA AATGCCACGC AGCCTTCTGC    23040
AGGAACTGAT GAAACGAATG AAGTCCCTCA AAACCAAGCA CCTGATACAA ACATTACAAT    23100
TACCAATGTT GTTTTCAATC CTGAAAGAAA TGAAATTAAT GGTACTACAT TACCTAATGC    23160
AACCATTACA GCAACGGTAG TCGGTGATGC TTCTGCACAA GCAGGTGTTT TTTATGCGGA    23220
TGCCAATGGC AATTTTACAG TAATTAGTCC CAGAGCGGGA GCGACTACTC AATTAATCGC    23280
AACCGTTGAT CAACGGAATA GTGCACCTGT CCAAATTGAT ATTCCAAGTT CAGGACAAGA    23340
AGCAGCGCTT TCTTTTAGCA ATATTACGAT TGATCCGAAA CAAGGGACAA TTTCTGGTAA    23400
AACAGCACCG AATGCAACTA TTTTAGTGTC ACGTGCAGAT GATGCGCGGG TGATTTTAGC    23460
AAGTTTTACT GCGGATGCCC AAGGGAATTT CACAGCCAGT AATTTAGTTC CCGGCACAAA    23520
AAATCGCTTA GATGTTACGT TAAATGGAGA AATAGGGACA CCTTACTTGT TTGATTTACC    23580
AAATTAAAAT AGAAGAGAAG TGATCAAAGC CACCAACGCT AGTTAGCGTT GGTGGCTTTG    23640
ATGGTTTAAG CTTCTTCTTT TTTATGATTG CGGATATGTA AATAGATGGT ATTTTTTGAA    23700
ATACCTAATC GATCGGCAAT CTTTATGACA GAATCTTTGA GATTAAAAAT TCCTTGATCA    23760
TATAAACGAA TAACAATTTC TTTATTTTGA TTCAAATGTG ATACATTTGG ATCGCTTTCT    23820
ACTGCTTGTT TTACTTTTGT CAGTGAAGTA GTGATTAATT CGTCAATATT GTCAGAAAAA    23880
TGTTCAGAAT TTTTAAGCTC TTGGGACATA TCATTTTGTT GTCCTGAAGG TAATAAGTCT    23940
TGCAAGTAAT CAATTAATGG CATTTCTAAA TGCATATTGA TACAGATCAA TCCAATAATA    24000
CGATCATTTT CACCGATAAC TGGAATGGTG GAAGATTTTA AAATAACCCC TTTTTGATTT    24060
TTATTGTAGT AGGGTTTCAC AGCGTAATCA TGTTCCGACT CTACTTGACT TAACATTTTT    24120
AACGCCATAT CAGTAATGGG GGCACCTTCt TGCGGTTACT GAAATGGCCA TTGATGACTT    24180
TGATAGCCGA ACTTTCCAAA TTTTCTAAAC TATGAAGAAT GACTTCATGG GTACTACCAA    24240
AGAAGTTCCC TAATCCATCA AGCATTAAGG CATGTGAATG TAAGATTTGT TTATCTACTT    24300
CACTTAAACG TAAATATGAC TCCaAGTGAT ATCCTCCCCT ACACTGAAAT TAATTTTAGT    24360
GTAATACTTT TTTAATGGAA AAGCAATAAG ATTCTTTACG TtAAAAAAAc nCaTtGACTA    24420
AAcTTTkGTT TAGTCGTATA CyCkGAATGT GGAATGAAAT TTTTGTTACT CATTGTACTT    24480
TAGTTCATCG GAATGCTCGC TCATTCTTTC ACTTGTTGaG CaATAAAAAt TTTaTTCAGG    24540
AGTACCATCC GTTATAAAAA AATATAGTTG GAGGCAAAGA AATGAAAACG TTATCAGGAA    24600
```

TCTTGATGTT AGTCTTAGTT TTATGTGGCA TGTCCCTATT CTCTTTTAAA GCACCGCGAG 24660

GGAAAAAATC TTTAAGTGCT TTGTCTGGTG CAGCTTGTGC CACCTTTCTA CCAGAGGCGT 24720

TTTTGCGGTA TGCAGTAGGC GGCGTTTTTC ATTTAGATTA TGTGTCGCAA ATTGGTGAAA 24780

CTATGGGTAG TCTAGGTGGT TTAGCAGCTG GAGCTCTTGT GCCGCTTGCT TTCGGGATTA 24840

GTCCAGTTTT TTCAATTATT TTAGGTCTGT CATTGATTAA AGTCAGCTTG TTACCAGCCT 24900

TTATTACCGC GTATCTTTTA TCTTTTGTTG TTGAGCAGAT GCAAAAACGT ATTCCAGAAG 24960

GGATTGACTT ATTAGTCGTT ATTTTAGTCA TTCCAGTCGC TGCTTCGTTA GTTGCTGGTT 25020

TTATTCAACC CGTTGTATTA GGTGTCCTAG AAGTTATTGG GGGCACGATT TTAAGTGCAG 25080

TAGATGGCAA TCCATATGTG ATGGGCGCAA TTTTAGGTGC TATTATCCCG ATTGTTGGTA 25140

TGACACCGTT AAGTTCAATG GTTTTAACTG CCTTGATTGG TTTGACTGGT GTACCAATGG 25200

CTGTCGGTGC GTTGACTTGT TACGGCAGTT CCATTGTCAA TGCGGCGCTA TTTAAAAAGT 25260

TAAAACTAGG CACAGCTTCA ACCCCGTTAG CTGTGGCAAT TGAGCCATTA ACACAAGTCG 25320

ATATCATCAG TTCCAATCCA ATTCCTATTT ACGCAACGAA TTTATTTTCA GGAATGGTTA 25380

GTGGCATTGT AGTGACCTTC TTTGGCTTAA AAGTACCTGT GACAGGAATG GCAACACCGT 25440

GGGCTGGTTT ACTTGTGACA TTAGGTAACA ATGCCATTCA AACGACTTTA CTAGCAGTAG 25500

CGATTATCAC AGTGGTCAGT CTGATGTTTG GATTTTTAGG TGCGCTGGTC TTTAAACGTT 25560

ACAGAATTGA AGCGGTGGAT AATACACCAG AATTTGAAGG aACGACGGAA GTAACCAGTG 25620

AAAAAGAAAA AGAAGCAGTA AAAACCAAAG GAGGAGTTAC GTATGAACAC AGCTATGGCC 25680

ATCAATAAAC CTAAAAAAGT GAACAAAGTA ATGAACTACA AAAGTTGTTT TGACATTATC 25740

GGTCCTATTA TGATTGGTCC TTCTAGTTCG CATACAGCTG GGGCCTTAGC GATTGGCTTA 25800

GCGGCAAGAA AGTTATTCGG CGGAACGCAG AAAAAATTGT GATTAAATAC TATGAATCTT 25860

TTGCAGACAC GCATAAAGGA CATGGGACTG ATTTTGCAAT TATTGGCGGA ATTCTTGGAC 25920

TAGCAGCAGA TGATGCCAAC GTAACACAAT CCATTGAACT TGCGGAAGAA CAAGGAATTG 25980

AGATTCAATT TTTAGAAATG AGTGAAGAAA GTCCAGTGAA GCATGCTAAT ACAGCATGTG 26040

TGACACTCAG TGATGCAACG CACGAAATTC ATTTAACAGG AATTTCTGTA GGTGGCGGCA 26100

CTATCGAGGT GAAATACATT GAATTAGACG GTTTCAATGT CCAACTACAT GGTCCTTTAC 26160

CTATTTTGTT AGTCATTAAT CAAGAAGAGC AGGCGATGCA AGCTTTTAAA GATACGTTAC 26220

AAAAAAATAA TATTCAAGTG AATGCGGTGT CTCGTTATGT GGAAGGGAAT CAAATTTTAT 26280

TCATCTTTGA CTTAGATTCG GCACCAATTT CCTCGGTGAA AGAACAACTG TTTTCACTAG 26340

ATGACACATC GAAAATAATC TTATTATGAT AGGAGTTTGA CATATGTATC TTTCGATTGA 26400

AGAATTTATC GATAAAGCTG AAAAAACAGG AAAAAAAATC TCTGATTTAA TGATTGAGCA 26460

AGAAATGGAA CGTAGCCAAA AAAGCTATGA AGAAATTTGG ACACAGATGG GTCAAAATTT 26520

GGATGTCATG GATGCTGCTG TCAAACGCTC GCAAGAAGGA GCAGGCGTTT TTTCTCCTAC 26580

TGGGTTAACG GGCGGTGATG CGGCTCGTTT AAAAAAATAC CGCGCAgcTG GCAAAACGTT 26640

ATCAGGTGAT TTAATGATGT CTGCAGTTCA AGCCGCACTA GGTACTAATG AAGTCAATGC 26700

AGCAATGGGC GTTGTTTGCG CAACGCCAAC CGCTGGTGCA AGTGGCACCT TACCAGGGGT 26760

TTTAACAGCG ATTAAAAATA CCTTGAATTT AACGCGTGAT CAACAAATTC GTTTCTTGTT 26820

TACCAGTTCT TTATTTGGAA TGGTCACAGC CAACAATGCG ATGATTGCTG GTGCTGTTGG 26880

TGGATGTCAA GCCGAAGTAG GCAGTGCTTC AGCGATGGCC GCCGCTGCAG CTGTTGAAGC 26940

GGCTGGCGGA ACACCACGCC AATCTTCTGA AGCTTTTGCT ATGGCTTTAG GGAATTTATT 27000

AGGTTTAGTT TGTGACCCAG TGGCTGGGTT AGTCGAAATT CCTTGTGTCA AACGAAATAC 27060

AGTAGGTGCA GGCAATGCGT TAATCGCTGC AGATATGGCT TTAGCTGGAA TTGTAAATAA 27120

AATACCAGCA GATGAAGTGG TGGAAGCCAT GAATAAAGTC GGCCGTCAAT TACCACGAGA 27180

ATTGAGAGAA ACGGGTTTAG GTGGCTTAGC TGGCACAGCG ACTGGCCAAC GAATGAAGAA 27240

TGAAATTTTT GaAAAAGTAA AATTCACCAT TGnTTAAAAT TTAATCATGA AGAGGaAtAa 27300

yGCTGGTTGA AAAATTCGCA ATGTGAGTGT CAGAAAGTAA ACCGGTnGGC nAGAGTTAnC 27360

(2) INFORMATION FOR SEQ ID NO: 165:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11427 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

CAGAAACTGA ACAAAGTGAA TTCCCTTTTG CGATTCACnA AAGTCACATT TTCAcGcAGG 60

tTTaAatCTt CCTGAAaaTT TtAAaGGTaT TTTTttGCTt GAaGAaaGTT ATTATaCTTT 120

AAATGGAAAA GAACGATTCA AATCAGATAT CTTTCTCTTT GAAGTAAATT CGTCAAGTAA 180

CGTTAAATTA AGTTCCATAA CTGTTCCTAA AAACTATGCT AATAAGAAAT ATGAAGAAAT 240

TGAATCTATT CCTTTTTCTG AAATGACCAT TTCAGAAAAA TTTGTTCCGA TTGTTTACAA 300

AAAAGAGAAT GAAACTTTTA TTGGTCAATC TGAAAGTAAA TTTACAGAGA AGAATCATTT 360

CATATTAAAA CAAACGATAT CTGAAAAAAA ATTGATAATC AAAGAAGAGA TAATGAGAGG 420

AAATAAACGG ATATTCGGAT TCGATAAACC TATCGAATAC GTAAGAGTTA AAGAAAATTA 480

AAAAAGCTTT TAAAAGAATA TAATTTAACA GTGAAGTAAT TTCTTTGTAA AATAGTAATT 540

ACTTCACTGT TATCATATTA ATAAAGATTT TCATCATTTC TGAAAATATC TTCTCTTTCC 600

AAGAAGAATG CTAATTCAGA TGTGTGTGGT TAAATTGATA GTCTGCAACA TTAAACATTT 660

GACTGATTCA CTTTCCAAAA TTCTATTATA CCAGTTAAAA TCGCTTAAAT AGAAATACAC 720

TTCTAGCAAC ATGTTCCATA GTAAAAGTAG CACTTGTTAA AAATAACTAA TCTGCAACTT 780

ACTATCCTTG AAAATAAAAT TCTCAAACAC TTTACAAACA ATTATTTTTT AGGAAGATGG 840

TTTGTTATTC TTATATTTCG AGAAAAATAA AAATAGCAAA AGATACATTA CGATACAACT 900

```
ATACTATCAA TACAAAATTA GATACCTCAT CTGCCCTTTG TTCAGTGATA AAAGCAATTG      960
AGTTAGGAAT AATAAATGTT CACCTTCAAT AGACTTATTT ATTCAGTCAG CTAAACGCTG     1020
TTAGATACAG CTATAGAAAT CTAATTATAG ACATAGTTAG ATAAACAACT ATCCTTTATA     1080
AAAATTGGTT CCATGTCTTT AGTACAAAAA GAGTACTATC ATTATTTAAT TTAAGTAAAT     1140
AACGGATTTC AGAACTCATC GTTGATTCGA CGAGCACTCT TATTCAATTG GATAAGCGTT     1200
ATCTCCTCTT TTTTAAAGTA ATTTGTATTC AGCAACCTAG AATATTGATA GTGCCGATTA     1260
TTTAGACAGC TATCCGGATA ACTTGGAATA ACTGAAACTA GAGTAAAAAA TCTAATAAGG     1320
TGCTTTTAAA ATATCATCAT TTAAGATACG AAATCGCATA CCCCGAAAAA TATCTGCTCT     1380
CTTTAGTTCT ACAATTTCTG ATATCCAAAT AGTTCAATTT GAGTTAACGC TGGAAATGGT     1440
GATTCATCTT CTGCTTTGAT CAGCTGTTTC AATGTAGCAG TCGTTACTGT TCGTGGAGAG     1500
AAGGTGAAGT ATTGCGTTGC TATCGCATCA CTCGTAGTGA ATATTTCTTC TGAACCATCG     1560
GAAAAGGCTA TGGTTACTTG TTTCCAATAA CTATCATGGG GAAAATCAGC ACGTAATACC     1620
AATCCAAGCT TATCGAGTTC AACTTCTCGG CCAAAATCAA TTGTTATTTC AGCATCTTTT     1680
TGCTGATTGA TCCCCCAAGA TTGATAGGGG TACGATCCAT GATCTCTGTT GGCCAGAATT     1740
CCATCAATTG CATTTCTAGC AAAAAAAGTC GCATCGTTTC GTGTTTCAAC ATTTGCATAA     1800
GCGTGGGGAT AAACATTGGC CACATCTTTT TGATCATGAG GATTGAGCGC CCAGTTACGA     1860
TAACTAGTCT GTTCTTCAGC AAAAGCCTTT CTGACAAAAA AGTAATGCTT CTCGTCAAGA     1920
AAAGCCGCTG GTGGATAGGC TTCTCGCAAT TTTTCCGTCG TAGGAATTTC ATAGACCCAC     1980
TCGCCTTTTG ATAAATACAC TAAAGAGGAT CCTAAAGCAT TGTCTAGTTG AACATAAACT     2040
AATTGTTTTT CTGCGGTTAC AACTTTAAAT TTATCTCCTG CTTGAAAAGC ATACTTTTTC     2100
GTTGCCAAAA CTGCTTCATC CACTCCACTG CTGCACATAA CTACTTCATT TTCTTGATTA     2160
ATAATTGATA ATTCTATTTG CATACTTTCT TCGCTCCTTT TACTTTTTCT TTAAAAACTT     2220
ACCCGCAATA AATAAAAGGA GAGGCGGCCT TAACCGTTTC CTCCCCTCTT TATTCTCTTA     2280
AGGTTGTTTA CCAATATAAG CTAAAATTCC GCCATCAACA TACAGAACAT GACCATTTAC     2340
AAAGTCAGAG GCCTGTGAGG CAAGAAAGAC TGATGGACCT GCAAGATCGA CCGGATCACC     2400
CCAGCGTGCT GCTGGCGTAC GACCGACGAT AAATTGGTCA AATGGATGAC GTTCCCCGTT     2460
TTCTTGCAGC TCTCTCAGCG GCGCTGTTTG CGGTGTCGCA ATATAACCAG GACCGATACC     2520
GTTGCACTGG ATATTGAATT GACCATATTC AGAAGCAATA TTTTTTGTTA GCATCTTCAA     2580
ACCACCCTTT GCCGCGGCAT AAGCACTAAC AGTCTCACGA CCTAATTCAC TCATCATTGA     2640
GCAAATATTA ATAATTTTAC CCCCGCCTTT CTCGATCATC TCAGGAATAA CCGCTTTTGA     2700
CATAATAAAC GGCGCATTCA AATCAACATC GATGACTTGT CGGAAATCTG TCGCTGACAT     2760
ATCTGTCATT GGAATTCGTT TTATGATTCC AGCATTATTA ACTAAAATAT CGATTGAACC     2820
AACCTCTTCT TTGATTTGCT TGACCATTGC TTGAACTGCT TTTTCATCCG TTACGTCACA     2880
```

```
GACATAGCCT TTTGCTTCAA TGCCTTCTGC TTTGTAATTT GCCGATCCTT CATCGACTGA    2940

TTCTTGAGTC AAATTATTGA AAACGATCGT TGCTCCTGCT GCCGCTAATG ATTTGGCGAT    3000

TTCAAAGCCA ATTCCATAAA CTGCGCCCGT AATCAATGCG ACTTTCCCAT CTAAGCGAAA    3060

CATATCCATA TTAAAATCCA TCTATTCTCC CACTTTCATT CGTTCTATTT TAGTTCATCC    3120

ATTGGAACCA TATCCATGTC GGTATAAGTG ATATTCTCTC CACACATCGC CCAGATAAAT    3180

GAATAGTTGC TTGTTCCTAC ACCTGAATGG ATCGACCAAC TCGGTGAAAT CACGGCTTGT    3240

TCATTCCCCA TTACCAGATG CTTTGTCTCA GCTGGGTCAC CCATCATATG GAAAATTTTT    3300

GTATCTTCCT CCATATCAAA GTAAACATAC GCTTCCATTC GACGTTCATG CGTATGACAC    3360

GGCATGGTAT TCCAAGAACT GCCCGGCTCT AAAATCGTAT AGCCCATTTG CAGCTGACAG    3420

CTTTCACAAA TATTGGGATG AATATATTGA TAAATTTTAC GTTCATTTAA TGTCAATGCT    3480

TCCCCAGTTT CCATTGGCTT AATCTCATCA ATACTGATTT TTACATTTGG GTACTTGTGA    3540

TGCGCCGGAG CAGAACTAAT GTAAAATTTC GCAGGATCAG CTGCATCTTC TGAAGAGAAA    3600

ATCACTTGTC GAGTTTCTTT CCCGACATAA TAGCCATCTT GCTTCTTCAT CGCTTCTTTT    3660

CGACCGTCGA TTTCAATAAA ACCCGGACCG CCAATATTAA TGACTCCTAA TTCACGCCGT    3720

TCCAAGAAAT AATCGACACC TAATTCTTTG TTCAAAATAA TCTCTAACGA CTCAGTCGTC    3780

GGTGTTACAC CGCCAAAGAT CATGCGATCA TTGTGGGTAT ACGTCAAACT GACTTCATTA    3840

GGTACAAATA CTTTTTCAAC TAAAAATTCT TTCCTCAATT CTTCTGTTGA ATAATGACTA    3900

ATATCTTTTG GACTGTGCGT ATATCTCGTA TCCATATTCT GCATGTTTAT TTCCCTCTTT    3960

TACTTAATAT TTAACGAACG ATTCGACCAG AACCTGCTGT CAAAAATTGT TCGATCTCTA    4020

TTTCTGAAAA TTGATTACAA TCCCCGTGAA CAGTATGCTT CAATGCTGAT GCAGCCGTAG    4080

CAAATTCAAC GGTTCTTTGC CCGTCAAAAT CTGAGATCAA TCCATGCAAG ATCCCCGCCG    4140

CAAAAGCATC ACCTCCGCCA ACTCGATCAA CGATTGGATT TATTTTATGA ACTTCAGACT    4200

CATAATACGT ATTCTCCAGC CATAAAGTCC CTACCAACTG ATTGTTGCTA GCCGAAGAAA    4260

CGATCCTTTT CGTTGAATAG AAGCATTTAA TATTGGGAAA CTGCTGCTGC ATTTGTTGAT    4320

AATAATAAAC GAATCCTGAA TTCATCCGGT CTTCTGGAAT CTCCAATAAA TAACGCGCAT    4380

CCAATTCTCC AGCTGAACAA TAATCGACAT AAGGCAAGAT TTCTTTTAAT GCTTTCCCAG    4440

CTGCTGCCTG ACTCCATAGT TTCCCACGAT AATTAATATC AAAGCTGACT AACACACCAA    4500

ATTTCTTTGC GGCTTTCACT AATTCAACCG TCCATTGACA CCATTGCACT GACAAAGCTG    4560

CTGTAATGCC TGAAATATGA AATAGCTCAA TCCCAGAAAA AAGCTTTTCA TAAGACCACG    4620

GAAACTCTGT TACTTCCGCA AAGCTTGACC CCGCTCGATC ATAAGTTACA CTAGCAGCAC    4680

GTTCACCGAT CCCCGTTTCC ATGTAATAAG TTCCTAAACG CGcTCCGCCA AGTTGTACAA    4740

AATCACTCAT CACACCGTTT CGACGCAAAT GTTTCACTGC CGCTTGACCT AGTGAGTTAT    4800

CAGGCACTTT ACTTGCAAAA GAAACCTGAT GACCAAAATT GGCTAATGAA ATAGCTACAT    4860
```

```
TTGCCTCGCC GCCGCCATAG TGACAGATCA ATTGCTGTGT TTCGTTAAGA CGAACACCTA    4920
TATCCGTTGA AAGTCGCAAC ATAATTTCGC CTAAAGTTAG TATGTTTGCC ATAATTAGTT    4980
TTCCTTTATT TCTCTAAATT TGGCCATATA TTGTTTTGCC AAAGCCGTTA CTTTACCAAA    5040
CTCACCAGTT TTAGCGTGAG CTAATAAGTT TCCGCCAACA CCCACTGCCG TCACACCCGC    5100
CTCAAACCAT GACGCCATAT TATCCAGACT AACACCGCCT GTGGGCATAA TATTCAAGTA    5160
CGGCAAAGGT GCTTGGAATG CAGAGATAAT ACTCGGACCA TAGGTACTTC CCGGAAATAA    5220
CTTCACGATA TCTACTCCGC TCTTCAAAGC TGTCTGCATC TCAGTAATTG TCATACATCC    5280
TGGCAAGTAA GGAACTTGAT ACAAGTTACA AATTTCAGCT GTCTCCCGAT TAAAGCTAGG    5340
GCTTACAATA AATTCGGCTC CGGCCATTAT AGCTAGGCGA GCCGTCACAG CATCTAACAC    5400
CGTGCCCGCG CCAATCACGA TATCTTTCTC GTCATGATAT GTTTCAACTA ATGACCGAAT    5460
GACTTCATCT GCTTGAGGGA CTGTAAAAGT CAATTCTAAT CCAATCATTC CGCCCTCTAC    5520
TACTGCATGA CTAATTTTCA AAGCTTCCTC TTTAGAATCT GCTCGCAATA CTGCGATAAC    5580
TCCCGCATTT TCTAAGCGTT TCAATACTTC AACTTTTTTC ATTCGTCTAC GCCTTTCTTA    5640
TAAAGCTTTA TGGAATCAAA TATACGAGCA CCAGATTGAC TATACCGCGC TACTTCTTTT    5700
TAAATTCCTT CTTCTGGCAG TTCTTCTGTT tCATTGTTTT TTTCAAAAaG CmTAATATtG    5760
CTCTTTCCAG CCTGTACGAC TGTTTCAGCA ATTTCAAATG AATCCGTGAC GCTATTTCGC    5820
ATCATCAAGC CTTCTAATAG TATTCCTAAG TTTGTGCCAC TCACAACAAA AACTTGATCT    5880
TTTTCCTCAG AGAATTCAAC ACAGGTTTTA AATGGGGTAC CACCCAACAA ATCTGTGAAA    5940
CAAATAGCTT GATCAAACAT TTGAATTTCT GCCGCTACTT TTTGGCGATA CTCATCTACC    6000
GTCTCCTCGG CTAAAAAGGG AATAATGATC ACATCTTCTA GTTCTCCAGC AATCATCGCT    6060
AAAGCACTCT GCATCCCAAT CGAAAATTCA CTATGCCCAG TAATTAATAA TTTCATCTCA    6120
TTGCTCCTTT ATTTAATCCG AATCTTTGCT TCCTTTTGAT CAATCAAAAC AAGCCGACCA    6180
TAAAAAAATT GGCCATATTC ACTCACTAAT AGTTTATTCC CTTTGACAAT ATCTTCCCGC    6240
ATAACAAAGA GATCAAATTT TTCCGATCCG GCGATTATCC GAAATCCCTT AGCCTGACAG    6300
AATTGCTCAT TTTCTCCTGT TTGACTCACT TTTATTGGCG TAATTTGGAT ATCCTCTAAC    6360
GGAGAAATGA CTGTTGCTTG AATTTCTTTT CCCGTTTTGT AACAAAACTT GTTTGATAAG    6420
CGTGGGTGCT CATTCAGCTG GTTATAAATC TCTGAACCTT TTGCCACTGA TTGCTGTGTT    6480
TGGCCTCCAG CAAAAAGGAG TGTGTACTTA TGCTTGTTCG TAGTTAAAGC AAACCGATGC    6540
GCTTCTTTTT GACAATTTAT CGACGGTGCC AAATTATAGG TACTCGTAAT TTCAGTCTCT    6600
TTCTGTCCTG CAAAGCTATC AATAATAACT ACGGAGTTGA TCGACTTTAA ATAGATGAAG    6660
cTGCGCTCAA AAATCATTGG ATTCTGATCC GCCTTATCCA GCCAGCCGCA TTCTGCAAAA    6720
AAACCGACAG AAAGTTCTTT TATTTGCTGA AATAAGGGTG TCGGTAATTT GTCATAACCC    6780
CACGTATCGG AAACTAAAGT ATGGGGATTT TCTGCGATAA ACATCGTATT GTGCGAAgCG    6840
```

CACTCTTTTA GCTGAAGTCG CTCTGATTTG TTGACATAGC TGTAACGACC ACTATCGGAA 6900

AATAAGTCAT CCCCTTGTAG TTGTAGTGTA AATCCACCTG TAGAAGCATG ACCATGTGCG 6960

CTCCCATGCA GACCGTTAAA AAGTGTAAAA TAGATATCCT CTGCTTTGTA CGCCATCAGC 7020

CCACTTGATT CGCCACGAAA AAGTTCTTTT GGCTTCATTG TTTCCCAGAT TCTTTCTTCA 7080

TAAAGATCCC CCGTCCAAAG CCTTGCCATA TTCGCAGTCA TGGAGGGTTC AAATATGAAA 7140

CCTAATTTGC GATAGATATC ATATACGTAA TGAAAGTTGA CATGATCACT ATCATTGATC 7200

GGATTTAGGA TATCTTGGTT ATCCGCCAAA TAATGGGTAG AGAAAATAGG TGTTTTTAAT 7260

TTCATGCGAA GATCTAATGG TAACTGGACT TCAAGATATT CAGAAATCTG CAATAGATAC 7320

ACGAATGTCA TCAAAACTTC GTGCTGGTAC AGCGGGCTCT GCTCCCAATG AATTCCATCT 7380

GAATAGAATT GTAGATCAAG CTGTTCAGCT AAACGAGACC ATATTAGATC CCTCTGTTTA 7440

CTGGTCACCA GTTCTGGAAG AAATAAATCA ATAGCTGCCA TTCCACCAAT TGCCAACACA 7500

CCCCAATTAC TGAGCCTGTA TTTATCGATA TAGGACCGCT CCAAATAGTC CAGATGGATC 7560

AGCAAGGCGT TGTTCAACAC ATCATCAATT CCTAATAGTC TGAAATCAGC GATTGGAATA 7620

TACGTCAAGC TTTTCATCCA GTTTGTTACT CGAATCCCAA CATCTAACGG ACGCCAAACA 7680

TCCCTATTCG TCGAATTTGG CTCACCCTCA TCGTTAATAA AATCAATAAG CAAGCTGTGC 7740

CATTTCTGTA AGTAACGTTC TTTTTTAGTA AGTGCATATG CTTGTGCCAG ATCTACAAGA 7800

AAGCTTTGAC GACTCAACAT AAAGAGCCAT TCTGGATCAT CATCAGGATA CCGATTCCAC 7860

GCATACTCTT TTAAAGAATA AGGAATGGAG CAAGCTTCCA TATCCATTGC ATCATTGTAA 7920

ATAATCTGAT CTTTCAGCAA ATAAGAAGTA CGTTCAAAAA GTGGCTTCAA AAATTGTGAC 7980

TGTTGCTCCA AGTACTCTGA AAGCCAATCT TGTTGAAAAA ATTTTAGTTG GTAGTCTTCG 8040

ATTCTTTGAA GATTCATCGT GTTCCCGCTT TCAAATTTTT ATTCGCCTCA GGCTTAAAAT 8100

TACAGTATAC TTCCTCATTC ACTCCGAAAA GTGACAAATA CGGAGACAAG AATAATTTTC 8160

CAAAAATTAA AAATCTTGTT TCTGATTACC CTCGTTTACT ATGCTAAAAT TCCTAGGACA 8220

GACGCTGCTA TCCCAATAGC AAATAGTAAG ATCAATAGTT TATACGTGTT CCAATTTCTC 8280

TTCTTGATCA AATAAAACAC TAAGATAGTT AAAATTACTG GAAGCATTTT AGGCATAATC 8340

TTATCTAAAA TATCCTGAAT AGCGATTACT TGTTTTTCAC CGGATTCTAC GTGATGCGAA 8400

TACTTTATGG CGATGTTGGC CTTAACAAAT GATGTTGCTA ATGCCGCTAT AACAGTTACT 8460

CCGACAATAT TAGCAGCCGA AGAAATTTTC CCGATCTTAT CACTCAATGT TTCGATGACA 8520

CTCGTTCCAA GTTTAAAGCC TAAATAGCCC ATTAATAACT TGATAATCAA CATCGAGATT 8580

AGCATTGAAA ACCAAAAGCC CATCGGTGCA AAACCCAAAC CATCCATTGC TAACCCGGCA 8640

AAGATTGTTG AGAATAAAGG AGCAATCCCA AATTGAGCGA TTGAGTCTCC AATTCCCGAT 8700

AATGGTCCCA TTAGCGCCAT TTTGATGCTG CGCGTATCCG TGACACTTTG TCCATTGTCA 8760

TACATTGCCA ACTGCATACT AGTAATGAAT GGCAGCATTT GAGGGTTGGT ATTATAAAAC 8820

```
TCTAAGTTTG CACTCGCCGT TTCTTTCAGC TTCTCTGGTT CATCTTTATA GATTTTCTTC     8880
AGCGCTGGCA TGATCGTGTA TAAATAACTT ACCCCTTGAT AAGATGTATA ATTAAAAGCG     8940
TTCTGTAACA AATACGACCG TAGTGCTGTC TTCATGTAGT CTTTTTGGCT TAATTTTGGT     9000
GTCCTAGATT CCATTACTAA AATCCTCCTC TTCCATCAAG GTGGTTTGGC TTTCGTCTCT     9060
TATGCTCTTT TCTCCAACAA TCGTTTTATC TTTTCCATAA TATTGAATCA ACGCAAAGAC     9120
AGTTCCTACC ATTGCGACAG CCATCGTAGG CAGCTCTAAA TAAGCTGCAC AAATATAGCC     9180
TAGTAACACA TAAGGTGTCA GACTCTTTTC GAGCATCACA GACATGATCA TTGCAAAACC     9240
TATCGCTGGA ATAAGCCCGC CAGCAACAGA AAAGCCATTG ATTAGCCAAT TTGGTAACGC     9300
TTCAACAAAT GCCTGCAATC CAGGACGACT CAAGGCCGCT AATAGTCCAA TCATAAATCC     9360
TGCAACGGCA AAAGCTATAT ATGTTGAATT AGCAATCAGA TTAAAACGTG TATATTTCCC     9420
TTGCTCAATT ACATTCTTCG ACCATTTAAC ATTCCCAGAA AATATAGTAT AGATCGCCGT     9480
TGTTACCAGT TGGAACAATA CTGCAAAGGG AAATGATAGT GCTAACGCGG ATTCAGGTGT     9540
TACCCCTTGA TGTTTAAGCG TAATGGCCAT AATCGTACCA ACGATTCCGG GTCCAACTGG     9600
ATTGGGTGGT ACGGTTCCGC CTGCTCCAAC ACCAAATCCC ATATAAGCCA ACTCGGCTAA     9660
AGCACCAAAT GTTAAAGCCG TTGGAATGTC ACCCAAAACA ACACCTACAA AAAAGGACAT     9720
AACCAGTGCA CGGTTTGTGT AAATTCCTAA TAGCTGCCCC GaATAACAAA AAGCTGTGAT     9780
CAATCCGATC AATATACCTT GAACTAATGA AATCTCCATC TGCTTCTCCT TCTCCTTCTC     9840
CTTCTAATTT AATTAGATAT TTTTTCCATC AATACATCTA AGTACTGTGA GCCATCATTT     9900
GAAAGTGGTG AAGTCTTGGT GTTGAAAGTT ACACCATAGT TATCTCGCAT CAAGCAGAGA     9960
GCTTGCTTAT CTTCTTTACC TAAATAGATA AATTGAGAGA TTTTCTCCTT GCCTTCGTCA    10020
GCATGAATAT TCCCAATATT TAACTGCTCC ATTGGAAAGC CCAATTTGCA TAGTTCTAAG    10080
GCATCATGCA AGTTTCCAAC GACTACAAAA ATTGTTTGAC TAGGGGCTGC TTTCCAAATA    10140
ACTTTTGCCG TACGTTCGAT CGTCCAAAAA CGAATATTTG TTTCTTTCGG GACTACCGTT    10200
TTCATCAGCG TCTGCTGTAA GCTATCCTCT GCCGCTTTAT CATTGGCACA AATAACTAAG    10260
TTCACTCCTA AACTCTTGAT CCATAGTTGA CCTTGTCCAT GTATCAGACG TTCATCTACA    10320
CGCACCATTT TTACATTGGG TTTTTGCATT TTCTTTCCTC CTGTTTTTAC CAGTACATTT    10380
TCCAATTTTT TTTGAATCGC ACTAAAGCTT CTAGATAGAA ATAATCACCC CAAATATTAC    10440
CTTCATTTAC GCCTTTACCA CTATGCCAAG AATAAACACC CTCATTCAGT AAAGCTGTAA    10500
TTCCGTCTGT TTTGGATTCA GTATAGTTTT CAGATAAGCT TCTTAGCATG GAATGctGCG    10560
CACACTTGTA AACCAACTTA TGACTGTTCG TCTCAGGTAA AAACGCGTCC ATTAGACTCA    10620
TACCACAAAC AGCAATCGcT GTCGCAGAAG TATCTCGGGA CTGTTCGGAA CCATCCATAA    10680
AAATCAGATC CCAATAACTA ACAAAATCCT TCGGcAAGCG ATTCAAGaAA TAATTCGTAA    10740
TCGCCtCAAA AATTTCTGTA TTCTCTTCCT CTGGACAGTA GGAGCGGAAC AATGCCAAAC    10800
```

CATaAATCAA CCACGATTGT CCCCTCGCCC AGCTTGAACT ATCGGAATAT CCTTGGCGGG 10860

TTTTTcCATA CAAAGGATCG CCCGTTTCAG CATCGAaGAA AAACGTATGG TAAGCAGATG 10920

AATCTTCTCT AATCGCATTA TCGATTGTCG TATGATAATG TTTAACGGCT ACTTCACGAT 10980

AACGTTTATC TTGAGTCACT GTTGACGCCC AAAATAATAA TGGAATATTA AGTAAGCAGT 11040

CTACGATCAA ACGATAGTTA TCTTGATTGC CCAATTCTCC CCAAGCTTGG ATAAATTCTC 11100

CCTTAGATTG GTAGCGCTGC AATAATTTTT CTGCTGCACT TATCGCGGCA TTCTTAGCAG 11160

CTTCATTTCC AGTCAACTTA TACGCACTTA CACAGGATAA GCTGTATAAG AATCCCAAAT 11220

CATGATGCTC GACTTCAATT TCATTTTCAA TTCGGTAAAT AAAACTCTTG ACATTTTGTT 11280

CTGCAACTTC ACGGTAATGC TTTTCCCCAG AAACTTCATA GCACAGCCAT AATATTCCTG 11340

TCCAAAAACC GTTTGTCCAT TCTGTATTAT CCATAGCTAt ATACGTAwTG TCCGTTGkTG 11400

ckGGTGTTGG GAATTaTCTC CCAGTTT 11427

(2) INFORMATION FOR SEQ ID NO: 166:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1669 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

CCCTGnGACT ATTTTTCTTA TCTCTTTTAT AAATAGAAGA GAAAGTATTC TGATTTTGTG 60

TTTTAAGCTT TTTATGCTAG AATAACGCCA GAACATAAGG AAAGGGACGA GwCAwTTGGA 120

ACCCATTTAT TTAGATCATG CGGCAACAAC ACCGCTTCAT CCAACGGTCA TTCAAGCGAT 180

GACTGAAAAT ATGCAAACGA CTTTTGGCAA TCCATCAAGT ATTCATCAGT TTGGACGAAA 240

GGCACACGGT CTTCTGGAAG AAGTGCGTCA AACGATTGCC GAGAGCTTAC AAGCCAAACC 300

CCACGAAATT ATTTTTAATA GTGGTGGCAC AGAAGGGGAT AACACAGCCA TCTTGGCAGT 360

TGCTTTTTCT CGTCAAAAGG AAGGAAAACA TATTATTACT ACGGCAATTG AACATCCCGC 420

TGTGTTACGA ACGATGGAAT ACTTAGAAAC GTTAGGCTTT GAAGTGACTT ACTTGCCTGT 480

GAATGaAAAT GGTCAGATTT CAATGGATCA ATTCAAAAAG TCTTTACGCG AAGAAACGAT 540

CTTGGTTTCA ATGATGTATG GCAATAATGA AATTGGAAAT CGATTACCGA TTGCTGAAGT 600

TGGTGCAATT CTTAAAAATC ATTCGGCGAT TTTTCATACA GATGCCGTCC AAGCTTATGG 660

AAGTGAAGTC ATTTTACCTC ATGAATTAGG GATCGACTTA TTAAGTATTT CCGCTCATAA 720

AATCAATGGT CCAAAAGGCG TAGGTTTTTT ATTTAAAAGC GATGCAATCC AATTACCACC 780

TCTTTTACAT GGTGGGGAAC AAGAAGAAAA ACGACGTGCG GGCACGGAAA ACTTAGCTGG 840

AATTATTGGC ATGGGTACTG CCGTTTCATT ACTAACTTCT GCAGAAAAGC AAGCAAGAAA 900

AACAGCCTAT CAAAGTTTTC AAACGATTAT TTTAAAAGCG TTAGAAGAAG CCAATATTGA 960

TTTTtCmATT AATGGkGaAC CAACCAATCG TTTAGCGCAT GTCTTAAayC TTCATTTAAA 1020

nGGAATcCCA GTGATTTGCT CTTAATGCAT TTGGACTTAA GAGGTATTGC GATTTCAACA 1080

GGCTCGGCTT GTACAGCGGG AACTGTGGAT CCTTCGCACG TGTTGACAGC GATGTATGGA 1140

GAAAATTCCT CAGCTATTAA AGAATCAATC CGGATTAGTT TTGGCTACGG GAACACACCA 1200

GAAGAAATTG CGACTTTTTC TGAAGTACTA GTTGCAGTGA TTCAACAATT GAAAAAATAA 1260

TCTGATTGCG GCAGCTACAG GTTTTCGCTA TAATGAAGAT AACAATGAAA ATAATGAGGT 1320

GGAACCCATG GCATTTACAA CAACGGCTGC TGTTGAAGGC TCAACAGTTT TTTACAAAGT 1380

ACATCCTAGT GCAAAACGCT ATACATTAAA AGATAACGGC TTTACTGAAA CAAAATCAGG 1440

GAACTTTCAA TTGATTCGTT CCTTAGATCC AACGCCTCAG AGAAATGAAG GCTTCAAATT 1500

GAAAATTACG ATCACTGCTG ATCTAAAAGA GTTGAAAATG TCCATTACAA CGGCAAATGG 1560

TTTAAAACCA ATGAATATTT TcAAGAATGA ACAACATGAA ATGAGTAAAG AAAAATACTT 1620

TTTCTTAATG GATGGCTTAA TTAGCCGTGG TGTATTAGAG AAAGTAGAA 1669

(2) INFORMATION FOR SEQ ID NO: 167:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19031 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

GGTGAGTGTT CTAGCAGTGC TGTTTGTGGC AGCTGCTTGT TCCTCAAAAC CAGAAGCGAA 60

AATTAATACA GAACCATACT CTGACCGTCA AACAATGCTT GGTACCTACG TTCAAATTCG 120

CATCTACGAT GATGGCAAAG AAGACGTCTT ACCGAAAGCC TTTGCCCGTG TCAAAGAACT 180

AGGCGATAAA ATCACCGTCA ACCAACCAGG TTCAGAAATT GATGAArTCA ACCAAGAAGC 240

AGGCGTGAAA CCAGTCAAAG TCTCAGATGA CTTGTATCCT TTGTTGAAAA AAGCCTATGA 300

ATACAGCAAA GACTCCCGCG GCGGCTTTGA CATGGCAATT GGCCCAATCA CCTCTATGTG 360

GCACATTGGC TTCGATGATG CCCGCAAACC AAGCCAAGCC GAAATTGATC AAGCCTTGAA 420

ACTTGTCGAT TACACGAAAG TCAAATTCAA CGACAAAGAG CAAACCGTTT ATCTTGAAGA 480

AAAAGGCATG CAACTAGATT TAGGCGCAAT TGCCAAAGGC TTCATCACCG ATGAAGTCGT 540

GAAAGTTCTG AAAGACAACG GCGTTACCAC CGCTATTGTC GATTTAGGTG GGAATGTTTA 600

CGTCTTAGGA CACAGCCCGC GTGGCAAAGA CATGGATTGG ACAGTCGGGA TTCAAGATCC 660

GAATAAAGCC CGCAACACAG TCTTAGGTCA AGTCAAAGAA AGCAACAAAA CCTTAGTCAC 720

ATCAGGAATT TATGAACGCT ACTTAAAAGT CGATGGCAAA ACCTACCACC ATCTCTTTGA 780

TCGTGAAACT GGCTATCCGT TTGATAACGA CATCGCGGGT GTCACCATTA TCACCGACAA 840

ATCCATCGAT GGTGATGGTC TCTCAACAGC CGTCTTCTCA ATGGGCGTCA AAAAAGGACT 900

AGAATATGTC GAAAGTCTGA AAGGAACCGA TGCCATTTTC GTCACAAAAG ATGACAAAGT 960

```
CTACATCAGC AAAGATATCG AAGGCAACTT TGAAATCGGC AAAGACTCTG GCTATACGAT    1020
GGGGAATCGG GCCGATTTGA AATAAGGAGC GACTAAAAAT GTCTTTAAAT GTGTTTTTAA    1080
AAGTAGTGGA AATTCAGACG AAGCTTGCCA GTCTCTTTCC CTTCATTATC GGGGTCTTGT    1140
TTTCAATGAC CTATTTCGGG GAAGTTCAGT GGGGCAACAC GCTGGTTTTC TTCATTGGCA    1200
TGGTGGTCTT TGATATGGCC ACTACCGCCA TTAATAATTA CATGGACTTC AAAAAAGCCA    1260
AATCAGACGT CTATAAATAT GAAGAAAATA TTATCGGACA ATCAGGCGTT TCACCACAAC    1320
TGGTCCGCAA CATGATTTTC GCCATGATTG CCTTCACTGC CGTTGTGGGT GCGTACTTGA    1380
CGGTTCAAAC CGGCTGGCTC TTCCTTGTTT TAGGCGGGAT TTGCTGCTTC ATCGGGATTT    1440
TCTACACGTA TGGACCGATT CCGTTATCTC GCATGCCCTT GGGGGAAATT TTCAGTGGCT    1500
TCACGATGGG CTTAGGCATC TTCGTCTTAA CCGTTTACTT GAACGTCGTC ACCAATCCGC    1560
CTTTTTACTT AACGGTCGAT TTTGCTAGCG GCGCCTTTCG CTTAGACGGC AATCTCTGGG    1620
CGGTGCTTGC GATTGTCGTG GCGTCCTTGC CCTTGGTCTG TTCGATTGCT AACATCATGC    1680
TCGCCAACAA TTTACGGGAC TTAGACACGG ATATTGAGAA CCACCGTTAT ACATTAGTCT    1740
ATTATATTGG GCGCCCATCG GGGATTGTGC TGTTTCAAGT GTTAGCGCTA sctGGGTATC    1800
TGGTGATTTT ATTCGGCTTC ATTAGTGGTA TTTATCAATG GCCTGTGTTA ATTACATTTG    1860
CCACTTTACC TGTCATTTGG AAAAATATTC AAACGTTTAA ACAAGAATTG CCGCAGCCGA    1920
AAAGTTTTCG CCATTCCATT AAGAACTTAA TGGTTTTCAA TGGCACCTAT GCTTTAGGGC    1980
TATTCATTAG CGCCTTACTT GGTTAGTCTT GCGAAAAAAT AGCGGGCTGG GCGAAAAAAA    2040
TACTGTTTTT TTTCAGAGGC CTGCGTCAAA ACAATCAAAC AAAGCTCAAA ATGTTAGTAT    2100
TCGTTGCTAA CATTTTGAGC TTTTCTTGTT AAAAAGAGCG GAATCTTAAA ATTTTGTCAG    2160
GGGATTTTGG TAAAAATTTG CCGAAATCGC CCTTTTTTTG AGCCGATTAT TAAATAACAG    2220
CTGTTGTAAC TCTTGTCAGA ATAGGAACCT TCAAAATTTT CGCATGGGGT GCGAAAAAAT    2280
CGTCAATTTC TGCTGGGGGT TTCTTTGTTA TTATAGAGGA GTAACAAAAG GTTGTTTTGT    2340
CTTTCTTGTG TAAAAGGGCA AGAAAGGCTA GCGAGTTAAA AGGAGGTTTT TCAATGAAAA    2400
AAAAGCGTTA TTTAATGATT GTGTGTCTAC TATCTTCTCC TAGTTTTTTT ATAAATGTTG    2460
AAGCGTCTGA TGGTGGTTCT AGTTCGGTGG GGATTGAATT TTACCAAAAT CCGAGAACAC    2520
CCGCTCCTAA AGATCCCCCA CCGAAAACAG ATGCGCCAGC TGCTGATCCC AAGGAACCAG    2580
CTGGTCCTCC GCAAGGAGAT CAACGAAgTG GTGGTTCGAC ACAGACCACC ACAACTGGCT    2640
CAACGCTCCC TCGTACAGGG AGCAAGAGTC AGGCAAATTT GAGCATTCtC nGnTTCGCCT    2700
TAATCGGTTT GGCGGGAATC GTACATAGAA AGAAGGGACG ACATGAAGCA AACTAAGTGG    2760
CAACGATTAG CAACCATTGG CTTGTGTAGT TCTTTAGTAA TTAACGCCTT TTCTGGTGTG    2820
ACGGCAGTTG CGGAAACCGT GACGATTGAA AGTAGTCCGA CCGCCGAAAG TAGTGCCAAG    2880
GAAGAGACGC AAGCAAGTAG CGTGAAGGAA GAAACAACGA AAGCCAGTAC GGAAAATAGT    2940
```

CAAGTAACAA CTGACACGAG TCAGGAAGAA GCAACGAAAG AAGCGGAGAA AGAAGAACCG 3000

CAAGCAGAAG TGGAACAAGC AGAAACACCA ATCATTCCTA AACCAAAAAA AATCAATATG 3060

AAGGCAACTT ATTCATTTTC TGCAGAAACT TATCAGTTTG GATTTGTGAA TGAATCAGGT 3120

CAATTAATAA ATCCAGATAT TATACCAATT ACGTATAGCT ATGCCAAAGG ATCATGGAAG 3180

ACAGATGGTT ATAATCGAAA GTGGACTAGT ATGGTTCAAG GGAGTGCTTC AACCGTAGGA 3240

AACTTAAAGA ATGTAATAAT GCCAGCAACT TCTGTAGTTA TGCCACCAGG ACCGTCATAT 3300

GAAGGAACTC AAGAGGTGTA CACAAACTTT TCAATTCGCA TACCAAAATA TTATGCATCA 3360

GCGAGTCTCT ACAATAGAGA AGGTAAAATT GATTCTACTT ATCCGTTACC TGCTATTGCA 3420

CTAGCAGGTA CTAGACCGCT ATCTTTGACT CAAAGTAGTG TAATTAGTGC ATTGGCGCTG 3480

ACCAGTAAAG GAGACAATGT TTATACACCA CGGGAAACAT TTTTTGGAGG AGATCCTGCA 3540

GGTGTAAAGT TTACTAATTT TTTGTATCGT ATAAATGACT TTGATGTGAA AGGTAATAAC 3600

ATAGGTTATA AGACTGTGAG TAGCCCAATC TATTACCATC TGACCAACCG CCGTGTCACC 3660

GAAAACTTCG TAGATACAAG TGGCGCCAAA ATCACGCCAC CAAGTAATTT CACCCAAGGG 3720

AAACAAACGG TCATTAACAG TGATCCTTAC ACGTTCCAAC AAAGTGGTTT TTTACCCGAG 3780

ACCTACAAAG TTGGCACGAA ATCTTACCGA TTCAAAGGCT GGTACAAAGG GAAAACCAAA 3840

ACCGAGCCTT TGGCCACCAC TAAAACACCT AGCTATAAAG TCACGTATGA TGACAATGAT 3900

GATTTGACGG TGGTCTATGA GGAGTTTTCA GGGTACGAGC TGCCTGCTTC GACCAATCAA 3960

TTTGGCTTTG TGGATGAAGC GACGAACAAA TTAATTGCCC CCGACCAAGT GCAGATGAAG 4020

TATAATCTTA CTTTAAATGA AAATAATAAA AAAACAGTAA TGAGCAGTAA CTTAACGGGG 4080

ACAGATACAG CGACACTGAA AAACTTGTCC GTGCCTGTCA ACTATTTTGA ACAATATCGC 4140

GTCAATACGT TTTATGGCGC GAGTGACATT ACGTTTACAT TGCCCAAACG GTACAAATCA 4200

ATCAATATTA CCAAATCAGA TGGCAAAACC GACCCAGCTT TTCCTCTTCC TAAAATCTAT 4260

AATATAGATC AAGTAGAAAT GTCACACATG CCTGTGACCA CTTATAACAA GTTGAAACAG 4320

CTGTCGGGCC AAACGTTTGG CTTTAATGCT TTAGCCGATC AACCTGAATT TTATACGAAA 4380

ACGTTATTTG GGACAGAGTC TGGCATCGAT GACCCAGTCA ATTATTATAC AATGAGTGGC 4440

CCTGTTTACT ATTATTTAGA AAACCGCAAA GTCACCGAGA aCTTCGTAGA CACCAACGGC 4500

GCTAAAATCA CACCGCCAAC AGGTTTCACC CAAGGTAAAA AAACGGTGAT TACAAGCGAC 4560

GCCTACACTT TCAAACAAGC AGGCACCTTA CCAGACACTT ACACAACAGG CGGTAAGACC 4620

TACAAGTTCA AAGGTTGGTA CAAAGGCAAG TCCATACTCA ACACATTGAC AACTACCAAA 4680

GCGCCAAGTT ATCAAGTGAC CTACGATGAC AATGATGATT TGAATGTGGT GTATGAAGAA 4740

GAAACAGTTA CGACAGTGTA TCCATCAGTC GATATGAACT TTGTGAATGA AAAAGGCGGG 4800

GCTTTCACAC CGGCGTTAAC TTTTAGTGGT AAGTACTATG CGCAAAGTAC GAGTGCGTAC 4860

TTAAGAACCG ATTTATATGA CGTGACCTCA AAAAATAATG GTAATGGGCA ATATACGGTA 4920

AGTATTAATA ATGGTAGTAT GCCATTGTCC CAAGAATTAT TGAAAAAATA TAATAATGGA 4980

CAACCAATCA GTGCTACCAA CAGATTACAG TTTAATGTTG ATAAATTAGC CATCGACCAA 5040

CAACTAAAAT ATGTTGACAG CATTCAATTA GACACAGCTC AAAGTAGCAA TCTGAAATCC 5100

TATAGATATG TGTACACGAA CAATAGCTCA CTGGTTTTCG ACCCAAATGT AGCACCAGCA 5160

GAGGTTGACC TTAGTTCAGA ATCTCTTAAC TTGCTTAATT TTGATTCAGA TGGCACCTAT 5220

TTTTCTAATG CAAATAATAG ACTTTTTTAC ACGCATTTAG GATATAGTGG CACACCAGGA 5280

GTTAACTATC TTCTCGTAAT GTTTCTTTTT AACGCCAAAc CTGCGGATAA GTCAAAACTT 5340

GTCTACAAAG TCACTCGCAA ACAAGTCACC GAAAACTTCG TGGATGTCAA CGGTGCCAAA 5400

ATCACTGCAC CAACAGGCTT CACCCAAGGT AACCAAGTAC CAATGAACAG TAACACCTTC 5460

AAGTACACAG CGGCAAAAGC TTTACCAGCG ACGTATACTA CAGGTGGCAA AGTCTATACG 5520

TTCCAAGGGT GGTATAAAGG GAAAACCAAG CCAAGTACGT TGAACAAAAC AACAACTCCA 5580

ACGTTCAATG CGACCTTTGA TGGCAATGAC GATATGACCG CCATGTATAA GGAAGAAATA 5640

CCAACAGCTA GTGTCACATT AACTCGACCA AAAGAAGTGA TTGATACGAA TACCAATGTA 5700

ATCTGGACAA CAACGATCAC GAATACTAGC AAAGCACCCT TACAAAATCT CACCTTGAAA 5760

AAAGGGCCCA ATTGGTCAGC TGGTCTGACG ATCCCGACCT TTATGGAAGT GACACCAGAA 5820

GGAGAAACGA CAAAaTCAAT CCCAGTAAAT AGTACACTTT GGACAGAGGG GGTTCCTTTA 5880

CCAAATGCCG TTCcTATCGG CAAAAAAGTT TCAGTTGCTT TCACAACTCG CGCAACAGGG 5940

AAACCAAACA CTGTTTTGAA AGCAGAAGTT GTAGTATTTG GTGGTATTAA AGATAGTACA 6000

GTGGATAACT TCGTGAGAAT TCGTCCAAAT GATCAAGAAG TAGTCACACC AACGACCGAA 6060

GGCTTCATCA GTGTGCCAAC CTTCGACTTC GGCCAAGTGG GCGTTGCAGG AACTAAGCAA 6120

CAACACAGCT TGAAACAAGC CGCGGATTAC TACGGTAACG GCACACGGAA TCCGTATCTG 6180

CGGATTAAGA AAACGCAACC CAATTGGAGC TTAACAGCGC AACTGTCACA ACCAAAATCA 6240

GCGACAGACA GCTTGCCTAC AGCGACCCGC TTATTATTAG GGGCGGCGCC TGTCTCTAGC 6300

TTTACCAATT ACAATCAACC AACCGAGTTG AAAAATACGG TCGGTACCAC GAGTGCCATT 6360

AGCTTAACAG CCAACAACAC AGCAACGAGT ATTATTGCCA ACAAGCAATT CACAGGTAGT 6420

AATGTTTATC AGTTGGACTT CACCTTCAAT AATGTCAAAC TTGAAGTGCC AGCCAATCAA 6480

GGTGTTAAAG GGCAACAATA CAAGGCCGCA GTTACATGGA ACCTAGTTAC AGGTCCTTAA 6540

AATAGTTAAG TTATCATGAG AAAAGCTCAA AATGTTAGTG GTTATCACTA ACATTTTGAG 6600

CyTTTTGGTT TAAATATTAA GGAAAAAACA GTTGGAGTTC ATCAAAAGAG CGTTTAAATC 6660

CTTGATTTTC GCTAAAATAA CACTTATTTT CAAAGGGAAC AGAGGCGCTC AGCCTTGAAA 6720

GGATAGGAAT CACTAAAAAA CGGCACTCAA TATGTCAAAA TTTGAAATTT CAAGCTGTGT 6780

GTTCTTTGGT AAAATAwATA wAAAAATGCT AGTTATCAGT ATCGATAATA ACAGGATACT 6840

GATTAAGAAA GGACTTTATA GAGACTATAG ATTGAATTTT TACATAGAAA GAAGGAGCAA 6900

GATGAAGCGA GTAAATTGGA AAAGATGGCT AGTTGTTGGG TTAAGTTGTT CTTTGTTCAT 6960

GGATTCAGTG GTTGGTGTGA CTGTGTTAGC GGAAACGATT ACTGGGGCGA CGGAGCAAGG 7020

AGTAGCAACA TCTCAGTCGA GTGACGAAgC GAGCCAGACG aCGCAAACAA CCGaAGAGTC 7080

ACAGGCAACG GTCGCTAGTG AAGCGAAAAC AGTACCGCCA CAGGAAACGG CAAGAATTGC 7140

TTCTCGAGCG ATTGGTTATT CTTCTGTGGA AGGGCGCGAG ATTCCCTTTT TCTTTGTGGA 7200

GGAAGACGGG ACGTTGTTTG ATCCCGACCG AATTACGATG GCGGTCAATC TTTCCACGTT 7260

TTCGTTTTAT GAAGAGAAAT TACAACGAAC CCCCCTTGAG CCCACCACTG TGAATGGCGG 7320

AAAGTTACTG TCTATTCCAA CGTCACCAGC TTTTAAATAT GATACAAATA ACCAGAATCC 7380

AAGTAATATT TATGGCGTTT CTGAAGTGTC GTTTACTATT CCTAAGGAGT ATCAAAGCCT 7440

GGACATTCGA CCAAGTACGT TTTATACAGG AGACACTACG CAATATCCAG TGCCAACGGT 7500

TTTTGCGAAC GTTGGGGGCA AAGTGACGAA CTATGTGGGC GCCAATGCGG AGACGGAATT 7560

AGAGTTAACC AATGAAAAAA TGCCCAATAA GCTGACGTTT GGTCCTAAAA AGACGTTTAA 7620

ATATACGGTA GCTACGGCAC CAGGAGGCGT TACGTATGCG CTGACCTATT TTTATGGAGA 7680

TGTCGGCGGT CCAACTAGTT CGCACCAAAG ACGAGGAACA GCGGGTCCTG TGTATTATTA 7740

TTTAACAAAG CGGCGTGTCA CGGAAAAATT TGAGAATCCC GCAGGCGGGG CGATTCCTGC 7800

GCCAGAAGGT TATACGCAGG ATAAGAAAAC CATTGTAACA GGGGAGGATT TTACTTTTAC 7860

CCAAGAAGGC ACCTTGCCTG AACGTTACAC AGGCAGTGAT GGGAAGACGT ATTTATTTAA 7920

AGGTTGGTAC AAAGGGAATG CGAAACCTAG CACGTTGGAA ACCACCAAAA CGCCTAGTTA 7980

TGCGGTGACC TATGATGACA ATGACGATTT GCATGTGGTC TATGAAGAAG CAGTGATGAA 8040

AACCTATACG TTGCCAGCGA GAGAAGCTTT GTTCGGCTAT GTTGATGAGC AAGGAAACTT 8100

GATTAATCCC GCCAAGTTTA AGCTAAGTGC GACCATGGGT GAAAGTGACG GAGCCACAGG 8160

GGAAATGACG ACTTTTCCCA CAATTGATGG AATCGATATG CCAGCAAGTC AATTAAAGAA 8220

ATTAGCCATC CCGCAAAAAG TCTACACACG CCCAGACGAT GGGACAATCG TAACTTATGG 8280

CCCGCAAGAA GTGAGTGTTG AAATTCCTAA GTATTACCAG ACGATTTCGA TTTCACCAAC 8340

TACTGCGTAT ACAGGGGATA AAACCAAGTA TCCAGTACCA AATGAAGTGC GCCGTGGCAT 8400

CGAAAACCCC GACAACATTG TTAGTAGTTT aGTGGGAAmC mCTGCGTATA ACTTGACCCA 8460

AAAAAGTGCC ACACGCTATA CTGCCCGCCG TTCTTACTGG AmGTGGGGCC CCACGAAGAC 8520

ACTTTACTCA ATGAGTATCT ATTCAGGAAC TGCTGGGGGC AACTATAATT TATCGACCCC 8580

TGATGGCACC ATTTATTATT ACTTAGAAAA TCGGCGGGTC ACTGAACATT TTGTAGACGA 8640

AAGTGGCGCA AAAATCACGC CACCAACTGG CTTTACACAA GGAAATCAGC TAGTGGTGGA 8700

CAGTGAAAAC TATGTCTACA CTGTCGCAAA AGCTTTGCCG AAGATCTACC AAGCTGGTGA 8760

AAAAACCTAT ATCTTCCAAG GCTGGTTTAA AGGCAAAACC AAGCCAGCAA CATTAAAGAC 8820

GACAACGACC CCAAGTTTTA CACCAACTTT TAATGATGAG GACGACATGA CCGCTGTGTA 8880

CCAAGAAGCG ATTCCCACCG CGGaAcTAAC GTTAACAGGT GCCGTTGACA TAATCGAAAA 8940

TGGCGCCACA ATGGATTACT GGGAGGCGCT ACTGAAGAAC ACAGGCGAAg cGCCGTTAAC 9000

CACCATTAAA ATCAAGCCAA CGGCAACTTG GGCGGCTGGC ATCGGCGCAC CCAACACGAT 9060

ATTTGTACAA GGAACGGGTC AAAACACCAA AGCTTTTCCT GTCACCAAAG AACAATGGAC 9120

GACCGGTGCA GGAGTGTCCA TCACGTTGGA TCAGCCTTTA CCAGCTGGCG GTCAATTAAA 9180

AATGAACTTA TTAGGAACCG CCGTTACAGG AAATCCTGGT CAAGTTTTAA CCGCTGATGT 9240

TGAAGTAACG GGCAACTTTG GCAGTTTAAC TGCCAAAGAT ACGGTCCGTA TTAAAGACTT 9300

AGATCAAGAA ATTACGAGTC CTGACGGCGA CGGCTTTATT AGTACCCCGA CATTTGATTT 9360

TGGTAAACTA GCAATTTCAG GAAGTAAGCA ACAATATGGT TTGAAGAAGG CCGCAGATTA 9420

CTACGGCAAT GGCACTCGCA ACCCTTATTT ACGCCTGAAT ACTAGCCAAG CCAATTGGAG 9480

TTTAACGGCC CAGcTATCGC AACCAAAATC AGCCACAGAC AGCTTGCCAA CAACGACCCG 9540

CTTGTTGCTA GGAACGGCCG CTGCTGCCAG CTTTACCGAT TACAACCAAC CAACAGAAAC 9600

CAGGACACCA CTTGGCAAGA CCAGCACCGT GACTTTAACC GCCGACAATA CCGCAACAGC 9660

GGTGGTCGCA AACCAACAGT TCACAGGCAG TGACGTCTAT CAGTTGGACT TCACGTTTGC 9720

TAACATCAAA CTAGAAGTGC CAGCCAACCA AGGTATGGCT GGCCAACAAT ACCAAGCCGC 9780

CGTCACGTGG AATTTAGTGA CTGGCCCCTA AAATAAAAAT GCCCGAGCCT TAGAAAAATA 9840

AGAGGCTCGG GCATTTTTAT TGATGGTCTT TTGATGGTCG TTGTGCTATA ATAGCATTGA 9900

AAAGAGAGAT ATGCTGTAAC ATACCTCTCT AGTGTAGAGC CGTTTAAGAC GGTGACCGAT 9960

TTTGTTACAA AAAATAACCG TACTCAGTCA AAGTAAACGG TTATTTTTTC TTGTCTTCTT 10020

TAACAACCTT TAGGATTCCG AAGATTAAAG TCGCGATAAA AGAACCGAAA CTAATCATCA 10080

GAGCTAAAGC TTCTGCGATA GACAAAAGGC CTCTCCTTTC ATAGATTTTA GTACTTACAT 10140

TCATAAGCAC CACCACCTTT CAGATTGGAT AGCCACCGCC ATAACTTCTC TACACTTTTA 10200

TTATACCATT TAGAACTAAA AAATAATTAG TTTTCAATAT ATTTTTTTAG GAAATTGTGC 10260

TATAATAGCA ATGAAAAGAG AGATATGCTT CAACATACCT CTCTGATGTA GAGCCGTTTA 10320

AGACGGTGAC CTTTTTAGTT ACAAAAAATA ACCGTACTCA GTCAAAGTTG ACGGTTATTT 10380

TTTATTGTCA TTTTTAAGCT TTAAAAGATT TTCCTGTTCC CTAATTGTTA AGAATAGAGA 10440

AGAAAAGCTC AAAATGTTAG TTCTGATTAC TAACATTTTG AGTTTTTTTC AATAAATATC 10500

AAGAAAAATC AGCGGAAATT AGAGAGAAAA TGGTGAATTC CACTGATTTC CAGGCTGTTT 10560

TTTACCAAAA AACAGTGGCT CCTCCTTGGT ATAAAAGGAA TCATTGGAAA TATCAAGTTG 10620

TCGTATCAAA ATTTAAAATT TCAGTGGGGC CTTTTTTTGA TAAAATGAAA ATCATACAAA 10680

AAATCATACA TAGGACAATA AAAATAACAA TAAAAAGTTA CATAAAAAAG GAGAATGGCA 10740

TGAATCAAAA AAGAAACCGT TTATTTGCAA TCGCTTTGTG TGGTTCTTTA CTAATCAGTT 10800

CATTAAATGG CATCCCAGCA ATTGCTGAAA CCATGGCGGC CCCAGATGTT CAAGGGGAGG 10860

ACCGCCAAAC AGTAGTCCCG TCAGAGGGGG TTCAAGGACG GGCGCCCGTT TCTTTTCCAG 10920

AGGCAGATAC GGAAGTnAGc GTACCGCCAG CTGATGAAGA GCTACCAGAA CAGACTGAAC 10980

AAGTAGTCGG GCAAGTTGAA GCGTCAGTGA CGAGTCCAAT AATAGTTAAT ACACCAGAAT 11040

CACTGCGTGC GCAgCGAGCA GTGCAGACAT TCGATATTCC AGAGAAGACG TACAAAATGT 11100

CGTTTTTAGA TGAAAGTGGT CGTTTTATTG ATCCAGCGAA AGTTTCGATT ACTGGGGATA 11160

TTATGAAGTT TACTATGCAA GGCGCTTTGA CGAAAGTCGG GGCCGTGGCC ACTAATAATG 11220

TGGGAACCAC GAAGCAATTG ACGGTGCCGA AAGTAGCTTT GTCTGGGGAT ACAGCCATAA 11280

AGAGTTATGA TTTTTCCGGC ATCCAAAATG AAAAGATTAC TTTGCCGCGG TTCTATAATG 11340

TTCCGAAAAT CACGCCCAGT GCGAATTACA CTGGACAAAT CTATCCAATG GCAGTGACGA 11400

AACAGGTGAA TGTTACAGGA GAAGTATCGA ATGAGGAAAC CACAGGCGCG AGGCTGAATA 11460

TGTTTGTATC AGGTATCCCT AACCAGTTTG TCTTTTCTCG TGCTAGTTGG GACAATGATC 11520

CAGCGGGAAA TATGTTTGCG CGTGGCTTTA ACATGGCTAC TGCAGGCGGC ACACCAGCTT 11580

ATCTGACTAC GGATGATCAC ATTTACTATT ATGTGCCTAA CAAACGGGTT AACATTTATT 11640

ATAATGACAT GACCCCGAAT GGCTTACCAG CGTATCCTCC AGGGTATAAT GCCAGTCTTT 11700

CAAAAACAGT AGTTGATTCA GAAAATTTTC ACTATAAAGC GCCAAAAGCT TTTCCAGAAT 11760

TCTTTAGTTC TGGGAATCGT CACTTCCAGT TTAAAGGCTG GTATAAAGGG CCTGCAAGAC 11820

CAGCGGATCC AAAAGCCGTT AAGTTAGAAA CCTCATTGAC GCCAGAGTTC GATGTCACTT 11880

ACGATGGTGC CGATAATTTA TATGTTTTTT ATGATGAAGT GAAAGAAATG ACCACTACTA 11940

TTCCAGAAGT TACCTATAAG TTTGGCTTTG TTGATGAAAA AGGGGCGCTT GTTGCGCCAA 12000

CGAATGTAGA CATTCAGGCG AACTTAACAA CTGTCGAGGA TAAGGTGGTC CAGAAGGTGG 12060

GGGCTATTAC AGGCGCTAAT GTAGGCAATT TAAAACAGTT AACCGTTCCT AGCCAAACAC 12120

TAACCTATGT ACCTAAAATT AAAGTTGCTT CTTCTGGGGT AACCGATTTT CGGTTGACTA 12180

TTCCCAAGCG CTATAAAATT CCAACAGTCA CACCTGGTTC ATTCTACACT GGCTCAACGA 12240

CCGCCTATCC TTTGGCGACA ACGCTGTTAA GGCATATTAG TGGTCAGGCA GATGAACGGA 12300

TAACAACGGA TGGCACTCGC TATAGTTTAT ATAATTTACC GACGCCTCAT AGCTATCGTA 12360

TGTACAGAAG TAGTTGGCAA CCTGACGTTA CCAACACGGG ATTTAGTGCA ACTTTAAACA 12420

TGGCGACGAC CGAGATTTAT CCAGCTTATT TCACGACCGA TAACACGATG TACTATTTCT 12480

TAGAAAACCG CCGCGTCACT GAACACTATG TTAATGAAGC AGGTGCTGAG GTACCGATGC 12540

CTACAGGCTT CACGCAAGGG AATCAAACGG TGATTGACAG TGATACTTAT CATTTTAAAT 12600

TGGCAAAGGA ATTGCCGTTT AGTTATTTTT TAAACAATAA AGCATACCGA TTTAAAGGCT 12660

GGTATAAAGG CAAAACCAAA CCTAAGGTAT TGGAAACGAG TCGAACGCCT GAGTATGATA 12720

CAACTTTCGA TGATAACGAT GATTTAACCG TTGTTTACGA AGAAATTAAA TACGGCGGAA 12780

ATGCGGTGAC ATTTGGTTTT GTTGGGGAAG ATGGGCAATT GCTTCAGCCG ACAGGGTTTC 12840

AGGTGACTAC GGATATTGTT GAAACAATCG ATGGGCTGAG TACGGTTTTA AGTAATGTTT 12900

CAGCTGTCGA CAGTACAGGC AACGTTAAAA CTTTGACGAT CCCGCAAAAG GAATATGTAT 12960

TGACTCCCCA GATGAATTTC TATGGAACAA AAAATTCATT GATTACGATT CCAAGACAAT 13020

ACCAAGAAAT TTCTTTTACC AAACCTGCCA ACTACCAAGG ACTAGATTAT CCAGTTGCAG 13080

GGCGTGTTGA GAATATGTTT AGTGGTAATC TGTATGTTGA GGGGGCACAA CCACCTTACT 13140

TTTCTGCTTC GAAAGTCAAA GCCAATCAAT ACAAAATAAT TGAGTTTCAT TGGAAAGAAG 13200

AGCCTACACA AAGTTTCTCT AGTTTGTACA AAGCTACGTT AATTCGAAGT AGCGTGGCAC 13260

CAACAGGCGT AGAAATGTAT CCTAATAAGC CAATCTATTA CTATGCGACC AATCGCCGAG 13320

TGACTGAAAA CTTTGTAGAT AAATCTGGCG CAAAAATCAC GCCACCTCAA GGATTTACGC 13380

AAGGGAATCA AGTCACAGTC ACCAGTGATC CTTTTACTTA TACTGCTTCA AAAGCTTTAC 13440

CAAGCGTTTA TGCGGCTGGG GCAAAAACCT ATAAGTTTGT TGGTTGGTAT AAAGGAACCA 13500

CCAAGCCAAC GACCTTAAAA ACAACGGCTA CTCCTAATTA TTTGGTGGAT TTTGATGATA 13560

ATGACGATAT GACGGCGGTC TATGAAGAAG AAACGCCGAC TGCTGCTTTA ACTTTGACCA 13620

GCACCAACCG TGTCGTGAAT AATGACGATA GTGTAGATTG GGTGGCAACA CTGAAAAATA 13680

CTAGCCTGGC ACCACTTAAA ACGCTGACAG TGAAACCAGC GACTACTTGG CCTGCAGGAA 13740

TTGGTACGCC GACGTCACTT TCTGTCCAAT TGGACGGACA AGCACCGAAA ATTTATCCAG 13800

TAACGGCGAC TACTTGGTCT GAAGGAATCA GCTTaACAGG TCTTGAAATA CCGGCAGGTC 13860

AAACAGCCAA CGTGACGCTT GTTGGCACCA AAATTTCAGG AACATCAGAC CAACGCTTGA 13920

CGGCAACCCT TGATGTTACA GGAAATTTTG CGACAGTGAG CGCCGCGGAT GCGGTACGTC 13980

TGACAGATAC AACCCAAGGA ACAATTACAC CAACAGAAGG CTTTATCAGC GTGCCAACCT 14040

TTGATTTTGG AAAAATGAAT ATTGCTAGTA AAACGCAACA ATCAGGCTTG AAAAAAGCGG 14100

CCGATTACTA CGAAAATGGT ACACGCAATC CTTACTTGCG AATTAAGAAA AATCAACCGA 14160

ATTGGCAGAT GACAGCGCAA CTATCTCAAC CGAAAGCAAC CACAGATAGT CTGCCAACAG 14220

CGACTCGTTT ATTGTTAGGT CCAGCGAATG TTTCAAGTTT TACAAATTAT AACGAAGCAA 14280

CCGAACAAAT CAAAGCAGTT GGTAAGACTA GTTCTTTGAG CTTAACGGCG AATAATGTGG 14340

CAACCAGTGT TGTCGCAAAC CAACAGTTCA CAGGTAGTGA CGTCTACCAA CTGGATTTTA 14400

CGTTTGAGAA CATTAAATTA GAAGTACCAG CCAACCAAGG AACCAAAGGA CAACAGTACA 14460

ATGCTGCGGT CACTTGGAAT TTGGTGACTG GTCCCTAAAA TAAAAATGCC CGAGCCTTGA 14520

AATAGGAGGC TCAGGCATTT TTTATTGATG GTCTTTTTAG GGTTGTTGTG CTATAATGAT 14580

AAAGAAAAGA GAGATATGCT TCCAACATAT CTCTCTGATG TAGAGCCGTT TAAGACGGTG 14640

ACCAATTTTA TTATTTAAAA ATAACCGTGC TTGGTCAAAG TAGACGGTTA TTTTTTCTTG 14700

TCATTTTTAA GCAATTTCAC AATCAGCGCA ATCAAAGCAA TGGTAAACAT ACCAAACCCA 14760

AGAATTGTCT GAATTGTTTC ATATGCTGAC AAAAGGCCTC TCCTTTCTGT GAATTTCGGA 14820

AATACGTGCA TACGCATCAC CTCTCTTTCA GATTGGATAG CCACCGCCAT AACTTCTCTA 14880

CACATACATT ATACCATGTG GGGCTAAAAA ATAATTAGTT TTCAATATAT TTTTTTAGGT 14940

AAGTGTGCTA TAATAGCAAT GAAAAGAGAG GTATGCGCCA ACATACCTCT CTAGTGTAGA 15000

GCGGTTTAAG ACGGTGACCT TTTGGATTAT TTAAAAATAA CCGTACTTGG TCAAAGTAGA 15060

CGGTTATTTT TTCTTGTCTT CTTTAACAGC TTTTAGGATT CCGAAGATTA ACGTTGCGAT 15120

TCTTGTTGCA AAGCGCATGT TCCTGTGCGC TTGCCGCCGC CATAACTTCT CTACGCTTAG 15180

AGTATACTAT GTAAAACAAA AAATAATTAG TTTCCAATAT ATTTTTTAGG CAATTGTACT 15240

ATAATGGTAA TGAAAAGAGA GCTATGCGCT AACATACCTC TCTGATGTAG AGCCGTTACT 15300

AACGGTGACA AAAATTAATA TTTTTGAAAT GACCGTCTAA CTAGGCAAAG TTAAGCAGCG 15360

CCCCTTTCAT TAAGGGGCGC TGCTTTTTAG TCGTTATTGC GATTGTTCAG TGTCTTTTGT 15420

AGAGAGGTTC ATTTGTTCTT TGATGGCATT TTCGACACGT TGGCGTTCAC TCTCTTCCAC 15480

AAGTTGATAA GAAATACCTT CAATCCCTGT CACGCCATCG CCAGTGTAAT CGTAGCCTTT 15540

TAATTGATCG GTTTCAATCG TTTTTAGTGC GGGGGTGTAT TTTTTTAGCA TTGCCATCAT 15600

TTGGTCAAAG GATAAGTCGG TTTCGCCATG TTCGCCGATT ACGTCCAGAA TTTCTTGATA 15660

TTTTAGCAAG CTGTTAAAAG AAAGTAGCTT GTTGGTTAAT TCGATTAAAA CTTCTCGTTG 15720

GCGGCGTTGA CGACCATAGT CGCCTTCGGG GTCTTCGTAA CGCATGCGCG CGTATTGTAA 15780

AGCTTCCCAG CCACCTAAAT GCTGTTTGCC TTTCGGATAA TGAATCCCTT CAGCGTCAAA 15840

AGCAAAGGCA TTGTCAACAG TGACGCCGCC CACCGCATCG CCAAGGTCCT TCATAGCATC 15900

CATATCAATT GAAAAGTAAT GATTGATTGG GATGTGCAAG ACATTTTCAA CCGTTTTAAT 15960

TTGCATATCG GCACCACCGA AAACGTGGGC CGCATTGATT TTATTGTATT CGCCATTGGT 16020

GCCAGCCATT TCCACATAGT AGTCACGCGG GATGCTAGTG ATGGTAGTTT TTTGTAATTG 16080

ATTATTCACT GTAACCACCA TTAAGGCGTT GGAGCGGTAA TCGGTTTTGC GTTTTGAGTC 16140

ATTCGCAATC CCCATAATTA AGAAACTAAC CGGTTCCCCT TTGGCTAAAT CGGGACTCTT 16200

CGTTGAAGAT TTTCTTGGTT TATACATGGT GTTCGCCGTT TTTTGCAAAT TGCGATACAT 16260

GCTGACGCCG TAAGCCGAAA CAGAAAGTGT ACCTAAAAAT AAAAGGGTTA AACTGAAGAC 16320

CACGAATAAA CTTTTTCTTT TACGAGAATG GCGGCGCCGG CGTTGTGTCG ATACTTCTGC 16380

TTCCTCTGGT TCAGGTGCCG GCAAATAACG TTGGACAATT TGAAATGAAA AGTTGATCCA 16440

GCTATAACTT AGGAGCATCC CAGCAATGAC ATCACTAGGA TAATGAACGC GTAAGATTAA 16500

GCGACCTAAC ACAATCAAAA CTAAATAAGT GAGTAAGAAA ATTTTGATGC CTTTGGTTTT 16560

TTCGACATGA AGATAGGTCA AAATGATCAA AATCGTACAG ACTAGCGCCA TTGCTAATAA 16620

TGAATGTCCG CTGGGAAAGC TAGCAGAGCT CCGGTCAGCT AAAGCGCCTA GGTTTGGTCG 16680

AGCCCGACCC ACCACGTGTT TAAAAACGCT CCCTAAAAAG CCGACAGCTA ATAGGTTACT 16740

GACAGCCCAA AAGGCTAGAA TTTTTTCTTT TTTGCGTAAA AGGTACACCA TCAATGATAA 16800

GAAGAATAGC AGCATTGGTC CAATCGTGGC AGTTTTTGCT AACAGATTGA CCGCTGTTAA 16860

AAGTGGGGCA TTTAATTGCC ATTTGATGTG ATAGATAGCT TCGTCCATTT GTTGAAACCA 16920

ATTTGGACTA GCCGCCACTA ACAAAGTGAG GAAAGTAAAA CCAAGCGTTG TTAGGGCATA 16980

AAGTGGAAAT TTTTTATTCG ATGTTTGCAT GAGAATCCCT TTCTTTTTCC CTTGAAAGAG 17040

TAGACTAAGC CAGCATCGAT CGTAAATAAG CGCTTGCGAC TAAACAGGAC ACAGCAAGCC 17100

AATCGGTTCT CATCGTCTGA GCTTTAGCAC TATCCAACGT AAAGATTTCT CTTAGCTATT 17160

TTTAGAAGAG CCTTAATTCG ACCATTCCTG TTTTACACAT TGGTGTCTTT CGACATTTCT 17220

GCGCAATAGC CTATGTTTGG ATAGGAGCCT CACCTAACAA GGTTATTTTA TCTAACGCTA 17280

GAAGTATAGA ATTTATGCAG CGGTTCTCAC AATGAAGTTT AAGAAAATTT ATTGTTTTTT 17340

AAATAAATTT AAAGGTGTGG GAGAAAAGCG TTATTTCGGT AAATTTATTT TAAAAACGGA 17400

AAAACGTTTT ACAGTATCGG TCTTTAACGA TATAATAAAA GAATAGATGA AAAAACGAGG 17460

AGGTGAACAG CATGCTAGTT AATTTTACAG TCAAAAATCT TCTATCTTTT AAAGAGGAAC 17520

AAACACTTTC CCTACAAACA GGTGCGTATT TACGAAAGTA TGTCAATAAT CGTTTgcGAA 17580

GTGAAACCTA AAAATTACCC AGAAAAACTA AGGTTACTGA AGAGTGCCAT TGTTTTTGGG 17640

GGAAATGGTT CAGGAAAATC AAATTTAGTG AGTGGGTTAA ATGCATTAAA AACATTAGTT 17700

TTAAGAGAAC AAGCTTCGAT TAATGATGCT TTATTATATG AACCTTTTGC GTTAGACCCT 17760

CTCACTAAAC AAAGCCCATC TGTTATGACC ATTGAATTTA TTAATGACAG TGTCTTGTAT 17820

CGCTATTCGT TAGCTAACAC AGCGCGCGAA GTAGTTAAGG AATCTTTAGA GATTTATGAC 17880

AAAACGCAGG ATGATTTTGT TTATTATTTT AAAAGAGACG GTGCGGAATC TTCCGTTTTT 17940

CCAGAAAAAT ATCAAGAATA CCAAACACAA ATTAAGTCTA ATTCGTTAGT TATCCATACG 18000

TTGGAATCAA AAAATGATCA GCACGCTACG AATGTCGTTC GTTGGTTTGC CAATCAGTTA 18060

GTTATTTTTG ATGGGTCTTT ACGGAATATT GAACGGCTGA GCAATGAAAA AGACAAAGAG 18120

AAATTCTTGA ACTTTTTGAA ACTTGCGGAT ATGAACATGG TTGACCTGGT TCCTGTGAAA 18180

GACAAAGAGG AAGATTACTC GAAACAAGAA AAATTAGCCG CTGTTTTAGA ACGAATTATC 18240

CAAGAAAAAG ACCCGACCAT TGAAATTGGT CGAGTCAAGA AAAGTCTGTA TTCGCTGTAC 18300

TCTGTTTATA ACCAATATGA TGAAATGAAT CAAGTGGTCG GACAAGAACG TATTCATTAT 18360

GATTTGGAAT CTTCTGGGAC GAAAAAATTA ATTGGCTTAG CGCTAAATAT TTTATTTAAT 18420

CCAGAAGAAA AAGTTTTTGT TTTTGATGAA TTTGATGATG CATTTCATCA AGAGTTGTCA 18480

GGAACATTGT TAGAAGCGTT CAATGCCATG GaGACGAATA CACAATTTAT CTTGACCTCC 18540

CATGAACTTC ATTTGATGGa TAATCAATTG AGAAAGGATC AAATTTACTT CACAGATAAG 18600

AATTATCGTG GCGCCAgcGA ATTATACGCT TTGTATGACT TTGAAGCAGA CCCAAAAAAA 18660

GGTCGCGGAG ATATCACATA TTATCGCCGT TATTTAAATG GTCTGTTCGG TGGTGTTCCT 18720

CATATTGATC GAAGTGAAAT CGTTCAGGCA GTGAAGGTAA AGGAGTAGAA GTTGGTATGC 18780

CAAAAATAAA AAGGGTAAAA AAAGGaAAGA AwTTAAAACC AAtGTCATGA TTTTTTGCGA 18840

gGAGaAACAG AAGAaGCTTA CCTTAGATTG ATTwAACGAA AaTATTCTGC AGTCmATATT 18900

wAAAGTAAaC TCmAAGTaAA cmTGtAGaCG TCmAGGrGgG CGTTGGTCGA ATACGCCGTT 18960

TCCTGTATTC GGAGTATGGG GAAAAGCGGA CGnGGATAAT TATGATTTGT TTTATGTTAT 19020

GTATGACAAG G 19031

(2) INFORMATION FOR SEQ ID NO: 168:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1563 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

GAAAATTTAT CTTTGCCATT AATACGCAAA ATGATTTTTT GGATGATTCT CGTTTCTGCT 60

TTAATGGGGA TTGGGTTGTC ATTTATTGTC GGCTGGCCGT TATTAATggG TCTTTATTGT 120

TACTTAATTG GTATTTTTTA TTCATCAGGA CCGAAACCTT TATCTAGTCT ACCATTAGGG 180

GAAGTTTTCT CAGGATTTAC AATGGGGTTC ATGATTACCC TTATTTGTGT CTACATTAAT 240

ACGTTTGAAG TTTTCAATTG GTCAGCAGCT AATTTATGGG GGATTTTCTT AGTCGCCTTA 300

CCTAACACTT GTTACATTGC TAACTTAATG TTAGCAAATA ATATCTGTGA CTTAGAGGAA 360

GATGAAAATA ACAAGAGATA TACATTAGTT CATTACCTTG GCAAAGCAAG TTCTTTGAAA 420

TTATTCGTCG GATTAAATAC AATAGCAATG TTAGCCATTT TGTTAGCTGT CGGTTTAGGA 480

TTAACCCCGC CAACAATGTT ACTCATGCTT TTAACTTTAC CTTTTGtACG GAAACAAACC 540

CAAGCCTTAC TGAAGGAACA AGTGAAGAGT AAAACATTTG TTTGTGCTGT GAAGATCTTA 600

GCTGTGGGAG CAACTGCACA AGTGCTATTT TTTGCCATAG GTTTGTGGTG GCTGTAAATT 660

TTTTATAGAG TGGAAGGGAA TCAAAATGAA TAAGAAACAT GTTGTTATTT TAGGTGCTGG 720

TTACGCAGGC TTAAAAACAT TACGCGAATT ACAAAAAGGC GCCAAAGACG TCGAAATTAC 780

TTTAGTGGaT CGTAACAACT ATCACTACGA AGCAACGGAT TTGCATGAAG TTGCTGCTGG 840

AACACAACCA GCTGAAAAAA TCACATACAA TATTATGGAT GTTGTCGATG AAAAAATGAC 900

CACCTTTATC CAAGGAACAG TTAAAACAAT TGATGCTGCA ACACAAACAG TTGCTTTAGA 960

AGATGGCCAA ACAATCAATT ATGATTACTT AGTTGTTTCT CTTGGGTTTG AATCAGAATC 1020

ATTTGGTATC CCTGGTGTGC AAGAACATGC TCTACAAATG GTAGATGTGA AAACGGCTTT 1080

AAACGTGTAT GAACATATCC AAGAACAAAT GCGTCAATAT AAAGCAACTC AAAACGAAGA 1140

ATTCTTAAAA ATTGTTGTCT GCGGAgcAGG CTTTACAGGG ATTGAATTAT TAGGTTCTCT 1200

TGTTGAAAAC AAACCAAAAT TTGCTGAAAT CGCAGGTGTC TCAGCAGATC AAATCCAAAT 1260

TTATTGTGTC GAAGCAGTTA CTCGCCTATT GCCGATGTTT AACGAAAAAC TTGCGAACTA 1320

TGGTGTGCAA TTATTAAAAG ATTCAGCTAT TCATTTATTA TTAGGTAAAC CGATTAAAGA 1380

AACAAACCAG GCGCAGTTGT TTATCAAGAT AACGAAGCAG GCGACTTGGC AGAACTTTCT 1440

GCTAAAACAA TTATTTGGAC AACGGGTGTT AGTGGGAGCC ACGTTGTAGG CGATTCTGGC 1500

TTTGAAGCTC GTCGTGGTCG TGTCATGGTT CAACCAGATT TAACAGATGC GAACCATAGC 1560

AAC 1563

(2) INFORMATION FOR SEQ ID NO: 169:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3168 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

AAGCGTTGAC ACTCGTGATT TTATTCAAAA CAACTACACA GAGTACAAAG GCGACGACAG 60

TTTCTTAGAA CCAATCGCTC CGAGCACTGA CAAATTATGG ACAAAACTAC AAGAATTATT 120

TGAAATCCAA CATGAAAAAA ATGGCGTTTA TGACATGGAT AGCGATATCC CAGCAACTAT 180

TACGTCACAT GAACCTGGTT ACTTAATCAA AGAAGAAGAA AAAATTGTCG GTTTACAAAC 240

AGATGTACCA TTAAAACAAG CATTTATGCC ATTCGGTGGT ATTAAAATGG CGAACAATGC 300

CTTAGTTTCA AATGGTTATG AAACTGACGA AGAAATGACA AAAATCTTCA CAGAATACAG 360

AAAAACACAT AACCAAGGTG TTTTTGATGC CTACACAGCT GAAATGCGTT TAGCACGTAA 420

AAACAAAATC ATTACTGGTC TTCCAGATGC TTATGGCCGT GGCCGTATTA TCGGTGACTA 480

TCGTCGTATC GCATTATATG GTATCGACTA CTTAATGGAA CAAAAGAAAA AAGATCATGA 540

TAACACTGGT AACAAAGAAA TGACAGACGA TGTTATTCGC TTACGTGAAG AAATTTCTGA 600

ACAATACCGT GCGTTAAATG ACTTAAAACA AATGGCTGCT TCTTATGGTT TCGATATTTC 660

TCGTCCTGCT GCAAATGCGC AAGAAGCAAT CCAATGGTTA TACTTTGGTT ACCTAGGTGC 720

CATTAAATCT CAAAATGGTG CGGCAATGTC AATTGGTCGT ATCTCAGCAT TCTTAGATAT 780

CTATATCCAA CGTGACTTAG AAGCAGGTTT AATCACTGAA TTTGAAGCAC AAGAAATGAT 840

TGACCATTTA ATCATGAAAT TACGTATGGT TAAATTTGCA CGTACACCAG AATACAACCA 900

ATTATTCTCT GGTTATCCAA TTTGGGCAAC ATTATCTATT GCTGGTATGG GTATTGACGG 960

TCGTTCATTA GTTACCAAAA ATGACTTCCG TATCTTACAC ACCTTAACAA ACATGGGACC 1020

ATCTCCAGAA CCTAACTTAA CTGTATTGTA TTCTTCTCAC TTACCAGAAG GTTTCAGAAC 1080

ATATGCAGCT AAAATCGCGA AAGAAAGCTC ATCTATCCAA TTTGAAAACG ATGACTTATT 1140

ACGTGAAAAC TGGGGCTCTG ACGACTGTGC AATCGCATGT TGTGTTTCTG CAACAGTTAT 1200

GGGTAAAGAT ATGCAATTCT TCGGCGCTCG TGCGAACTTA GCGAAAGCTG TTCTTTATGC 1260

AATTAACGGC GGGGTCGACG AAAAAACTAA AATGCAAGTA GCACCAAAAT ATCGTCCAAT 1320

GACTGGTGAC AAATTAGATT ACCATGAATT CATGGaACGC TATAAAGACA TCTTAGACTG 1380

GTTAGCTGAA TTGTACGTTA ACACATTAAA CATTATTCAT TACATGCATG ATAAATATGC 1440

ATATGAAGCA CCACAATTAG CTTTAATGGA TACTGATTTA CAACGGACAT TCGCAACTGG 1500

TATTGCCGGa ATCTCTCATG CGACTGATAG TATCATGGCA ATTAAACACG GTGAAGTAGA 1560

AGTAATCCGT GACGAAGATG GTATGGCAAT TGATTATGTA CCAACAAAAG AATTCCCAAC 1620

ATACGGAAAC GATAACGAAG AAGCCGATGC AATGGCTAAC TGGATTTTAG ATTACTTTAT 1680

GACACAAATT AAACGTCAAC ACACATACCG TAACTCTAAA CCAACGACTT CATTATTAAC 1740

AATCACTTCA AACGTTGTTT ATGGTAAAGC AACAGGGAAC ACGCCTGACG GACGTCGTGC 1800

TGGCAAACCT CTTGCACCAG GTGCGAACCC AAGTTATCAA GACGGTAAAT TCTTAGGTGA 1860

GAAAAATGGC TTATTAGCTT CATTAAACTC AACAGCTCGT TTAGAATATA CAATCGCTTT 1920

AGATGGTATT TCTAATACAC AAACCATCAA CCCTAATGGT TTAGGTAAAG ACGACGACAC 1980

AAGAATTAAC AACTTACGTA ACGTATTAGA TGGCTACTTC GATAAAGGTG GTTACCACTT 2040

AAACGTAAAC GTATTTACAA ACGAATTATT ATTAGATGCA CAAGCACATC CAGAAAAATA 2100

TCCAAACTTA ACTATCCGTG TATCTGGATA TGCCGTGAAA TTCCGTGACT TAACTCCTGA 2160

ACAACAAGCA GACGTTATTT CAAGAACTTC ACACGACAGA CTATAAAAAA CATTTACATT 2220

ATTTAACTGA AAGGGTGACA TTTAACGTGT CATCCTTTTC GTTTGAAAGA AGAGGGATCA 2280

TTATGACAAC ACCAGTTACA GGTAGAATTC ATTCAACAGA AAATTTTGGT ACCGTTGATG 2340

GCCCAGGTGT CCGTTTTATC GTATTTACAC AAGGGTGTCG CATGAGATGT CAATTCTGTC 2400

ATAATCCAGA TACTTGGAAA ATTGGTTCCG GTGGTCGTGT GGTGACGACT GACGAAGTAT 2460

TAGAGGAAGC ATTACGTTTT CGCTCTTATT GGGGCGAAAA GGGTGGTATC ACTGTCAGTG 2520

GTGGTGAGCC GCTATTGCAA ATGGACTTTT TGATCGATTT ATTTAAAAAA GCCAAAGCAC 2580

AAGGCATTCA TACAACCTTA GATACGTGTG GCAAGCCTTT TACTCGGGAA GAACCCTTCA 2640

TTAGTCAATT TGATGAATTG ATGAAGTATA CAGATTTACT CTTATTTGAC ATCAAACATA 2700

TTGATAACGA GCAACATAAA CTGTTAACAA CGCAATCAAA TGATAATATT TTAGAGATGG 2760

CTACCTATTT ATCTGAAATT GACAAACCTG TTTGGATTCG CCACGTATTG GTGCCTCAAC 2820

GTAGTGATTA CGACGAATAT TTAATCCGTC TGGATGCCTT CATAAAAACA TTAAACAATG 2880

TAGATAAAGT TGAAGTGCTT CCTTACCATA CGATGGGCAA GTACAAATGG GAAGAACTAG 2940

GTATTCCTTA TCCATTAGAA GGAATTGAGC CTCCTAAAAA TGATCGTGTT GAAAATGCGA 3000

AAAAACTACT ACATGTCGAG GACTATCAAG GTTATCTAGC ACGTTAACAA ACGAAAAAAT 3060

AGAAAAGAAG ACAATTTTTT TGTTGTTTCT TTTCTATTTT TTGTTATCAT TAAGAGTGAA 3120

TCAAAACCAA GGAGTGTGTA TTATGTCAAA AATTTTAGTT TTTGGACA 3168

(2) INFORMATION FOR SEQ ID NO: 170:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1282 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

ACCATCTTAA TTTGACAATG AATGACACTG GTTGAGAAGT CGATTCCAAT ATGGAAGGCT 60

CTGTCAGCTA GTTTTTCACT CTACCAACAA AATAATCAAC CAAAATAACA ACGGAAAGCC 120

AAACAAGAAA CATACTAATC ATGCTTCTtC AAAGAAGCAC AAATAAAAAA GTGATGACTC 180

GAGAAGAAAA TCTCTATGGA GTCATCGCTT TTTCTTATAT TTTAGTTAAA AAATTATTTT 240

ACTAAAGGTT GAAGTAACTT GTTCCAGAAA CTAATAAGTG GGCCAGCACC AAATGCACAA 300

ATAATTGTAC CAATCCCAAT TGGACCACCA GCAATAAAGC CAATTAATAG GACCAACAAA 360

TCTTGTGCTT TTCGAACAGA AGCGTAAGGT TTGTTGGTTT GTTGACAGAT GGTTGGCGCA 420

ATCGCATCAT AGGGGGACTC GCCTTGATCT GTAGCCATGT ATAAGGCGCA gcCAAATGTA 480

AAAATCAGGA CTCCGACGAA TAAAGAAAGT AGTTTGCCAA TTGTTGAGCC TCCTAAACTA 540

TTGGCATAAG GTTGAAGCCA AGTAGAAATC TGTTGAATCA AAATTCCTAC TAAAAGCATA 600

TTTAAAATAG TTCCTAGACC AATGGCTTTT TTGTTTTTTA AATAGACATA GAATAAAATA 660

AACGCATTGA CAATAATTTG AAAAATGCCT AATTGCATAC CAAACATCGA AGAAAGTCCG 720

ATATTCATAG CAGTGAAAGG ATCAACTCCT AAACCTGCCA CTCTTAAAGT GGATGCGCCA 780

ATCCCGATAA TAACTAAACC AATTAAAATT AAAAGAACAT TTTTGACTAG TTTCATGGTA 840

AGCCTCCTGA ATAGATTTTG TGTGTGATAG TACAAAGTCT GCAACAAAAA TTCATAGAGA 900

ATTTTGTTGC AGACTACTTT TTAACCAGCT CCGTGTTGGA ACGAAGGATA TAGAGACATC 960

CCGCCATCTA CAAAGAGTGT TGTTCCGGTT ACGTATGAGG ATTCATCAGA AGCTAACCAA 1020

GCGGCCGCAG CAGCTACTTG ATTTGGTTTA CCAATAATAC CCATAGGAAC CATGCTTGTT 1080

GTTTGTTTTA ACTGTTCGGG ATCGGAAAAT TTTTCAGCAT TAATAGGTGT GTTAATCGCT 1140

CCAGGCGCAA TACAGTTGGC ACGAATACCA TATTCCGCAT ATTCCAAAGC GATTGTTTTC 1200

GTGAACATTT CGGTTGCGCC TTTACTTGCC GCATAGTGGG CGAACGTTGG CCAAGGAATC 1260

TTTGGTGTAC GGAGACATAT TA 1282

(2) INFORMATION FOR SEQ ID NO: 171:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11871 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

nnAAATAAAG AAATCATAAA ATAACTTTAA AAGTCCTCTT TTAAAGGAAA ATGTCTGTAA 60

AATAGAGAGA AATGGTTTTC TTTCTCGTCA AGTAATGATA AAATAAAACG GATAGATTGT 120

GCTTCTTGTA CAAAAAGGAT ACAAAAATAA AAAGAGGTAA TTGAAAGGAA CACAGATACC 180

AATGGCAAAT TTTTTGAAAA AAATGATCGA AAACGATAAA AAAGAATTAA GACGATTAGA 240

AAAAATCnCC GATAAAATTG ATGCACATgC AAGTGCAATG GAACAACTAA GCGATGAACA 300

ATTACGTGAA AAAACAGATG AATTTAAAGC CCGTTATCAA AAAGGGGAAA CCTTAGACGA 360

ACTATTACCA GAAGCGTTTG CGGTTGTCCG TGAAGCTGCT AAACGAGTAC TAGGTTTATT 420

CCCGTACCGT GTCCAATTAA TGGGTGGGAT TGTGCTTCAT GACGGGAACA TTCCTGAAAT 480

GAGAACAGGT GAAGGGAAAA CCTTGACTGC CACAATGCCT GTATACTTGA ATGCACTATC 540

TGGCGAAGGA GTACACGTTG TTACAGTTAA TGAATACTTA GCAACCCGTG ACTCTAACGA 600

AATGGGAGAA TTATATAATT TCTTAGGCTT AAGCGTTGGC TTAAACATTA ACTCTAAATC 660

TTCTGATGAA AAACGTGAAG CGTATAACTG TGATATCACT TATAGTACAA ATAACGAATT 720

AGGATTTGAC TATTTGCGTG ATAACATGGT TGTTTATCGC AGCCAAATGG TTCAACGTCC 780

ATTAAACTAT GcAATTGTCG ATGAAGTGGA TTCAATCTTA ATTGATGAAG CAAGAACCCC 840

ATTAATTATT TCTGGACAAG CTGAAAAATC GACAGCACTT TATACGCGTG CCGATAATTT 900

TGTCAAACGC TTAAAAGAAG ATGAAGATTA CAAAATTGAT ATTCAGTCTA AAACAATTGG 960

TTTAACAGAA GCAGGTATTG AAAAAGCTGA ACAAACTTTT GGATTGGATA ACCTTTACGA 1020

TATTGAAAAT ACAGCATTAA CGCACCATTT AGATCAAGCG TTACGGGCAA ACTATATCAT 1080

GTTACTAGAC ATTGATTACG TGGTTCAAGA CAATAAAGTC TTAATTGTTG ACCAATTTAC 1140

TGGTCGTATC ATGGATGGTC GTCGTTATTC TGATGGCTTG CATCAAGCGA TTGAAGCAAA 1200

AGAAGGCGTC GAAATTGAAG ATGAAACAAA GACAATGGCG ACGATTACGT TCCAAAACTA 1260

CTTCCGGATG TACAAAAAAC TAGCCGGGAT GACTGGTACT GCTAAAACAG AAGAAGAAGA 1320

ATTCCGTGAA ATTTACAATA TTCAAGTCAT TCAAATCCCA ACCAATCGTC CAATTATCCG 1380

TGACGATCGT CCAGATTTAC TTTATCCAAC CTTAGAAAGT AAATTTAATG CAGTGGTGGA 1440

AGATATTAAA GAACGTTACC ACAAAGGGCA ACCAGTGTTG GTGGGGaCTG TTGCGGTAGA 1500

AaCGTCAGAA TTACTTTCTG ATAAATTAAA CGCAGCTAAA ATTCCTCATG AAGTATTAAA 1560

TGCGAAAAAT CACTTTAAAG AAGCCGAAAT CATTATGAAT GCCGGTCAAA AAGGCGCAGT 1620

TACAATTGCA ACAAATATGG CTGGTCGTGG AACGGATATT AAACTTGGTT TAGGTGTACT 1680

TGAACTAGGT GGTCTAGCAG TTATCGGGAC AGAACGTCAT GAGTCACGTC GTATTGATAA 1740

TCAGTTGCGT GGACGTGCAG GACGTCAAGG GGATCCAGGG GTTTCACAAT TTTATTTATC 1800

TTTAGAAGAT GACTTAATGA AACGCTTTGG TTCAGAACGA ATTAAAACAT TCTTAGAACG 1860

AATGAATGTG CAAGAAGAAG ATGCAGTTAT TCAAAGTAAA ATGTTTACTC GCCAAGTAGA 1920

GTCTGCTCAA AAACGTGTGG AAGGAAATAA CTATGACACA CGTAAAAATG TCTTACAATA 1980

CGATGACGTG ATGCGCGAGC AACGTGAAGT TATTTATGCG CAACGACAAG AAGTAATTAT 2040

GGAAGAAAAT GACTTGTCTG ATGTATTGAT GGGCATGGTA AAACGGACGA TTGGCCGAGT 2100

GGTGGATAGC CATACACAAT TAGAAAAAGA AGAATGGAAT TTAGATGGGA TTGTCGATTT 2160

```
TGCAGCCTCA ACACTTGTCC ATGAAGATAC GATTTCTAAA AAAGATTTAG AAAATAAATC    2220
AGCTGAAGAA ATCAAAGACT ACTTAGTGGC ACGTGCGCAA GAAGTATTTG AAGAAAAATC    2280
TCAACAATTG AATGGCCAAG AACAATTGCT AGAATTTGAA AAAGTAGTCA TCTTGCGTGT    2340
TGTCGATACT AAATGGACAG ATCACATTGA TGCAATGGAT CAATTACGCC AATCTGTCGG    2400
GTTACGTGCG TACGGTCAAA ACAACCCATT AGTAGAATAT CAAACAGAAG GTTATTCTAT    2460
GTATAACAAT ATGGTGGGTT CAATCGAATA TGAAGTAACA CGTTTATTCa TGAAATCGGA    2520
AATTCGCCAA AATGTCCAAC GTGAGCAAGT AGCTCAAGGA CAAGCGGAAC ATCCAGAAAC    2580
AGAACAAGAT GCAGCTGCAC AAAGCAACAC AAGTGCGAAg CGTcAACCTG TGCGTGTAGA    2640
TAAAAAAGTG GGCCGTAATG ATTTATGCCC ATGCGGAAGT GGCAAAAAAT TCAAAAATTG    2700
TCATGGAAGA AACGCTTAAT TAAAGAAAAA ACGCAGGGTG GGGCAATTTG TCCCGCCTTT    2760
GCTCTTTTTA CAATAAAAAG AAAGAAGGAA TCATTTTGGA AAACCCAGAA ATTCGTACAT    2820
TATTAGACGA AATGACACAA AAAATTACTA GCTTCAGGAG GTCTCTTTGA CTTAGACCAG    2880
TTAGAAGAAG ACATTGCCGA GGCGGAAAAT CGGATGGGAG AACCTGGTTT TTGGGACAAC    2940
ACAGAAACGG CTCAAAAGCT AATTAATGAA ACCAATAGTT TGAAAGAAAA ATATCAACAA    3000
TTTCATCAGC TCTCTGAGGA AATTGACGAA TTAACGGTGA TGTTTGAATT ACAACAAGAA    3060
GAATATGATG CAGAGATTCA AGAAGAATTG GAAGAAAGAA TTCATCTCTT GCAAGAACGT    3120
TTAACAACGT ATGAGCTATC CTTACTACTA AATGAGCCAT ATGATCATAA TAACGCCCTA    3180
ATTGAATTAC ACCCTGGTGC CGGTGGTACT GAATCACAAG ATTGGGGAAG TATGCTTTTA    3240
CGAATGTACA CACGTTGGGC AGAAAGTCAC GGTTTCCAAG TTGAAACGTT GGATTACCAA    3300
GCTGGAGATG AAGCAGGGAT TAAAAGTGTC ACCTTATTAA TCAAAGGTTA CAATGCTTAT    3360
GGTTACCTAA AATCTGAAAA AGGTGTCCAT CGTCTTGTGC GGATTTCGCC CTTTGATTCC    3420
GCTAAACGCC GTCATACTTC GTTTTGTTCA GTTGATGTAA TGCCTGAATT AGACGACACA    3480
ATAGAAATTG CGATCAATCC AGATGACTTA AAAATTGATA CCTATCGAGC AAGTGGTGCC    3540
GGTGGTCAGC ATATTAATAA AACAGAATCG GCTGTTCGAA TCACACATAT TCCAACAGGT    3600
ACCGTTGTCG CAAGCCAAGC ACAACGTTCT CAATTAAAAA ACCGTGAGCA AGCAATGAGT    3660
ATGTTAAAAG CCAAATTATA TCAATTGGAA ATGGAAAAGA AAGCCCAAGA AGCAGCAAGT    3720
TTACGTGGTG AACAAATGGA AATTGGCTGG GGCTCACAAA TCCGTTCTTA CGTGTTCCAT    3780
CCGTATTCAA TGGTCAAAGA CCATCGGACA AATTACGAAA CCGGCAATGT ACAAGCTGTG    3840
ATGGATGGAG ACCTTGATGG TTTCATTGAT GCCTACTTAA AACAGCGCTT AAGTTAAGCT    3900
CTTCATTAAA TTGAAATAAA AATGTAACTT AATGCAAGAA GTTTGCAATA TTGACATGTT    3960
ATAATGAGTA GGATTGAAGA AAAAAGAGAT GGAGAGAACG CCATGATTGA AATGAAAGAT    4020
GTAATGAAGA AATATTCGAA TGGTACAACA GCGATTCGGA ATATTTCAGT AGAGATTGAC    4080
CAAGGCGAGT TCGTCTATGT GGTTGGTCCC TCAGGTGCAG GGAAATCAAC CTTTATTAGA    4140
```

TTAATGTATC GTGAAGAAAA AGCAACAAAA GGTAATTTGA CAGTTGCAGG TCATGATTTA 4200

ATGAAAATAA AAAATAAAGA AGTTCCTTAT TTACGTCGTG AAATCGGCAT CGTTTTCCAA 4260

GACTATAAAC TATTACCAAA GAAAACTGTG TATGAGAACG TGGCCTATGC GATGCAAGTT 4320

ATCGGACGGC GTCCTCGCGA AATCAAAAAA CGCGTGATGG AAGTTTTAGA TTTAGTCGGT 4380

CTAAAACATA AGGTACGTGT TTTTCCAAGC GAACTTTCTG GTGGGGAACA GCAACGGGTA 4440

TCCATTGCAC GAGCAATCGT AAATACACCG AAAGTTTTAA TTGCCGATGA ACCAACTGGG 4500

AACTTAGATC CAGAGAATTC ATGGGAAATC ATGAAATTGT TGGACCGGAT TAATGCACAA 4560

GGAACAACAG TCGTGATGGC AACGCATAAT AGTACAATTG TAAATACGAT TCGTCACCGT 4620

GTTATTGCCA TTGAAAATGG TCGCATTATC CGTGACCAAG TGGAAGGAGA ATACGGCTAC 4680

GATGATTAGA ACATTCTTTT CTCATTTATT TGAAAGTATC AAAAGTTTAA AACGGAACGG 4740

TTGGATGACT GTTGCCTCAG CAAGTGCGGT TACAATTACA TTAGTTTTAG TTGGTATTTT 4800

TATGGGCGTT ATTTTTAATG CAACGAAACT TGCTGATGAT ATCGAAAAAA ACGTGACAGT 4860

GTCTGTCTTT GTCGATATTG GTACAAAACA AAATGAAATG AAAACATTAG AAAAACAACT 4920

TAAAGGATTA GACAATGTTG AAGATATTTC TTATTCAAAC AAAGATCAAC AATTGAAAAA 4980

AATCCAAGAA CAAATGGGGG ATGCCTGGAA TCTTTTCGAA GGCGACAGTA ACCCATTATA 5040

CGACGTGTAT TATGTCCGTG CAAAAACACC AGAAGACACG AAAGATATTG CGAAACAAGC 5100

AGCTAAATTC CCAAGTGTTT TTAAAGCCGA TTACGGTGGT GTAAACTCAG ACAAAATCTT 5160

TAAAATCGCT CAAACCGTTC GAACTTGGGG CTTAGCAGCT GCTGCATTAC TTTTATTCGT 5220

TGCAGTGTTC TTAATTTCCA ACACTATTCG GATTACGATT TTATCTCGTC AAAAAGAAAT 5280

CCAAATTATG CGTCTTGTAG GAGCCAAAAA CAGTTTCATT CGTTGGCCGT TCTTCTTAGA 5340

AGGTGcTTGG aTtGGGtTGA TTGGcGCAAT CGTGCCAGTC ATTATTAkGA CGTTAGGCTA 5400

CCATCAAGTT TATAACATGT TCAACCCGCA GTTATTGCGT TCGAATTATT CATTAATTCG 5460

TCCAGAAGAT TTTATCTGGA AAGTCAATCT ATTAATGATT GCAACAGGAA TGATTATTGG 5520

ATCACTAGGT TCAGTTATTT CAATGCGTCG TTTCTTGAAA ATTTAAACGA AAAAAAGCTC 5580

TCCATTTTGG AGAGCTTTTT TTGTTTTAGT TTTAAAATAC TTGTTATTTA ATTTAAGAGT 5640

TTTCTCAGTT GTACCAACAA ATTGATAACT TATAATTTTC AAGTCCTACA TATAATGGTA 5700

AAATAGAATG GATTGAAATT AATTGGAGGA ATAATGAATC GATGAAAAAA AGATTGCTAT 5760

TATTTATTGG TTTGGCAAGT ATACTTACTT TGACAGGATG TGCAAAATGG ATTGATCGTG 5820

GTGAATCCAT CACAGCGGTA GGCTCATCAG CTTTACAACC ATTAGTAGAG ACAGCGAGTG 5880

AGGAATATCA AAGCCAAAAT CCGGGAAGAT TTATTAATGT CCAAGGTGGC GGAAGCGGAA 5940

CAGGTCTGAG TCAAGTCCAA TCTGGCGCGG TAGACATTGG TAATTCTGAT TTATTTGCAG 6000

AAGAGAAAAA GGGCATCAAA GCGGAAGACT TAATTGATCA TAAAGTTGCT GTCGTTGGGA 6060

TTACACCAAT CGTTAACAAA AATGTCGGTG TCAAAGATAT CTCAATGGAA AATTTAAAGA 6120

AAATCTTTTT AGGTGAAGTA ACAAACTGGA AAGAACTTGG CGGGAAAGAC CAAAAAATTG 6180

TTATTTTGAA TAGAGCGGCC GGTAGTGGTA CGCGTGCGAC TTTTGAAAAG TGGGTCTTGG 6240

GAGATAAAAC AGCCATTCGT GCGCAAGAAC AAGATTCCAG CGGCATGGTT CGTTCCATTG 6300

TTTCTGATAC ACCAGGAGCG ATTAGTTATA CCGCATTTTC ATATGTTACT GATGAAGTAG 6360

CTACGTTAAG TATTGATGGT GTTCAGCCAA CAGATGAAAA TGTAATGAAC AATAAATGGA 6420

TTATTTGGTC TTATGAACAC ATGTACACTC GTAAAAATCC AAGTGATTTA ACCAAAGAGT 6480

TTTTAGACTT TATGTTGTCa GATGATATCC AAGAACGTGT GATTGGTCAA TTAGGGTATA 6540

TTCCTGTTTC GAAAATGGAA ATTGAACGGG ATTGGCAAGG AAATGTCATT AAATAATTCT 6600

CTTAACACAA AATTTACAAA CTTAACATAA AATTTACATG TTATTAAAAA CAAATTGATG 6660

TAAAAAAGGT ACACTATTTT TATCTGTAAA AGTTGGATTG TTTGAGGAGG AAATATTCTT 6720

GGAAGATGTA CAGAAAAAAT TATTAACAAA ATCGAAAAGA GCCAAGATGG AGCAACGAGG 6780

GAAATTTATT AGTTTCCTAT GTATTGCCTT AATTGTAGCA GTGGTTGTTT CAATTTTCTA 6840

TTTTGTTGCA AGTAAAGGAC TTGCTACATT CTTTAAAGAT AAGATTAATG TGTTTGATTT 6900

CTTATTTGGA ACAAACTGGA ACCCAAGTGA TATCGGTGCA GACGGTAAAC CCATGGTTGG 6960

GGCTTTACCG ATGATTGCGG GTTCCTTTAT TGTAACGTTT TTGTCAGCGA TTGTTGCGAC 7020

GCCATTTGCG ATTGGAGCGG CTGTTTTTAT GACAGAAATT TCACCTAAAA AAGGAGCAAA 7080

ATTTTTACAA CCAGTTATTG AATTACTTGT AGGGATTCCC TCTGTTGTTT ATGGATTTAT 7140

TGGTTTATCT GtCATTGTTC CTTTTGTCCG TTCAATCTTT GGCGGAACTG GCTTCGGTAT 7200

TTTAGCTGGT ACGTTTGTGC TATTTGTCAT GATTTTACCA ACGGTCACTA CGATGACCGT 7260

AGATGCATTA AAAGCCGTGC CTAGACATTA TCGTGAAGCG TCGTTGGCGC TTGGTGCTAC 7320

GCGTTGGCAA ACGATTTATA AAGTTGTCTT GCGGGCAGCT GTGCCTGGTA TTTTAACTGC 7380

GGTTGTCTTT GGTATGGCTC GTGCGTTTGG TGAAGCCTTA GCAATCCAAA TGGTAATCGG 7440

AAATGCAGCT GTGATGCCAA CTAGTTTGAC AACACCAGCT TCGACACTAA CAAGTATTTT 7500

GACAATGGGG ATTGGGAATA CAATTATGGG AACAGTTGAA AACAACGTTC TTTGGTCTTT 7560

AGCCTTGATT CTATTGTTGA TGTCTTTACT ATTCAACATT GTTATCCGAA TTATCGGTAA 7620

GAGAGGAGCC TTAAAATAAT GAATGCAAAA AAAGCAGATA AAATCGCAAC AGGTATCTTA 7680

TATGCCGTAT CTGGCGTCAT TGTTTTAATT TTGGCGGCCT TACTGATTTA CATTTTAGCT 7740

AGAGGAATTC CACATATTTC GTGGGAATTT TTAACACAAC CATCCAGAGC CTATCAAGTT 7800

GGTGGTGGGA TTGGTATTCA GCTGTTTAAC TCACTATATT TACTATTGAT TACAATGATT 7860

ATTAGTATTC CAATTTCATT AGGTGCTGGA ATTTACTTAT CGGAATATGC GAAAAAGAAT 7920

TGGTTAACTG ATTTAATTCG TACCTCGATT GAAATTTTAA GTTCATTGCC TTCCGTTGTT 7980

GTCGGGCTGT TCGGCTTCTT GATTTTCGTT GTACAAATCG GATATGGCTT TTCAATCTTA 8040

TCAGGTGCGT TAgcATTAAC ATTTTTTAAC TTACCATTAT TGACTAGAAA TGTTGAGGAA 8100

TCTTTGAAAG CCATTCATTA TACACAGCGT GAAGCGGGGT TAGCCTTAGG TCTTTCACGT 8160

TGGGAAACAG TTAAAAAAGT TATTTTGCCA GAAGCATTAC CAGGAATTTT GACAGGTGTT 8220

ATTTTGAGTT CTGGTCGGAT TTTCGGTGAA GCCGCTGCGT TAATTTATAC AGCAGGACAA 8280

AGCGCACCTG CCTTAGATTT CAGTAATTGG AATCCATTGA GCGTTTCAAG TCCAATTAGC 8340

ATTTTCCGTC AAGCGGAAAC GTTGGCTGTG CATATTTGGA AAATTAATAC CGAAGgAACG 8400

ATGCCAGATG GCGCTGCGGT CTCAGCAGGT GCCTCCGCTG TGTTAATTTT AGTAGTTTTA 8460

CTTTTCAATT TCGGCGCTCG TTTTATCGGC AACCGTCTTT ATAAGAGAAT GACATCAGCT 8520

TAATGTTAGG GGAATTAGCA TGAAAGAATA CAATTTAAAT GATACCCATT TACTTCAACT 8580

AGATAGCCAA AAAGACCCGA TTGCGTTACA CACGGAAGAC TTACATGTCT TTTACGGGGA 8640

CAATGAGGCT ATCAAAGGTG TTGATTTGCA ATTTGAAAAG AACAAAATAA CCGCCTTAAT 8700

TGGACCTTCT GGTTGCGGTA AATCCACGTA TTTACGTTCG CTAAATCGTA TGAACGATGG 8760

CATTGCCAAT TCACGTGTTA CTGGAAAAAT CATGTATAAA GATGTCGATG TCAATACGAA 8820

AGAAGTTGAC GTCTACGAAA TGAGAAAACG CATTGGCATG GTTTTTCAGC GGCCCAATCC 8880

ATTTAGTAAA TCTATCTATG AAAATATCAC ATTCGCATTG AAACAGCATG GTGAAAAGGA 8940

TAAGAAAAAA TTAGATGAAA TCGTTGAAAC AAGTTTGAAG CAAGCCGCAC TTTGGGATCA 9000

AGTGAAAGAT AACTTAAACA AAAGCGCCTT AGCTCTGTCT GGTGGGCAAC AACAACGTTT 9060

GTGTATCGCT CGGGCAATTG CGATGAAACC AGATATTTTA TTGTTAGATG AACCGGCAAG 9120

CGCGTTAGAT CCGATTTCAA CAGGAACGGT TGAAGAAACG TTAGTGAATT TAAAAGATGA 9180

TTACACGATC ATCATTGTGA CACATAACAT GCAACAAGCG GCGCGGATTA GTGACTATAC 9240

AGCTTTCTTC TATATGGGTA AAGTCATTGA ATACGACCAT ACAAGAAAAA TCTTTACACG 9300

ACCAAAAATT CAAGCAACGG AAGAcTACGT TTCAGGACAT TTTGGTTAGG AGGAGCAAGA 9360

TGGGCAAAGA AATTATTTCA TCAAAAGATT TACATTTATA TTACGGGAAA AAAGAAGCAT 9420

TGAAAGGCAT TGATTTAACT TTTAATCAAG GAGAATTAAC TGCCATGATT GGTCCTTCTG 9480

GTTGCGGGAA GTCAACGTAT TTACGTTGCT TGAATCGTAT GAATGATTTA ATCCCAGACG 9540

TAACTATTAC TGGAAGTGTC GTCTATAAAG GAAAAGATAT TTATGGACCG AAAACAGATA 9600

ACGTTGAATT ACGTAAAGAA ATTGGCATGG TTTTCCAACA ACCTAATCCA TTTCCGTTTT 9660

CAGTTTATGA AAATGTGATT TAaTGgCTTG CGCCTAAAAG GAGTCAAAGA CAAGCAAGTC 9720

TTAGATGAAG CAGTTGAAAC CAGTTTGAAA GCTGCTGCCG TTTGGGAGGA TGTCAAAGAT 9780

AAACTTCACA AAAGTGCACT TTCACTTTCT GGTGGACAAC AACAGCGGGT TTGTATCGCT 9840

CGTGTTTTAG CTGTGGAACC GGATATTATT TTATTAGATG AACCAACTAG CGCCTTGGAC 9900

CCAGTTTCTA GTGGTAAAAT TGAAAACATG CTGTTAACAT TAAAAGAGAA ATATACCATG 9960

ATTATGGTGA CACATAATAT GTCACAAGCA TCACGGATCT CTGATAAAAC GGCCTTCTTT 10020

TTACAAGGTG ATTTAATTGA ATTTAATGAC ACGAAAAAAG TATTTTTAAA TCCAAAAGAA 10080

AAACAAACAG AAGATTACAT TTCTGGTAAA TTTGGTTAAA AGGGGGAAAA AGAATGTTAC 10140

GGACACAATT TGAAGAAGAA CTATTGAATC TGCACAATCA ATTTTATGAA ATGGGTATGA 10200

TGGTTAGTAG TGCCGTCCAT AAATCCgTGA AGGCATATAT TAATCATGAT AAAAAATTGG 10260

CTCAAGAAGT GATTGATCGC GATATTGAAA TTAATGATAT GGAAGTCAAA CTTGAGAAAA 10320

AGAGTTTTGA AATGATTGCG TTACAACAAC CAGTGACAAC AGACTTACGT ATGATTATTA 10380

CTGTTATGAA AGCTAGTTCG GATTTAGAAA GAATGGCGGA CCATGCCGTT TCAATTGCTA 10440

AATCAACGAT TCGCTTAAAA GGAGAAACAC GTATTCCTGA AaTCGAAAAA GAAATTTCAG 10500

ATATGTCTGA TTATGTGAAA AAAATGGTGG ATAATGTTTT AATCGCCTAT GTTAAAACGG 10560

ATCAAAAAGA TGCACGAATG ATTGCACAAA TGGATGATCG TACTAATGAA TATTTCGATA 10620

GTATTTACAA aCATGCCGTA GAAGCnATGA AaGCTAaTCC aGAAcGGTCA TTAGCGGrAC 10680

TGrTTATcTA CATGTGGCAC AATACTTGGA AAGAATTGGC GACTATGTCA CAAATATTTG 10740

TGAGTGGATT GTTTACTTGG CAACTGGTAA AATTACCGAA TTGAATAGTA ATCGTGACGA 10800

ACAATTTTAA GAAAAACAAA TAAAAAAAAC TAGGAGACGG CCTTTTTTAG ACCAATTCCT 10860

AGTTTTTTTA GCGTAAAATG AACTGATTTT ATAAAACTAG TCCTTACTAA TGAGAAGAAA 10920

TCGTTCTTAA GTCCTATGAC AAATGATCAA AAACAAGGCA TAATAATGCA TGTAAGGAAA 10980

ACACTATTGA GTGAAAAATT TGCTAAGGAG GCAATATCCA TGAAAGAAAG AGAACGCGTA 11040

TTAGAATTAG TGAAAAAAGG TATTCTAACG TCAGAAGAAG CGTTAATTTT ATTAGAAAAT 11100

ATGGCAACTG AAAAAGATGA AAAACAAATC GAAAAAGCTG CTGAAAAAGT TGATACACAA 11160

AATATTGGAA CAACAAATAA AGAAGATCAA GTCGCAGATT TAATGAATGC ATTAGAAAAA 11220

GGCGAATCAG AAGGACCTAC TGTTGATTCG TTTGAAGAAA ATACACAGGA CAGTGCAGAA 11280

AAAGATCGTG AAAACTTAGA AAGAATTCTT GATGAGTTAG CAACAAAAGC CAATCGTGCT 11340

TCGGCTGAAT TGGACGAAGT CAACGCAGAA ATTGCCGGCA TCAAAGAAGA AATTAAAGAA 11400

GTCGCAGAAG AAATTGGAAC ATTAGATACA AAAGAAGAAT TAGATGCATT AACAGAAGAT 11460

GAACAAGTTC AACGAAAAGA CTTGCACGTT TTACTTGCAC AATTAGAAGA AAAATTAGCG 11520

ACTCAAAGTA CTGAAAAAAC AGCACTGGAA GAAGAACTAA AAAACATTCG CAAAGaACAA 11580

TGGAAgGTCA ATGGAtGtAC AAAAGAAAAA GTTtCTTCTC AATTCTCTGA AGAGTGGAAA 11640

GATCAAGCCA CAGACACCTT TAACCAAGTC GGCGgCAAAG TTGCCGAAGT TGGTGGTCAA 11700

GTGGGAGAAT TCTTGAAAAA AACATTTAAT TCTTTCAGTG ATACCATGAA TGATAATGTG 11760

GAATGGAAAG ACATTAAAAT GAAAGTTCCT GGTGTGGCAA CAACTAAGTT TGAACATGAG 11820

TTTAACTATC CAAATCCACA AGCAAGTTTA ATTGATGTCA AGGTAGCAAA T 11871

(2) INFORMATION FOR SEQ ID NO: 172:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8084 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

AATGGGTAAC AGATTTAAGA TGGAAGATGG TCTTTGAGCA GAGTAAGTGA GCGAACAGTG 60

AACTTACTAC GGGGCGCCCG TTACAGCCAA CGGTATGCGA AGAATCAAGT ACCGTATGAG 120

GTTATTATTG AGAAATAATA GCGAATTAAG GTGGTATCAC GAAATGACAA ACTTTCGTCC 180

TTTTTGCTGT AATAGCAAAA GGATGGAAGT TTTTTTGTTT TTATTAAGCA AATTCGTTTC 240

ATAAAAAGTG CGAAACCTTC AAAATAAGAA AGTAGAACAG AAAGAAGAGG AGAACAAAAT 300

GGCATTAATT GAATTACGTC ATGTAAAAAA AGAGTTTTCC GGTAAAGCGG GTAAAGTCAC 360

GGCTTTAAAA GATATTGATT TAACAGTTGA ATCAGGTGAT ATTTATGGCA TTATTGGcTA 420

TTCAGGTGCA GGAAAAAGTA CGTTGGTTCG CCTATTGAAT GGACTTGAAA CACCTACAGA 480

AGGTGAAGTA GAAATTCAAG GACaAGATAT TGCATTGTTG CCAAATAAAG AACTACGAAA 540

CTTCCGTAAA AAAATTGGGA TGATCTTTCA ACATTTTAAT TTATTATGGT CACGGACGGT 600

CCTAGAAAAT ATCATGCTCC CATTGGAAAT TGCTGGCGTT CCAAAACAAA ATCGGAAAAG 660

CCGCGCAGAA GAGTTAATCA AGTTGGTTGG TTTAGAAGGA CGGGAAACCG CCTACCCAAG 720

TCAACTCTCT GGTGGGCAAA AGCAACGAGT GGGGATTGCC CGTGCCTTAG CTAACAATCC 780

TGACATTTTG CTTTGTGATG AAGCAACGAG TGCTTTGGAT CCGCAGACAA CTGATGAAGT 840

GCTAGAATTA CTACTAAAAA TCAACCAAGA ATTAAATTTA ACGGTTGTAT TAATTACCCA 900

CGAAATGCAC GTGATTCGTA AAATTTGTAA TCGCGTAgcT GTGATGGAAT ATGGTGAAAT 960

TGTTGAAGAA GGTAAAGTGA TCGATATTTT CaAAAAGCCT CAAACAGAAA TTGCGAAACG 1020

CTTTATCCAA CAAGAAGCGG ATAAAAACAT TGAAGAAACG GAACTGGTTG TTGAAGAAAT 1080

GTTGGAACAA TATCCGAATG GAAAAATTGT TCGCCTGCTT TTCCACGGTG AACAAGCGAA 1140

ATTGCCAATT ATCTCGCATA TTGTCCAAGA ATATCAAGTA GAAGTTAGTA TTATTCAAGG 1200

GAACATTCAG CAAACAAAGC AGGGTGCAGT GGGTTCTCTT TATATTCAAT TATTAGGCGA 1260

AGAGCAAAAT ATTCTAGCAG CCATTGAAGG ATTACGTAAA CTTCGTGTAG AAACAGAGGT 1320

GATTGGAAAT GAATAATCGT GGCTTAGCAA GTTTCTTTGA TTTTTCCCAA ATTGATATAA 1380

GTAAACTACA AGAAAAAACG ATTGAAACAC TTTATATGAC AGGGATTTCA ATCATAGTGG 1440

TCGCCATTTT AGGGCTGATT TTAGGACTAC TACTTTTTGA AACGTCAGGC AGAAAAGGGA 1500

TTTTGAGCCG ATTACTGTAT TGGCTAGTGG CCATTTTTGT AAATATTTTC CGTTCGATTC 1560

CGTTTATTAT CTTAATCGTT CTGTTAATTC CTTTTACAAA AATTATTTTA GGGAGTATGT 1620

CTGGTGTCAA AGGCTCCTTA CCAGCATTGA TTATTTCCGC AGCACCTTTT TATGCTCGCA 1680

TGGTCGAAAT CGCTTTTCGT GAAGTCGATA AAGGCGTGAT TGAAGCAGCC AAAGCAATGG 1740

GGGCCAATCG TTTCCAAATA ATTTGGAAAG TTCTATTACC AGAAAGTCTG CCAGCGCTTG 1800

TTTCTGGCTT AACCGTTACA ACCATTTCTT TAGTGGGCTA CACTGCAATG GCAGCTGCTA 1860

TTGGAGCGGG TGGTTTAGGA AGTTTGGCGT ATCAAGATGG CTTCCAACGT GGTCAAAATA 1920

CAGTTACCTT AGTAGCAACG ATTTGTATCT TAATTATTGT CTTTGCGATT CAATGGCTTG 1980

GGGATACAGT CGCTAAAAAA ATAGATAAAC GATAAATGGA GGGAATAAGA TGAAAAAACG 2040

TACATTATGG TCAGTAATTA CTGTAGCAGT AGCTGTCTTA GTTTTAGGGG CTTGCGGCAA 2100

TAAAAAGAGT GATGACTCGG TCTTGAAAGT TGGAGCTTCA CCAGTTCCAC ATGCAGAGAT 2160

TTTAGAACAT GTAAAACCTT TATTAGAAAA AGAAGGCGTA AAATTAGAAG TGACGACTTA 2220

TACAGATTAC GTGCTACCTA ACAAGGCGTT GGAAAGTGGC GATATCGATG CCAACTATTT 2280

CCAAcATGTG CCGTTCTTTA ATGAAGCGGT TAAAGAAAAT GATTATGACT TTGTGAATGC 2340

AGGTGCGATT CATTTAGAAC CAGTTGGGCT TTACTCGAAA AAATACAAAT CGTTACAAGA 2400

AATTCCTGAT GGTTCAACGA TTTACGTTAG CTCTTCCGTT TCAGATTGGC CACGCGTATT 2460

AACTATCTTA GAAGATGCTG GTTTAATCAC GCTGAAAGAA GGGGTAGACC GGACAACTGC 2520

TACTTTCGAT GATATTGATA AAAATACTAA AAAGTTGAAA TTCAATCATG AAAGTGATCC 2580

AGCAATCATG ACCACTCTTT ATGACAATGA AGAAGGGGCT GCGGTTTTAA TTAACTCAAA 2640

CTTTGCCGTG GATCAAGGAT TAAATCCGAA AAAAGATGCG ATTGCCTTAG AAAAAGAAAG 2700

TTCACCTTAT GCCAATATTA TTGCGGTTCG TAAAGAAGAC GAAAACAACG AAAATGTAAA 2760

AAAATTAGTC AAAGTGTTAC GTAGCAAAGA AGTCCAAGAT TGGATTACGA AAAAATGGAA 2820

CGGCGCTATT GTTCCAGTCA ATGAATAAAG AGAAAAAGAC AATAGAATAA AAACTTGAGG 2880

TCGGGACAGG CGTGTTAAGC TTCAAGAAAT AAGAAGAAAG TTCTGAAAGA TGCTTTTTGT 2940

CTCTCAAGGA CGTTCAGCTT ATTTCCGAAG AAGCTGCTCC TGTCCCGTTG TTTATCTAAA 3000

ATTTTAGGGC TAGGCAGATA CATCATGTTG ATTTACTTGC CAAGTCTAAG TTTTTTTCTT 3060

TTAAAAAAGG ATTATAAATG ATAAAGTTAC AAGGAATTTC TTAATTATCA GGAAAATTAA 3120

GGGACTATGC ATAGGCATTT GCATTAGTTA TTGTTATAAT AATTAAGATA GATTTTTTGG 3180

AATGGAGGGA AAAAGATGTT ACGTTTTTTT CCGCAAAAAA ATAAATATGT AGAAGAAGCA 3240

AGCCAGTACG CCTTTGATAA AGAAGGCCAA ATTCCGCAGC ACATTGCCAT TATTATGGAC 3300

GGAAATGGCC GGTGGGCACA GAATCGAAGA CTGCCGAGAA TTGCTGGACA TAAAGAAGGC 3360

ATGGATACGG TCAAAAAAAT AACGAAGCAT GCTAGTCATT TAGGCGTTAA AGTTTTAACA 3420

CTTTATGCAT TTTCAACGGA AAATTGGAAG AGACCAACAG ATGAAGTCAA TTTCTTAATG 3480

CAATTGCCTG TTGATTTTTT TGATACATTC GTGCCTGAAT TAATTAAGGA AAATGTTAAA 3540

GTCAATGTAA TGGGCTACCA AGAGTTTCTA CCAAGCCATA CGCAAGATGC TGTTAAGCGC 3600

GCCATTGAGC AAACCAAGGA CAATACGGGC ATGGTTTTGA ACTTTGCTTT GAATTATGGG 3660

GCGCGTGCGG AGcTACTAAC TGCCATGAAA CAGATTGCTG CAGAGGTTTC AGAAAAAGCA 3720

TATACGGCAG ATGAGATTAC AGAAGAAACG ATCGCAGATC ACTTAATGAC TGGCTTTTTA 3780

```
CCCACGGAAt TCGTGATCCA GAATTGTTGA TTCGAACAAg TGGGGAAGAG AGAATTAGTA   3840
ATTTCTTGCT ATGGCAAATT GCTtATAGCG AGTTGTTTTT TACAAAAGcA TawGGCCtGA   3900
TTTTTCAGGA GACACTTTAG AAACCGCGAT TGCtTCtTTC CAAawTCGTA ATCGTCGTTT   3960
TGGTGGTCtA AAAGaAACTm CAGAmACAGA AGGGAGCGAT CCGCAGTGAG ACAACGTGTC   4020
ATCACAGCTA TCGTCGCATT AATCATCTTT ATTCCAATTA TCTGGTATGG TGGTTTTGCA   4080
ATTGAGATTG CCGCTGTTGC TTTGGCAGTA GTCGGCGTGT ACGAGTTATT CCGTATGAAA   4140
GGATTAACGT TATTAAGTTT TGAAGGCGTT TTATCAGCAT TAGGGGCCGT TTTTTTAGTT   4200
ATCCCGACAA TTAAATGGTT AAACTTTTTG CCGGAAATGC AAACAAATTT CATTTTATTT   4260
TATGTCACTG TAATGTTGTT GTTAGGTGCT TCGGTAATTT CTAAAAATAC ATATACGATT   4320
GAAGAAGCTG GCTTTCCAGT TATTGTCAGT TTATATGTAG GGATGGGGTT CCAGAACTTT   4380
ATTAGTGCCC GTTCCGCAGG TTTTGAAGTC TTGTTATTTG CTTTATTCAT TGTCTGGGCA   4440
ACGGATATTG GTGCATATCT ATTTGGACGC CGATTTGGTC GTCATAAATT GATGCCAGAT   4500
GTTTCACCCA ACAAGACAAT AGAAGGTGCA CTTGGTGGTA TTTTATCAGC CGTCGTTGTA   4560
GCAGCGTTGT TTTTAGTTTT CACAGCTAAT AAAGGGCTAT TTCCTTATCC AATGCCGGTG   4620
ATGCTCGTAT TGACGGTTCT GTTTTCTATA GTAGGGCAAT TCGGTGATTt GGTTGAATCT   4680
TCTATCAAAC GTCATTATGG CGTGAAAGAC TCTGGTAACA TCTTGCCAGG ACATGGTGGT   4740
ATTtTAGATC GTTtTGATAG TTTATTATTT GTTTTCCCAA TTATGCACTT ATGTGGATTA   4800
TTTTAAATTG ATTAACAAAA TTTAAAAAAG TATCGTCAAA TAAGCnAAAA TCCAGAAGAG   4860
AGAcTCTTCT GGAtTTTTGC yGTTTrTTGG GTTAGTTTAG GTACCTTGTG AGTTATTTAr   4920
TTTATAAGAG TAATTGTATA AAGATGGCTA AAGTTTCTAA AAAATTAAAT AGGTGATCGG   4980
ATGTTAAATA TCTGAACATT CCaGTGTTAT TCGATTGTTA AAAATTAAAT TATmACAmTC   5040
GAATAACACT GGaATccyTa TTTCmATaAA GCCaTTGACG TTTrGCATAG ATAAATTTAT   5100
ACTTAAAAGA AGAAAAGCAA AATAATGTAT TTTACATGCT TGTTCaATCG TGGTGAGAGT   5160
TATTCATAAA TTTAAGCAAG TAAAGAGAAG ATAATTTAGA GATTTGTGAA CGCGTTCTTC   5220
ATACAAAATT ATAGATTAGG ATTATATTTT TGATAATCTA ATGTAAAGGA TTAATTCAAT   5280
GAAAAAATTT TTAGAGAATA GTATACATCA AGCCATAAAA GCTTGTTTTT TCTTTTTAGG   5340
TAAACTGCCT AAAAAGAAGT TATTTATTTT TGAAAGTTTT CATGGAAAAC AGTATAGTGA   5400
TAATCCACGT GCAATTTTTG AGTACATAAG AGATAACTGT CCTGAATATC AATGTATTTG   5460
GGCAGTCAAA AAAGGATATG AGATGCCTTT TGTGGAAGAG AATGTTCCTT TCGTCAAACG   5520
ATTAAGTTGG AGATGGCTAT GGTTGATGCC AAGAGCTCAG TATTGGGTTT TTAACACACG   5580
GATGCCAGCC TGGATGTATA AGAATCAAAC AACAATCTAT ATACAGACGT GGCATGGGAC   5640
CCCTCTAAAG AAACTGGGCT TAGATATTGA AACAGTCAAA ATGCCAGAGC ATGAAACAGA   5700
GAGCTATAAA CTAGAAGTCA TTCAAGAAGC AAAGCGTTGG GATTATCTCA TCTCACCTAA   5760
```

TGCATACAGT ACAAGAGTGT TTCGACAAGC CTTTCATTAT GAAGGGGAAA TATTAGAAAT 5820

AGGTTATCCA AGAAATGATT TGTTAGTTTG TGAAAATAAG GACAAAATAG CTTTTTCTGT 5880

TCGCAATCAG CTAGGTCTTG CAGCTAACAA AAAAATTATT CTCTATGCGC CAACCTGGCG 5940

CGATGATGAG TCTATCCGAA AAGGAGCCTA TCGTTTTACC AATCATTTGC CAGTTGAACA 6000

ACTTTTAGAA ATTGATTCTT CTATTGTAAT ACTTACACGA TTACATTATT TAATTGCAGA 6060

AAGTTTTGAT ATAAGTGCAT TTGGTTCACG TGTAATCGAT GTTTCTTCTT ATCCAGATAT 6120

CAATCAACTT TATTTAATCG CTGACTTATT AATTACAGAT TATTCATCGG TTATGTTTGA 6180

TTTTGcATTA ACAAAAAAAC CGATGCTATT TTTTATGTAT GATAAGGaAA AGTATCAAAA 6240

TAGTACAAGG GGATTCTATT TTGACCCAAC AGCAATTTTG CCTGGGAAAA TTGTGACAGA 6300

TACGGATAGC CTAGTAACGT CTGTAAAAAA CGTATTGTTT TCTAAAACGG AACAGGAACA 6360

ACGTTTATAT GTGCAGTTTT TTCGTGAATT TTGCCTTGAA AATCAGGATG CCACTAGACA 6420

GCTGATTTCA GAAATTTTAA AGAGATAAAA AGCCGGAGGG TGTAGGTTGG AAAAACAAAA 6480

AGAAGTCCAA AAAGAATATG AAGAGTATCA ATATATGTTA GATACTGAAA CGTGGAAAAA 6540

GGAGTTAAGT CAATTTAAGC TGTCCTCTTC AGaAAAAATA TCAAAAAATC yTAAAGTCTC 6600

aATCATTATT GCTAATwACc AATAATGCca CCATATCTTA AAAAGATGAT GGATTyTCTT 6660

GTTCAACAAA CCATTGGGAT TGAGCAATTG CaAGTGATGT TTATTGATGA TTGTTCCACA 6720

GATCATTCGT TaAAAGTAAT AGAGCCATAT CTTGAAAGAT ATCCGAATAT TGAAATTTAT 6780

CAGCTGCTTG AAAACACTGG AGGAGCTCAC GGCCCAAGAA ACGTAGGGAT TGTAAATGCA 6840

CGTGGTGAGT ATAGTGTCTT CTTAGATGCA GATGATTGGT ATGATTTAAA TGCATTGAAA 6900

TATTTATCAG ACCTACTAGA TGACTCTCAC GATGATTTTG CTGTTTCTGG kCTAATACAG 6960

AGTACGGACG GCCACTTAAT GTTAAAAAGT AAGCCATATT ATGTGGATGG GTCATTTAAA 7020

AATCGGTCTA TCCAAGAATT GCCAGCTGAA TTTTATGGTT GGTTAGGACC GCAAGCTATC 7080

ATGTTAAGAA CTTCACTGAT ACACAATAAT AATTTACACT TTGTnAATCA ACGGGTGGCT 7140

GGATGATGTG CTGTTCTTTA TGAAGCCAGC GATTTTCAAA ACCATTACAC AAGGGGAAAG 7200

GkTaACGACT TATTTAAATC GAGATGCAGA TaATGAAAGT TTAAGCmAAT CAtTAATCGG 7260

ACATTaTGCT TyCTgGTTGC GTGCGTTAAG TTATaTCAAT CaACAGTTyC CAGATGATTk 7320

GTCtAAAGrA CgGATGTTAG CACGCCGTTT AGAATGGCTA ATTTATGACT TTTGTATTCG 7380

GAGGAATACT GGGTATCCAT TCAGTAAAAG GCGATTACAA GATTTTAAAG TACAGATGGA 7440

TCAGTATTTA GGTGAATTAA ATTTTGATCC TAGCCCGTAT TTTAGAAGTG ATGTGCGCCA 7500

AATCATTTGG ACATTTTTAC AAACTAACGA TATTGATGGT TTGTATTGGT TTATTCAATG 7560

TCAGTCCATT CGTTGGGTGT TAGTCAATCG CTTTGGTTAT GTTTCAAAAA CAGAGAGAAG 7620

CTATTATTAT CCCAGATTAT TGAAACGATT CCCAGCAGTT CGCATGAATG CTTATGCAGA 7680

AGCAACGCAA CGTGAAGGCA ATACGCTTTC TTTAAATGTG TATACACACC AAAAAATTAC 7740

AGGATTTGAA ATAAAAGAGC TGAAAAATCC TTTTGAAACA AGACAGAGGC TGGCGTTTAA 7800

AAAAATTAGT GAAACTAGTT ATCAGGTAGA GTTGCCAGAA AAATTTTCaA ATAAAGAAAG 7860

TAGATTTACG ATAGTTTTTG ATAATTATTT AGAAATTGGG GTTAAAGATT TTAGTAAGTT 7920

TTttGCTTAA ATTTtGTCCG TTtAAATCAG GATATTTTGA GTATACmACy AwAAAtGATt 7980

AGTGrGGGrA ATTkGTTTak GGAAAATGTG AAAGTGTCAG TCATTGTACC CGTCTATAAC 8040

GTAGAAACCT ATTTAGAAGA GGCATTAATG AGTTTGAAAn ACCA 8084

(2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12438 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

TTnCnCCCAA AATGGwTTGk CCGrAGtGCy TAmCTGGaGG AGrATGCyTC CTTtCyTATC 60

CGGAATTTTA TTtGGTCCAc CATGGACCTA TCGTGCTTCa CGTCGGGTGA CGGATAATTT 120

aGCgACTGGG GAAGAGGAAA AAAATAGGAA ACAGCACTGT CATAAGCATA GAGGAAGAAA 180

AGGCCATGCT GCCTATGCAG AACATCTTTA TCAAATAATT GATCATTCGA AATTAACCAT 240

TGAAACGCCA AAAGAATTCA CATGGTTCAT GCAGACGGCG ATGCATGCCG TTTTCTCAAC 300

AATTGTTGAT GGGTATCGTC AACAAAGAGA AAATACTGCC TATGATTCAA CTGAAGCAGT 360

CAAACAATTA AAAATGAAAT TAAGTTGGTT GAAAAATGGC GCTTATAAGT AGCAAAAAGG 420

AGGAGAAGTT AGTTGATTAA ACTAGTCAAA CGGATGTCAT TATGGTCGGT CTTTGCAGCG 480

GTTGTATTTA TGATCGTACA AATTATCGCA GATTTATATC TGCCCACATT AACATCTAAT 540

ATCATTAATG ACGGTGTCGC TAAAGGAGAT ATTAACTATA TTTGGAGTAC AGGAGTTGTC 600

ATGTTAGGCT TTTCTTTGAT TAGCATTGTC GCTTCTATTG GAAATACCTT TTTTGCCACC 660

AAAGAATCAC AAAGTTTAGG AAAAAAATTA CGGACTGATA TTTATCGAAA AGTTGAAAAT 720

TTTTCTAATA ATGAATTTGA TAAATTTGGG ACAGCTTCGC TAATTACACG AACAACCAAT 780

GATGTCAATC AGATTCAAAT GGTTATGCAA ATGTTTTTAC GTTTAATGAT TAACGCACCT 840

TTAACGTTAA TTGGTGCAAG TTTTATGGCG TATAACAAGG ACCCACAATT AACGAAAATT 900

TTTCTATATG TTTTACCAAT TATGGCTGTT TTAGTGGGAG GTATCATGTT TTTAGCGGTT 960

CCGTTATTTA AAAGCATGCA AAAGAAAACA GACCGTTTAA ATTTAGTTTT TCGTGAAGGA 1020

CTGACAGGAG TTCGTGTGAT TCGGGCCTTT GGCAAAGCAA ACTATGAAGA ACAACGTTTT 1080

GATGAAGCCA ATAAGGATTA TACACAAACG GCGATTAAAG TGAATACGAT TGTCGCTTTG 1140

ATGATTCCTT TAATGACACT CATTATGAGT GGTACCAATA TTGCAATCAC CTGGTTTGGC 1200

GGTCATTATA TTGCGGAAAT GCAACTAGAG GTAGGTAACT TAATTGCCTT TATGACTTAT 1260

GCCATGCAAA TTCTGATGAG TTTCATGATG TTATCCATGA TTTTTGTCAT GGTACCGCGG 1320

GCGCAAGCTT CTGCTGATCG GATTAATGAG GTGCTAAATA CTGATTCAGA AATCAAAGAT 1380

GTACCCAATC CAGAACTACT CTCTCTAAAA GGCGACAAAG CAACATTGGC GTTTGAACAT 1440

GTTAATTATC GTTACCAGCA TGCTGAAAAC TTAGCACTGG AAGATATAGA TTTTTCAGCA 1500

AAATCTGGGG AGACGGTCGC GATTATTGGT GGGACTGGTT CTGGGAAAAC AACGCTGGTT 1560

AATCTTTTAC CACGCTTTTA TGATGTTGAA TCTGGCAAGA TTTTGCTAAA TGGTAAAAAC 1620

ATTAAAGATA CGTCACAGCA TAATTTACGA GAAATGATTG GCTTTGTACC GCAAAAGGCT 1680

GTTTTATTTA CAGGAACGAT TCGTGAAAAT ATGCAATACG GCGCCCCAAA TGCGACCGAT 1740

GAAGAAATTT GGCAAGCCTT AGAAATTGCG CAGGCGAAGG CGTTTGTTTC AGAATTGGCA 1800

GAAGGTTTGG ATAGTCATGT CGAACAAGGC GGCGGTAATT TCTCTGGCGG ACAACGCCAG 1860

CGGTTGGCTA TTGCTCGTGC TTTAGTAAAA CCAGCAGATG TCTATGTTTT TGATGATTCC 1920

TTTTCTGCAT TGGATTTCAA AACAGATGCT AATTTAAGAA AAGCCTTAAA AGAGCAGATG 1980

ACGGATGCAA TTGTCGTTTT AGTAGCCCAA CGTGTAAGTA CGGTTATGGA AGCATCGACA 2040

ATTTTAGTAT TAGACGAAGG GAAATTAGTT GGTAAAGGAA CGCATGAAGA GTTATTAGCG 2100

AATAACCAAA CGTATCAAGA AATTGTACAT TCTCAATTGA GAGAGGAGGA CCTTGCATGA 2160

GTAGTCGACC AAAAGGCGGA CCAATGGGCG GTGGACCAGG ACGCAACATT GGCGCAAAAG 2220

GGCCAAAACC AAAAAATTTC TGGAAAACAG TGAAACGTTT ATTTCGATAC ATGTCTAAAC 2280

GAATGCTGTC AATTATTGCT GTCTTAGTAT TAGCAATTGC AGCCGTCGTT TTCCAAATTC 2340

AAACACCAAA AGTATTAGGA CAGGCAACAA CTGAAATTTT TAAAGGCGTC ATGAAAGGCG 2400

CTGCTGAAAT GAAACAAGGC TTAAAAATAA CATCTTTTCC AATTGATTTT GACAAAATTG 2460

GTCAAATTTT ATTAATTGTT ATTGCCATGT ACCTTATTTC TGCTGTCTTT AATTTCTTGC 2520

AGCAAGTAAT TATGACACGT GTTTCACAGC GAACAGTTTA TGAATTACGT CAAGAATTAG 2580

AAGCGAAAAT GAACAAAGTT CCAATTTCCT ATTATGATAC TCATAGTAAT GGGGACATTA 2640

TGTCACGGGC AATTAACGAC ATGGATAATA TTGCGAGTAC TTTACAGCAA AACTTAACCC 2700

AATTAATTAC CAGTATCGTG ACATTTGTGG GTGTACTTTG GATGATGCTG ACCATTAGCT 2760

GGCAGTTAAC ACTAATTGCC TTAGCAACAG TGCCATTAAG TTTAATTGTT GTGATGGTTG 2820

TGGCGCCTCG CTCGCAAAAA CATTTTGCGG CTCAACAAAA AAGCTTAGGA TTATTGAATA 2880

ACCAAGTCGA AGAAACtACG GTGGTCACGT AGTAGTAAAA AGTTTCAACC ACGAAGAAAG 2940

TGATCAAGAA GTATTTGAAA AAGAAAATGA AAAATTATAT CATGCTGGTC GCAAAGCACA 3000

ATTTATTTCA GCGATCATTA TGCCTTTAAT GAACTTTATC AAAAATCTAG GCTACGTGTT 3060

TGTTGCAGTC CTTGGTGGCG TAAAAGTAGC GAATGGTATG ATGGATTTAG GGGATGTCCA 3120

AGCATTTCTT CAATATACCA ATCAATTTTC ACAACCGATT ACTCAAATCG CTAATTTAAT 3180

GAATACAATC CAAGCTACGG TGGCTTCGGC AGAACGTGTC TTTGAAGTAT TAGATGAAGA 3240

AGAAATGGTG GATGAACCTT CTGGCATACC AGTGGAAACA GATAGTCCTT ATCGTGTTTC 3300

```
TTTTGAACAT GTTGCTTTTG GCTATTCACC AGAGAAATTA TTAATGAAAG ATTTCAATTT    3360

AAATGTTAAA CCTGGGGAAA TGGTCGCAAT CGTGGGCCCA ACAGGTGCTG GGAAAACAAC    3420

CCTAATTAAC TTATTGGAAC GCTTCTATGA TATTAGTAGC GGCAGCATCA AATATGATGG    3480

CGTAGATACG CGCGATTTAT CTCGCGAAGA GTTGCGAGCA CACTTTTCAA TGGTTCTTCA    3540

AGATACTTGG TTGTTCACTG GAAGTATCTA TGACAATATT CATTATGGTA ATGAGCAAGC    3600

GAGCGAAGAA GAAGTGATCC GTGCGGCGAA AGCAGCCCAT GTGGATGATT TTGTCAGAAA    3660

ATTACCAGAA GGCTATCAAA CGATTCTAAA TGAAGAAGCC AGCAATATTT CTCAAGGTCA    3720

ACGACAATTA ATTACAATTG CTCGAGCATT CTTAGCAAAT CCAGACGTTT TaATTTTGGA    3780

TGAAGCCACC TCAAGTGTGG ACACTCGAAC AGAAATACTG ATTCAAGCAG CAATGAATCG    3840

TTTATTGGAA AATCGGACCA GCTTTGTAGT CGCTCATCGC TTGTCGACTA TCCGTGATGC    3900

CGATACGATT ATCGTTATGG CAGAAGGTTC AATTGTAGAA ACAGGAACCC ACGATGAATT    3960

AATGGCGAAA AACGGTTTTT ATGCTGACTT ATATAACAGT CAGTTTTCAG AAGAAGTAGC    4020

CTAAAAGGCG AAgcAAAAAA ATAGTAATTC ATGATGrGGC TTCAGGAACC ATAGTCGTAA    4080

AGACTGGTTC CTGAAGCCTT ATTTTTTTAA CAAAAAGTTA GCGAAATTGA AAAATTTCTA    4140

TTTTCTTTTG AGAAAAACAG GATATTACTT GATGAAATAG CAAGAGATGT ATAAAATGGG    4200

AGAGAAGAAA ATGGAGGTAA CAGAAAAAAT GACTGAATTT CAATTAGAAA AAGCTGCGAC    4260

ATTAAAAACA tTGCCATTCG TGCGAAAAAA GTTGGCGAAA TGGTCGACtT TTATAAAAAA    4320

GTAaTGGGTT TTGTTTTAAA GTCAGAAGAA AATAACCTGT CTATCTGGGG TACACGTGAA    4380

GCAGGAACGC AATTATTAAT TTTAGAAGAA ACTAGAAAAG CCGAAGATTT TCATAATGAA    4440

GAAAAGCAAA TGGCTTATTT CTCAATTAAA GTTCCTACGG AAAAAGAATT TTTACAAATT    4500

GCTCAACGAG TGCTTGAGCA AGACTACCCA ATTGATGAAA GTTTCCAAAT TGGCACACGA    4560

CAAAGCATGT TCATTACTGA TCTTGAAGGT AATCAATTTG AGATTTTTCA CGATGAAGCA    4620

ACAGCAAATT CGACTTCTGA AAAGCAACCA ATTGTGCTTA AAGATCTTAT CAGTGAGGAC    4680

TTAGAGCCAC ATCAAGGGTT AGCCGCTGGG TCATATTTAG CGCATGTGCA ATTAAAAACA    4740

AACAATCAAA AAGAAATAAA AGCATACTAT GAGGAAGTTC TGGGTTTAAA ACGCAATGAA    4800

AAAGATCAAT TTGTTTTAGA AGATGGCAAA GCAACTATTG GATTCCAAAA ACCTGAAACA    4860

AGTGAAGTGG ATCAATTACC TGATCCACAC TTAGGTTTAG ATTTCTTTAC CATTAAGCTC    4920

TCTGATCAAG AACATATTTT AGCAATGGAG CAACAATTAA CGGCAAAAAA TCAAGAGTTC    4980

TTTATTGACC AGAAAAAAGC GATTGTCACT GTGTTTGATC CAATTGGCTT GGAATGGTGG    5040

TTTGTCTTAA AATAAACGCT TAGACAGACA AAATAAGCTT GTTTTAGCAA GTCTGGAATA    5100

AAGTCGGAAT AAAAGTTTAA CGCTACTTTT GTTCGGAACT TATTCTAGAC TTTTCCTTTA    5160

ATATTAAAAA TAGAAGGTGA AAAAATGTAC GAAAAAACAA AGTGGAAAAT TCAAAAAATG    5220

ATGAAAAAAG GTGTTTTTTC GGGGGCTAGT TTTTGTTTCA TTGAAGGTGA AAAAGAGGAG    5280
```

AGCAATTGTT GGGGACTGGC ACAAGTAAAA CCCACCAAAG AACAATTAAC CTCGGCCATG 5340

CTGTTTGATG TGGCTTCTTT AACCAAAGTA GTCGGCACCA CCACCGTTAT TTTGCAATTA 5400

GTGGAAGAAG GAAAAATTAT TCTAGATCAA TCATTACAAA CTTATTATCC CTCTTTTCAA 5460

GATTCTAATA TTACAATCCG GCATTTGTTG ACACATACGG CAGATTTACA AGGCTATATT 5520

CCTAATCGTG ATCAGCTGAA TGCCCAAGAA TTAAAAGACG CCTACAATCA TTCATTTCAC 5580

GCAGGCAAAG CAATAGGTAA AAAAGTCGTC TATACAGATG CAGGGACTAT TTTATTAGGC 5640

TTTATGTTAG AAGAAATGTT CCAACAATCA ATGATAGAAA TTTTATCTGA GCGAGTATTG 5700

CTTCCTTTAG GGATGAACGA AAGTACGTTT CTACCAAAGa ATCCTTTAAA CTGTGTGCCA 5760

ACAGAATTAC ATGAACAAAG AGGGTTGATT CGAGGTGCAA CCCATGATCC TAAGGCATTC 5820

GTTTTACGGG AACATGCAGG GAATGCTGGC TTGTTTAGCA ATGTCTATGA TTTAACAAAA 5880

TTTGTTCGAA TGTATTTAAA TCGAGGCAGT TATCACAATC ACCAATTCTT GAAAAAAGAG 5940

ACAATTGATT TGTTACTAGT TAACCAAGTG CCAGCTGCAG ATAAACCACG TTCTCTAGGC 6000

TGGGATTTTA AATATGATGT AGcTACGCAA CGGCCATTAT TATTCCATAC AGGCTATACA 6060

GGCACTTTTT TGCTAATTGA TGTGCAACAA CAATCAGCGC TTATCTTTTT ATCTAATCGA 6120

GTCCATCCAG AGGACCACCG GAACACTTAT ATCGAAGAAA GAGATCAGTT GTTGGCAACC 6180

TATTTAAAAG AGAAATCATC AGTTTCAGAT GAAATGACTA GTTTTTAGTG ATATAATGAG 6240

AGCGAATTCA ATTTGGGAGG TCACAGTCAA TGCAAGAACC ATTATTTTTA CAACCTGTTT 6300

TTCAAGAAAA AATTTGGGGC GGCAATCGTC TACACACGCT ATTTGGTTTC GATTTACCGA 6360

GCGATAAAAT TGGTGAAGAT TGGGCAATCA GTGCACATCC ACATGGCGTT AGTACTGTTT 6420

TAAATGGCGA ATTTAAAGGA AAAAAATTAG ATGAATTGTG GGCAGACCAT CAAGAGTTAT 6480

TCGGCCATGC AGGTGGCGCT GTCTTTCCAT TATTAACGAA AATTTTAGAT GCTGAAGATG 6540

ATTTATCCGT GCAAGTTCAT CCTGATGATG CATATGGCCA AAAACATGAA GGCGAGCTAG 6600

GCAAAACTGA ATGTTGGTAC ATTATTGATG CAGAACCTGG CGCTGAAATC ATTTATGGAC 6660

ATCACGCGAA AACAAAAGAA GAACTTGCAG AAATGATTGA AGAAGGTCGT TGGGACAATC 6720

TTTTAAGAAA AGTACCAGTG AAAAAAGGCG ATTTTTATTA TGTACCAAGC GGCACTATTC 6780

ATGCCATCGG CTCGGGTATT TTAATTTTAG AAACGCAACA AAGCAGTGAT ACTACGTATC 6840

GGGTTTATGA TTATGATCGA ACAGATGATC AAGGAAAGAC CCGGGAATTG CATATTCAAC 6900

AATCCATTGA TGTGACAACC GTTCCGGCGA AGGATCCTGA CCTCTCTATT CAACAAGAAA 6960

ATCAAGGACA ATCAAGCATT GTTACTTATG TAAAAACAGA TTTCTTTAAT GTCTATGAGT 7020

GGCGTGTCAG TGGGATTCTA AAAGTCAAAA AACAAGCACC TTATACATTA GCAACAGTTA 7080

TTGAAGGTGT GGGTCGATTG ATTACAGAAG ATgCTgCAAA AGCAGATGTG GCTACTTTTG 7140

ATTTGAAAAA aGGCGATAgC TTTATCTTTC CGACAGATAT TCCGAGCTGG CGTTTTGAAG 7200

GTGATCTAAC GATTATAGCT TCAGAACCTG GCAGTCGCTA AAAAAAGAGG AGATCATCAA 7260

```
CAGATGATTT CCTCTTTTTT AGCATTAAAA CAGTTTATTT TGTTATAAAG GAACGTTTAG    7320
GTAAGGCAGC GATAACAGCT GTGACGATGA TGACTGATAG CAGACGATCC CCGTAATCTG    7380
TTATCAATTG AACCAACAGG ACGCTTGCTA ACTGATTGAA GCCTAAACCA TATAAAAATT    7440
GGACAATTAT GCTGGAGCCA GCAGAAGTAA TTCCACCAAA TAAACAAACT GTTATTAAGG    7500
AAGAAATCAC TGTTCCTGGT ATCGTTAACC ACAACGTTTT AAATGGTAAT TTGGTTACTT    7560
TTAACTGCTG ATGAAGGAGC AAACCTGCTA AAATTCCTGT GCAGAGTTGG ACAGGTGAAT    7620
AATACAGAGA AAAGATATCC GTTGTCAATC CAGTTAGTAG ACTTGTTAAA ATACCTGTCA    7680
GTGCGCCAAA GAGTGGGCCT AATAACGCAC CTGTAAAAAt AGTTCCCGCA CTATCAAGAT    7740
AAATTGGTAA ACGAAGGAGT AACGCAATGG TACTGCCAAG GTAATTTAAA GCAACACTAA    7800
AAGCGATAAT TGTAATGATT CGTGGTGTTA TTTTTTTCAT GGTTATCCCA TCCTTAAATT    7860
TTTGAAATCT CTTTCGATCT CATCTATTTG TACATTTAAA AGAATGGATA GAAACTTCAC    7920
GAAAAACTGT TGACTATCAA CCGTAGTCAA GACATGACAA TTAGCTGGTT GCTGCTGGAA    7980
ATGATATTGA TCGATCATTG ATTGCCCCAT AGCAATGCCG TCAGGAACAA TTTGTCCGTA    8040
ACTGGAAAAG CCTTGGCAGA GCGTAGGATC AATAAAATAA GCAACCGCTA AGGGATCGTT    8100
GATAACACAC CCTAAAATAC GTTCATATTT CCAATGAAAA TCAAAATAAA ATTGAGTAAT    8160
CTCAGCGAGG TAATTCCCCA TTTTGGGGTT GAGTCGTTGA CAATAATTGA GAATTGTTGG    8220
TGTAAAAACA ATTTCTCGTG TGACATCTAA GCCGACCATC TCGATTTTTG CACCTAATTT    8280
CTCAAAAACC CAAGCTGCTG CATCAGGATC ACACCAATAA TTAAATTCTG CAACAGGAGA    8340
GCAATTGCCG TGGCTTTTAA AGGTACCACC CATTGAGACA AAACGTTGAA GATTTCGCCC    8400
CAGAAGAGGT TCTTTTTTTA AACCCAAAGC AATATTTGTT AAAGGACCAA GGGCAATGAT    8460
TGAGCAATCG GTAGCTTGTT GAAAGGTTTG TGCTAAATAG TCAGCAGCGG AAAGCGCTTG    8520
TGCTTGGATT TTGCTAGTTC TGGGGAAATA TGTTTCACCT AAACCATCCA TGCCATGTGT    8580
ATCTTGTGCA CTAACAAATG GCCGTTTTAA AGGGCTAGCA GCCCCTTGGT AGACAGGAAT    8640
ATCTAAGCGA TTAACTTTTT CCAAACATTT AAAAACATTT TCCACGCCAA TCTCCACAGG    8700
GACATTTCCA CAAACAACGG TAATCGCTTG GACATCTAGC TCTGGTGAAG CAAGTGCATA    8760
CAAGAGAGCA AGCGTGTCAT CAATTCCAGG ATCACAATCA ATAATTACTT TTTGCATTTT    8820
TTCCAATCCT CCTTTTTCAT TAACTGCAAA TGATAACTCT TGTTAAAAAA GCTATGACTT    8880
GAGTTAGTAT TGAGAGATAC TTAGTTTTTT ATTCTGGGAA GTTTCCGAAC CAGTCCGAGA    8940
AATTCTCGAT TTTAAATGCA CCTCTTTATC TTACTATATT TTGTAAAGAT GTAAATAGTT    9000
GTCGGATAAA TGAAAAAGAA TCTAGCCCTA TATGTAGGAA CTAGATTCTT CTTTTATTTA    9060
TAAAGCGTTT CAGGAATAGC GATCGGTTCA CCATATTTTT TCTCTAAAGC GGTCCGAATT    9120
AACCGAATCC CTAGATTTTT GTTGCGTGCT AAGATTGGAC CGTGGAAGTA GGAGCCAAAC    9180
ACATTTTTAT AGACAACGCC TTCGCCATTA TCTTGACCGT TATTGCCTTT GCCTTGTTTG    9240
```

ACTGTTCCTA AAGGTCGTTC CCCTTCGCCT AAAAAAGTCA TGCCGTTATG ATTCTCAAAA 9300

CCATAATACG TTTCATCAAA TTCTTCATTA TAAATTTCAA TATCACCAAT GAAACGATTG 9360

TTGTCTTGGC TCAAAGTGTA ATGATCtAAT GCACCGATTC CTTTAATTTT TTCGCCTTGT 9420

GCGCCGACAT AGTAATGTCC TAGAAGTTGA TAACCACCAC AGATAGCTAA CAAGACGCCG 9480

TCATTTTCAA TATAAGAAGT CAAAGCTTCT TTTTTATTCT GAATGTCTTC TGAGATGATC 9540

AGTTGCTCGA AGTCTTGACC GCCACCAACA AAGACTAAAT CATATTTTTC GGGATCAAAC 9600

GGTTCGTAGA TACTAATAAT TTCTGATCGG AAGGTAACGC CCATTTTTTC GGTTAAATAT 9660

TTAAGCATCA ATAAATTTCC GTTATCTCCG TAAGTATTTA ATAAGTTGCC ATATAAATGA 9720

CAAACGACTA ATTCTTTAGA CACCATTCAT TCCCCCTTGA ATATACCCTT GGGCTGTTAA 9780

TGATTTTCTT AATTGCAAAA CAGCAGTATA GGTTGCCAAA ATATAGACAT GATCTGTCGG 9840

CAACGTTTTA ATTTCTGAAA TTACTTGCTC TAAATCAGCA ATTTCTAGTA ATTTTTCTTC 9900

AGGAATACCC GCAACTTTTA AACGCAAGGC CATATCTTCA TGGCGATCAC CGCCAGCAAT 9960

TACTTTTGGA ATGTCCATAT CAGCAAAGGC TTCATGATTA CCATCCCAAA TCCAACTGAC 10020

ATCAATTCCA TCTGCATAAT TAGCATTTAA TAATGAAACC AGCGAGAAAG AGTAAGGGGC 10080

TAAGCCCATC ATATCAATGA CCTGATTTAA ACCAACGGGA TTTTTTACTA AAACTAACGT 10140

ACATTCTTTA TCGCCAACTT TAATGGTTTC TTGACGCCCA AAAACTTTTT CATCATAAGC 10200

TAAGCCTGCT CGGATTTTAT CTGGTGCCAC CTGATAATAT TCTGCTACCG CTGTGGCCGC 10260

TAGCGCATTA TAAACATTAT ACATCCCACC AACAGCAATG CCGTATTCTT CGCCATCAAT 10320

CACAAAATCA GCAGATGTAT TATCCATCCG GACCATTTCC GTTAACTGCA CATCCAATTC 10380

AGGGCGTTTG AAGCCACAAT TTGGGCAATA GTAATCACCT AAATTAGCAT ACGTAATCAT 10440

TTTATAATGA AGAATATGAT TACATTTTGG ACATAAAACC CCATCTGTAT TGTAATGAGC 10500

CAATTGCTCT TTCGGTGGTA AATGATTAAA TCCATAGTAT TTGCGAGGAT TTACAGTTTC 10560

TTTTGAATTA AAAATTGGCG AATCACCATT GCATAAAATA GTTGCTTCAG GAGCAGCAGC 10620

GGCGCCTTCC ATAATTAAGC GATACGTCGT ATAAATTTCA CCATAGCGAT CCATTTGGTC 10680

GCGGAAAATA TTCGTAAAAA CAAATAATTT TGGTTGAATA TATTTAGTGA CACGGCTTAA 10740

ACTTGCTTCA TCAATTTCTA ACACCGCAAA TTTCTTTTGT CCTTTTTTCG CCTTGGCATT 10800

TAAAAAAGTA GAGACAATTC CTTGCTCCAT GTTAGCACCA GTTGGATTTG TTAATACCTC 10860

ATCAAATTCT TGGCGTAAAA TGTTAACAGT TAAAGCAGTT GTCAATGTTT TGCCGTTCGT 10920

TCCAGTGACT ACGACAATTT CATAGTCTTT TGCCAATGTA TCTAAAATTT TAGGATCGAT 10980

TTTTAGGGCT AATTTCCCTG GGTAGCTACT GCCGCCTTTA AAAAATGTTT GTAAAACCCA 11040

TTGCGAaGTT TTTCCTGCAG CAATTGCTAA ATGACTTCGA ATCCCCaTTT GTTGTCCCCC 11100

AATTTCTATA TCTTTAATTT TTCGTTATTT TAAATTAAAC TCAATCTATC ATACCATAAC 11160

TAAAATAAAA AATTAATTCC tTTTACTTTT TCTTACTATT ATTTAAAAAG TCTTTTAGAA 11220

AAAGTAGAAT AGTGAAGGAA AAAAGAATAG GGGCTGGGAC AAAAATCACT TTGGATTTTT 11280

GTCccAkGyw TAAAAAAcTG aTaAACGGCG GGAACAGAAG CAACACTCCG CGTTGCTTCG 11340

ATCTCAACAA CTTCTAAGGT GTTTTCTCGA CATGCGTCTC GTCACAGTAC AGCCTTCAAA 11400

AGAGCTACAG CTCTAAGAGT TGCGATGTTT CACTGTGrAA CATGAACAGA TGAAATCATC 11460

GTACCTACTT GGTGAAGGTC CTTCGGAAAT AAGCCGAAAT TGTTGAGAAT CACGAAGAAC 11520

GTAGTGrTTC GATGATGAAC AATACCTACT TGGTCTCCAA AAATTAAAGA ACACCGGGTC 11580

CTTCAATTCT AGCATGTTCC ATGCTTAGAA ATTGATGCGA ACGAAACGGA GTGTAGTGTA 11640

TCATGCAACA TCAACTCGTA CCTCGTTGTG TTTTATGACA ATTTTCGGAA ATTCCTTCTT 11700

ATTGCTGAGA AACACTATTT GCAAAGCAAA TAGATGTTGA ACAGTACCTA CGTGGTTCTT 11760

GGAGTTAAAC ACTTCTGTCC CGACCTCCAT CGATTTCCAT TTTTATTTCG CATCAATATA 11820

ACTGTATAGC GTATATTaGA AATATTGAAG AaGTTAGAAA CTTTGTCCCC AGATTTAGTG 11880

ACTAAGAAGG CGCCTTTATT GTTTAAAAAT TGAATACACT TGATTTTATC GTCCyTCGTC 11940

ATTAATGGGA CCGGCTTACC AACTAATTTG ACTGATTCAT GAATTAAATC GTCTAACAAC 12000

TCATTAACAT CCTGAGGAAT GTAATCAGGG TCTTTATCTt CTTCAGGTTC TGCAGTGGAA 12060

ATCAAAGATT TTAAGTTTGA CTCTGCAGCA ATCAATGTAG TGATGTCGTA GTTAATAGCA 12120

AAAATACCAT CGAGTTGACC ATTCTCATCT TTAAAATACA TTGTACTAGA TTTTAAAATA 12180

CGACCATCaT GGGTACGGGT TAGATAGTTA ATATGATCAT TCAACTCACT AGGGTCTTTT 12240

TTTAAGGCTT CTAAGACTAC TTGAGAAGGT CCATCGCCCA GATGACGAGA AGAGACATGA 12300

CaTTTTCAAT GGAAACAATA GAGTTGTTAA TATTGTCTTC ATTGATTTGG TGGATAACGA 12360

tTCGCAGTTT TCCCCAAATT GACCAGCTAA ACCTTTAGCA ATTTGTGTTA AAAAAGTAAG 12420

TTTTTCGTCG GATAACAT 12438

(2) INFORMATION FOR SEQ ID NO: 174:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3496 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

GATACATCTC ATCCCCTAAC AATGAACAGC CAATTGCTTC AAAATGTACC CGAATTTGAT 60

GTGTTCGTCC TGTATGGAGC TGAATGCGCA CAGCTGCTGC TwGGGCTCCT CTTTTTGCGA 120

GCCAATATTC TGTACTAGCA AATTGACCTG TTTCGATAAC TTGGCGTTTT AATAAAGAGG 180

ACAAGTCGCG CCCAATTGGC TTTTCAATCA ATTGATGCTC GCTTAACTGC TCAATCGCTC 240

CACCAACAAG GGCATAATAA ATTTTTGTCA CCTTTTTTTC TCGTAATTCT TGGTCAATTA 300

AAGCATGAGC AAAGCCGTGT TTCGCAAATA ACATCAGCCC AGAAGTATCT CGATCTAAAC 360

GAGTCACGAC ATGAATAACT TGATTCACGT AATTTTGACG CTTGTAATAG CCCTTAACAC 420

GATTCGCCAT CGTTCCTGAA GGATGATACT GAGCAGGTAT CGAGGCAATG CCAGCAGGCT 480

TATTAACTAC TAAAAAATGA TCATCTTCAA ACACAATTTC AATCGGTTCA TCATCTGTyA 540

ATAATGTCTC ATGTTCCGCT TCATCTGGAA TCGTCATTTT AACGACATCG CCAGTTTGTA 600

AAGGATAGAT GGCATTTTGA ACGGTCCCAT TCACTTGAAT TTTACCACCC TGAAATTTCA 660

CCTTTGCTAA TAACCCTTTT GAAACACCTT GTTCCTTTAA AAAATATTTC ACTTGTTGTG 720

TTGTTTCTTT GTCATATGTC CATTGAAATT CCATCGTTTA ATTAATCCTC ACCAATAAAG 780

GCATCTTTCA CTCGATGCCA AAAATGCATG TGCCGATAAG AGGCAAAATG AATCCGTTCG 840

TCTGCAATTC GATAGCAGAC AGAAGCGATA TTTTCTTGAT AAATATCTAA TTGGTCCACC 900

GTGACAAAAT AATCATCACT TTCCTGTAAT TTAATCTCCA GCCACTCCGT ATGTGCGATT 960

ACAATCGGTG ACCCCAATGT TCGAAAGACA CGATTGTTTA AGGAGGCGAT TTCGGTTAAC 1020

TGAAACGCAT TAATGCTAGG ATGCAAAACA GCCCCACCAA TACTTTTATT ATACGCAGTA 1080

GAGCCGGTTG GGGTCGAAAT CGACAACCCA TCACCACGAA ATCGTTCAAA CAGTTCATCT 1140

TTAATAAACA CATCGCCCAC CATCGTCCGG TTTCCTCGTT TAATTGTTGA TTCATTTAAT 1200

GCTAAAAAAT GTTTATCTGG TTTCCCATCT CTAAACCGAA TCCGCACATC TAATAATGGA 1260

TAACTCGTAC TTTTTTCACG ATGAATACAC AAACTCTCAA CTAGTTCTTT TAATTCATAA 1320

TCACGCCAAT CTGTATAGAA ACCCAAGTGT CCGGTATGTA CACCTAAGAA ACTTACTTCG 1380

TTTAACAGAT GATTAAATCG ATGAAAAGCG GAAAGAAGCG TACCATCTCC GCCGACTGAA 1440

ATCACAAGTT CAGGTTGACG ATTGTCTATT TGAATTTGGT TTTGCTCTAA TAGTGTTGTC 1500

AGTTGTTTTG TGACTTGTTT CGATTTTTCT TCCGAGTTAT GAACAATCGC CACTTTCATA 1560

AGCAGACTCC TTTACTGATT TTCTGGTGTT TCGTTTTTAA TTTGCTTATA GTATTCGTTT 1620

TCTAGCAATC CACGGCCATG AGAGAAGTAA TGTTGTGCTT CTTGAATTTC TTCACGAATA 1680

GATGACATTT CTTCATCTAA TAAATAGGCT GCCTCTGCCG CTCGCTGAAG GCGTTCTTTC 1740

ATTTCTTCAG GAAAGACACC TTGATATTTA TAATTCAAAG AATGTTCAAT CGTTGCCCAA 1800

AAATTCATGG AGAGTGTCCG AATTTGGATT TCGGCCAAAA TAATCTTTTC CCCGGTGACT 1860

AATTGCACAG GATATTCCAC TACCACATGG TACGAACGAT AGCCACTTTC TTTTTTATTT 1920

TCAATGTAGT CCCGTTCTTC CACAATTTTT AAGTCTTTAC GATTTCTTAA AAGTGCGACT 1980

ACCTGGTAAA TGTCTTCCAC AAATTGGCAC ATAATCCGCA AACCAGCAAT GTCTTGCATT 2040

TCTTCTTCAA TCCGATTTAA AGGAATATGC CGAATAGCCG TCTTTGTCAA AATACTATCT 2100

ACTGGTTTCA CACGTCCTGT AACGAATTCA ATTGGTATGT GTTTATCTTG CTCCCGAAAT 2160

TGTTTACGAA TACCTCGTAG TTTTACTTTT AATTCGCTGA CCGCCTGTTC ATATGGCGCT 2220

AAAAACAGCT CCCACTCTCG TTCCATGGTA GATCCCTCAA TTTCCTTGAA TTTCATTCAC 2280

TGATTATTTT ACCACAGTCC TCCCTAAAAT AGGGTGTTAT TCGCTATTTG ATTTCCAAAC 2340

GCCTTTCATT AGAAAAATTT TTCACAGAGA AGACCACTTT TCTTGACGGA AACCTATATA 2400

ATATGTGAAA ATTGGAGTGA TTTAAGAAAG AGAGAGAACC CCATGAGTGA GCACTTAGAA 2460

ATTGAATTTA AAACATTAGT TAGTCCGCAA GATTTCAAAC GATTAATCGA CCATTTTGCT 2520

ATTCAAAAAA CGGACTTTTT TACGCAAACA AATCATTATT TTGATACGGA CGATTTTCAA 2580

TTAAAAGCAC AACGAATGGG CTTAAGAATT CGCGTATTAG CAGATCGTGG TGAGTTAACC 2640

TTAAAAGTTC CTGCGCCAGA AGGATTATTA GAAATCAaTG ACCCTTTATC GCTTGAGACG 2700

GCGAACCACT TTATCAAGAG GAATCACCTC CCTACTGAGG GGGCTGTGGC GAAGAAATTA 2760

CAAGAATTAG GTATTGAAAT CGCCTCGATC CATTTAATTG GTTCATTAAA AACAGCACGT 2820

GCAGAAAAAC AAATTCCTCA AGGACTACTT GCGTTAGACG AAAGTTGGTA TAACCAACAG 2880

CACGATTTTG AATTAGAGTT AGAAGTTACT GAAGCTGAAA GCGGcAAGCA AGCGTTTCAA 2940

ACATTAATGG CTGATTTGAA CATTCCCATT ACGCTTGCAC CTAACAAGAT TCAACGCATG 3000

ATGCGGACCA CTGCCCCTTA GTATTCGAAG CACTGACGTT CTGGCACCTG TATTTGAAAA 3060

AGCATGGCAT TGGAGGAAGT ATGAATGGTT GAAATTTATT TGTTCGTCAA TCCTTTAGGC 3120

GGCGTCTGTT TAGAAATTGA AAAAGAAATT ATACAATTGT CTGTTAACGA CAAGAAGAAA 3180

ATACAATTGC GCTTTATTCC TTTATTAAAT ATGAAAACGA TCAATGAATT TTTGAGTCGT 3240

CAACATATTC CGATTAATGA CATTAAACGA CGGAATCGAA TTTTTGAAGA TTTATACTCA 3300

GCCGCTCTTG ACTATAAAGC AGCACAGTTA CAAGGCCGAA AAAAAGGCCG ACAATTGTTG 3360

ATTGGCTTAC AAAAAGCAGT AGCTGAAGAC GGTCTAGCTT ACTCTCCTGA GTTATCAGAA 3420

GAACTACTTC TAGCGGCTGG CGGGGATATT GATATGTATC GTAAAGACCG ACAATCCGAT 3480

TTnGTCAAAG AAnCTT 3496

(2) INFORMATION FOR SEQ ID NO: 175:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 902 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

TTTCACAAAA GAGCAAAACC GCATGTTTTT GCGGTTTAGT GATAATTTTC ACAAACAAGA 60

AAACGGTTTC TGTTCCACAC TTCTTCTGTA CTATTTTTCT CCCAACTGTT ATATTCTATT 120

TTGAAAATAT TATATAACAG TTCTATTAAT TTGAGCCCCC TTATCCCCGA ACTGGCTAGT 180

AATCTGATTT CCGAACTTTG CGTTTTCATG AAAATTATCT TGAAAAAACC AAAAAATAAA 240

AACCGTCAGC CGAGTAACCT ACTCAGCATG CGGTTTTTTA TTTTTTAGAG AACAAAGTTT 300

TTTTCGTTAA TTTTTATTTT TTTTGGCTCG TTTTTTTGCC AAATACTTTA CCTTAGGTTT 360

TATCACGAAC ATTCATTGCT TAAAATGACC AAATTTTCGT AAAACGAGTA ATACAGCTTA 420

GCAAATTCTT TTCTCTTTCA GTAGATTTTA CTAAATCCAT TCCTTGGTAA TTTTAATATA 480

TAAACGTTTA GCTACTTCAA CAGTTAACAT GTACGCAATG ATGATTCCAA AGAACCATGG 540

CCAGTAGTTA GCCGGAAGTT TCACAAAGTC AAAAACCTCT CTTATAGGCG TTGCTACAAT 600

CAAGAAACCG AGTAAGATTG CGCCTAAACT ACTTAGCATC ACAGGCATTG ACGCACGACT 660

TTGAATAAAT GGTATTTTCC GCGTCCGGAC CATATGAACA ACTAACGTTT GACTAACAAG 720

TCCCACCATA AACCAACCTG TTTGGAATAA ATGTTGAGTT GCCAAACTAT TGGCACTAAA 780

CACAAACCAC ATGACAAGAT ACGTTAGAAT ATCGAAAATA CTACTTACCG GGCCAATACA 840

TACCGTAAAT TTCGCTAGTC CGTTCGTTTC CCACCGAACA GGGCTTAATA ACTCTTCTTC 900

AT 902

(2) INFORMATION FOR SEQ ID NO: 176:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3568 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

TCTGGTGAAG TCCGTGTGAA TGAATCCTTG TTAACCGGGG AATCTGATGA GATTACCAAA 60

GTAGCAGATA GCCAACTTTT GTCAGGTAGC TTTATTGTGT CAGGTTCTTG TTATGCAAGA 120

ATTGATAAAA TTGGCGCTGA GGCCTATGTT CATCAATTAA CGCTTGAAGC GAAATCAATT 180

AAAAAAGGGG AACAATCAGA GATGGTCGGT TCCATTAACC GTTTGGTGAA ATGGGTAGGG 240

ATTATTATTA TTCCAATCGG CTGTGTTCTA TTTTTCCAAA GTTATTTCAT CAACCATCAA 300

GGGCTTCATG ACAGCATTGT CTCCATGGAA GCTGCCTTGA TTGGGATGAT TCCGGAAGGg 360

TTGTACCTTT TAACAACCGT TGCATTAGCT GCCAGCACAA TGCGCTTAGC GAACCnGGGC 420

GTTTTATTGC ACAACATGAA AAGTATCGAA TCCTTAGCAC GTGTCGATGT GTTATGTGTA 480

GATAAAACAG GAACAATTAC TGAAAACACG ATGGAAGTCC AAGAGTTGGT TCCTGTGGTG 540

AGTATCGAAA ATGATGGTAC AGACTTAACC AATGTCGAAA AATTAATTGG CGATTTTTGC 600

CGAACAATGT CAGCTGACAA CGATACGATG AAAGCGATGA AAGAATTTTT TGTTACTAAT 660

AATCAGAGGG AAGCCGTTAG TTATACCTCG TTTTCGTCTG TCGAAAAATT TAGTAGCGTG 720

ACTTTCCCGG AAGCGACGTA TATTCTTGGG GCGCCAGAGA TGATTTTGAG AGACAATTAT 780

GAAATGTATC AATCTGAAGT GGAACATTAT ACAAGTCAAG GCTATCGCTT ATTGGTGTTC 840

GGAAAGTATT TAGGAGAATT TCAGGAAACG TTGGCAGCAG AGGTTCAACC TCTAGGCTAC 900

ATTCTTCTTT GGAATCCGAT TCGTAAAGAA GCCAAAGCAA CTTTTGAATA TTTCGCTGAA 960

CAAAATGTGG CTATTAAGGT CATTTCTGGC GATAATCCTT TAACAGTATC CAATGTAGCA 1020

CAAGCGGCAG GCATTATTGG TGCTGAAAAT TATGTGGACG CGCGAACATT AACTACAATG 1080

GAAGCCCAGA CAGAgCTTTA GAAAAATACA CCGTTTTTGG TCGTGTCACG CCAGAACAAA 1140

AGAAACAATT TGTATTATTG TTGAAAAAAT TAGATCATAC AGTGGCGATG ACAGGGGATG 1200

GAGTCAATGA TATTTTAGCC ATGAAAGAAG CAGATTGTAG TATTGCCATG GCTTCAGGAA 1260

ATCAAGCTAC AGCACAAGCC TCGCAAGTTG TGTTGTTGGA TTCTGATTTT TCAACGATGC 1320

CGGAGgTTGT CTTTGAAGGC CGCCaAGTAG TTAATAACAT TGAACGCTCT TCAAGTTTAT 1380

TCCTTGTAAA AAATATTTTC TCCTTGCTGA TGTCTGTCTT TGCTATGATT TTTGCCGTGA 1440

CATATCCTTT ACAACCGACA CAAGTCACAT TGATTAGTTT ATTTACAATT GGGATTCCTT 1500

CGACATTTCT AGCTTTGGAA CCAAATCATC GTCGCATTGA AGGGAAATTT CTTTTCAATG 1560

TGTTATCAAA AGCAATTCCT GGTGGGTTAA CGGATATGTT AGTAGTCGGT GCCTTACTAA 1620

TTTGTGGTGA TATTCTTGCA CTTCAAAAAA CGGACATTTC AACGACTGCA ACATTGTTGC 1680

TAGTTTCAGT CGGCTTTATG GTTCTTTATA AAATTAGTTC ACCAATGAAT CGTTACCGCA 1740

AGCGCGTAAT GATTGGCTGT TTAATCGGAA TGGTTATTAC AAGTATTTCC ATGAAAAACT 1800

TGTTCTCGTT AACTTCGGTA TCGCCAACAG CCTTGTTACT TTTAGCCATT CTTTTCTTTG 1860

CAGCAGACTC AACTTTCCAA CATTTATCAA CCATTTCTGA AAAAGTTCAA TTATGGTTTT 1920

ATAAGAAAAG ACATTAGGAG CAAAACACCG AGCCAAAATT ATTTTGGTTC GGTGTTTTTT 1980

GTTTTTTATC GTTGGGAAAA ATAAAAACAC TTGCACCTAG TTAATCCTTG ATAGTATAAA 2040

CCTAACCAAT TATTTTAGAT AGGAAGGATT GaATAAAAAT GaAAAAAAGA AGTTTTGTCT 2100

ATTGGGGaAT GTTTTTCATc TTTTGTCGAC GCTGATAACG CCTTTTCTGA AACTTGAAAC 2160

AGGATATGCG CaAACGGraC CAACTAGTAC CAGTGrAACC AATCAAATTT nCTGtACGCC 2220

AAATGTAGTG CCGCGTAAAC AAGTGGGCAA CATTGTGACA GCGATTCAAC TAACGGACAA 2280

AGAGGGAAAC CCACTTGGGA CCATTAATCA ATATACAGAC ATCTACCTCC GAATTGAGTT 2340

CAACTTACCG GATAATACCG TTAATAGTGG GGATACGAGT GTGATTACTT TACCCGAAGA 2400

GCTACGTCTC GAGAAAAATA TGACCTTTAA TGTAGTAGAT GATACAGGTA CTGTGGTTGC 2460

AATCGCACAG ACCGATGTAG CAAATAAAAC GGTCACGTTA ACGTATACAG ATTACGTAGA 2520

AAACCATGCC AATATTAGTG GGTCTTTATA TTTTACAAGT TTGATTGATT TCGAAAATGT 2580

CGAAAATGAA TCCAAGATTC CTATTTACGT CACGGTTGAA GGGGAAAAAA TATTTGCTGG 2640

CGATCTTGAT TACCAAGGTG AAGGCGATGA TGTTAACGAA AAATTTAGCA AGTATTCGTG 2700

GTTTATCGAA GACGATCCAA CCGAAATTTA CAATGTACTA CGGATTAATC CAACAGGGCA 2760

AACGTATACT GATTTAGAAG TGGAAGATGT GCTAAAAACA GAAAGTTTAA GTTATATGAA 2820

AGACACTATG AAAATTGAAC GAGGACAATG GACACTGGAT GGAAaTGCTA TTTGGCAaTT 2880

CACGCCAGAG GAaGATATAA CTGATCAGTT GGCTGTCCAG TATGGTCCTG ATGACCGTAA 2940

TTTTTCGGTG CATTTTGGCA ATATTGGGAC TAACGAnTAT CGTATCACCT ATAAAACCAA 3000

AATCGATCAT CTTCCCGAAA AAGgTGAAAC CTTCACAAAT TATGCAAAGT TAACAGAAAA 3060

TCAAACGGTG GTGGAAGAAG TCGAAGTGAG TCGTGTCTCA CAAACAGGTG GTGGAGAAGC 3120

TAATGGTGAA CAATATGTTG TTGAAATTCA TAAAGAAGAT GAAGCAGGCC AGCGTTTAGC 3180

AGGTGCAGAA TTTGAACTCA TTCGAAACTC AACCAATCAA ACAGTGGCTA AAATCACAAC 3240

AGATCAAAAC GGAACAGCGA TTGTTAAAGG CTTGTTAAAA GATAACTATA CTTTAGTAGA 3300

AACCAAAGCA CTACCGGTTA CCAACTTTCG CAAAACAAGA TTCCAATAAC ACCTGAAGAC 3360

TTTGGTAAAA ACTTcAGTTG CACTGAAAAC TGTTGTGAAT CATAAGATTT CATATCAaCC 3420

AGTCGCTGCt TCATTTTTAG CTGGAAAAGT TCTTTTAGGA AAACCACTGA AGGATGCTGA 3480

ATTTCAATTT GAATTGTTAG ATGAGAAAGG CACAGTACTA GAAACGGTTA GTAACGATAC 3540

TTTAGGGAAA ATTCAATTTT CACCATTA 3568

(2) INFORMATION FOR SEQ ID NO: 177:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2206 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

CATGTTTATC TAAAAGTTTT GCCCAGAACA CTTCTACATT TTTTAACATC CTTTGCTCTC 60

CTTTTTATTT TTACTTCACG CAACCTTCTT TCACTAATTA TACATAACTG ATTGCTTTTC 120

TCCTATACTA TTTTTACAAA ATGAAAACTA AAATAATGTC AGTAGTATTT GGACAGAAAA 180

AAAAAGAGAG CGTGGGACAA AAATCACTTT GGATTTTTGC TCCACGCTCA AAAACTTATA 240

AACGGCGGGA ACAGAAGCAA ATCCTTCGGA AATAAGCCGA AATTCTCCAA AAATTACAGA 300

ACAATTTTCG GAAATTCCTT CTTATTTCTC GGAGTTAAAC GCTTCTGTCC CGACCTCATC 360

TTTTTATTTT GCTTGTCGTT CTTTGATTTT CGTGAGTAAA GATTGAACAT GCTGGCGTGC 420

CGAAGCTTCA TCTTCTTGAA GCATCTTTTC CGATAAAATA AAGGAGCCGC CCACTGCTAA 480

TACTTCTTTG GCTTCCACAT ACTCTATAAA GTTTGTTTCA TCAATCCCAC CAGTTGGTAA 540

AAATTGAACA TCATAAAACG GACCGCTTAA ACTTTTAATG GCTTTTAACC CGCCGTAAAT 600

ATCGGCTGGG AAAAATTTAA CCACTCGTAA GCCATACGCC AGCGCTCGCT GAATCTCCCC 660

AGGCGTAgcT GTTCCTGGAA AAATAGGCAC CTCATTTGCC AGACAAAAAT CAATGACTTC 720

TGGAACAATC GCTGGGGAAA CAATGAACTG TGCACCGTTC GCTACTGCCT CTTGGGCTTC 780

TTTCAAGGTT CTAACAGTTC CTGCGCCTAC TATTAAATTC CCTGAAGCAG CAAGTTGTTT 840

GATGGCAGGC AAGGCCAGCT GACTCCGAAA AGTGACTTCA ATAAACGGTA ATTCGTTTTC 900

TACTAAGACT GATTCTACTT GTGGCAAATA GCGTAAATCT GTTGCAGTAT ATAGTGGTAA 960

AAGTTTACAT GTCGCTAATT GTTGATACAG TTGATTTTTT TCAACTTTTG TTTCCATTCT 1020

GTCACTTCCT ATTTTGTTTG GCTTTGGTTG CTTCGTACAG TCCTTGAATA TAGGTCAACC 1080

CCATTGCTCG ATCATATAAA CCATAACCAG GCATCGCTTT TTCATCCCAA ATTGCCCGCC 1140

CATGGTCAGG ACGAATAACG CCTTCATACC CTACGTCTAC CAGTGCTTGC ATTAATTCCG 1200

CCATATCTAA CGAACCAGCG ACACTAGGAT GCGCCGTTTC TTCAAAACGA TGCTCACCTA 1260

AATACTTCAC ATTGCGGAAA TGAACAAAAT TAATTCTGTG ACCTATTTCA CGAATCATAG 1320

TAGGCAAATC ATTCGTCGGG TCCGCACCGA GGGAACCTGT ACAAAATGTA ATCCCATTTG 1380

CTGGTGAGTC CACCAAGGAA AGGATCCGTT TTAAATCCGC TAAATTTTTA GTGATTCTTG 1440

GCAAACCAAA AATTTCCCAA GGTGGATCAT CGGGGTGAAT CCCCATTTTG ATATTTTCTT 1500

CTTCACAGAC AGGAATCACA CGTTCTAAAA AGTAGCGCAA ATTTTCGACC AAATCTTCTT 1560

CCGTTACGCC GGCATACATG GCTTTCAATT CTTGAAATTG TTGCAAACGC TCCTCTTCCC 1620

AACCTGGTAA CCGGAATCCT TTTGACTGAC TGTGAATTAA TTGATACATG TCTTCTGGCT 1680

GCATGTTCTC TACCACAGCT TGATCAAATA ACAACGATAA GCTACCATCT TCATTCTCAT 1740

ACGCCAAGTC CGTTTTAGCC CAACCAAAAA TAGGTTTGAA ACTGTAGCAC ACTAAGGAAA 1800

TTCCACATTT TCCCAAGTTA CGTAACGTTT GGCGATAATT ATCAATATAG TGGTCTCGTT 1860

GATCCGTTCC CGCTTTAATG GCATCATGAA TAGCGACACT TTCAATTCCT AAAAGTGCTA 1920

ATCCTTCTTG TTCCACCGAT TGTTTAAGTG CCTGAATTTC GGCTACTGTC CAAACATCAC 1980

CTGGTAGTTT ATTTAACAAG GTACCGACAA CACCAGTAAT TCCTGGAATC TGTCGAATAT 2040

GTTTTAACGG AATGGCATCG CCTGCCGCTC CGTACCATCT AAATCCCCAT TTCATTGGCT 2100

TTCACCCTTT ACTAACATTC TTAAAAATCA TCGTCCTCAT CCTCTTCGTC CTCAAAGGAA 2160

GTCATAGAAA TATCCCATGC nTGCACACTT TCTTTGCCTT GCGCAA 2206

(2) INFORMATION FOR SEQ ID NO: 178:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1887 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

CCATCCATTA ACCAGATTTT TGCCACGAnC TTGnCTGTTA AACGCTCAAG TCGCmtTTTC 60

CCgaATGTGG GTAAAAtGTG TTTTTTCaTA CGTCGTTCAG TAGCTATGAA TGTGGATTCC 120

CTAACTGTTT TTTTGTATTC GTCCAACCAC ATATAATAAA CTTCTTCAAA AGTGGTTAAA 180

CGAGTATGCT CGTTAGCTAG ATTTCCATTA TCAAAATCTA ATTTTTTTTG ATTAAGCTTG 240

AGCTGTGCTT CTTTTTTTGT ATTACAGtTT CTGATAGTGA CATTAATTTG TTTTCCAGTT 300

AAATAATCTA CGCCTAAATA GGCAGTTACT TTCCAGTATT TTTTCCCTTT TTTTGTATAT 360

TGTTTAAAAG TTGCCATTGC TTATCCTTTC CACTTGGGCA AGCGAATAGA AGGAATGACA 420

AATTTCTAGC ACCTCCTTAT TAGATTTTAA AGCCCCTAGC ATGAATCGAA CACGCTAGAA 480

CTCACCAGAG AGGGGGATAA TAAAGCATAT TTTTTTGTTA AGCAGTGCGC GTTACTTTTT 540

ATTCTCTAAT TTTCTAATAG CTTTTTCTAA ATCATTTATT TTTTTTATTA GCATATCTAT 600

TTTTTCATCT TTTTCATCAT TGTTATTAGA TTCATTACTT TTATTAAAAT ATTCAGTAAT 660

AGTAGAAGTT AACATGCCAA CAAACCCAAT TCCCACAATC ATCAAAATGA TTGCAGCAAT 720

TCTTCCTAAT GGAGTAGTTG GaGAAATATC ACCaTACCCG aCAGTTGTTG TTGTTACCAA 780

AGCCCACCAA AACGCATCAA TGTATGGAAC GTTTTCTGCA TATGAGTAAA TCATTGCTGA 840

AATAACAATG AGAACCGAGC TTAAATAAAT CACGTTTAAA AATCCGTTAG TATTTAAAAA 900

TGATTTAGTG TTTCTTGTTA ATTTGCCAAC CACACCTATC GCTCTTGTTA GCTTTGCTAG 960

TCTAGCTATT TTAGCTATTC GAAACAACCT AGCGATTCTA AAGAAAGAAA AAATAGCATC 1020

AAAAGGAATT ATCGCGATCA GATCAAAAAT ATTTTCTTTA AAAAATTTGA TTTTATTTTT 1080

TGAAATAAAA AATCTAACAA TGTAATCAAT TGTAAATGtG aTTAAGaTAA TACTATCAAT 1140

AACGTTAAAT GGTGGATTAC TAATATTAAT AACATTTGAn AAATCAAGAA TAACTAACGC 1200

GATTGAGATT AATGCTAAnA CAACAATAGA GTAATTATAG AATTTTTGAT TTATTTTCAA 1260

TGAAGTTCTC TCACTCTCTA AATGrGGyTT AmGTCCGTGT TCCCAGCACG GACTTTTTTA 1320

TTTCAAATAA ACTTCTTGTC CCATTTTTAA GTtGTAATGA GCTATAACAT TTGAGTAATT 1380

GTATTGTCCT TCATATTTTT CGATTAAGCT TCTAAACATA TATTGTTCTG CTTCAGCTTC 1440

CATCTTAGAA CGAAAAACAG GAATTTTATA CAATGCCATT ATATCCACAT GTTCTTTTAC 1500

ATGCTTTAAT TCGTGATATA TTGCTTCTTC TTGTTCTGAT GGTGTTAAAT TTTGATTTAC 1560

AAATATGATA CCGTAGGTAG GGTCGAAACA TGCGCGTTTG TTCAAAGTAG TAAAAACTAA 1620

CTCCACATTA TATTCTTCTA CCAACTCTTT GATACTTTTC ATATAAGCAC AACCTTTGAC 1680

TTATTTCCCG AATCTACCCT TTAAATATGC ACGGATAACT TCTCTGTCAT GATCATCAAG 1740

CGGTTCACCG TCAAAACTCA TGACGTTATC CAGTACATCA TCTAAATCGT CAGATGGTTT 1800

AGCACCAGCT TGGTTTGGAT TTTCAGTACG ACCTAATAAA TAATCAACAG AAACATTAAA 1860

ATAATkGGCT ACAGCTTCAA GTTTATC 1887

(2) INFORMATION FOR SEQ ID NO: 179:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19024 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

TGTAAAAAGG nACnACAGCC ATTGGTACCT TAAACTACCT GGTTAGTTCA ACGATTTTAT 60

CCATAGTTTA AAACATCCAG GATACATGCG tTTCGTkGCG CATTCCaGgC CAGCTGTTTA 120

ATATCGTAGT TTGGATCAGT AGGTTGCGCA TTGACGCCGC GCTTAATTGT GAAAATCAGT 180

TTTGGGAAAA TAGCTGTTCG TTTTTCAATC CCTAAACCAT TAATTCGGAT TTGTAAAATC 240

GCCCGTTGAA TTTCTCGCTC AAACCAGTTG GTACCTAAAC CAAATCCCAG AGAAGTAAAA 300

GGGGTTTGTC CATTGGAAGT AAATAAGGTA TTAATTTCAT ACTCAAGACT TTGCATGGCA 360

TCAAAAATAT CTTTTTTCGT TTTCTTCCGA GCGAATGCTT TTTGGCGTTC TTCTCCTTCA 420

ATCCATTCTT GCGCATCTGC TAAATGTTTT TCATAATTAC GTTCCGCAAA AGGCGCTAAT 480

AATTCATCAA TCCGATCGGC AGAGCAACCA CCATATTGAC TGGAAGCTAC ATTGGCAATA 540

ATTTGAGAAA TTTGTGCAGT GGCCGTTTGA ATTGATTTTG GTGATTCGAC ATCAGCGTTC 600

CCAATTTTAA AGCCATTGTT TAGCATGCCT TTAAAATCAA TTAAACAGCA GTTTGTCATG 660

GGAGTATATG GATGATAATC TAAATCGTGG TAATGAATGT CGCCTTTTTG ATGGGCATTA 720

GCCACGTGTG AAGGCAACAT CTTTAAACCA ATTGATTTTC CAACAATCCC AGCGGTTAGA 780

TCACGTTGCG TATTAAAAAC GTCACTGTCT TTGTTGGCAT TTTCGTTCAC GACAGTCTGA 840

TCTTTATTAA TTAATTTGCC GATAGATACA TTAATATCAG TTGCTTTACT ACGCTCAAAA 900

TCACGTTGGG TTCTGTAATG AATGTATTCT TCGGCTAAGG CATATTCGTT GTTGGATAGC 960

AGAGTATGTT CCACAATATT TTGAATTTCA TAAATTTTTA CGTTATTGGT AAAACGTTCG 1020

GCAATTTCTC GATTGACATC AGCAACAATC GATTGGACTT TTTGATGTGT CAAAGGAGTG 1080

ACTTGGTCAT GTATTTTTTG CTCAGCTTTG ATCAAAGCGT CGTATATTTT TTGATCATCG 1140

AAAGTCACTA AACGACCATC TCGTTTGACG ACTTTAATTG TTGAGAGATT AACATCCTCG 1200

GCTTTTTCAT CAGTAATCTT CATCATTTTA TCCAGCTCCT TATATCTCGT TAGTAACTAT 1260

CCTAATACAA AGGTCGGCTA AATTCAATCG ATATATTGCG TATTTTTAGT AAAATATCAC 1320

TAGACAAACA CTATATATTG TGCTTGCGCT GTTTTGAAAA AATATGAAAA CAAATGAAGA 1380

AAAGGATTGA GCAAATTTTT TTGCGTGAAA AAAATTCTTT GTCCAAATTA AAAAACAGCA 1440

AAGAGTTTTT TGCCAAATTG CTAAAAGCAA CACTAAAAAA TAACAATTAT TAAATAATAA 1500

TTGTTTGCGG ATTTATTCAC GAAATGCCAA TTATTCTTTG ACTACCGAAC GATATTGTAA 1560

GATAAAACGA TGTTCATTCA GATTTTTTCT TGGTAAAATA AGAGTAATAA CAAAACGTTG 1620

GAGCGAGAAA AATGAAAAAA ATTGTAGAAG TAAACCATGT CAGCAAGACA TATGGTCGTT 1680

TAAATAATAA AACAGAAGTA TTAGATAATA TTTCGTTTAC CGTTGATGAA GGCGAATTTG 1740

TGGGGATTAT GGGTCCCAGT GGTGCTGGGa AATCGACACT GTTGAATGTC TTATCCTCGA 1800

TTATTTTGCC AACCGCAGGA ATCGTGCGCA TCGCCGGACA AGATATTTTA AAAATGCGAG 1860

ATAATCAATT AAGTGATTTT CGAAGAAACG AAATGGGGTT TATTTTTCAA TCGTTTAACC 1920

TAATTGATAC GTTAAATGTG AAGGACAATA TTCTGTTGCC TTTAGCTGTA GAAAAAGTTT 1980

CATTAGAAGA AATGGATAAA CGTTTGTTAC ATGTGACTTC GATTTTGGGG ATTCAGGAGC 2040

TTTTATCTGC TTATCCACCA GAGATATCTG TGGGACAAAA ACAACGGGTA GCAGCTGCTC 2100

GTGCTTTGAT TACGAACCCT AAAATTATTT TTGCAGATGA ACCAACTGGC TCGCTGGATT 2160

CAAAATCAGC AACAGAATTG TTGAAATATT TAGCTGAAAT TAATCTACAT GATGATGCCA 2220

CAATTTTGAT GGTGACACAT GACCCGTACA CGGCCAGTTT TTGCAATCGC ATCTTATTCA 2280

TTAAAGACGG TGCGATTTTC TCAGAAGTTG TTCGTCAAGG ATCACGTAAG GAATTTTTTA 2340

```
ATCGTGTGAT AGATATGCAA GCAACGATTG GTGGAGGCGG GCGCGCAAAT GATTTTTAAA    2400
CTATCATTAA AAAATTTAAA AGCTCATGTA CCTAATTATC TGGTTTATTA CATTAGTATG    2460
ACTTTTGCAG CGGTCGTTTA TTATTGTTTC CGTGCGATTG CTTACAACCA GCCTCTTGTC    2520
GCAGGTGCAG GCCGTGATAT CGAAATTAAG CGATCTTTAG GACTCGGTAG TACCTTAGTC    2580
ACCATTATTA TTTTGGGCTT TATGTTGGCA GCAAATCGTT TTTTTATAGG AAAACGTAGC    2640
CAAGAAATTG GCTTGTATCG ATTAATTGGC TTGAGAAAGA GCCAAGTTTC ACTAATTTTT    2700
TTAGTAGAAA ATTTAGTTTT AGGCTTTGTT TCACTAATTA ATGGAATTAT TTTAGGAGTG    2760
ATTTTCACCA AGTkGTTTtC AATGATTTTA GCCAAAGTTA TGTTTTTAGA AGTCCCCAGT    2820
TTCTTCTATA TTTCTTGGCG TTCCGTTATT GAAACCGGTG TAGCTTTTGC ATTTATGATT    2880
kTGATTGTCT GCTTAAGAAG TATTTGGATG ATTTATCGTT ATCCATTGTC GCAGCTTATG    2940
CATCAAGAAA AAATTGGTCA AATTAATTAT AGTCGTTTAA CGAAACGGCG GCAGCTTCTA    3000
GGTATTTTAG GACCAATTTG TTTATTAGTA GGTTATTTTA TCGCGTTTGA TTTTCGTAAT    3060
TTAGCGACAG GTATTGTGTA TAAACAATTA AATGATTCAT TGGATATTTC GATTCCAATG    3120
ATTATTACGT TGACGCCTTT TATCATTTTG GGTCTATGTG TGGTGGGAAC TTATTTATTT    3180
TTCCGTTATA CCATGTATTT TATCATTCGC TTATTTAATC ATAGAAAAAA ATGGTATTAC    3240
CAAGATTTAC GCATGATTAC TTTAGGCAAT GCCAAACGCT ACTTATTTAA AAGTAGCAAA    3300
ACACTCACTG GTTTGACCTT ATTAATTGCG ACAGCCTTGA GTGGGATCGG CGTGATGGTT    3360
TTTGTGTATA CGATTAGTAT GCACACGGTC AACAGTGATG GCCCCGTGGA CTTTTTAGTT    3420
TCACAGGAAA GCTATCCTAA ATTGAAGAAA GTTCTCGATG AAGCGAAGGA CACCAAAATA    3480
AAAAATGAAG TGACACTTTC TTATAAAATT ACTGGAATTG AACGCAGTTT GCGAATTGGT    3540
CAAGCGCAAG AGGAACAAGA AACGATTGAA GCAGTTAATG TTTTATCCTT CTCCAACTAC    3600
CGCAACTATC AAAAAATTAA TCCTTATTTA AACGATTTAC ATTTAAAAAA TGACCAAAGT    3660
GTTATCTATA TGGATAGCTT TACCAATATT TTAAGTGGTA TGTCTCGTTA TGAATCGAAA    3720
CTGCAATTTG TTGAAGGGGA GAAGGCACAA CTACAACAAG TGCTGCCTAA CTATCTAGGC    3780
AATTTTTTAT TGCAATATTC CTTACCAACG ATTGTCGTTT CCGATGCCCT TTATCAGAAA    3840
GTGACAAAAA ATTCCATTCA GTATCAGATT AACGCCGTAA ATGTTGACAC GAAAAGCCGT    3900
GAAAAACTTT ATGAAGCCGT AAATAAACAG ATTGCCACAG AATGGCAAGC GCCAATTTTG    3960
TATAAGTATG AAAAAAATAA CCAGCAGTTG AGTGGTTTTG CCAAACAAAT GCCTAACGAA    4020
GAGGTAAAGA ATACGCAAGA TGATGTCACA GAAACTTGGC GGCTAAATAT GAATGATCGc    4080
TACGCCTCTT TACGCTATGA ACGAAAAATT AACGGCTTGA CTTTATATGT CTCCATGTTT    4140
GTGGCAATTT TAGCGTTGTT TATTACGGGC AGTATTTTGA TGTTACGACA ACTGTTTTCA    4200
GCTGTAAGTG AGCGTCAAGA TTACGTGACT TTAAAACGGA TGGGTGTTTC TAGTAAAGCG    4260
ATAACTAAGC AAATTTATCG ACAGAATAGT TTGATTTTCT TTCCGCCGAT GGTCATTGGG    4320
```

ATTTTGCATA CCACCTTTGC TATTTATTTG TTAAGCCAAT TTATTAAGAG TCCAGGATAT 4380

TTGTTAGTAT ATCTTTCTTG TGGAATCTTA ATAATTGTTT ATCTAATTTT TTATATTTTA 4440

ACGTCTGCCA TTTATGCGCG GATGATTCGA CACATCCATT TTTAAAAAAC TGTAAAAGCC 4500

GTTGTGAAGC GGTTTTTACA GTTTCTTTAT GCGAAAAAAT TATGAGAGTT CCGTTTATGA 4560

AAATCATGAA AATCTAAGAA GTGACGAAAT TGTTAGTTTT TTTATTAAAT TGTAAGAAAT 4620

TTCAATGGAC TTTGCTGAAA TATAGCATAA AATAATGGAT AGCGATTAAC CTTAGGAGGT 4680

GAATGATAAT GAAAACAACT TTTCAGAAAA TTCTCAATAA CGAAAGTCTT AAATTCATCG 4740

GTAAAACGTT ATTTTATTCT TTCATTGTCA TTGCATTAAT TTATCTTTAT CATTACAAGA 4800

ATATCAATGG AGGAACCTTC ATTTATAACG AGTTTTAAGG AGGCGGCTAT ACATGGAAAA 4860

AGTAATTAAT ATGATTCAAA CGATTGATGA ATGGGCACGT AAAGAACCCC AACGTCCTGT 4920

TTATTTAACC GAAGAAAAAG TCTCTACGTA TGGAGAGTTA AAAGAAAAGT CAGACAATCT 4980

AGCCGCTTAT TTAGCTGAGC TGAAAACGGA CAAGTCAGCG ATTGTCGTAT ACGGAGAATT 5040

AGATTTTGAA ATGATTGTGA GTTTTTTAGG CGCATCGAAA GCTGGTTTTA GCTATATCCC 5100

CATTGATGCA CATACACCAA AAGAACGCAT TGAACTAATT CTAAACGTAG CCAAACCAAC 5160

CGCAGTGATT GCTGTGCATG AATGGCCTGA ATTAGCCACC GAAGTGCCAG TAATTACGGC 5220

GGAAGAATTA ACAGAAATGA TGATGCATGC ACCCCGGCAT GCGCCTGCGT TGATGCCTGT 5280

CACAGGAGCA AGTAATTACT ATATCATTTT CACTTCAGGG ACAACTGGCG TACCAAAGGG 5340

TGTGCAAATC AGCCATGATA ATTTAGTTAG CTTTACGAAT TGGTTGTTGC AAGATTTTGG 5400

TTTAGAAGAA GGGGCCCGCT TTTTAGCACA GGCACCTTAT TCGTTTGATT TATCGGTGAT 5460

GTCTATTTAT CCCGCATTGG CGTTAGGTGG CTCATTAACT CCGTTACCCA ATGAAATCAT 5520

TAACGATTTT AAACAATTAT TTACGCGTTT GCCACAATTA ACAATCGACG TCTGGGTTTC 5580

AACACCGTCC TTTATTGAAT TGTGTCTAAT GGAACCaAGC TTCGATGGTG AACATCTACC 5640

AGCACTGCGC ACGTTTTTAT TCTGCGGAGA AGAATTACCA AAACCAACGG CTGAGAAATT 5700

AGCCGCTCGT TTCCCAACTG CGCATATTTA TAATACGTAT GGGCCGACAG AAGCGACTGT 5760

AGCGATTAGT GCCATTGAAA TCACTCAAGA AGTTTTAAAA AGTGTCCAAC GTTTACCAAT 5820

TGGGTATGTG AAAGAGGATA CGCAAATTTA TATTATGGAA GGCATGTCAA AATTACCTGC 5880

AGGTGAAATT GGTGAAATCG TCATTGCTGG ACCAAGTGTT TCCAAAGGCT ATCTAAATAA 5940

TCCTGCGAAA ACAGCTGAAr CTTTTTTCCa ATTAGATGGC GTACCAGCTT ATCGCACAGG 6000

CGATGCTGGA AAGTTAGTCG ATAACCTATT GCAGTATGAG GGACGCTTAG ATTTTCAAAT 6060

TAAATTGCAT GGCTACCGCA TTGAGTTAGA AGAAGTGGAT CATCACTTAA CGAATGTTTC 6120

CTACGTGAAA CAAGCAGTCG TTGTGCCGAA ATATCAAGGA AATAAAGTCC AACAATTAAT 6180

TGCCTATGTC GTTCCGCAAG CACATGAATT TTCAAGTGAT TTTCAATTAA CTAAGGCAAT 6240

TAAACAAGAA TTAGCCACGT TGACAATGGA CTATATGATT CCGCAAAAAT TTGTATATGT 6300

TGAACAATTA CCTTTGACAA GTAATGGTAA AATTGATCGG AAAGGGTtGA TGAATGAGGT 6360

GAAtGCCACa tGATGAATTn TCCTCATATG ATyCCCTATA AtGCACCGTA TtACTTTGTT 6420

CtGCTAAtTG CGGCCcTTGT TGCCgATGAT TTTAACTTTA GCCaTTAAAG GGACACGTTG 6480

GCCTTGGTAC CAAACACTCG TGACACTGGT CTTTCTATAT ATCAGTTTTG GTGGCGAATT 6540

TTGGCaACAA GGAGTCGCTC TGATTGTATA TGTTATTTAT CAAACGTTAC TAACGTGGGG 6600

CTATGCGGCT TATCGAAAAA ATAAAAATGC AGGTTGGGTC TTTTATCTGG CTGTCTTTTT 6660

AGCTATATTA CCGCTTGTTT GGGTGAAAGT TTCTCCCTTC ATGACAGGCA AAACAACCTT 6720

ACTTGGCTTT TTAGGGATTT CTTATTTAAC CTTTAAAGCG GTCCAAGTGG TCATGGATTT 6780

ACGGGATGGC GTAATGAAAG AATATCATCC GTTTCGTTAC ATTCAGTTTC TCTTGTTTTT 6840

CCCGACTATT TCTTCAGGAC CCATCGATCG TTATCGGCGC TTTGAAAAAG ATTTGAAAAA 6900

CCCGCCGAGT GCGGAGAAAT ATTTAGACTT TTTAGAAAAA GGAATTTTTT ATTTATTTCT 6960

AGGATTTTTA TATAAGTTTA TTATCAGTCA TTATCTTGGT GGTGTCTTTT TACCACATGT 7020

GGAAAAAATG GCCTTGGCAC AAGGTGGTTT ATCCTGGTGG ACAGTCGGCT ATATGTATAA 7080

TTACAGTTTG TATCTATTCT TTGATTTCGC AGGCTATAGT TTACTGGCCG TTGGCACAAG 7140

TTATTTAATG GGCTATGATA CACCAATGAA CTTTAACAAA CCGTTCCTAA GTTGGAATAT 7200

TAAAGAATTC TGGAATCGTT GGCACATGAC CTTATCTTTT TGGTTCCGTG ATTACATTTA 7260

CATGCGCTTA ATGTTTTTCT TAATGAAAAA GAAAGTCTTT AAGAGTCGGA TTGTGACTTC 7320

TAATATTGGT TACTTTGCGT TGTTCTTAAT TATGGGCATT TGGCATGGAT TAACGTGGTT 7380

CTACATCGCT TATGGGCTCT ACCATGCAAC ACTAATTTGT GTAACAGATG CTTGGCTACG 7440

TTTTAAGAAA AAACATAAAG ACAAGATACC ATCAAACAAA TTTACACATG CATTTGCAGT 7500

CTTTTTAACT TTCCAAGCGG TTTGTGTCAG CTTCCTGATT TTCTCAGGAT TTTTAGACAA 7560

ACTATGGTTT AAATAAATAA AAATAAAAAT ATAAAGGGGC AATTTATAAT GAACATTCAA 7620

GAAACAGTAT TAAACATTTT AGAAGATATT ACAGGAACAG ACGAAtTGTG AACAATCAAG 7680

ACATTCAATT ATTTGAAGAA GGATTATTGG ACTCTTTAGC GACGGTTCAA TTATTAGTAG 7740

AAATTGAAGG ACAATTAGGA ATCCAAGTTC CTGTATCAGA CTTTGACCGT GAAGTATGGG 7800

GAACACCAAA ACAAATCATT CAACAAGTGG AAGCTTTACA ATAATGTCTA GAAATAAACG 7860

ACTAATCAGA ATCTTAGGTC CAGTGCTTTG TTCAGCGGTC TTGGTTGCAG TGTTCTTTTT 7920

TGCACCGTTC AGAATCAATT TAACTTCCGA AAAAACGTTG AAGGAAGCAT CAACGTCAAT 7980

GGCACCGAAC GTTTTAAAAG GCAATGTGAT TAAAAACAAA GCAGTGGCTT CTGGTAAATA 8040

TGTTCCTTTT TTTGGCTCTT CTGAATTAAG TCGATTTAGT GCTTTCCACC CTAGTGTGCT 8100

CTCTGAAAAG TATCAACGGA ACTACCGACC GTTTTTACTA GGAGAAGCGG GCACACAGTC 8160

ATTAACCCAA GCAATGGTCA TTCATTCGAT GGGcGATGCG ATTGcCAATA AGAAAGCGGT 8220

CTTTATTTTA TCTCCGCAAT GGTTTGTCAA AAAGGGTGTA CCTAATGATA GTTTTGGTGC 8280

ACACTATTCA CAATTACAAA CCTATCAATG GTTAGCTAAT TTAACCGAAT TAACCAGCGG 8340

TGACCAATAT CTAGCACAAC GTTTGACTAA ATTTCCAGTT GTTCAAAAAG ACAAAGTCTT 8400

AATGGAAACC TTGGCAAATC TGCAGGCAGG GCAACTTCCC CAACGCTCTC AACGAGATTA 8460

TTTTATAATG AATCTCCGTT TCTTGAATCG TGAAGATGAA TTGTTTAGCC AAATTGGCAT 8520

GGTTAGCAGA GAACCAATTG TTGAAAAAGA TATGAAACAA TTACCTGCTA CTTATAATTT 8580

TAATGAATTA GATCAGCTAG CTGGAAAAAT TGCGGCAAAA GCCATTAATA ATAATAAATT 8640

TGAAATTAGT AACGGTTTTT ATCGCCAACG GATTAAACCA GTTTTACCGA AATTAGCTCA 8700

TTCACAAAAG AAATGGGATT ACCGTTTTTC TCCAGAATAT GGTGATTTCC AAGCTGCGTT 8760

AGAACAGTTA GCAGAAAAAA ATGTCGATGT TTTATTTGTA ATCCCACCTG TTAACAAACG 8820

TTGGTCGGAC TATACGGGCT TGTCACAAGA TATGTTGCAA CAAGTAGCAC GCAAGCTGAA 8880

ATATCAACTT CAAGAGCAAG GGTTTACTAA CATTGCTGAC TTCTCAACAT GTAGTAATGA 8940

ACGCTACTTT ATGGCGGACA CGATTCATTT AGGTTGGCGT GGTTGGCTCG CTGTTGATCG 9000

TCAAGTAGAT GAATTTATGA AACAACCTGC GTCTAAAAAG TTAGCGTATC AAATTGATGA 9060

TCGTTTTTAT CAGACAGATT GGCAACAACA AAATCCATTG GTCTTACCTC AATTTTAAGA 9120

AGGCAAGATG GATCAAGTAC CAATTATCTG AAGCGGACTA GAAACGCTTC GTAACAATGA 9180

AAATAGCCTG TACCTGCCGT TGAATAACGA GCGTGCAGGC TGTTTTTTCG TTTATTTTTA 9240

GTGGAGAACG TTGCTCAATT TATTTTTTTG TAAGTATTTT TTTAGTTCGT GGTATAATGG 9300

ATAAAATTGG TTTTTGAGGG GAGTTATTTT TATGGATGCA AAAGAAGAAA AAATGTTAAT 9360

TGATTTAACA AGTGCTAAAG GGATTGCTGG GAATGAAGAC GAAGTACGTT CGTTATTTAA 9420

AAAGTATGCA GCACCGTATG CGGATAAATT TTTATATGAT GGCCTTGGCA GTATTATTGC 9480

AAAAAATGTT GGTGATAAAG AAGGACCAAA AGTCTACATT TCTGGGCATA TGGATGAAGT 9540

CGGTTTCATG GTTACCCAAA TCACAGAAAA AGGCTTTTTA AAATTTCAAA CCGTTGGTGG 9600

TTGGTGGGGG CAAGTGATGT TGGCCCAACA AGTTCAGATT AAAACCACGA CTGGTAAGGT 9660

TTACCATGGT GTCATTGGGT CTAAACCACC GCATGTGTTA ACAGCGGAAG TTCGTAACAA 9720

ACCCTATGAG ATTGCAGATA TGTTTATTGA TATTGGTGCA AGTAACGATC AACAAGTTGC 9780

AGAATGGGGG ATTCATCCAG GGGATATGGT GACACCATAT ATTGACTATC GTCGCATGAA 9840

TGATACAAAA TTTTTATTGG CAAAAGCGTG GGATAACCGA ATTGGTACGG CCGTTTCATT 9900

AAAAGTTTTA GAAAATTTAG CGACAGAAGG GCACCCCAAT ATTTTATTTG CAGGCAGTGA 9960

TGTGCAAGAA GAAGTCGGGc tACgCGGTGC GCAAACAAGT ACGCATCTTG TGAATCCAGA 10020

CATAGCGTTT GCTTTAGACA CAGGAACAGC TGGTGACACA CCAGGAATGA CACCTAAAGA 10080

AGCCGACTCT GTTTTAGGTG AAGGACCACA AATTTTGATT TATGATGCAT CGATGATTCC 10140

GCATAAAAAA TTACGTGATT TTGTTATTCA AGTGGCGGAA GAAAATAACA TTCCCTTCCA 10200

ATATACAGTG ATTACAGGGG GAGGTACGGA TGCTGGTCGC ATGCATTTAA CGCGCAACGG 10260

GATTCCATCA TTAGCCATCA CGGTGCCTGT TAGATATCTG CATTCACATA CCTCAGTTAT 10320
TCATGAAGAT GATTATTTCA ACACAGTCAA ATTAGTCACA GAAGTGGTGA AACGTCTAGA 10380
TAAAGAAACA GTCGCTAAAT TGACTAGTTA TTAGTTTCAG AAAGGATGGA ATGAATGAAG 10440
AGTTTAGTAA CAAATTCGCA ACATCGTGCC AATGCTCGCA GAgCTCTTGA TGGGCAATGG 10500
GGGATCATGG CTTGGTTAAC TTTCTTAGGA GTGGTTTTAC AGAGCTTCTT AACATCGATT 10560
GTGCAAAATT TATTTTCTGG TGAAAATCAA GTTTTTCAAA GTAGTTTTGC CGCAGTATTA 10620
CTGCAAATTT TTGTTTTATT CGCGTTAAGT TATGCGCTAT ATTACGCAGC CTTGAAGGTG 10680
TTACGTGGTG AAAAGGTGCG GGTCAATATT TTAATGAGTG TTTTTCAAGG AAAATATTAT 10740
GGTCCGCTAT TTGTGGTCAA CCTACTTCAA AAAGTACTTG AACGTGTAAT TGGTTTGTTA 10800
TTTTTGTTAC CAATTTTATT TGGGGCAGGG ACAACCTTGT ATTTCAAATT GATGTTTAAT 10860
ACAGTTACAC CAGAGGAAAT CCAAGCCTTT TTCTTAAATG ATTTTTCCTT TTTCCTAGCA 10920
TTTGGCGTTG CCACTATCTT GATCTTTTTA ATTGGTTTAT TTGTGAGTGG CGTTTTTCAA 10980
TTTGCTGTGT GGCTACGTTT TGATAATCCA GAACTGCCAA TTATGATGGC ATTGAAACAG 11040
GCGTTATATT TAATGAAAAA TCGTTTTTGG CAATATTTAT GTTTACAATT TTCTTTTATT 11100
GGTTGGTATA TTGTGGGCTT TTTAGCTATT GGTATCGGTC TATTATGGGT GATTCCTTAT 11160
AACTATGTAG CGTTACTAGT TTTTATCAAA CTGCCTTAGA AGAAAAAGGT TTAGCCGAAG 11220
AATAAAAAAG AAGCTTGAAT CCTTCGTTAT CTCGACAAAG GGTTCAGGTT TTTTCTTTGG 11280
TCTTTTTGTG AAAATTTAAG CGTTTTAACC TCGTTTTtCC GTTATTTTAT GGTAACATGG 11340
AAGTATCGTA TTGAAGGAGG TTGTCCACAC ATGAAATTGA CTGTCTTAGG CTGTTTAGGT 11400
GCTTATCCAT ACAAAGGAGA AGGAACGAGT TCCTATTTGT TAGAATCTGA AGGCTTTCAT 11460
TTGCTTATTG ATGCAGGCAG TACGACATTA GTGCGTTTAG AAAACTATCT AGATCCATTG 11520
AAATTAGATG CTGTCATTCT GTCGCATTAC CATCATGACC ATATCGCAGA TTTAGGTGTT 11580
TTACAATATT ATCGTCAGCT TTATCCAACG CTGGAACCGA CGCCTATTTT GCCAATATAT 11640
GGCCATACAC AAGATCAAAT GCATTTCAAA GAACTAACGC TTCCTAACGT TTCCGAAGGG 11700
AAAGAGTATT TTGAAGCAGA AGAGTTGGCA GTAGGACCAT TTTTAATTAC TTTCATGGAA 11760
ACGATTCATC CTGTCACTTG TTACGCGATG CGGATTGTTG AAAAAGCGAC TGGCAAAGTC 11820
TTTGTCTTTA CAGGAGACTC TGGTTACTTA GCAAGTTTTG AAGACTTTGC AAAAGACGCC 11880
GATTTATTCT TAGCGGATAC GTACTTATTT GAAGGTAATG AACGACATCA TGCACATTTT 11940
ACGTCAAGAG AAGCGGGTGA AATTGCCAAA GCAGCACAAG TCAAACAATT GATGTTGACC 12000
CATTTACCAC AGCATGGGTC ATTGGAACAA CTACGGGAAG AAGCGCAAAA CTATGCTGGA 12060
ACGGAAATTC CAGTGACATT AGCCCAACCA GATTTAACAG TTGAGATTTA AAAAAGTAAC 12120
TATACCAAGG GAGAGAACCG ATTGTGATTT TTGTACCGAA TGAAAATAAT GACCCACGAG 12180
TGAACTTAGC AATTGAAACA TATTTGTTAA CAGAAATGCC TTTAGATGAG CCCATTCTTT 12240

TATTTTACAT TAACGAGCCG TCAATCATTA TTGGCCGTAA CCAAAATACA ATTGAAGAAA 12300

TTAACAAAGA ATATGTAGAT GAGCACGGCA TTCATGTTGT TCGTCGCTTA AGTGGCGGCG 12360

GCGCCGTTTA TCATGATCAT GGTAATTTGA ATTTTAGTTT TATTATGCCA GATGACGGCA 12420

ACTCTTTCCG TGATTTTGCG AAAGTTACCC AACCAATTAT TCAAGCATTG CACGACTTAG 12480

GCGTTGAAGG TGCCGAATTA AAAGGGCGCA ATGACCTAGT GaTTAATGAT ATGAAGTTTT 12540

CAGGAAATGC GATGTACGCT ACCAACGGCC GCATGTTTGC ACATGGAACC TTAATGTTTG 12600

ATAGTGATAT TGATGAAGTT GTAAATACAT TAAAAGTACG GAAAGATAAA ATCGAATCTA 12660

AAGGTATCAA GTCTGTTCGT TCaCGGGTAA CAAACATTAA ACCATTTTTA TCTGAAGATA 12720

AACAAGAGAT GACAACGGAA GAATTCCgCC aAGAAATTCT TTTGAAAATT TTCGGCGTAG 12780

ATTCGATCGA CCAAGTTAAA ACGTATGAAT TAACTGATCA AGACTGGGCG GCTATTAACA 12840

AAATTTCAGA GCAATACTAT CGCAATTGGG ATTGGAATTA TGGTAAATCA CCTGCATTTA 12900

ACTTAGAACG TCGCCATCGT TTTCCAATTG GTTCAATTGA AATGAAAATG AACGTGGCGG 12960

ATGGTGCCAT TCAAGAAATT AAAATCTTTG GGGACTTCTT TGGGCTTGGT GAAATTAAAG 13020

ATGTTGAAGA TATTTTAACC GGCGTGAAAT ATGATAAAGC TTCTTTAGAA GAAGCCATTG 13080

ATCAAATTGA CGTTAAAAAA TACTTTGGCA ATATCGAAAA AGAAGATTTG CTTGGGTTAA 13140

TTTATTAAAA AAAAAAGACA GCCGTTTCAA ACGGTTGTCT TTTTTCTAGC TAAAAAATAA 13200

TTTTTAGCTT TTAGAAAGGT AGTGCTATGT GCAAAATCGT CGAAACGATT TTTATCaGGG 13260

AACTATCAAA TAGCTCGCCt GCATATGTAC TTGAACCAAG TCAAGTCCGG GAATTTGGTT 13320

ATGTGCTTCA CTATAAATCA CTATCTGTAA AGAGTATACA AGATTTCCTT TCTTTTGTGA 13380

TATCAAATTT TTACAGACAA AGTrGTAAAA TTTGACGTTT CTTATTTTrG CAAGGTTATA 13440

GGCTTGTTTT GGGGCTTGAA nTGGCAAAAA TTAATAGTGA TTTAnGGGAA ATATTnACAT 13500

ATGTAGAAGT ATTTTTATAA GAATAACGAG AAAAAATCAA TAGGTATGTT CTTGAGGGCT 13560

ACTTTTTGAC GTTTTCTAAT TAAAmCTTAw TAmaaAATAG CGGAAACaAA GnAAAAATGT 13620

GCTTACAAGA AGTAATCCGC ACTTAGCTAG GAAGTTAAGA ATGTGCATGT TATTTTCAAA 13680

CAAATAATAA GAATCCTATT TCAATCCTTT ACAGAGGAAC AAGGACTTCG TATAATAAAG 13740

AAACGAAAAG GAGCGAAAGT GATGAGAAAC GAAATGATGC ATCGACCAGT CGTTAAACCA 13800

GAATTGGTTG AGTATATGCG TACAAAACAA AAGAAACTTC CCGGAGAATT AGGTGCAGTG 13860

GAGGCGGAAG CCAATGAAGC AGGGGTGCCG ATTATTCCGC ACGAAACAGT AGTTTTTCTA 13920

CAATTTTTGC TAGGACAACT AAAGCCCAAA AATATTTTAG AAATCGGGGC AGCAATTGGT 13980

TTTTCTTCAA GCTTGATGGC CCAATTTGTT GGTGAAGATG GTCACGTGAC CACGATTGAT 14040

CGTTTTGACG TAATGATAAA AAAAGCGAAG AAAACCTATC AACGCCTAGA GTTAGAAGAA 14100

AAGGTCACTC TTTTAGAAGG ACAAGCAGCT GAGATTCTAC CAACGTTGGA AGGACCTTAT 14160

GACTTTATTT TTATGGATAG TGCCAAATCA AAATACATTG AATTTTTACC TGAATGTTTA 14220

CGGTTGCTAC CAGTTGGCGG CGTTTTAATG GTGGATGATG TATTTCAAGC TGGGACAATT 14280

TTAGACCCTG CTGAGGAAGT ACCGAAAAAA AATCGAGCGA TTCATCGTAA ATTAAwCCAA 14340

TTTTTAGATG TAGTCATGGC TCACCCTGAT TTAACTTCTA CTTTAGTTCC TCTTGGTGAT 14400

GGAGTTATTT TAATTACCAA AGAGAAAGAA ACGATTATTT TAGATAGTTA AAGAATGTCT 14460

TCACAAAAAG AAAGCGAATC ATTAAAAAGC AAGTTCTTTT AATGATTCGC TTTCTTTTTT 14520

ATGTTATTTA AAAAGTTATC TCGCTATTTC TAATTGAGAA TGATAATCAG AAAATTCTCA 14580

ATAGGACAGT ATTATGCGTT AATAAAACCT CGTTTGTTTA TAATTATTAT ATTGACGGAA 14640

TAATTATAAA CTGTTATCCT TAAAGAGTAA AAAAACAGAG GAGGAACGAC AATGAATTTA 14700

ATTAATCaAA AACTTTTTGA CTTTGAATGT GATGCCTATC ACGATGGTGA ATTTACACGA 14760

GTATCAACAG AAGATATTTT AGGAAAATGG AGTATTTTCT TCTTTTATCC AGCTGATTTT 14820

TCTTTCGTGT GTCCTACGGA ATTAGGGGAT ATGCAAGAGC ATTATGCGCA TTTACAGGAA 14880

TTAAATTGTG AGGTTTACTC TGTTTCAGAA GATAGTCATT ATGTGCATAA AGCATGGGCT 14940

GACGCTACAG AAACGATTGG AAAAATTAAA TATCCAATGT TAGCCGATCC CAACGGTCAA 15000

CTTGCTCGTT TCTTTGGCGT TTTAGATGAA GCTTCTGGAA TGGCTTACCG TGCTTCATTT 15060

ATTGTGAGTC CAGAAGGGGA CATCAAATCT TATGAAATCA ACGACATGGG TATTGGTCGT 15120

AATGCTGAAG AACTTGTTCG TAAATTAGAA GCTTCTCAAT TTGTAGCAGA ACATGGTGAT 15180

AAAGTTTGTC CAGCCAATTG GCAACCAGGA GAAGAAACAA TTGCGCCAAG CTTAGATTTA 15240

GTTGGTAAAA TTTAAATTCG CTTTGTAAAT AGAGGTCAAA GAGTTGGATA AAACAGAGAA 15300

AATGTTTTGT TCAGCTCTTT TTCTGAAACA CAGAAGAAAG AGGAGAAGCA AGTGACAGAA 15360

GAAATCTACG ATTTGATTAT TATTGGTGGC GGCTCAGCTG CTTTATCTGC AGGTATCTAT 15420

GCAGGTCGCG CAATGATGGA TACCTTAATT ATTGAAAAAG ACAAAATTGG CGGACAGGTA 15480

ACGACGACTT CTGAAATTGT GAACTATCCA GCGATTCGTC ACACAACAGG TCCAGAACTA 15540

ATGGAAGAAA TGCGCATTCA AGCACAAGAC TTCGGTGTAG CTTTTACAAA CGACGAGATT 15600

ATTGATGTTG ATTTTAGTCA AACGATAAAA ACTGTTCAAT CGGCGAGTCA AACCTATCAA 15660

GCATATGCAG TATTGATTGC GACAGGGGCC TCTGCTCGTA AAATTGGTTT CCCTGGTGAA 15720

AGTGAATTTA CAGGACGCGG CGTGGCCTnA TTgTTCTACT TGTGATGGGG AGTTTTTCCA 15780

AGGCCTAGAT ATTTTTGTGA TTGGTGGCGG TTATGCAGCA GCGGAAGArG CCGTTTATTT 15840

GACACGTTAC GGTAAATCAG TAACGATGAT TATCCGAGAA CCCGATTTTA CATGTGCAAA 15900

ATTAACTGCG GAGGCGGCTA AAAATCACCC GAAAATTAAG ATTGTTTATA ACATAGAAGT 15960

GAAAGAAATT ACGGGAGATG ATTTTGTCCG AAAAGCTGTG TTTGTCAATA ACCAGACGGG 16020

AGAAGAAACG GTTTATGAAG CGCCAAAAGA CAGTACATTT GGTCTGTTTG TTTTTGCTGG 16080

AAATAAACCA AGTACAGAAA TTTTTGAAGG AAAAATTGCC TTAGACCGTG GGTATGTGCC 16140

AACAACTGAG AATATGGAAA CCAATATTCC TGGAGTTTAT GCAGCAGGAG ATTTACGAAT 16200

AAAAGAATTA CGTCAAATTG TGACGGCGGT TGCAGACGGT GCAATTGCCG CCACTCATGC 16260

GCAACGTTAC GTGACAGAAC AAAAAACACA GGCGGGCCAA TCGATTGTAA CGAAGCGAAT 16320

GACTGAACGA TTAGCTAATC AATCTGCCCC TGAAACAAGC AGTCAGCAAC CAAAAGAAAA 16380

ACAGCCAGCG AAAGTTACTG GTAAGCATCA GTGGTTCCCA GAAAGTATGC GCCAACAATT 16440

AAGCGGTATT TTCGCTAAAC TAACCAAAAA AGTGACATTG CTTCAGTTTT TGGATGCAAG 16500

TGATGAAAAA TCATTGGAAC TTCAGTCATT CTTAACAGAA TTTGCAAGTC TGGATCAGAA 16560

GATTACTTTA GAAACGATTT TGAAAGATAC TGAGCCAGCC AAAGAATTGC TGTATGGGAT 16620

TGAAAAAATG CCTAGCGTTG TGCTCTTAGA TGCGGCAGGA AACTATACAG GAATTAAATT 16680

TAGTGGTATT CCTAGTGGCC ATGAAGTAAA CTCGTTGGTT TTGGCGGTTT ATAATGTCGG 16740

CAGCGAGGGG CAGCCGTTAG AAGCAAGTCT GCAAAAGAAC ATACTTGCAT TGCCAAAAAG 16800

AAAAATAGAA ATCTTTGTTT CACTAACATG TCATTTCTGT CCCGATGTCG TGGCGGCTTG 16860

CCAACGTATT GCATCAATTA ATCCTCATGT GGaAGCAGAA ATGGTGGACA TTAGTCTATT 16920

TCCAGAGTTG AAAAAAGAAA AGAAAATTAT GAGCGTTCCa GCAATGCTGA TTGATGGCGA 16980

ACAAATGATT TTTGGTTCAA AAACAATGAC CGAAATTATT GAAGCCTTGG CTTAAAAGAG 17040

GCGGGAAAAG TAATTAGCCT AAAGAACCTA AGTCATTCAA AAAAGTGAGA AGTACGTCGA 17100

GATGACGTAT TTCTCACTTT TTTCGCATAG TTAGTCAAGT CTCTCTTTTG ATTAGGCGAT 17160

TGAGTTTCCT AAAAAGAATT AAAAAGACAC AACAGCAAAT GAACAATAGG CTTCCTGTTA 17220

GTGGTAAAAA GATACCTTGC ATGTAGCTAT CGATATGCAT TAAACCAGGT ACCTCGCTGA 17280

AAAAATCATG GATAAACTGT TGAATGAGCT TATTGGCTAG AAAAGCAGCT AAAGGCAAAA 17340

TAAAAATCAA ATAATAGATG GGTTGTAAGG CATAGTTAAT TAAaAATTTG TAATTTTGTC 17400

GTACCTAAAA GAGCTAAGAT CAGTAAGTCG TTTTTTTTGC TCTAAAAAAG TTAGGAAGAA 17460

AGTTCGGCAA ATATAACTAA TCAGGAAGAA AAAGAGTGTT CCAGCGctGA TTCCAGCCAA 17520

ACCTCTGAAA ATTTTCAAGG GAGGAATCAC TGCATAATTA TCAAGAAGAA TCATTTGTGA 17580

AAGCCATCGA GAAAACAACA AAGTCAATAA TAGAGCAGAA AAAGTCAGAA AACTAATAAT 17640

AAAGATAAGA ATCAACGTAC TAAATTGGTT TTTTGAGCGC CACGCATTTT TGAGAGAGTC 17700

TGTGAATAAG TAAAGAATCG AATACATCAT CGCGTCCTTT CTGAAAAATT TTTGTTTTAG 17760

GGGAAAAGTT ATGTAAAAGA AAAGACGTTA TCTCCTGCCA ACTAGCCACT GAAAGAGATA 17820

ACGCCTTCGC TGGAAACAAG AACTGGATTA TTTCACCAAT ACTAATTTTT TTAAAGGTTC 17880

AACCCCAAAT CGTTGTTCTT GCTGTAAAAG AATTTCGGCA CAAGCTCGGC TATCTTCTAA 17940

AGCATCATGa TGATTATTTA AAGTGATATT GAGGTTTTCA CAAACgGTAT TTAATTTATG 18000

ATTAGGAAAT TCAGGAAACA GTTTGCGACT CGTTTTGACT GTACATAAAG AAAGGTAATT 18060

AGGTTGTTCT AAGCCATAGT AATCTAAACA GCCAGCCAAA ACACTACAAT CAAAACTTGC 18120

ATTGTGAGCG ACCACTAACC GATTTGGTTG AAAACAATGC TGAATTTGTT GCCAGACTTC 18180

AGGGAATTTG GGCGCATTGA CAACATCTGA TTCATGAATG CCGTGAATTT GCGTATTTCG 18240

CCAGAAAAAA GGTGTTTCTG GTTTGATTAA CGTGTAAAAT TCATCCACAA TCCGACTATT 18300

TTCTACTTTA ACAATTGCAA TGGAACAGGC ACTATGTTTA TCATAACTAG CTGTTTCAAA 18360

GTCAAAAGCG TAAAAGTTCA TTAAAAAAAA CTCCTTTnCA ACGCGTTGAG TAmCACTTAA 18420

AAATCAAGTA CAACTTCCAC TGGACAATGA TCACTACCAA GAATATCTGT ATGGATTTTA 18480

GCACTGACTA AACGTTCATT CAAGGATTCT GACACAACGA AATAGTCAAT CCGCCAACcT 18540

GsGTTATTTT TTCGTGCATT AAAGCGATAA CTCCACCATG AGTAGATGCC CTCAGCTTCT 18600

GGATAAAAAT AACGGAATGT ATCAGTGAAA CCACTGTCTA ATAACTCAGT AAATTTTTGC 18660

CGTTCTTCAG GCGTGAACCC CGCATTTTTT TGATTGGTTT TCCAATTTTT TAAATCAATA 18720

TTTTGATGAG CTACGTTTAA GTCACCGCAA AGAATGACGG GTTTTTCTTG ATTTAATTCG 18780

TTTAAATAGG CACGAAACGC ATCTTCCCAT GTCATTCGAT AAGCTAAACG CTTTAGTTCA 18840

GCTTGTGAGT TTGGTGTGTA GCAAGTGACC ATGAAAAATT CTGGATATTC TAAAGTGATC 18900

ACGCGGCCTT CTTGATCGTG AGCCTCGATA CCTAAGCCAT AGCGAACACT TAACGCTTCT 18960

TCTTTAGGCG AAAAATTGCT GTGCCAGAAT AGCCTTTTTT nTCGGCGAAC TCCAnTACTG 19020

AnGA 19024

(2) INFORMATION FOR SEQ ID NO: 180:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31517 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

GTGACAAtGA TTTATTTTGA TAATAGTGCA CTACACCGAT TTACCCCCAG GCTTTAGATA 60

CATATGTAAA AACCAGTCAG CGGATTATTG GCAATCCGTC TAGCCTACAT GACTTAGGCA 120

GCCAAGCGAA TCGGCTCTTA CAACAAGCAC GGAAGCAGAT TGCGGAgcTA ATCAATGTTT 180

CAGCAAATGA AATCTATTTT ACAAGTGGCG GAACTGAAGG AGATAACTGG GTTTTAAAAG 240

GAACGGCAAT CGAAAAACGA GAATTTGGTA ATCATATTAT TATTTCTGCA GTGGAACATC 300

CAGCAGTTAC AGAAACAGCG GAACAATTAG TCGAACTAGG TTTTGAACTT TCTTATGCGC 360

CAGTGGATAA AGAAGGTCGT GTAAAAGTCG AGGAACTACA GAAATTAATT AGAAAAGAAA 420

CGATTTTGGT TTCTGTTATG GCAGTGAACA ATGAAGTGGG AACGATTCAA CCAATTAAAG 480

AAATCAGCGA GGTTTTAGCG GAGTTTCCTA AAATTCATTT TCATGTAGAC GCGGTTCAAG 540

CAATTGGGAA AGTTAATTAC TCAGAGTGGT TAACGGATCG GGTGGATTTT GCAACGTTTT 600

CTGCACATAA ATTTCATGGA CCACGAGGCA TTGGTTTTAT GTATTGGAAA CAAGGAAAAC 660

GTTTAGCGCC GCTTTTAACA GGAGGCGGGC AAGAAAACAA CCAACGCAGT GgGACAGAAA 720

ACGTACCAGC AATTGTGGCA ATGGCCAAAG CGTTGCGTTT GCATTTAGAG AACGAACAAC 780

AACGGCCGCA ACATGTGGCG ACATTACGGA gTTATTTATT AAAGGCGCTA GAAGAATTTC 840

AAAATGTGAC TGTTTTTTCA CAAGATAATG AACATTTTGC GCCACACATT TTATGTTTTG 900

CTTTGAAAGG GATCCGCGGC GAAGTGCTAG TCCATGCATT GGAAGAAAAA CAGATTTATA 960

TTTCGACCAC GAGTGCTTGT TCTAGCCGGA AAAAAATGGC TAGCAGCACC TTATATGCCA 1020

TGCATGTGCC AGGGGAATTA GCGACTTCCG CTGTGCGCAT TAGTTTAGAT GAAAGCAATA 1080

CAATGGCAGA AATTGAACAA TTTATGATTG TTTTCAACCA ACTTTATCAA AAATTTTCTC 1140

GTGTAAATTA ATAaAAAATA AGTTGGAGCC TAGAAATAGG CTCCAACTTA TTTTTGCTCT 1200

CTTTATCGAT TGAAGCGGTA GCCaCTTCCC CAGAGCGTTT CAATTGGAGA GTCAGAAAGA 1260

TTGGCTTTTT TTAATTTTTC CCGAATCCGT TTAATGTGAA CAACGACAGT CGAAACCTCA 1320

GTATCTAAGG CATCAAACCC CCAGACACTT TCAAAAAGTT CTTCTTTATT CCACACACGG 1380

TTAGGATGTT CCATCAAGAA AACCAGCAAT TTGTATTCTT TGCTTGTAAA AATAACTTCA 1440

TTTTCTAAAA CAAACACTTT GTGGGCGGCA GTGTCGACAG CGATTTCATT GATTTTTAGT 1500

AATTGTGGTG GTTGTTCTGC TTTGCTTAAC AATTGATAAC GGTTCATATG GGCTTTTGCT 1560

CGAGCGACTA ATTCATTGGG ACTAAACGGC TTGATTATAT AATCATCCgC ACCTAAgCCA 1620

AGCCCACGaA TTTTATCAAT ATCTTCTTTT TTAGCCGAAA CAATCATGAT CGGCGTTTGC 1680

TTTTTTTTTC GAATTGCACG ACAGATTTCA AAACCATCCA TAGTTGGCAG CATCACATCT 1740

AAAATAATTA AATCAAAGGG CTCATTTAAA GCTGCTTCCA ATCCTTTTTT TCCATCATGT 1800

TCAATTGTGA CAGTCATATC GTTAATTTCT AAATAATCTT TTTGCAAATC AGCAATGCTT 1860

GGATCATCTT CAATAATTAA TACATGTGGT TTCATTTTTC ATCCTCCAAC AACGGCAGCG 1920

TAATTATCAC GTTGGTGCCG TCTCCTTTTT TACTTGTGAC AGTGATTGTT CCCTGATGAT 1980

AGTCAATAAT TTGTTTGACA ATACTTAATC CCAAGCCGCT TCCTTTAACT GTTGGGGTTC 2040

GAGATTTGTC TACACGATAG AAGCGTTCGA ATAAATAAGG TAATTCTTTT TTATCGATAC 2100

CAATGCCGTT ATCTGTAATG GTTAAAACTA AGTCTGTTTG ATTATGCGTT AAAGAAATCG 2160

TAAAGGCTAG TTGTTCTTTA GTGGGATCCG CAAATTTAAT GCTATTTTGA ATTAGATTGG 2220

TAATCACGCG ATTCATCTGC GTGGGATCCA TTTGAACGTA TAAAGCTTCG GTGGGAAGTA 2280

CAGAAGTGAT TACGAGTTCT TGTTCCAAAC GGTATTCTTC AAGAATATGT GCAATAAAAC 2340

GAGTGAAATT GGTCTTTTCC ATTGTAAAGA CAGCACGGTC TAGGTCTAAT TTAGAATACA 2400

AAAATAATTC TTCAATTAAA TCATTGAGGC CTAAACTTTT TTCATGAATG ACCGTTAAAT 2460

AGCGCTGTTT TTTTTCTTCA GTATTGGCAA CGCCATCCAT CAAACCTTCA ACATAACCAA 2520

TAATAGAAGT AATTGGTGTT TTTAAGTCAT GAGAAATATT TGCAATTAAT TCTTTCCGAT 2580

TAGCTTCATA TTTTTCGCGT TCCGCATTGG CTTCTTCTAA ATCTTGCCAC ATTTGTTCAA 2640

AACTAAGTTG GAGCTGTTTG ACTTCTGTTG AGACGGTTTG GTGGCTAGTT GGCGTAAATG 2700

GATTTTCTTT TTGGGCATCC TTACCCAATG TTTTAGTCGC TTTTTCTAAT GCTTCTAATG 2760

GCTCGATAGT GGTTTTGGTT AAGCGCTTAT TAATGAGCCA AGCTGCCGCA ATTGCTACAA 2820

GGATAATTAA CAAAATCATC CAGATAATCC AGCGTGTGAA GAATTCAAAA AGGCTGCTTT 2880

CGCGTTTCAA GACAATAAAG CTCCCGTTAG AACCATCTAA ATAATGAAAA TCGGATTTGA 2940

CATAATGATA TAGACGCCCT GCATTATCCA GTGTGCCAGT TGGCATAATA TTATTCATTT 3000

CGTAGTTTGG AATATGTACA GACAAAGATT TTTCTACTAA ATTATCCGAA TAATAAGGAA 3060

ACCGTGCATT CTTTCGAATA ATCACACTTA ATCCTTTTGC TTCAATTGTT TGAATCGTTt 3120

CCTTTAATTC TTTACTCAAA GGTGTATCTA GTAATTTCGG TGATTTTTTT AGCAGTTGAT 3180

CTAAGGCCAA ATAACTTTCT TCTTCGTTAG CGGTCAATGG TCGTTGCTTA TTCATCATTC 3240

GGTAAGCTTG GGGCAAACTG GGGACAGTTC CTAATGTGAT GTAAGAAGCT AAGGATAAAA 3300

CAGCTAAAAC AGAGGCTAAA GTAATAATGA TGGCACTAAT ATAAGAAATA AAAAAGCGGC 3360

GTTTAATGGT CAAAAGCGTC CCTCCTAGTG GTTTTAATCT ATTTCTTACT TATAGTGTAT 3420

TACAAGTGAA CAAAAAAAAG CTAGTATTGT TATTAACTTT TTTTCATTTG TTTATAATTT 3480

GTTCATTTTT CGGCAAGTAG GAGTTTATAA GGCTTTGCTA TTCTAAATAT ATCAAAAAGA 3540

AAAAAGGGGA TTACCAACCA TGAAAAAGAA AAAAGTTTTT AGTGCGCTTA CCTTATTAAC 3600

CTTTAGTACG TTGTTGATTG CAGGCTGTGC TGGCGGAGCc AACTCTGcAA CAGATAAATC 3660

AAGTGCAGCT AGCTCAAGCA CTGCAGTCTC TAGTTCAGCA GAAGCAGCTA AAGAGCAATC 3720

AAAAGGACAA GAATTAACAG AAATTTTATC CAGTACTGAT TGGCAAGGCA CAAAAGTTTA 3780

CGACAAAAAT mATaATAATT TaACAGCAGA AAATGCTAAT TTTATTGGTT TAGCAAAATA 3840

TGATGGTGAA ACAGGTTTTT ATGAATTTTT CGACAAAGAA ACAGGTGAAA CCCGTGGCGA 3900

TGAAGGCACA TTCTTTGTGA CAGACGATGG CGAAAAGCGT ATCTTAATTT CGGATACACA 3960

AAACTATCAA GCGGTGGTCG ATTTAACGGA AGTGACGAAA GATAAATTTA CCTATAAGCG 4020

AATGGGTAAA GATAAAGACG GGAAAGATGT AGAAGTCTTT GTAGAACATA TCCCTTATTC 4080

TGACGAGAAA TTAACCTTTA CGAACGGCCG TAAAGATTTA GAAACAGAAA CTGGCAAGAT 4140

TGTTACCAAT GAACCTGGGG ATGACATTTT AGGGGCCACA TTATGGAATG GCACGAAAGT 4200

TTTAGATGAA GACGGTAACG ATGTTACTGA AGCAAATAAA ATGTTTATTA GTTTAGCGAA 4260

ATTTGATAAT AAAACAAGTA AATATGAATT CTTTGATTTA GAAACGGGTA AAACACGTGG 4320

AGATTTTGGT TACTTCCAAG TAATTGATAA TAACAAAATC CGTGCTCACG TTTCAATTGG 4380

TGACAATAAA TATGGAGCTG CATTAGAATT AACAGAATTA AATGATAAAC GTTTTACGTA 4440

TACACGAATG GGTAAAGACA ACAATGGCAA AGAAATTAAA GTCTTTGTAG AACATGAACC 4500

ATATGAAGGA GACTTTACGC CAGACTTCAC GTTCTAATTA GTCAATCACG AGTGGATGAA 4560

GCGGTCAAAA GGAAGTAAAG GATTCTTTTG ACCGCTTCAT TTTTAACAAA AAAATAGTTT 4620

ATAGTTGTTT TTAATCGTTT ATAAATTGTT CATTTTTCTG ACAGTGGTAG TTTAGAATCG 4680

TTTGTTAAGA TAGGTTTATC AAAGAAAAGG AGCGATGCTT TATGAAAAAG AAAGTATTAA 4740

GTTCGATTAC TTTAGTAACA TTAAGTACGT TACTTATAGC AGGTTATGCA AGTCCAGCAT 4800

TTGCAGATCA TGCAGCCAAT CCAAATAGTG CTACAGCAAA TTTAGGCAAA CATCAAAACA 4860

ATGGCCAAAC AAGAGGCGAC AAGGCGACTA AGATTTTATC TGGCACGGAC TGGCAAGGAA 4920

CCCGTGTTTA TGATGCTGCT GGTAATGATT TAACGGCAGA AAATGCTAAT TTTATTGGTT 4980

TAGCAAAATA TGATGGTGAA ACCGGTTTTT ACGAGTTTTT CGACAAAAAT ACTGGGGAAA 5040

CCCGTGGTGA CGAAGGAACA TTTTTTGTGA CAGGTGATGG CACAAAACGA ATTTTAATTT 5100

CGCGGACACA AAATTATCaA GCCGTAGTGG ATTTAACCGA AGTGaGTAAA GACmAATTTA 5160

CTTACAAGCG TTTAGGGaAA GATAAACTGG GGaATGATGT TGAAGTTTAC GTGGAACACA 5220

TcCCtTATCA TGGGAAAAAA TTAGCTTTTA CAAATGGACG TGAAGCATTA ACCAATCAAA 5280

CTGGCAAAAT TGTGACAAAT AAATCAGGGG ATAAAATTTT AGGAACAACC TTGTGGAATG 5340

GCACAAAAGT CGTAGATAAA AACGGTAATG ATGTGACAGC GGCCAATCAA AATTTCATTA 5400

GTTTAGCGAA ATTTGATCCA AACACAAGTA AATATGAATT TTTCAATTTA CAAACAGGTG 5460

AAACCCGCGG CGACTTTGGG TACTTCCAAG TGGTGGACAA TAACAAGATT CGGGCCCATG 5520

TATCTATTGG TACGAATCGT TACGGCGCGG CGCTAGAATT AACGGAACTA AACAATGATC 5580

GATTTACGTA TACTCGAATG GGTAAAGATA ATGCTGGTAA TGATATTCAA GTGTTCGTGG 5640

AACATGAACC TTACCAAGGC ACATATCATC CAGCCTTTAC TTTCTAAAAG TCTCTTGTCT 5700

TCTCTCTTAT TTAACAAAAG AACGTCCCCC AACGTTCTTT TGTTTTCTCT TCCTCTATAA 5760

ATAAATTGTT ATTAGTTATA ATTAAATATT ATTATTAAGA TAATTATGAT GATTATTATT 5820

TAATAATTTG TTCATATTGT GGACAGGTGG AGTTAATAAC TACTTGCTAT CCTTAAATTA 5880

TCAAAAGGAA AGGGAGTAGT CGTTTATGAA GAAAAAAATC GTTGGTACAA TTACGTTGTT 5940

GGCTTTAAGT GCGTTATTAG TTGGTGGAGC AGGAGGGGCT TTGACGGCAG AAGCATACGT 6000

TCCTCAAAGC GTAGACAATC CCAATAATTT AGGGGATTTA CCTGAGTATT TACGTTCAGT 6060

TGGTATTAGA CAAGATGAAG GATTATCAGA AAAAGATTGG GCTGGAACAC GCGTTTATGA 6120

TCGAAATGGG AATGACTTAA CAGATGAAAA TCAAAACCTA TTACATGCAA TCAAATTTGA 6180

TGCAACCACT AGTTTCTATG AATTTTTTGA TAAAGAGACT GGAGAATCAA CAGGAGATGA 6240

AGGAACCTTC TTTATGACCG CTGGTATTAC AGATGTTTCC CGTCTTGTAA TTATTTCTGA 6300

AACCAAAAAT TATCAAGGTG TATACCCACT TAGAACTTTA TACCAAGATA CTTTTACGTA 6360

TAGACAGATG GGGAAAGATA AAAACGGAAA TGATATTGAA GTTTTCGTAG AAAACAAAGC 6420

AACCTCAGGA CCAGTTTATG GTCGTCCGCA GCCATACCCC AATAATCGTC CCAGAACACT 6480

AGAATTCACG AATGGACGCC GTGCCATGAC AGAACAAACA GGCCAGATTG ATGTAAATCG 6540

ACAAGGGGAT GAAATTATTG GTAAAACTTC CTTTGATGGG ACACCGCAAC TTCTTTGGAA 6600

TGGCACAAAA GTAGTGGATA AAGATGGCAA TGACGTAACT TCGGCCAACC AAAACTTTAT 6660

CAGCTTAGCG AAATTTGACC AAGATAGCAG CAAATATGAA TTTTTCAATT TACAAACTGG 6720

TGAAACTCGT GGCGACTATG GCTACTTTAA AGTAGGAAAT CAAAATAAAT TCCGTGCCCA 6780

TGTTTCCATT GGAACCAATC GCTATGGCGC TGTCTTAGAG TTAACAGAAT TGAATGATAA 6840

TCGTTTTACG TACACACGAA TGGGTAAAGA TAACGAAGGA AACGATATCC AAGTCTATGT 6900

GGAACATGAA CCATACCAAG GAACTTTTAA TCCTGAATTT ACCTTTTAAG AACAAAACAT 6960

ACGAATGGAA ACCAAAAAAA GTTTCGACTT CTTTTGGTTT CTCTCTTTTT AAAATAGAAA 7020

AACAGAGAAA GCGTTTTAAT CGACAAATAA AAAAAGATTC GTTAGATTAG AACCACTGGA 7080

GAAAAATCTC ATATTTCTCT CGAGGAAAGG AAGTTGAGCA CAATGAACAA AAAAATTTTA 7140

ATGGGGCTAT TAAGTGTCGT GACCATTCCA TTACTTGCTG CGTGTCAAGG AGGAGAAACA 7200

CCTTCCGCAG CGTCAAAAAA TAGTCAAACG GTGACTACTC AAAGTAGTGC AAAAACTGAA 7260

AGCACCAGTA CAACCCGTTC GGTAGCTCAA ACAACATCAA AAGAGGAAGT GAAAGAACCG 7320

ATGAAGACCT ATGAAGTGGG TGCGCTTTTA GAAGCAGCCA ATCAACGAGA TACGAAGAAG 7380

GTCAAGGAAA TTTTACAAGA TACTACTTAT CAAGTGGATG AAGTCGACAC AGAAGGCAAC 7440

ACACCGCTCA ATATCGCTGT TCACAATAAT GACATTGAGA TTGCAAAAGC GTTGATTGAT 7500

CGGGGTGCCG ATATTAATCT GCAAAACAGC ATTAGTGATA GTCCCTATCT TTATGCGGGA 7560

GCGCAAGGAC GTACGGAGAT TTTAGCGTAT ATGTTAAAAC ATGCGACCCC AGATTTAAAT 7620

AAGCATAACC GTTACGGTGG CAATGCGTTA ATTCCGGCAG CTGAAAAAGG ACATATTGAC 7680

AATGTGAAGC TCTTGTTAGA AGATGGACGA GAAGACATAG ATTTCCAAAA TGACTTTGGC 7740

TATACAGCAT TGATTGAGGC AGTGGGGTTA CGTGAAGGGA ACCAACTTTA CCAAGATATT 7800

GTAAAATTGT TAATGGAAAA TGGTGCGGAT CAATCCATTA AAGACAATTC TGGTCGAACA 7860

GCAATGGACT ATGCCAATCA AAAAGGTTAT ACGGAAATTA GTAAAATTTT AGCACAGTAC 7920

AACTAAGTAT TGTGTCCTTT CGGGAAGGAA TAAATACAGG TAGGAGTGAA GGAAAGCATG 7980

AATTATTGGT TAAAAAATGT TCGAATTGAA ACAGAGTATC AAAAAGAAGG TGACTGGTTA 8040

TCTGGGACAG TAACCAAAGA AGTGGCAATT GAAATAACAG ATGGAGTATT TACGAAAATC 8100

ATCCCCAATG AAGAATTGGT AGAAGAACCT GGGTTGACAA TTATTGATGG TCACCATCAA 8160

TTGATTTTAC CAGGTTTGAT TGAAAAACAT TGTCACTTTG ATAAAAGTAA ACTAGGTGTT 8220

CCGTGGTATC CGGTGACGCC AGCGAAGTCA ATTGTGGAAC GTGTCGAAGC AGAAATCCCT 8280

TATTTAGATA GTTTAGAACT TTCTTTAACA GAACGCGCCA ACCATTTAAT TGACTTAGAA 8340

TTACCACATG GCGCGACAgc CTTCAGAACA CATGTCGACG TGGAACCGAT GACAGACTTA 8400

CGTTATTTTG ATGAAGTCCA AGCGTTGGCC CAAACGAAAC CATTTGCGGT AGAAATTGTC 8460

GTCTTTCCAC AGCACGGCTT ACTGCGTTCA GATTCCGTTG AATTAGTTGA TCAAGCATTG 8520

GCAAAAGGCG CCGATTTTAT TGGTGGCGTG GATCCTTATT CACTAGATGG TGATTACAAA 8580

AAATCATTGG CTGAAACATT CCGCTTAGCA GATAAACATG GTGTCGGTGT GGATATTCAT 8640

TTACATGACC GTCACGAAGC TGGGACAACA ACGATTAAAG AAATTATCCG TTTAACGAAA 8700

GAATATGGCC TACAAGACAA AGTATTTATC AGCCATGCCT TCGGGTTAAA TGATTTCATT 8760

GGTGAAGAAC GTGCCGAAGT TTATGATGCT TTAGCAGCAG AAAAAATTCA TATTAATTCC 8820

AGCGTGCCAA TCACACCTAA TACGATTCCG CCAATTATGG AATTGTTACG TCATGGAGTC 8880

AATGTCCATC TTGGTTGTGA CAATATCTAC GATTGCTGGT CTCCCTATGG TGATGGTTCA 8940

TTACAAGAAA AATTGGCTCG TTTAGGTGAA CTTTTTAACG TGAAAGATCA AGATGCTTTA 9000

ACCCAATTAT TAGGCTTAGT GACAGATGGT GTCACAACTT TAGACGAACA AGGGCAAGTA 9060

AATTGGCCGC AAGTTGATGC TGAAGCAACT TACTTACTCA CTGAAGCAGA ATGTGCGGCG 9120

GCTTTTGTCG CTCGTCAGAC ACCGGTGGAA ATCAGTGTGG TAAAAGGAAC CACACTTTAT 9180

CAAAAATAAC AAAAAGAAGC GAGATTTCTC GCTTCTTTTT TTGTTATTAA ATTTATTTGG 9240

ATTAGACGCT TAATTTTACA ATTGAAAACG ACATCCATCT TATATTTCAT TGATGAAATT 9300

AGTGAGAAGA AGTATGATTT AATTTAGCTC TGATTACTTT TGATACATCG TTTTTCAACT 9360

GATCAAAATC TGATCTATAA GTAGCAATAA AAACCAATAA ATCAACGCCT TCTTGCGTAT 9420

CTAATTGGAA ATCATAACCG TTTAATTGTA AAAATAAGTC GGTTGCCATT ACAGCTGTCC 9480

GTTTATTTGC ATTGTAAAAA CAATGCTTTG TTGCAATGTT AATCATTAAA ATAGCTGCTT 9540

TTTCTTCAAT TGTTGGGTAC AGGTCAACGT CAAAGACGAA TTGTGCTGGT TGGTTGCTAG 9600

CCATTTCTAA TGCATTACGA TCTTTGATGC CAATTTGTTC TGCGGGAGAA TACGTTTGGA 9660

TCAAGAAAGT ATTCATTTGA AATAGATCAT CCGCTGATAA ATAATAAATA CTTTTCATTT 9720

AGTAACCAAT TCTTTGTACA GCTCATTATG TTTTTGTTTT GAGTTCTCCA TCATTTGTTG 9780

TAATTTTTTT TGATGTTCTG ACATATTTTT TTTAACAATA ATATCTTTTA ATTTTTCyTC 9840

GTCTACATAA ACGGTGTCTG TCGCAGTCGC ACCAATGCTT TCTAGAAACT CTTTTGACAA 9900

TGTCATAACA ACaGAATTCC CAACTTTTCT AAGTTTTCTT TCTTTGATTT CCACTAGTAA 9960

CCCaCCTTTC GTATATATTG TATATACAGT ATATCATGGA TCATAAAAAA GTTTCTATCA 10020

CAAGTCTAAT ACTTTAATAA AATTTCTTGA TTGCTTATCG TTCCTTCAAC CAGCTATCCA 10080

ATTGCTCAAA TAAAGTGATG ACCGAAGGAT CTTTGGTTTG TTTTAAAAAA TCATGTTCTA 10140

AATAGTAAAC CGCTTTAAAG GTACTCTCAG GAATGGTCCG TCCGATTTTT TTGCTATAGC 10200

GGAACGGCAC TTCTTCGTCT GAGCTGCTAG CTGTACTGAA ACAAGGTGGA AAAGTTTTTA 10260

ATGTTTCGTC TGACAGTGCG TAAGCAGACC AGTCCCCGTT TTCTGGTAAT CCGTAAAAGT 10320

GCGGCAATAG CGCTTGCTGA ATACTGTAAT GATACAGCAG ATAGCGAGAA AGAAAAGGAT 10380

CATCCCAAAC AGGCTTCGTT TGGTCGATAG CAGCTATTTC TTTAGCGGAA ATGGCTTGCT 10440

TTAATAATTT CCGTGGCTCT TTAATAAATT CTAAATCTGT ATAGCCGTAA AAATTGACTA 10500

AAAATTGTGG CGTCAAATTT AGTGTTTGTA ATTGTTTGGT TAACTGTAAC ATTAAATAAC 10560

CGCCAGCGGA ACGCCCGCAC AAACCAAAGG ACTGGTTTTG AATAATCTCT TCGTTTAACA 10620

ATTGGAAGGT TTCCGTTAGC GTCCTTAAGA TGTGATCGAT TTTTGTATTT GGAGCCAACA 10680
AATAGTCCAG TGCTAAAACG GTATAGCCAT TGCTTGTAAA CAGCTCCTTC AGTTCTTCTG 10740
GTAAATCACT TTTGGTTCCA TAAATCATCC CGCCACCGTG TAAATAGACG ACATAGTTCG 10800
TTGGTTCAGT CGTAGTGGGG TAGATAGTAA CTGTTGCGCC ATTAGCGAGT GTTTGATTGT 10860
TTTTCAAGTG CACGTCTCCT CATCTTTATA ATGTATGTGG CAATTTCATA ATTTAACAAT 10920
TGCAAGGGAT TGGCTAAGTC AATATCAAGA AGTTGTTCCA CTTTGTTTAA ACGATAACGA 10980
ATCGTTTTTG AATGAAGGAA CATCGCTTCA GAGGTCTGTT TATAGTTGCG ATTATTTTGG 11040
AAAAAACTGT ACAAGGTTTC AAACAAATCG TAGTTATTTT CGGCTAATTG ATAAAGTGCG 11100
CGAGGAATTA ATTCGTCTAA ATTTTCCAAT TGGTCTtCGC GGATAAAaTT TTTAAATACC 11160
CCAATATCGG CTAAGGTGAC AATTGGTCCG TTATAAAAAG TTTCATTAAA GCGCAGAATG 11220
TCCAGACATT CAATCAGTAG CTCTTTAATC CCTTCTCGTG ATTTTAAACT GCTGACGGCA 11280
ACGGTAAGCG TGTCGTTTTC TTCTAAAAGT TCAGCTAATA AACGACTTAA TTGCAGCTTT 11340
GTAATTTTTC CATCACTTTC ATTGAAGTTA AAAAGGACCG CCGAATAATT TAGCTGATCA 11400
AAGAAGATGG ATTTCACACG CAGGGTACGA AGTAAGGCAA TAATTCGTTC TTTCATGAGT 11460
TGTGTATTGG TGTCTTTGGT AGAAAAAGCA ATCGCTTGGT AAGAATCTAG CTCTTGCATT 11520
TGCACTTCAC GAAGCAAACT GTTTAGCTCG TCAGGATTTT GCGGTGTATT TTGTAAAATA 11580
GCATCCGCTA AGTTGTTCAA ACGAGTATAC CGTTCTTTTT TTAACAAATT TTCGGTATTG 11640
AATTTTTCTT GTAAGACATC AATCGCATTT TCGATAATCA CTAAATCCGT TTCTTTAACC 11700
TTTTCATCTT GTAAATAAAC CAGAAGTAAA CAATTAGTGG AATAACTATT GAAAATAGAA 11760
ACTTCTAGTG CCAACTGTTT TTTGGGGTTG GCATGAGAGT AAAGTGTTAA AAGAGAGTAA 11820
TCGTTTTTAG TGAATTCACT GTTTTTTAGT TTTTCTCTTG AGAGCACAAT CGCATCATGG 11880
GGTAAGTCGC CATCGCTAAT TTCCAGATGT TTATCCAACA TTTTTAAAGC GAAGGGTAAT 11940
TTAATGATGT CATAAAATTC TTGCATAATC GTCTCAACCG TTGGGTAGTG GCGCTCTAAT 12000
TTGGTGAAAC GTTGGCGCAC ATCGTAATAC GTACGTAAGA CATGCGACTG ATAATTCAAC 12060
AAAGGCTCGT AAATCGTAAA TAGAATTTTT TCATAGCTAA TGTCTTGTGT GATTTTAATT 12120
AAAGGAATTT CATAATTGAA ACATAATTCG ATAAACCACT CTGGAATCAT TTTAATGAAT 12180
CGATCTGTTT TCAAAATTAA GCCGCTGATG CCAATTCGGC GCATCTTTAC GAAAAAATTT 12240
TCCAACTCTT TTTCGGGTAG ATCATGCAAT GCGTAGAAAG AAGTCAAAAT TAACTGATTC 12300
TTTCTACTCC AGTTCTCAAT GTCAATCGCT TCTAAGACCA TTGCCGAGTC AACTTCATTG 12360
TCTAAACCAA TCTCTTTCGT CAAAATTTTA CCTTTTTTTA AGTCACCTAT CGATAATAAC 12420
TCTCTTAATT TCATCATTTT CTCCTTTTTT CCTCTTATGG ATAAAAATTA TACCATTTTT 12480
TTAGAATTAT GTCCGATGAA AACACTTTCT CGTTTCGATA CAATAGCATT GTAAGGAATA 12540
GATAGGAGGA TGAACAATGT TACACACAGT AATTTTGAAA AACAATTACC AAGACTCCAT 12600

CAACTTAATG TTATTGACCA ATAAAATCAA TGCATTAGAT GGTGTAACCA TGAGTCAAAT 12660

CATGATGGGA ACAGATGCCA ACAAAGACAT CCTTAACAAC ACGAACTTGT TAACGGATGA 12720

GGCCAATAGT GCATCTGCAA ACGACATGAT GATTGTGGTT GACAGTGAAA AAGAAAACAT 12780

CATGGAAGAA GTTATGCCAG CGATTGATGA ATTCTTAGAT GATTTATCAG CAAAGGGAAC 12840

TAGTGAAGAA GCACAAGCAG CAACTTCTTG GTCAGAAGCG TTTGATCTTT TACCAGAGGC 12900

GAATGTAGCG CTATTCAGTA TTCCTGGTGA ATATGGTGCG CCAGAAATGG AACGTGCCTT 12960

AAAAAATGAT TTACACGTGT TCTCTTTCAC GGATAATGTG TCAATTGAAG ATGAAGTACG 13020

TTTGAAAAAA TTAGCGCATG AAAAAGGCTT ACTAATGATG GGGCCTGACT GTGGAACTGG 13080

TATTATCTCA AGTATTCCAA TCGCCTTTAC AAATGTTGTT TCTCCAGGAA ACATTGGCGT 13140

TGTTGGTGCT TCAGGAACAG GGATTCAAGA AGTTACAACG ATTATTGACC GCTTAGGTGG 13200

CGGAGTTGTT CACGCCATTG GTACTGGTGG TCGTGACCTA AGCGACAAAG TAGGCGCTAT 13260

CACAGTGAAA GATGCTATTG TGGCCTTAGA AAATCACGAA CCAACCGATG TGATTACAGT 13320

TATTTCTAAA CCACCTGCAA AAGAAGTTCG TGATGAAGTG GTTGAATTAT TACAAAGCAT 13380

TTCAAAACCA GTCGTTGCGA TTTTCTTAGG AGAAAAACCA ACAAGTCACG AAGGCAAAGT 13440

TTATCTAGCT CATACATTAG AAGAAACAGC TAAAATTGCT GTTGATTTAG CAAATGATGT 13500

GGCCGTGAAG AAAAATTACT TTGAAGCATT AGCAAAACCA GCTGTACCAA CATTACCAGA 13560

AGATAaAGTT GTAAAAGGCT TATATTCAGG TGGTACATTA GCTTCTGAAG CAGGTATGTT 13620

AATTTCAGAA GCTTTAGATC TTGGCGGTTT AGTGAAAGCA GAAGGTTATG TCTTGAAATC 13680

TCACGGTTAT GAAGTCATTG ACTTAGGCGA TGATATGTAT ACACAAGGTC GTCCGCACCC 13740

AATGATTGAC CCAGATGTTC GTATCGAAAA AATTCGTGAA TACGCACAGG ATGAAAAAAC 13800

GGGGATTATC TTATTCGATG TGGTGTTAGG CTATGGTGCT CATGAAGACA TGGTTGGTGC 13860

ATTGTTACCA GCAATTGAAG AAGCGCGTGC CACAGCCAAA GAAGCAGGAC GTGATCTTTA 13920

CTTTGTTGCA ACTGTCTGTG GGACGACGAA AGATCCGCAA AATTACCAAT CGTCTGTTGA 13980

TCGTTTGAAA GAAGGCGGCG TTTTAGTAGC AGAAAGCAAT ACTAAAGCTG TTCAATTAGC 14040

TTTATTACTA AAAGGAATCG AAaTTTCAGA AGACGATAAA GAAGTGGTTG CTTACAACGG 14100

ACCmACTGTC GATGTrCCrA AACCAGGCGA AAAaGTAATG GAACTATTaA CAACCAAACC 14160

AAGAATCATT AACGTTGGGT TGCAAAGCTT TACAGAATCA ATCGTTGATT ACGGTGGCGA 14220

AACTGTTCAA TTTAACTGGC GTCCACGTGC TAACGGCAAT AAAAAAATGA TTAAAATTTT 14280

GGATGCTTTG GAAGATTACA GTGAACAAAT CGAAGCTGAA AATCACAAAG TAACAGACAA 14340

AATTAAAAAT GCACAACCAT TCTTAGTAGA TGTGGTACCT GCGAAATCAG TGATTCCAGA 14400

ATTAAATGAT GACGCACAAA AAACATTACT TCATGCTGGA CCACCAATCC AATGGTCTGA 14460

AATGACTGGA CCAATGAAAG GTGCTTGTAT CGGAGCTGCA TTATTCGAAC GTTGGGCAGA 14520

CAATGAAGAA GATGCTTTGA AGATTTTTGA AGCTGGCGAA GTACGTTTCA TTCCTTGTCA 14580

CCATGTAAAA GCAGTAGGAC CAATGGGCGG TATTACTTCT GGTAATATGC CAGTCTTTGT 14640

TGTGGAAAAT CGCTTAGAAG GTAACGAAGC ATACTGTATC TTGAATGAAG GTATCGGTAA 14700

AGTATTGCGT TTTGGTGCTT ATTCACAAGA AGTGGTTGAT CGTTTAGATT GGATTAAAGA 14760

TGTTTTAGGA CCAACTATTG CGAAAGCCTT GAAACTTTCT GAAGAAGGCT TGAACTTAAA 14820

CGTATTGATT GCGCGCTCTA TTACAATGGG TGACGAGTTC CACCAAAGAA ATATTGCCGC 14880

ATCTTTGAAC TTCTTAAAAG AATTGTCTCC ATTAATTATC AAAACAGATA TTCCAGAAGA 14940

TCAAAAATAC GArtAATCAA ATTCTTAGCT GACACAGATC AATTCTTCTT AAATGTCATG 15000

ATGGCAACTG GAAAAGCAAT TGTTGATGGT GCCCGTAAAG ATACCAACGG AACAATTGTC 15060

ACAACGATGA CTCGTAACGG CGTGAACTTT GGGGTCCGGA TTGCAGAAAC AGGCGATCAA 15120

TGGCATACCG CACCAGTAAA TACACCAAAA GGTCTTTACT TCACTGGTTT CAGCGAAGAA 15180

GACGGAAATC CTGATGTCGG GGATAGTGCA ATTACTGAAA CAGTAGGTGT TGGAGCAATG 15240

GCTATGGTAG CTGCACCTGG CGTAACACGG TTTGTTGGTG CAGGTGGCTT TGAAGATGCG 15300

TTAgCAACCT CAAATGAAAT GGCAAAAATC TGTGAAGGAA ACAACCCAAC ATTCTCAATT 15360

CCAACTTGGG ATTTCCAAGG AACTTGTTTA GGAATTGATA TCCGTAAAGT GGTGGAATTA 15420

GGAATTACAC CAATCATCAA TACAGGGATT GCACACAAAA ATGCCGGTGT TGGTCAAATT 15480

GGTGCAGGGA CAGTTAGAGC ACCTCTAGGC TGTTTTGAAA AAGCCCTAGA AGCGTATGCT 15540

GAGAAACTAG GCATTACGGT TGAATAACCA ACTAGTCATT AGTGACTACA TTTTTCCTAT 15600

CCACCAATAT GGCAAGATGG GAACCATTCA CAGTGTTTTT GAACGTTCAT TTAATCTGAA 15660

AGTCCAGGAT CAATTAATCA ATGTTGCAAA CTTCCATAAT TATTTATCAA GCTTTGGTAT 15720

GTTTTTGCCA GATCAATTGT TTCAAGAAAT CTTTCCGTAT GTACAACAAG GTAATAAAGT 15780

CAAAATCACA GAAAATCAGT TGACCTTTTA CAGTACAGTC GGCGTGAAAA CGATCCAACT 15840

CACACCTGCT GAGGTCGTTT CTTTAAACAT TACTCATTTC AAGTTAGAAA AAGACCAATT 15900

ACGCTTGTTG CGAGACCGCT TATTAGCGAG AAACTTAGAA CGTAGAATTG GCCTGCCTTT 15960

AGACGAACGA GCGAAgcATA TTTTCAAAAA AATGAGCCAA AAACAAAAAG TTTGGACACT 16020

ATCGGAATGG CAAGAAGTAA CGAATTATCT GATTGGTCGG GGCAAAGGGT TAACCCCGAG 16080

TGGTGATGAC ATTTTAGTGG CGTATCAAAC AATTCTGTTC ATACTTGCTG ATGAAAGAGC 16140

TGCGGCGTTA GCCgcTAcTC TTTCGGCAGC AAATTTATCC ACAACAGATG TTAGTAAAGG 16200

CTATATTGCA AGCTCGGCTA AAGGATATGT AAATTCATTG TTGTATCAGT TACTCTTGGA 16260

TTTGGAGAAC CATCGGGATA ATAGAGTTGA CGAAAATATT GACAGAATTA TACAAATCGG 16320

ACATTCCTCT GGTAAAGATA TGTCATTTGG AATGTTACTA GCTTTACAAT CTGTAGAGTT 16380

AGATGAGAAG TGAGGAGAAG ACATGGAGAA TAGTAAAAAA AATCAAGCAG TGCAAGTGAG 16440

CAATAGTTAC TGGATTAAAG TCGTAGCCAT TTTCTTCGTA GGCTGGATTT TAATGTACGC 16500

AACACGAACT ATTTTTAATC CAGTGATGGG GGTCATTGGA GAACAATTTG GTCTAAGTAA 16560

```
CACACAACTT GGTTTAGCCA ATAGTATCTT CTTTCTAACC TATGCTGTTG CACAGATTCC   16620
TTTTGGGATG ATTGGCGATA AAATTGGGCG GAAATTAGTC ATTGCTGTTG GGTTTATCGT   16680
TCTTGCTATT TCTACTTATT TTAGTGGATT AGCGACAACC TTTGTTATGT TTTTAGTCAT   16740
TCGTGCAATT GCTGGTATTG GTCAAGGAGC TTACTATGGC CCGCAATATG CGCTATCGAC   16800
AGAAGCGATT CCAGCAAGTA AACGGACCAT CGGTAATGCC ATCATCAATA GTGGGATGGC   16860
CTTCGGTACA TCAGGTGGGT ACTTACTTTC AAGTAAGTTG GTTTTAGAAA ATGGTGAACA   16920
TTGGAGCAAA CCATTCTTTA TCATGGCGAT TCCAACGTTT ATCGTAGGAA TTCTTTTCTA   16980
TACAATTTTG AAAGAAAAAG TGATTCGTCC TGGTGAAGAA GCTGCACGAG CAGCTGCTGA   17040
AGAAGGACCG CAAGAAAAGA TCTCATTAAA AGAAATTTTC TCAAATCGTA ATTTATTAGC   17100
CGCATTTATT TTATGTTTCA CAAGTATTTA TGCAAACTTT GTCATTATTA CTTGGTTGCC   17160
ACAATTTTTA ATTGCTGAAC GTGGCTTTAC AGGAGCGAGC GTTGGTTTCA TTTCATCACT   17220
TGTACCATGG GCCTCTATTC CAGGCGCCTT AATCTTCGCT CGTTTGAATG ATAAAACAGG   17280
TGCGACGAAA AAATTAGTCT TTACATTAGT ACCTTTAGCA ATCCTATCTG TCTTTGCGAT   17340
TGCTTTTGTA ACAAATCGGA CATTGTTAAT TTCTGTTTTA ATTCTTTATG GGTTAACAGG   17400
GAAATTAGCG TTAGATCCAA TTATGGTGAC TTTTGTTACC AAACATGCAC CAAAAGCGGC   17460
CTTAGGAACA ACGTTAAGTG CCTATAATTT TATCGGGATG TCAGGGTCAA TCCTAGCACC   17520
GTATGTCACA GGATATTTAG CAGATACAGC AGGCTCAATG CAAGTTGGCT TCTATCTTTC   17580
TTGTGTACTA TTAGTTATTG GTTTACTAGC TTTTGCTTTC TTAGCTAAAG ATGAGAGTAA   17640
ACCGAAATTA GGTTAATTAT CCCTTCGATA ATGACTGAAA CAATCAAAAT GrGCAAGACA   17700
AAATCTTTCA TTTCGTAAAC AAATTATTTC TTATCCGTCG ATACTGAAAA GATTTTGTCT   17760
TTGCTTTACT AAAAAGGAGC GAaTCAATCA ATGAGTTTAA ACGTAATTGC CCTAGGTGGG   17820
AATGCTATTT TAGATACGGA TCCAACAGAT GAAGGCCAAA AAGCCGTCGT TAATCACGCG   17880
GCCAAATATA TTGCAGAATT TGTTGCTAAA GGGGAACAAG TCATTGTCTG TCACGGAAAC   17940
GGACCACAAG TAGGGAATTT GCTTTTACAA CAAAAAGCAG GCGAAAGTGA AAAAAATCCA   18000
GCGTTGAAAT TGGATACTTG TGTTGCGATG ACACAAGGAA GCATCGGTTA CTGGTTACAA   18060
AATGCTTTAA CTAATGAGTT TGAAAAACGT AACATTGCTA AACCAGTTAT CTCTGTAGTT   18120
ACGCAAGTAC GTGTGGACAA AGAAGATCCT TCGTTTAAAA AACCATCCAA ACCAATTGGT   18180
CCATTCTATA CAAAAGAAGA AGctGACGCA GAAGCAGCCA AAGATGGCTC TACGTATGTG   18240
GAAGATGCTG GCCGTGGCTA CCGAAAAGTT GTTCCATCAC CAATGCCTAA AGAAATTGTT   18300
GAAAAAGAGG CGGTTCGTGC TTTAGTAGAG GCGGACGTTT TAACTATCTG TTCTGGTGGC   18360
GGGGGCATTC CGGTCGTAGC AGAAGACGGA CAATACGTAG GTGTGGAAGC GGTTAACGAC   18420
AAAGATTTCT CTGCACGTGT TTTAGCTGAA AACGTTGATG CAGACCGTTT GATTATTTTG   18480
ACAGGCGTTG ATAACATCTA CATTAATTAC AATCAACCAG ATCAAAAAGC ATTAGAACAA   18540
```

```
ATCAGTGTAG CCGAAGCAGA AGAATACATT AAAGAAGGTC ACTTTGCTGC TGGTAGTATG    18600
CTACCAAAAA TCGAAGCTGC CTTAGACTTC GTAAAAGGTG ATGACAAACG GAAAGCAATC    18660
ATTACTTCTA TCGAAAACTT AGAAAACATC GATAAAGAAG CAGGAACTGT GATCAGTCAA    18720
AAGGGGTAGT TTACCATGAA AAAACCGTCC ATGATCACCC AAATTGCAAT TGCTGTTGTC    18780
GTAGGAATCC TCGTAGGACT CTTGATTCCT GCTAGCGGCA ATTACTTGAA AATAGTTGGA    18840
GATGTCTTTT TGCGACTGAT GCAAATGGCT ATTCCTATCT TGATCTTAGG GCAAATTGTC    18900
CAAGCAGTTG GCAGTATTAA TCCAAAAGAA TTAACGAGCC TTGGGGGTAG AACAATCGCT    18960
GTTTTCGGGA TTTCCTCCTT AGCTGCCGCT TTATGGGGCG TCCTAATGGC AGTCATTTTT    19020
AATCCAGGAT ACGGTGTTAA AATGACTGGC TTCCAAGATG CTTCTATTAA AGCGCAAGAA    19080
ATCTCAATCA CTGATACGAT TTTAAATTTC GTTCCGAAAA ATATTTTTGA TTCACTGACT    19140
CAAGGATCGA TTATTCAAAT TATTGTCTTT GCATTATTTT TTGGTTTAGC ACTGAATAAA    19200
TACTTACAAA GCCATCCAGA GACACAACTC TTCCAAATAA TTGTGGACTT TAATGAAGTT    19260
ATCATTACAG TCATTCGATA CGTCATGTAT CTAGCACCTC TGGGGATTTT TGCTTTAATT    19320
GCTTCAACAA TTAGTCACTT AGGCCTTCAA ATTATTTTAC CGTTAGTGAA GTATCTATTG    19380
GTCTATGGAT TAGGAACTAT TTTATTTTTG GGCATCTGGA TTTTGGTTAT CACATTGTAC    19440
TGTAAAGTAA GTCCACTACG TTTAATTACA AACATGAAAA ATATGTCTGT GATGGCCTTG    19500
GCGACAACTT CTTCAGCAAT TACTTTACCA GTAGCTTTGG AAGAAACAGA AGCAAAAATTA    19560
GGTTTAAGTA AGCGAATTAC TAACTTAGTG TTGCCGkTAG GTATGTCATT AAACAGTAAC    19620
GGTTCAGCGA TGCATATGGC TTTTACTGTA ATGACTATCG CtCAGATGTA TCAGCTAGAG    19680
TTTGATATTA CGAAAATGAT TTACTTAGCA ATTACAGCAA CGTTTGTCTC ACTAGCAAAC    19740
GCAGTTGTTC CCGGTGCGGG CTTAGTCTCA CTGGCAGTGA TTGTyCCGCA AtGGGTTTAC    19800
CCATTGAGAG TATtGCAaTT TTgCaGGCGT CGAATGGTTT GTTGGCATGT TACGAACAAT    19860
ttTAAATGTC AATTCAGATG TTTACTCTGC TATTCTAGTT GCAAAATCAG TGGATGAAAT    19920
TGACTATACT GTCTTTAACA GTAGCAATAA GTAGTGAAAA AAGGGGAGGA AATGCAAGGA    19980
CGAAGAAGAA AATCACTTGT TTTCTTTCCC TTTTTTAAGA TTAATGATAA AGGAGCAAGC    20040
GCAATGGAAG AAATGGTTGT CGTTACACCG GAAAAATTAG AAAAGCTGAT GAAACAAAAA    20100
TTAGAAGCTG CTGGTTTACA TTCTGAACAT GCCGAAGAAG TTGCTCGACA TTTAACTTTT    20160
GCAGATGCTA GCGGTATTCA TTCACACGGG GCGGTTCGTA TGGATTATTA TGCGGAACGG    20220
ATTGCAAAAG GCGGGATCAC GATTGACCCC GAATTATCAT TTGAAAAAAC TGGACCGTCC    20280
ACTGCCATTT TTCATGGCGA TAATGGCGTG GGTCAATTTG TTTGTAACGA AGCATTAGCT    20340
GCTGCAGTTG ACCTAGCCAA AGAATCAGGG GTAGCATATG TTGGCGTTTC TCAAACGAGT    20400
CATAGCGGGA CGTTATCGTA TTATGTTAAA AAAGCAGCGG AACAAGGCAT GGTTGCTCTT    20460
TCGATGTGTC AATCTGATCC AATGGTGGTA CCGTTTGGTG GTACAAGCAA TTATTTTGGG    20520
```

```
ACAAATCCAA TCGCTTTTGC AGCACCTAGA GCAGGTCATG AACCAGTCGT GTTTGATATG    20580
GCTACGACAG TACAGGCGTG GGGGAAAATT TTAGATGCAC GTTCAAGAGA TGTTGAAATT    20640
CCCGACACAT GGGCAGTAGA CGAAAAAGGC CAGCCAACGA CTGATCCTTA TGAAGTACGA    20700
GGTTTATTGC CGATTTCAGG ACCAAAAGGA TATGGACTAA TGATGATGGT CGATGTCTTA    20760
GCTGGTTCGT TGTTAGGCTT ACCTTTTGGC AAACATGTTA GCTCCATGTA TGCGGATTTA    20820
ACGGAGAAAC GGAATCTGGG TCAGATGTTT TTAGTCATTG ATCCAAGCCG CTTTACAGAT    20880
TCTGAACGAT TTAAAGAAAA TATTAATCAA ATGGTGGAAG AACTTCATTG TGTGCCTACA    20940
GCAGAAGGGT TTCAACAAGT TTACTATCCA GGAGAAATTG GTCAGCTAAA CTATCAAAAA    21000
GCCATGTCAG AAGGCATTGA AATTCCACAA AGTATTTATG ATTATTTAGT TAGTGAGATT    21060
GTACATTACG ATAAATATGG CGGGCAAGGT GCCTTTGCCA AATAATGAGG AGCGCAGAGA    21120
GTATTCGCTA ctGCGTTGCT CTCGCCTTCT TTCTATTGGA GGAAAGTAAG ATGAAAAAAT    21180
CGTCAAAAGT AGCACTTATT TTTGCATTCT TAGTTAGTAT TTTTGTGCCA GTTGCGGCCT    21240
TTGCTGATGA GATTGGGGAT GTAGTTGCCC CGACAGGAAT CATGGCATTA GTTGTGATTA    21300
TTCCATTGAT TGTTGTTTTG GTTCTATTGT TTATGAAAGT CGATATGATT ATCGCTGGTT    21360
TAATTGGTGG CGTCTTAGCG ATGCTGATTG GTGGTATTGG ACTAGAAGAA GCGAATAAAC    21420
AGATGTTGGA AGCTATCCCG ATGATGTTAG GAATTACCGT TCCCATCATT AATTCAGCTG    21480
TTGCGATGGC GGTCTTTAAA GCAGGTAGTT ATTCAGCTGC ATTAACTTTA GCCAAACGAG    21540
GCACAAAAGG AAAAGTCGAG TATGTTTCAG CTTTTATTGT TATCTTATTA GCCGCAGCCA    21600
CATATATGTC AGGAATTGGT GGCGGAAGTG CGATGGTTAT TGCGCCATTG GCCTTTGCTG    21660
CAGTTGGCGT TGTTCCTGAA TTAATTGCAG CAATGTCTTT AGCTGCTGCG GTaTCATTTA    21720
CAACATCACC CGCTTCATTA GAATCAAGTA TTGTTTCAAA ATTAGGCGAT ATCAGTGTCG    21780
GTTCATATGT TTCAACCATG CGACCTTACT GGTTATTCTT CGTAGCATTA GCGATTGTGT    21840
TAGCCTTTTG GGGAACGAAA CATAGAAATG TTGGCTTTAA AGAGTCGGCA GATGATGAAT    21900
TTGATAAAAA GAGCAATGGT GAACTATTTA AAATTACCTT GCCAGCCATT TTCTTATTGT    21960
TTGCTGTAAT TTTTGGACCA ATTGTGAATG ATTTAATTGG TTTTGCCTTC TTTACGCCAC    22020
TTGTATACAT GATTTTAACA TTAGTTTTAA TTTTCTTATG TACAGATTTC TCAATGAACA    22080
AATCTGTCGA AGCGATGGTA GATGGGTCTA CATATATTTT GACTCGTTTA TTCCAAGTAG    22140
GGATTTTCTT AGCCTTTATT AATGTTATTG CACAAACAGG AACGTTTGCT GTAATCGCAG    22200
GTGTTGCTGA AAATGCACCA GCCTTTATTG TCGTGCCAGT TGCCATTTTA ACAGGGATTT    22260
TAATTGGGAT TCCAGCTGGG GCCtACGTAG GTTCTGTTTT AACTTTAGTG CTACCAGTTG    22320
CTGTATCGTT GAACTTCCCG CCGTTAGCAT TAGGTTTTGT AGCAATGGGT GTTGGTTTAG    22380
GTAGTCAAAT GAGTTTCGTT AATATTACTA TGCAAGCTTT ATCTTCAGGC TTCCAAATTC    22440
CAATTTTGGA TGTTGTTAAA GGAAACGTGA AATGGTTAAG TATTGCCTCT GTACTATTAT    22500
```

TAGTTATCGG TTTAATCTTT GGATAAAAAA CAATTTAGGG AGTGGAAAAC ATGGATTTAT 22560

TAATTAAACA AGTTCGTCTT CAAGACGGTG AAGCATTACA AGATGTAGCT ATTAAAGAAG 22620

GTAAAATCAT TGAAATCGCC CCAGTTATTA CTGGCGATGC CAAAGAAGTC ATCGAAGCAG 22680

ATGGTCGTGT TTTAATCCCA GGTTTTGTAG AAAGCCATTT ACATTTGGAC AAGGCATTAA 22740

TCGCTGATCG TAAACCAAAT AAATCAGGTA CGTTGAAAGA AGCAATTGAA GTAACCGCAG 22800

AATTAAAACC AACATTTACC GAAGAAGACA TTTACGACCG TGCCAAACGT GCTTTAGAAA 22860

TGTTAATCGT TCATGGGACC ACGGCGTTAA GAACGCATGC AGAATTTGAC CCAGCGCAAG 22920

GTTTTACTGG CTTCAAAACA ATTATGCGTT TGAAAGAAGA ATACAAAGAG TTAATTGACA 22980

TGCAAGTAGT TGCGTTTCCG CAAGAAGGAA TTTTTAAAGC CCCAGGAACA GAACAAATGA 23040

TGTATGAAGT AATGGATATG GGCGCAGACG TTGTCGGCGG TATTCCTTAT AATGATGCAC 23100

CAGCGAAAGA ACATATTGAT TTAGTGTTTG AAATTGCTAA AAAATATGAT AAAGATGTAG 23160

ATTTACACCA AGATTTTGCC GATGAAGCTG ATGATATGAG TATTGAATAT TTATGTGAAA 23220

AAACCATCGC TGAAGGCTAC CAAGGCCGTG TTTCAGTGGG CCACTTAACG GCATTGCATG 23280

CCTTACCTGA AGCAGAATTG GCGCCTATTA TTGCAAAAAT GGCGGAAGCA AATATTAGCG 23340

TGATGGCTTT ACCAGCAACG GATTTACATT TAGGGGCTCG TGGTGATCAG TACAATGTTC 23400

GACGTGCGGT TACGCCAATC CGTAAATTAC GCGATGGTGG CGTGAATGTC TGTATTGCTA 23460

CGAATAATAT TCGCAATGCT TTTACACCAT ATGGCACAGG GGATATTGTT CAAACAGCGA 23520

TGTTAGCCAT TCCTGTTGGG CACCTAGGTG GAGCGGATGA CTTAAAAACA GTGCTACCAA 23580

TGGTAACAAC AAGTCCTGCT CGTGCGATGG GACTAGAAGG ATACGGTGTC GCAGTTGGGA 23640

ATAAAGCAGA TTTAGTCTTA TTAGACACAA AAGTGGTTGA ACATGCGATT ATTGATATTC 23700

CAGAACGGTT ATATGTTATT AAAAACGGCC GTGTTACAGT CAAGACAGAC AAAAAAGTCG 23760

AAATCATTCG TTAAGAAAAA AAACAAAAAG CTAGGAAGGT GAAAATATGA TTAATTTATC 23820

CTGTTTTAGC TATGATTCTC TATTGAAAAG TTtATTTTTA GTTTATCTTT GGTCGATTTT 23880

CCTTGATATA TATATAATCT TCAGGAAATG CTCCTGTCTG TTTAAAATTC GGAATTTCAG 23940

AAATTATTCT TTTTTGCTAA ACTAGTTATA ACTGGTTAGT AAAAAGAAGT ATTGGACGAA 24000

TAGTGCTGAA ACAAGTGAAT GATCCAAAGT ATAAAGAAAA ATGAAACCGA TACCACAAGA 24060

CAAGACTGTT GCTGAAAATG GTAAGTGATC ATTCTGCTAA GTTTAGTGTC TTAATGTTTA 24120

GTTAACTAAG AATTGTTGGA TTGTTCTTTA GAAAGAAGGG ACAATATGAA GCGAAGTAAA 24180

TGGAAAGAAT TGATAGTAAC GGGCATCTGC CATATATTAG TATTCCCCAT ACTAATACAG 24240

ACAACTGTTT TTGCAGAAAC ATTACCAAGT ACAAAACAAG TAAGAGAAGG AACCAATCAT 24300

TCATTAACAG CAGAAAAAGC CGAAAGTGAA CAACCACAGA CAAAGGATAA ACTACATGAT 24360

GAAGAAACAC TGGCATTGTC AAAAAGTGAG TTAATCGATA ATGAGGCTAA TGTTACAAGT 24420

CAAACGATTA GAGAAAGAAT TGAGACGCCT AACCTAACTT ATCGTTATGG ATTTATTAAT 24480

```
GAAGAGGGGC AGCCAGTAAA CGCCAATGAG ATCCTTCTAC AGTATCATAG TTGGCAAGGC   24540
AATTCCCCAG ATGGCATAAA TGTGTGGGAA GGTGAAAGTC AACCAGTGAC AGCATCTACA   24600
GTGGCTAATT TAAAAGAAGT GGTAATTCCA AGTGAGAAAG TAGCCGTCTA TTCCGACATG   24660
TCAACGGTGC TTGCAGCGAG TAATCAAACA TTTTTTTTAC CAAGATATTA TACTTCTTTA   24720
AGCTTATACA ATAAGAAAGG GGAAATTGAT CCCAATTATC CGCTGCCAAC TATTTCCGAC   24780
GCATCAGGAA ACCAATATCC AACAACAATT TCGCAATTTG AATTGGAAAA AATGTCTGCA   24840
CAACAATATA GTCAGAAAAC AGGAGTAACG TTTAACATTA GCGAGAGTCA AAAACTAATC   24900
GTTCCTTTGT ACAACCAAGT GAAGGTTGAT TCATCGAATC AATCTGGGCT ATTGAATTAC   24960
TTTAAATTTT CAGGGCCGGT TTATTATCAT GTTACCAATC GCAAAGTGAC AGAACATTTT   25020
GTGGATACTC AAGGGAAACC AATCCCTCCA CCACCGGGGT TTAGACAAGG AAAGCAAACA   25080
CTTATTGAGC GTGACCCTTA CACCTTTAAA CAGAAAGATC TTTTGCCAAG TAGCTATGAA   25140
ATTGACTCAA AAACGTATCA ATTTCAAGGA TGGTATAAAG GGAAAACGAA ACCTGAAAAT   25200
TTAGAAAAAA GCGTAACGCC CAGTTATGAT ATTACCTATG ACGACAATGA TGATTTAACT   25260
GTTGTCTATA AGGAGATACC TCAAAAAAAT TATACATTTG AGGATGTCAA TGGTGTTGAA   25320
ATTGCACCAC CATCTGATTT TATTCAGGAT CACCAACAAC CAATAACTAC GGATGGCTTT   25380
CGCTATTTAG CTGGAAAAAA ACTGCCACAA CAATACAGCG TTAACGGTAA AACTTATTTA   25440
TATCAAGGTT GGTATCAAGA TAAAACmAAA CAAGAGAGCT TAGAAAAAAC gAAGCGACCC   25500
ATAAACTCCC CTGTTTTTAA TGAAATGAAC GCTATTACAG CAGTGTATAA GGAAATAACT   25560
GCAAAAGCTG AAATGCAAAT AGAAGGACTA GTCAAAGTCA TGCCAAGTGG TTATATACAA   25620
ATTTGGCAGA TTATGCTTAC AAATGTGGGA GAAGTACCGT TAAAAAAAAT AAACTTAAAG   25680
CCAGCAAGTG GTTGGTCACC AGGTCTAGCT CGGCCAATCC AAGTCACGAT TCGTGTTGGA   25740
TCTGAACCAA ACAAAATTGT TCCTATTACT GATGAAAATT GGCGAGTTGG CATTACTTTA   25800
AATACGGAAG TGCCTATTGG TCAGACAGCA ACTATTATGA TGACAACAAT TGCTACAGGT   25860
GAACCAGATC AAGTGTTACA AGCGGCTGTT GAAATGAATG GAAATTTTTC TGCTGTTCAC   25920
GCAGCTGATA CTGTCAGAAT CCAACCTAAA AATCAAGAAA TTGTGGCACC AGATGAGGAA   25980
GGTTTTATCA GCACACCAAC TTTTGATTTT GGCAAAGTCG CCATTTCTAG CAACACGCAG   26040
CAACATGGTT TAAAGCAGGC AGCAGATTAT TATGAAAATG GTCAGGAAAA TCCATATTTA   26100
CGTTTGAAAA AATCACAACC CAATTGGGCA CTAACTGCAG AACTATCCCC CTTTGAAGGA   26160
AGAGTGGATC AACTATCATC AATGACAAAG TTATTGTTAG GAACAACCAA TGTTTCAGGT   26220
TTTATTCAGT ACAATCAACC AACGGAAACT AAAGTTGCTC TTGGCAAAAC AACCGCTATT   26280
CAATTAGTTG CCAACGGTGT AGCTAGCCAT ATTGTTGCCA ATGGTCAGTT TGACGAAAGT   26340
GATGTTTATC AATTTGATTT TTCTTTTGAT CAAATCAAAT TAGAAATTCC AGCAAATCAA   26400
GGTAGAAAAG ATCAAACTTA TCAAGCAATG GTGACTTGGA ATTTAGTGAC AGGCCCATAA   26460
```

```
GTCCATAAAA AATAAGTTGA ATAGTTTGAT CAAAAACATT GCATAGAAAT AATAAAAGGC    26520
GGCTCAACGA AGCGGCCTTT TTCTTGATTC GTTACCAAAA TATAACGAAT TTTGTAACAA    26580
AAACGTTTAA TAAAAAGAAG GAAAAATGAT TAATTTTACG AAAAAAAGCA TATATCGACA    26640
AAAAAGAAGC AAAATTTGAA TAAAAAAAGA AAAAATATGA AAAAAAATCG AAGAATTGTT    26700
TAACAATTGA AAACACGTGA AAACTAGTCG TAACAAGGTG TAAAAGCCTT ACAAAAGGTA    26760
AGTTTACCTA TTATGGAGAA AGAAGGAGTG ATCTGAAATG TTTTTGTAAC ATTTCTATTG    26820
CTTAATTGAA ATATGGGTGT GGTAGAATGT TTTTGTCGTA AGAAAAGATT AAGATATTTT    26880
AAATAAACTT TAGAAaTTAA TTCGGAGGAA TGACAAGTGA AGAAACGTCT ATTTGCATCA    26940
GTATTACTAT GTTCATTAAC GCTATCAGCA ATTGCTACCC CAAGCATCGC TTTGGCGGAC    27000
AATGTTGATA AAAAAATTGA AGAAAAAAAT CAAGAAATTT CATCATTAAA AGCAAAAACAA   27060
GGGGATTTAG CTTCACAAGT ATCTTCTTTA GAAGCAGAAG TATCTTCAGT ATTTGATGAA    27120
AGCATGGCTT TACGTGAACA AAAGCAAACA CTAAAAGCAA AATCAGAACA ATTACAACAA    27180
GAAATTACAA ACTTGAATCA ACGTATTGAA AAACGTAACG AAGCAATCAA AATCAAGCA     27240
CGTGATGTTC AAGTTAATGG ACAAAGCACA ACAATGCTAG ATGCAGTTTT AGATGCGGAC    27300
TCAGTTGCAG ATGCAATCAG CCGTGTTCAA GCTGTTTCAA CAATCGTAAG TGCCAACAAC    27360
GACTTAATGC AACAACAAAA AGAAGACAAA CAAGCCGTTG TTGATAAAAA AGCTGAAAAC    27420
GAGAAAAAAG TGAAACAACT TGAAGCAACA GAAGCTGAAT TAGAAACAAA ACGTCAAGAT    27480
TTACTTTCTA AACAATCTGA ATTAAACGTA ATGAAAGCTT CATTAGCATT AGAACAATCA    27540
TCAGCTGAAA GTTCTAAAGC TGGCTTAGAA AAACAAAAAG CAGCTGCTGA AGCAGAGCAA    27600
GCACGCTTAG CTGCTGAACA AAAAGCTGCA GCTGAAAAAG CCAAACAAGC TGCTGCAAAA    27660
CCAGCTAAAG CTGAAGTGAA AGCAGAAGCA CCAGTTGCCT CTTCATCAAC AACAGAAGCA    27720
CAAGCACCAG CAAGCTCAAG CTCAGCAACT GAATCAAGCA CGCAACAAAC AACTGAAACA    27780
ACTACACCAA GTACAGATAA TAGTGCAACA GAAAATACTG GCTCTTCTTC ATCAGAACAA    27840
CCAGTACAAC CTACAACACC AAGCGATAAT GGAAATAATG GTGGCCAAAC TGGTGGTGGA    27900
ACAGTTACAC CAACACCAGA ACCAACACCA GCGCCTTCTG CTGATCCAAC AATCAATGCA    27960
TTGAACGTTC TACGTCAATC ATTAGGTTTA CGTCCAGTAG TATGGGATGC AGGTTTGGCA    28020
GCTTCTGCAA CTGCTCGTGC AGCACAAGTT GAAGCAGGTG GCATTCCAAA TGATCACTGG    28080
TCTCGTGGAG ATGAAGTTAT CGCAATTATG TGGGCGCCAG GTAACTCAGT AATCATGGCG    28140
TGGTACAATG AAACAAACAT GGTAACAGCT TCAGGAAGCG GTCACCGTGA TTGGGAAATT    28200
AACCCAGGTA TTACGCGTGT CGGTTTTGGT TACTCAGGTA GCACAATCGT AGGACACTCA    28260
GCCTAATCAG CAATGATAAA AATAGACGTA AGCAGCTTCT GAAAAGAAGC TGCTTTTTTC    28320
ATTACTATTT CATCCGTCTC CGTTGACAAA TTTGCTAAAC GTGGTATTCT AATTATAATC    28380
TAATAATTTC CGTTGAATAG GAAAGTAAGT ACAAGACAGC TGGTAGAGAG ACTTTGGTTG    28440
```

GTGAAAAAAG TTAGATGTCT GTATCGAAAA ATGGCCTAGG AGGAATTTTT GTCGAAACGT 28500

AGGCAAAAAC GTAATGTCTT TCGTTAACAA GACCGCTGAT TATGATTGTC AGTTGAAGGG 28560

AAAGAGTCAT CTTTCCGAAA AAAGGTGGTA CCGCGATAAT AATCGCCCTT TTACTAGTTA 28620

CGGCTAGTAA AAGGGcGTTT TTTTATAAAT AAAAATAGGA AAGAGTGAGG CAAATGAGTA 28680

GTCAGATTCC TTTTACAGTT GATCATGTGG GAAGTTTTTT ACGTCCGGAA ACAATTAAGA 28740

AAGCACGGAA AAAGGTCCAA GCAGGTGCTT TATCTCAAGC AGAGCTACGT CTTATTGAAG 28800

ATGAGGAAAT CAAAGCATTA GTAGCAAAAC AAAAAGCTGG CGGCTTACGG GGAATCACAG 28860

ATGGTGAATT TCGTCGCGGT TTTTGGCATA TTGATTTCTT AGAAGGATTA AATGGTGTGA 28920

CAGGCTATAT CCCAGAAACG GGGTATAACC AAACCTTTCA TGGAAAAGCA GCTCCGGCCT 28980

ATAATATCCG CGTAGTGGAT AAAATATCTT TTCCTACGCA ACATCCATTT TTAGAGGACT 29040

TTTCATTTTT AAAAGAGGTG GTGGAAGTTG GTGATGGCAC GGTAGCCAAA GCAACGATCC 29100

CTAGTCCGAC CATGATTTTA CGTCAAGAAA TTTTATCTAA CGATGGCACC TCGCGCATTC 29160

AAACAATTTA TCCAGAGATT GGTGCTTTTT ATAAGGATTT AGCACAAACA TATAAAGACG 29220

CAATTGCGGC CTTTTACGAG AGAGGTTGCC GTTATTTGCA ATTTGATGAT ACAAATTGGG 29280

CGTTTCTAGC GGATCAGACA AAACAAGAAG AATTACGAGC AAAAGGAATA CAGCCAGAGG 29340

AAATTGCACA AGTGTGCACC ACGATTATTA ATGAAGCCTT AGCTGGCAAG CCAGAAGATT 29400

TAACGGTAAC CACACACATT TGCCGTGGGA ACCATGCTTC GTCTTGGTTA TTTTCAGGTG 29460

GTTATGAACC GATAGCAAAA GAATTATTTG CTACGAATTA TGATGGTTTC TTCTTGGAAT 29520

ATGACTCAGA TCGTGCAGGG GACTTTTCAC CATTAAGACA TTGGCAAAAT AATGGCAGCA 29580

AAGTCGTTTT AGGTTTAGTT ACTTCTAAGT TTCCTGAATT AGAAGAGACT GAACAAGTGA 29640

AAGCACGAAT TGAAGAAGCC CAACAGTTTG TTCCTTTAGA AAACTTAAGT ATTAGTCCGC 29700

AATGCGGTTT TGCTTCAACG GAAGAAGGAA ATCGTTTGAC GGAGAAAGAA CAATGGCGCA 29760

AAGTTCAATT ACTTCAAGAA ATTGCTAGAG AAGTATGGTA AAAAAACTGG GACATAAGTC 29820

TAAACGGCTT ATGTCCCAGT TTTTTCATTG TATTATTTTA AATATTGATC TTTAGTAAAA 29880

TTCGTGTGTT CATACTCGCC GTTGTCACCA TAAAAGCGGT TATAGCGTTG TTTGAATTTG 29940

CGGAATAAAT AAATGACTGT CACTTTAACT AGGCAATACA TTGGAATCCC AAAAATAACG 30000

CCCATTAAAC CAAACAGTTC ACCCATAACT AACAAGACAA GTAAAATAGT AATCGGATGG 30060

ACTTTTAATT TATCGCCCAT CACACGAGGA ATAACAAAAT CACCATGAAT CAATTGGATG 30120

GCAAACCAGA CGATAACGAA TTTTACTACC ATAAATGTGG AATCCTGAAA GGCAATAATC 30180

AATCCTGGAA TAAAGGCAAT AAATGGCCCG ATATAAGGAA TAATACATAA AATACCCGCA 30240

GCTAAAGCTA AAATGCTACC AAATTGTAAA CCAATTaATA GAAAAGCTGG cCAATACATA 30300

ATCCCTAAAA CCAACGAAGC AATGACTTGA CCTTTTAGAT AATCGCCAAT TTGAATGTTG 30360

GCAATTTGGC TATATTCGTC AAAATCAGGA CGAAAAGCTG GGGGAATGAT GCCTTTTACA 30420

AAGCGATaAA ACTTTTCTTT GTCTTTTAAA AGGAAAAAGG CGATGATAGG TCCTGTAAAC 30480

AAGGTTAGGA AAGTAGAAGA GACTGCAGAA AAGACGCTTC CCAGACCTTG TGCGCCTGTT 30540

TGTAGGTAAT CCCCTGCAAA GTCTTTAAAG GAGTTTGCAA GTTTGCCCCA GAGCTGATTA 30600

ATCGACTCAT TAGCAGATTC TAAATCTTTT GCAAAAGGTG TTTTAGCCAT AAATTCTTGT 30660

ACTGTTTGGA GCGTGCTTTT CCAAAAACTT GGGAATTGTT TGATTAATTC TTCCGTTTGG 30720

TCTTGGATAA TTGGGaAAAG TTGCAGTCCA GCTAACACTA GTAATGCAAT AATTAAGACA 30780

TAGAGAATGc GCAATCGCCC aAACACGGGG AATTTTTTTC TCAAGGCGTT TCACCATTGG 30840

ATTAAGCAAA TAAAwTAGAA TAATACCAAA AAGTACAGGG CCTAAGACGG CACTAAAAAT 30900

AACAAAAATT GGTTTAAAGA TAAATGAAAC AGTGTTGAAT AAAAAGATAG TTACAGCAAT 30960

TAGACAAAGG ATAGTTAATG TGTAAAGTAA ATTTTTCCCA CCTAAAAAGC GGATCCACTT 31020

GGTACTTTCT TTATCCATAA GTTGCTCCTT TCGTTAAGCT TTATTTCTAA AATGACCAAT 31080

AATTTCAATG ACACCCATGA AAATCAGGAT AAATCCAAAG ACTTGCATCA AAACTAAGAG 31140

ACTTTTGAAG GGATTAAAAA CTAAAACGGC TCCAGCGATT AAAACCAGAA TACTGTAGAT 31200

AAGCGCGCCT GTTTGACGAG TATTTAAAGC AAAAGTsAAT TGAACGAGGC CGTTTATAAC 31260

AATGCTAATC CCTAATAAAA ACGGCAGAAT CGAAACGATC ACTGGTGCAA AAAATAAAAA 31320

GGCTACTGCT AAGACGACAA AGATTAAACC TGTTACGAAA CCAAGTCCCC AGCTACCCGT 31380

TTGTTTTTTT ATTTTGTAGT CATCATAAAT ATTAATTAAC CCTAAAAGCA TCAAATAACC 31440

AGCAATTAAA TAACCGACGA ACTGGAAGAA AGCACCGGGA TTAAAGACAA TGGCACTTCC 31500

GATAACGATA aAGACGG 31517

(2) INFORMATION FOR SEQ ID NO: 181:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11879 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

AGGTAATAGA CAGTGCAAAA ACAATGATTA GAAAGGTAAA TTCCATCCTA TCATTCCCTT 60

CTATATTAAA TTTTTATTGA CTTACCGATC AAAACATTTA AGTAATTTTT TCTTTTTAAG 120

AATAATTTTC TTTTCAACTT CTTCTAACTC AGAACTGCGT TCGCAAAGTT CCAATTGAGC 180

GCGTCgTGCT TCTAAACGAA GAAGTTCTCC TTTAATTAAT TCAATTTTTT TGTAGCGTAA 240

GAAATTTTTm TTTTGCATAC GTTCTGCAAA GTGTTGTCGT TTTTCTTCAG GGATTGCTTG 300

ATATTTCnTT TTCAAGTCAG CGACACACTT TTCAAATTCT TCACTCACCG AAACAACTTC 360

CTTCCmAAAT TTAATCkGcT AwTTATwACA aATTTTACCA AAAGTACAAG CAAAATTGCT 420

TTCTTTTTTc AAAAGAAATT CAGACAAAAC TtAACgCGGG AAAAGCTGAT TTATAGGAGC 480

TGAACATGGT TAATCTTCAT TAACAATAAT AcAATTTCTG CCcTTCATTC ATCCGTGTTG 540

```
AAAGAAAAAG AACCATTTCT TCTAGTGGCG CAGAAAAGCG TAAAACCACC CCTCTTCTGC    600

CTTTTTTTAA ATACCCTAAA CGAATACCGT TCGGCTTTTG AACAAGGACA ATCTTTTCGA    660

TATCTTTAAA GGCCGTGCCC CGATTGTCAA AAGGATGCGT AATAATACAG TCCTCCGCAA    720

AACCTCTTGC ATCGTAAAGA AAACTGAAGA TAATCAAAAT ATATAACACA CTTCGAATCA    780

TTTCAATTTT ATCTCCTGCC GAAAATGCAT TGGCCAGAAT CATTAAACTA AAAAAAGCAA    840

CTAAATATTT TGTTGTAGGa TGTTTTGCAG CCGGaATTAC AATACTTTTA CGTACGAAGT    900

AAAAGTAACC ATACATAACG ACGACAAATA AAAATAGCAA TAACGTAAAT CCATCGATAA    960

TCACAAACGA CTCCTTTATT TCTTtTTCTT TTGAATATCT GTTAATATAT CATTTATTTC   1020

TTTAATTtCT TTTTTkGATT CGTTAATCTT TTkGATAACC GCCTCACGGT TTTTTCCAGG   1080

CGGTAAGTCA ATTGaTTCTT GGTAGTAGTT TGCTAAAGCT TGAACTTTTT TCATTTTTTC   1140

GCTTTGCAAC AAGGCTATAT TTCTTTTCAT TTAACCGCCT CCTCTTCCTT AGTATACATT   1200

ACTTTGTAAG GAATTGCTTA ATTTGCAAGA TATTTTCTAA TAAAAAGGAG TAATTTTACA   1260

TTTTCTTATA AACAAAAAGT GTTCGGCTGA GTAACACCAG CGAACGATAA AATAATAGAA   1320

TAAAAGATGA GACAGAAGTA GTTTACTTTT GTCTCATCTT TTATTATTGA GCTGAAATCG   1380

CTAAACCAAT TGCTTTTTCC CCTAGGTGGG TCCCGATAAC TGGACCAAAA TGACCAATTT   1440

CAATTTCTGC ATTTGGGTAT AGCTTTTGTA ATTTTGCTTG TTCTTTTTCA GCAACGATGC   1500

GGTTATTGGC ATGAATCACA TACAGTTTAA CTGGTGCTTC AATCCCTGCG TTTCGTTCAC   1560

CAATAATCTT TTCTGCACGA GCAAAAGCTT TCTTTGTTGA ACGGATTTTT TCAAATAATA   1620

CAATCTTTCC ATCTTCAAAA GTCAAGATAG GTTTAATCTT CAATAGTCCA GCGATCAAGg   1680

nTGCGCCATT CGTTAAGCGA CCGCCACGAA CTAAGTTGTT CAGATCTTCT ACAATTAGAT   1740

ATGCATACGT ATTGTCACGA ATTAAATCTA ATTTGGCAAA AATTTCTTCT AAGCTGGCTT   1800

TTTCTTCTGT TAAATCTAAA GCAGCTTCTA CCATGTGTCC CATTGGCATA CTTGTAATTT   1860

TTGAGTCATA TGGATACAAA GCAACGCCTG GGATTTCATC GGTAAGTCCG TGCAATGTAT   1920

GAaCAAaaCC mGAGATtCCT GaAGAAAGAT GAaTGCTGAT GATGGTGTCG TACCCTkGTT   1980

CAGCGATTGA TTTGTAAAGC TCTAACACTT CTCCTAAAGC AGGTTGTGAA GTCGTCGGAA   2040

ATTCTTTACT ATTATTTAGC AATGCATAAT ATTCATCTGC TTCAATGTCA ATGCCTTCGT   2100

TGTATATTTT TCCATCTAAA ATGACTGGGA TGGGAATTAC AAAAAGATTC GGATGATCTT   2160

TAATGCGCTC GGGTAAATAA GCTGTACTAT CTGTCACAAT AGCAATTTTC ATTTTGTAAC   2220

CTCGTTTCAC TAATTTTCCA CGTTGTTTCT AATATAGCAT AAAAAAGGAT ATTTTAAAAk   2280

GCTTTCCTTA TTCGTTCAGG AATTTTTCGA TGGCGTCAGc ATTTTTTTGT TCATCTACTT   2340

CTAAAATACT GCCTGCATAA GAGGATTCAC CAAAATTCCA AGAATCTGGT ACCGGAACCG   2400

ATAATGACTC AACCCCTGCT TTTCCTTTAA TCAACAATGA TGGTCCATTC GTTAACATGA   2460

AACTAACAGG CACATCTGTC GACATATAGC CGACTAATTT CCCAAGTGAT TCTGGGGTTC   2520
```

```
TTAATAAAGT CATTGGGTTT TTCATTTGGC TCATTACAGC TGACATAACT TGTTGTTGGC    2580
GACGAACCCG ACCAAAGTCG CCTTCTTCAT CCATCCTAAA GCGAGCGTAT TGAAGTAAGA    2640
CATGTCCATC CATGACCTGT TGCCCTTTTT CAATATCAAC ACCATCTAAA TTCAGTGATT    2700
TTTCTGCATC GATTTTGACA CCTTTTGGAA ACATAGAGTC AACAATTTTT TCAAATGATT    2760
GGAAATCTAC CTTAGCATAA TATTTCGTAT CTAAATTAAA GTTTTGTTTT AACGTTTCAC    2820
GAACCAATTC AGCACCGCCA TAAGCGTATG CGGCATTAAT TTTATTCGGC CCGACACCAG    2880
GAATGTCCAC GAATGTATCA CGCATAAATG AAATTAATTT CGGTTTTTTT GATGGTCCAT    2940
TTAGTTGGAG AACCATTATT GTGTCGGCTC GTCCAGCATC TTCCCCACGT GTATCGCTGC    3000
CAAGAATTAA AATATTCTTA GCCCCGTTGC TACTTTTGAC ACCATTAAAT GTTTCTACTT    3060
TTTCTTGAGG CAACGAGTCA TCATGTTCTG CTGCAGATTT TCCTTTTAAA AACATGACTG    3120
TACTATAGGC AAATAACAGA ATCAATAGGA TCACTAGCCA TTTGAAAAAG CGTTTTGTTT    3180
TTGATTTTTT ACGTTTCTTT TTTGGTTTCT TGTCTTTTTT TCCGCGGGGT GGTTGGTCAG    3240
AACCGTTGCC ACCACCTGAA TCTGTTTGAT GTCGCCTCTC ACTTCTGCGA CTGTCCTCAG    3300
TATAGCTGTC TTCGTAAACT GGCkCTTGAA GGCTTTCAGC TAGTTCTTCT GGTTCTTCTT    3360
CTCTATGTTC ACCCATTGAT TTTCGGGGAT TAAAGGTCTT TTTATGCTCT GCTGGTCTCG    3420
ATTTTkCATG AATATGTTTA TAACGATCTA CACGACTCAT CGTGTCACTC CTTTACCTCT    3480
GTGTTTTATC ACCTTTCTTA GGATACCTCA TTTCACAGGA AATTCCTATT ACAAATATGT    3540
AATTTAATAA AAAATGAAAC TCACTTTATC TAATAAGTAA GCTTCATTTT CATTTTTAAC    3600
TCTTTAAAGT GTGGACAAAA CATAGATCGG GTAATCCCCT AAAATTTTTA TTTCTGCACC    3660
AATGAGTTCC AGTTCTGTTA AGGCTGCTTC AATTAATTTT TCTGGTTGAT CTTTCACTAA    3720
GTCCATTAAA AAGAAGTACT CTCCTAGCTT TGTTTTCAAC GGCCGCGATT CTATTTTGCT    3780
AAGATTAATC CCTCGCCAAC TAAACACGCT TAATACTTTG TGTAAAGAGC CAGGAACATT    3840
ACTTGGCATC GTAATCGCCA GTGTTATTTT TTTCTCAGAT AGCGGGAAAG AAATCGCTAA    3900
ATTTTCAGAA CCAAGAACCC AAAATCGGGT TTGATTTACC GATAAATCTT GTATATTTTT    3960
TTCAACAATG GTCAAATCAT ACATTTCCGC AGAAAGTTTT GGTGCAATAG CTGCAAAAGG    4020
TAATTCTGGA TGTTCTGCAA TGTATTTGGC GGCGTAAnCT GTTGAAGGTG TTGCTTCTAA    4080
AATCGCTTCT GGAAAGTTTT TCTCTAGAAA CATCTGCGAT TGAGCTAATG CTTGCGGATG    4140
TGATAAAATT TTCTGACTTT GTTGCCAGAT CGCTTGATTC TCTTTTGCCA CCATTAATTG    4200
TTGTTGAATC GGTAAAACTA ACTCTGCTTG GACAGGTAAC TGCGCTTGAT GATACAAATA    4260
ATCTATCGAT GCGTTAACAG TTCCTTCAAT CGTGTTTTCT ATTGGGATAA TGCTCCATGC    4320
CACTTCTTGC TGTTCAATTG CTTTCAAGCA AGCTGGAATC GATGCGTACG GCATCAACGT    4380
AGCTTCAGGA AAAGCAGCCA ACGTTGCACT GTACGTAAAG GAACCAATCG GACCTAAATA    4440
ACCAACTTTC ATCGTTTCTC TACTCCTTTA ATTGTTGCAG AATTTCATTG ATAATTTCTT    4500
```

CTGGCGAGCG ATTGGTCGTA TCAATTGTCA TCTTCGCACA TTCTTCATAA AAATTTTTAC 4560

GAGACTCAAA CAAAGTAATG ATTTCTTTTG AAGAACGTTC TATAGCTAAG GGCCGTTGGT 4620

TTTCAGTATC TTCTGTGATT CTTTTTAACA GCTCTTCTGG TGTCGCATGT AAATAAATCA 4680

CTTGCTGAAA AGATTTTAAT AAGCTACGAT TTTCTGGTCC GACAACAATC CCGCCCCCAG 4740

TTGAAAGGAC GGCTGTATTT TTTGACAGCT CCTTTAAAAG TTGGGTTTCC TGTTCTCGGA 4800

AAGCTGCTTC ACCATATTTT TCGAAATAGT CAGGAATTGA GCGTCCTATT TTTTCAATTA 4860

ACGCTGTATC TAAATCAAGA TGAGGCATCT TCAGTTTATT GGCCAAACTT TGGCCGATAG 4920

TTGTTTTACC CGCACCCATG AAAcCAATTA AAACAATGCT TTCCATGGTT CGTTATCCTC 4980

CTTCGTTAAC AACTTAAACG TTCTAAGTCG TCAAAAAATG CTGGATAAGA AACTGAAACT 5040

GCTTCAGCCT TATCTAATTC AACAGTGCCT TCTTTTACAA GTAATGCAGC AATTTGTAAC 5100

ATCATCCCGA TACGATGATC CCCATAACTG GTAACTCTTC CACCATGTAA AGAAGTTGGT 5160

CCATGTATAA TTAAGCCATC ATCAGTAGGC GTGATGTCGG CGCCTAAAAT TGTTAATTCT 5220

TTCGCTACTG CATCAATCCG ATTGGTTTCT TTGACTTTCA ATTCTTCTGC ATCTCGAATG 5280

ATTGTCGTGC CAGTAGCCTG AGTAGCTAAC AAAGCAATAA TCGGTAACTC ATCAATTAAA 5340

CGTGGGATAA TAGCGCCACC AATCTCTGTA GCTGTTAATT GACTCGTTTT TACAAGTAAA 5400

TCGCCAGAAT GATTGGCCTC ATCTTCATTT AAAATAGTGA CGGAACCGCC CATGTTTTTA 5460

ATCACATCTA AAATACCTGT CCGCGTTTGA TTTAAGCCAA CATTTTTCAG AAGTATCTCG 5520

CTATCTGGGA CTACTAAACC CGCAACTAAA AAGAAAGCTG cAGAAGAGAT ATCACCAGGa 5580

ACTACCACAT TTTGACCTGT TAATTGTTGC GGTCCAGTTA ACATAATTTT TTTACCGTCT 5640

ACTTCAAGTG TCCCACCAAA TTGTCGAATC ATCTCTTCTG TATGATCACG TGTCTTTTCT 5700

TTCTCAACCA CTACAGAAGT GCCCTCGGCT TGCAAAGCGG CGAATAAAAT AGCCGATTTA 5760

ACTTGAGCAC TTGCAACAGG CATTGTGTAG TCAATCGGTT GCAAATTTTG AGTCCCGCGA 5820

ATAGAAATTG GCGGAAACTC CGTTTGCTGA ACCCCTTGAC ATTCCGCTCC CATTTGGTTT 5880

AAAGGAAGCA TTACACGATT CATTGGTCGT TTGGCAATAG ACGCATCACC AGCTAGGCGC 5940

GTTTCAAAGG GACAGCCAGC TAAAATGCCC AGCATCAGAC GAATTGTTGT CCCTGAATTT 6000

CCAACATCAA TTGTATTCTT CGCCTTTTTT AAGCCTGCAA ATCCTCGCCC CTCAACGGTG 6060

ATTGTCGTCC CGTCATCTTC AATGTTCACA CCTAAAGAAC GAAACGCCGC TAAGGTACTT 6120

AAACAATCTT CGCCTCTTAG AAAATTTGTA ATCGTCGTTT TTCCAGAAGA AATCGCTCCA 6180

AACATAATAC TTCTATGGGA AATCGATTTG TCGCTAGGAA CCATCAGAGT CCCTTGTAAA 6240

TGCTTCACAT TTGTACGTAG TTGCACCCTC ATCACCTCTT ATTTTTCGTA ACACGTATAG 6300

TTCGTTTTTT TAGATAAAAT TTGTTTTGCA GCTTGGCAGT CTGGTTGATT TTTAAAAGAC 6360

AATTGAAGAA TCCCATAGAT TTCTTCTCTC GTTTCTAAAA TTTTAATATT TGTAAGAGAA 6420

AGGTCCGCTT CGCCTAAAAT TTGCGTAATT TCAGCAATGA TTCCTGGTTG ATCTGGTACA 6480

TCAACAAACA GATCATAGAA AGCTGGGATT GCGCCTTCTT TATGAATAGG CAATTGAGCA 6540

CGTGTTTCTT TGGCCTTATC AAAAAAATTA CGAATAGCTG GGGCATTTTT TTCTGTTAAC 6600

CATTGGCAAA CAGTAGTCAT CTCTTTTTGC CAATTTTCTA GTAAGTCCAA TAATACTAAG 6660

CGATTGCTTA ATAAAATATC CGTCCACATC GTTGCATCAG AGGAAGCAAT TCGAGTAATA 6720

TCTCTGAATC CTCCAGCCGC TAGCTGCTGC GCTCTAGGGT AAGTGGTATT CAGTTGCTGA 6780

CTTTCGTTCA CTAACGCTGC GGCAACAATA TGTGGCAAGT GACTTAGAGC ACCAGTAATT 6840

TCATCATGTT CCTGTGCAGG CATCGTAATA AACTTCGCAT GCGTTCCTTT TAGTAACGTT 6900

TGTAACTCCT GAATCTGTTT GTTTTTTTCG CCATGGTCAT CGGTAAAAAT ATAGTAGGCA 6960

TTTTCAAACA AACGTTCATC AGCGGCTGTT ACGCCGGATT TATGTGAACC AGCCATTGGA 7020

TGACCACCAA TAAAATTTTT CATGTTAAGC TTTGTTGCTA ACTGATTAAT TTCCACCTTG 7080

GTACTACCCA CATCTGTGAT CAGAGCAGTT TCTAGTGATA ATTGGTTTAA TTCTTCTAGT 7140

TGTACTAAAG TTGCTTTAAC TGGGGAACAA AGAAAAATAA TCTCTGCTCG TCTTGCCCCA 7200

CTTGTTAAAG ATTCAGCTAT CTCATCAATT AGACCCGTTT TCTTAGCAAA TTCAGTTGCC 7260

TCCGCTTGAT TATCGAAACC GATAATCTCA CTGTTTGGAT GCCCTTTTTT GATACACAAA 7320

GCCAGTGAAC TCCCGATTAG CCCTAATCCT ACGATTAAAA TACGTTTCTT CATTTTTTCC 7380

CTCCATATAC TTGCGACTTT CTTCCATTAA AAGTTTTGAG TATATAGACG ATAACGTTTC 7440

ACTGCTTCTT TCATTTGTTC AAATGAGTCA CTATCAAATT TTTCGAGCAT AGCCTTTGCT 7500

ACTTCCGTTG CAACAACGGC TTCACAAACA ACGCTAGCGG CCGGTACCGC CGTGCTATCA 7560

GAGCGCTCAA CACTGGCCTT ATAAGGCTCT TTTGTATCAA TATTCACGCT TTGTAATGGT 7620

TTATAAAGGG TTGGAATAGG TTTCATGACA CCACGAACGA TGATTGGCAT TCCGTTGGTC 7680

ATTCCTCCTT CAAAACCGCC TAAATTGTTG GAAGTTCTGG TATAACCAGT ACTTTCGTCC 7740

CAAACAATCT CGTCCATCAC TTGACTACCA GGGCGTTGTG CCATTTCAAA TCCAATGCCA 7800

AACTCAGCAC CTTTAAAAGC GTTGATGCTT ACAACTGCTT GCGCAATTTT GGCATCTAGT 7860

TTACGATCCC ATTGGACATA GCTACCTAAG CCAGCTGGAA CGCCACCCAC AAGTACTTCT 7920

ACTACCCCAC CAATTGTATC GCCATTTTTC TTGGTTTGGT CAATTAGTTC TTTCATTTTT 7980

TCTTCTACGG AAGGATCTAA TACGCGAACG GCAGATTGTT CAGATCGTTC TTGAATTTCA 8040

CGAATCGTTA AATTTTCAGG GATCGTAGCT TCAATCCCAC CTAATACCGC GACATGCCCA 8100

GCGACTTGGA TATCCAGTTC AGCTAAGAGT TTTTTAGCAA CCGCACCAAT CGCCACACGC 8160

ATCGTTGTTT CTCGTGCCGA AGACCGTTCT AAAACATTCC GTAAATCATC ATGTTGATAT 8220

TTCATGCCAC CGACTAAATC AGCATGTCCT GGACGTGGTT TGCTGACGCG GCGGATTTTC 8280

TTTTGTTTTT CAGGAACTGG CTCTACTGAC ATCACGGAGG TCCAATTTTT CCAGTCTTTG 8340

TTTTCGACAA TCAACGTTAC TGGTGAGCCA AGTGTTTTAC CATGCCGAAT ACCCGAAGTA 8400

ATACGTACTT GGTCTTTTTC AATTTTCATC CGTCCCCCAC GACCGTAACC GCCTTGACGT 8460

CTTGCTAATT CTCGGTTAAT CTCTTCGCTA CTTAAAGGCA AGCCGGCTGG TAAGCCTTCA 8520

ATAATAGCAG TTAATTCAGG TCCATGTGAT TCGCCTGCTG TAATAAAGCG CATCTTAAAT 8580

TCCTCCTTCT CTGTTTAAAA ATGCTCGAAT TTCTTCCGTT GGTAAAGAGA CAATTTTCGC 8640

TTGACCAATT TTTTCTAAGA CAATGATTTT AATCTGTCCC CCACGTGTTT TTTTATCATG 8700

TGTAATTGCT TGATAGAGAC GTTCTTCTGA CCAACGATCT GTGGTTACTG GTAAATGGAA 8760

CTTTTCTAGC ATGGTGACCA ACTCTTGGGT AGTCCCAAGT GGGGAAAGCC CTTGTTGTTC 8820

TGCGACTTGA GTTATTTGAA TCATTCCTAA AGAAACGCCT TCACCGTGAG CAATCACCCC 8880

GTAACCAGCT GTATTTTCTA ATGCGTGGCC GATCGTATGC CCAAAATTCA GAATCAAACG 8940

TAAGCCCAAA TCTAATTCAT CTTCTTCGAC GACATCACGT TTAATCTTGC AACAGGCCGT 9000

GATAATTTCT TCTGCATGTG CCACTAAATC TTGTTCATTT TCCAAAGAGG ATAAACGGTG 9060

CCACAATTCA ACATCAGCGA TAGCTGCTGA TTTTACAATT TCTGCAATTC CTTCACGCAC 9120

ACGTCTAGGT TCTAATGTTT TAAGTGTATT AGGATCAATT AAAACCCCAT CTGGTTGGGC 9180

AAAAGTACCG ACAAGATTTT TGGCTTTTTT AGTATTAACC GCTGTTTTAC CTCCAATGCT 9240

ACTATCTACT TGTGCCAGTA AGGTTGTTGG TACTTGCAAA AAGTGAATAC CGCGCATATA 9300

GGTTGAAGCG ACAAATCCAG CTAGATCTCC CACAACGCCT CCACCTAAAG CAATTAGACC 9360

ATCACTTCGA GTCAATTGCT GTTCCGCTAA AAAATCATAA AGCAGTTGAG CTATTTCGAG 9420

GCTCTTACTT TGTTCACCTG CCGCTACAGC AATCAATGAT GTTTCAAAAC CAGCAGCTTG 9480

CAATTCTTTC TCAACAGcTG CGCCATATAA TTTATTCACC GTTTCATCGG TAATAATCGC 9540

TACCCGTTGT GGCTGCCACA GGCTACGTAC CCAGGTGCCA ATTTTATCTA AGGCACCTGT 9600

TTCGATGGTT AAATCATATG AATGTGTAGG TAACGTTACG GTTAATTTCA TGTCCTCTAC 9660

TCCTCTGTCT TTACTAATCT TCTAAAGCAT TAATTTCTTT CATTGCTTTT TCGATGATAT 9720

GAACTTCTTT CATCATACGT AGGTAAGTTT TTTCATTCAA GGATTGTGGC CCATCTGACC 9780

ACGCATTCGC TGGATCTGGA TGAATTTCTA CAATCAAGCC GTCCGCACCT GAAGCAACAC 9840

CTGCTCGGGC CATTGGCGGT ACTAAATCCC AGATACCAAC ACCATGACTC GGATCAACAA 9900

TAATTGGAAA ATGACTTAAT TTTTTAATTA AAGGCACCGC ACTTAAATCA AGTGTATTGC 9960

GCGTAGcggT TTCGTACGTA CGAATCCCAC GTTCAATGAA GATCACATTG AAATTTCCTT 10020

GCGCACAATG TATTCAGCTG CGTTTAGCCA TTCATCAATC GTACCAGCAA TCCCGCGTTT 10080

TAAGCCAATA GGTTTTCCAG TTTTACCAAC CGCTTGTAAT AACTTGAAAT TTTGCATGTT 10140

GCGCGCACCG ATTTGTAAAA TGTCACTGTA TTTTGCGACA AGTTCTAAAT TCGCTTCATC 10200

CATCACTTCT GTAATGACTT GCATATCTAA TTCATCAGCC GCTTGGCGAA TGTATTTTAG 10260

TCCTTCTTCT TCTAGTCCTT GGAAAGCGTA TGGCGACGTT CTAGGTTTGA ATGCACCACC 10320

TCGTAAAATT GTTGCACCTC CAGCTTTAGC AATTCGCGCA CATTCGCGAA TCTGATCTAA 10380

GCTTTCGATT GAACAAGGGC CCGCCATCAT GGTCATACTG CCGTCACCGA TTTTAACACC 10440

GTTCACATCA ACCACTGTAT TCTCTGGATG AAATAAACGA CTCGTTAGTT TGTAAGACAA 10500

CGAAATGCGG ACACTTTTCT CCACGCCTTC TAATGCTGTT AACTCTGCTT CAACTAATTT 10560

TTCTGTATCA CCAATCAAGC CTATCACTGT TTTTCCATTA TCTTGACTTG TTTGAGTAGT 10620

TAAGCCTGCA CCTGTTACTA AATCAATGAC TTGTTTCATT TGCTTTTCGG TTGCATTTTC 10680

TTTCATAATT ACGATCATTA TATTCTCTCC TCATGCCTTT AAACGGCGCT CTTTTTATCT 10740

ATTTTCAATT TTACGTTTTA CGACTGATAC AGGCATTTTT TGTCCTGTCC ATTGTTCAAA 10800

AGCAGCAGCG CCTTGATAAA GTAGCATGTC CaAGCCGTTG CTTGTTTCTG CACCACGTAA 10860

ACGGgCTTCT TTTAATAACT GTGkTTCyCT GGGATTATAA ATAGCATCAT ACmCAAATAA 10920

CTTCGGTTGA ATCATTGCAT AATTTTCTAT AGGACtACTA TGCGCATGTG GATGCATACC 10980

CACACTCGTT GCATTAACTA ACAAGACGcT TTCAGCAACA TCTTTTGCTA GTTTTTCAGT 11040

ATCCGCTAAA TCGTTTAAAT GAATTACACA GTTGGTTCGT TCAGTGAAAT TTGCCAGTTT 11100

TTTTTGACTG TCGTTAAAGC CCGCGCTTTT CCTGTTGTAA ACGGCGATTT CTTTCACGCC 11160

ATCTAAAGCA GCTTGGGCAA TGATTGATAA GGCTGCACCA CCTGTTCCTA AGATGGTCAT 11220

TTTATTCTGA AAAACGTCAA CCTTTTTCTC TTTCAAACTC CACATAAAAC CAGTACCATC 11280

CGTGCTGTCT CCGTAAAGTT TTCCATCTTT GTTGACAATC GTATTAATTG CGCCAACCAG 11340

TTGAGCGGTA GGACTTAGTT CATCCATATA ACTTACAGCC GCCACTTTAT TGGGCATTGA 11400

TAAGTTGGCC CCTAACATGT GAAACGTTCG AATCGCTTCA ATTGCTTGTG GCAATGTCTC 11460

TTGTCCAACG TCAAACGCCA GATACCGAGC ATCAACTCCT AAATTTTGAA ATGCTGTATT 11520

ATGAATCAAC GGTGAAATAC TGTGTTGGCT GGGTTTCGCG AATAGCCCAG CTAAACGAGT 11580

GGCTCCAGTT ATTTCTTTCA TCCTGTGTGC TCCTTTACGA AATTTTTTTA GTAAGTTTCT 11640

GAAATCCTAA TTTTTGGCCA AACAAAAAGG ATATTCACAG TAAATTTTAA GAAAAATCTA 11700

CAGTGAATAC CcTACAAAAA AGCAATATAA AAGCCACTAT AGATTTTCTT AAATCTACAT 11760

GGCTGAAAGT CATCTTAAAT TTCCCTAGCC ATTATAGATA CAAACGTTTG TTGTCTGTAC 11820

CCATAATGTT ATGAATACAA ACCTATCTAA AATAGCTAAA GTAAAAGTTA TTCAGTTGT 11879

(2) INFORMATION FOR SEQ ID NO: 182:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1187 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

CATGGCAAAT TTTAGTATAT AGTCCTAAGG ACATTTTACA AAAAGAATTn AAAACGAAAC 60

nCATGaCGGG gTaTGACCCm ATTGAAGTAG aCGrATTTTT AGACAaTGTT ATTAAAGACT 120

ACGAAGCTTA TAATAAAGAA TTATTAAGCT TGCAAGAAGA AAACAGCCGC TTAATGGCAA 180

AATTAGACCA ATTATCaAAA GCCCAACCAA CGCCTCGTGT TGCTCaAGAA GtACCAAAAA 240

GTGCAGCAGT GACCAACTTT GATATTTTAA AACGTTTATC AAACTTAGAG CGTGAAGTTT 300

TTGGTAAAAA GTTAGATGAA ACACCTTCAA CACCTGTAAC GCCTTCTGCA CCGTCAATGC 360

CTGCAGAACC AGCGAATCAT GATGTTGATA ATGCACAAAC ACGTCAATTT TAACAAGACA 420

ATTTAAGTCA AGTTTTCACG AGGaACAAAA AGCATTTATT AGAATAGCGG ACACTTGTTG 480

TAGTGTGTAA ATTTTTGCTG TTCGTCTAAA CAGATTTGTw ATTTTCGGGT GATCGCGGTT 540

TGCTTTTGCA AGCTGAGGAA AGTCCATGCT CGCACAAGCT GAGATGCTTG TAGTGTTCGT 600

GCTTAGCGAA ATCATAAGCT AAGGTACTCT TTTAGAGTAA CGGCAGGAAA AATGACTAAG 660

GTTTCGACTA TGTCAAAGTA TCCTTGAAAG TGCCACAGTG ACGAAGCGAT GTGGGAAACT 720

CCATCGGTGG AACGCGGTAA ACCCCTCGAG CGAGCAACCC AAACAATGGT AGGGGCGCTC 780

TTCTAAAGGA AATGAACGAG TAGAAGAGGC AGAGTTTACT CTGCAGATAG ATGATGACCG 840

TCACCAATTT TTCCTGAAGA ATTGGTGATA CAGAACATGG CTTATAGAAA ATTACAAGTA 900

ATTCAGGAAG ACCTTCCAAC GTAGTTGTGA AGGCTTTTTT TACGAAAAGA ATAAAAAAAT 960

GTtAAAATAA AGAAACTGAG wTAAAaGTAG TAAgCTGTCT GTTCTAAAGA GGCCGGAACA 1020

GGACCGTTtA GTcAGAGAAA GCGGACAATT wTAAAAAmCA aGCgGAccAG TAGTCsGCAG 1080

GACTGTTAGC TCCAGACmaT wAAAAAGrgT CmAGAAGGAG TTTTaTgaCG AAAGAAAAAm 1140

CaTyCAACyT GTTGCaCGGC TGCTAGTGGC TTAGAAGCAT AGTTGGn 1187

(2) INFORMATION FOR SEQ ID NO: 183:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4178 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

AAATAACTAA AACAGTTAGA AATAATGAAT TTCCTCGTTG CTGTTGGCGT TTTGCtAAAC 60

GaGAACCACC ACCATTTTGA TCAGGTTCAT AAATTGATTG TTCCCATGGC TCACGAGCTT 120

GGTTTTTCTT TTTGTCCTTT TtACTCACTG CTAATTTCCT CCTAAACTTC TCTCGCTTGT 180

TATTATATCA TAGCTTTTGC CAACTGCCGA CTAAAAAAAT GAATTAATCA CGTTCTTTAA 240

ATAATTTTAG AAATTTAGAA CGCCAACCTT CTTCATTTTG ATTGCTGGTC TTGTTCGATT 300

TTACTATTGA ATCTTGGTAA ATGTCAAAGA ATAACGCACA ATTATCACAA CAAGTTTCCG 360

GTGATTGTTT GACAATCGGT TCTTGAAAAT AAGTTTGAAT CAACGTTCTT CGACAGGTTG 420

TGGTTTCAAT ATACGTCAGC ATCCCTCTTA ATTGTGCTTG GCGATCTTTT TCTTTCTCCT 480

CTAAACGATG CTCTAGCTCT TCAAAAGAAT AATCCTTGGC TAAATAGCCT TGGACCCATT 540

TTTGTTGTAA CTCGGTGGCA TTTTCTAGCA AACCTGCTTG TTCTTCTCCT TCAATCAGCG 600

ATTTTAAGGT CGCTCGATTC GCTCTTGTCT CTTCTAAGAA AAACCGATGA ATAAATTCAT 660

```
CCCCTTTTTT ATACAATAAT AGCGCCCAAC TTTTTTGACC ATCACGTCCA GCTCGCCCGA      720
TTTCTTGCAG GTAATTTTCT AAGCTATCAG GCAAATCAAA ATGAATCACT CCACGAACAT      780
CAGGTTTGTC GATTCCCATC CCAAAAGCAT TCGTTGCACA AAGAATATCG ATTTGATTAT      840
CAATAAATTG TTCTTGCAAG CGACTACGTT CACTAGCCTC CAAGCCCCCA TGATAATAGG      900
CTACCTTAAA ACTGGTTTTT GCCTGAATAA GTTGACTGAT TTGTTCTGTT TTGTTGCGCG      960
TGGCGCAATA GATAATGATT TTACCAGCCG TTTTAGATAA GTAATCTAGT AACACTTCTT     1020
CTTTTTCTGA CGTTTCCTTC ACGAACAAGC CAATATTTTT TCGATTCACT GAAGAAAGAA     1080
CTTCTTGATA GCTGCCATGA GAAAATAATT GTTTTATAAT TGCGTGCTGT ACAACGGGTG     1140
TCGCTGTGGC AGTTAACGCC AAGGTCAAAG GAAAGTCTAG CTGTTTTTGG ACCAGATCTA     1200
ATTTACTATA TTCAGGGCGA AAATCGACGC CCCACTGATA AACACAATGG GCTTCATCCA     1260
CCACAAATAA TGCAATTTTT TGGCGTTGTA ACACACTAAG TACTTCTTGC TGCAAAAGCA     1320
TTTCTGGGCT TAAAAACAGA TAATCGTATT GGAACATTTT CGCTAAAATA TAACGCTTTT     1380
CTGAATACTG TAACGTACTG TTTAAGGCAA CCGCACCTTT AATTCCTTGT CTTTGTAATT     1440
GACGCATCTG ATCTTCCATT AACGAAAGTA AGGGAGAAAC AATTACCGTT AGTCCCTCTA     1500
ACTTCTGGCC CACAAATTGA TAACACAATG ATTTTCCAGT CCCTGTTGGC AAAACAGCCA     1560
ATGTATTCAT ACCGTCAAGT AAGTTTGTAA TCACTTCTTT TTGACCAGGC TTAAACGCTG     1620
CGTAACCAAA TTGAGTGAAT AATTCTTGTT CTAATGTCAA CAACACTCAC TCCTTTTCTT     1680
GCCAAATTTG TTGGATTTGA TATAACCGAA AATCTAAATA AGAAAGCGTC GCTTCTTGCG     1740
CTCGCCAATC CCGATACACC CACAGCCGCG GATCTTGTTC TGTTTGAGCC AAGTTTTGGA     1800
GCATTGCCTC CTTTTCAAGG TAAACAGACG CCTGCAGCTG CCCTTGAATA GCCAACTCTA     1860
ATAGATGGTC GTGAATGGTA CTCGGTTTTA TTTTTCGCTG CTCAGCAATT TCTTGAAGTG     1920
TCTTACCTTC CGTTAACAAT CGTTGGGTTT CTTGCATACT TTGGTTTAGA TTTTTTTCTA     1980
GATACGGCCA TACAAGCGTT TGTAAAGCAG GATAATTTCC TCCGTACATT ATCAACTGCA     2040
GTAAATGATG AACCGCTTGC GTGTGCCACA AACGCTGTTC AAGAGCTGTC TTATGGCTTG     2100
TTAATTGCTG AGGAACTTTG CCAGTTTGCT GATAACCAGA AAGTTGCGCA ACCAAATAGT     2160
CGCTTTCTGC TTCAGGTAGT AACTCAAACC CTCGCAACAG TTCTTCTTTG ATTGATTGAA     2220
CGCGCTGCTC TTTCTTGTCT TGTTGTAACC ATCGTTTCAA ATATAGTTGC GGAATCGGTG     2280
TTGATAAAAG TGGAATATAC TGTTTTTCCT CAAATGATAA ATATGAAGTC ACTTGCACGG     2340
CAAATTGAAT GAGTTGCCAA CTTTCTCGAT CACTTCTGCC AAAACGATAA AGATCAATTG     2400
CAGGATAATC CAGCAACGAA AAATGGTGCT CCTTTAAAAA TAGTTGCCCT TCTTTTGTTA     2460
GTTGAATCTC ATTTGTCTCT TCGTGATACT GAATTTGTTT CGCCTTCGAA AGTTTTTCAA     2520
GAAATTGAAA ATAGGCTTGT TCGGAAATCG TTGGCCACCA GCCAATGAAC CGTAAACAAT     2580
CATATAAAAA ACCATAAATC AAAACAGAGC TAGTCCGTTT GCCTTTTAAA AGATGATACA     2640
```

AAGTTGATGC ACGTACCTTG TATCCATGAG ATAATAACGC TAAAATAAAA GTTGCTTCCA 2700

TAGTAAACCT CTTTTAGATA GGATGTGTCA CTGTTTATGC AATCACGTTT AGTTCCTGAA 2760

AAATGTATTG CTTGTGGACT TTGTCAAGTT TATGCGCCTA ATATATTTGA TTATGATGAT 2820

CATGGGATTG TCCTTTTTAA AGACGAGCCG CATGCTCGAC AACAATTTAT TCCAGAAAAC 2880

GAACGCGAGA ATGTGTTAAA AGCGTACCGC AAATGTCCCA CGCGTGCAAT AGAAATAAGC 2940

AAAAATGAGT GCTAAATAAT TTAGCACTCA TTTTTGCTTA GAGTATCTTT TCGAGCCAAG 3000

CAAGTGTTTG TTCATCTATC GAAACAAACT GTAATTGTCC TTCTTTTAAC GGTTCTTTGA 3060

CTAAATGAAT TTGATCATTT AGGCCATTGA TCCaTTCATA ACGAAGACCA TTTTCTTTTG 3120

CTTTAGAATT TTTGGCTTCT AAATAATAAT TGACATGGTA ATTTTCTTTA CCATCCAATT 3180

GTAGCGTATT AATGGTCACT TCAAAATCTG GTGAACCATC GGTCGATTTC TTTTCACGTA 3240

ATCGTTCAAG TGCTTCTATA AAAGAAGGAA ATTGAGGATT ATTTTCCCCA CCAACGAGAT 3300

ATTTTTTAGC AAATTCTGTA CTATCCGTAT GACTTTTCAC TTGTTTGGCA AATGCTTGAT 3360

AAAAGGTTTC TAGCGCTTTG CGGGCATCAT AGTCGCTAGC TAACGGAAAA TTCAATTGTA 3420

GCGTTGGATT ATCGGAGAGC GCTTCGACAA AATCATTTTC ACCAATCGTT TCTTTTGATG 3480

TTTGAATTTC TTCTCCATTA ATCGTCTTTT TAAGTTGAAT CGTCATCCCT TTGGACCAGA 3540

ACAAAGGGCC AATCGTTTTG CTTCCATTCG TTAGCGTACC AcTTTCCGAT CATTAATATA 3600

AATATCTGCG TCTGGCAGAT TGCTTTCCAC AACAAAATTC ATTCGTTTAA AATCTAAAGG 3660

AATACTAGAA ATATAATTTT CCTCTTCTAA TAAACGGTAC TCTTCTGTGA TAACAGGTTC 3720

GCCGGTGCTA TCTTTCACTG TGGCTGTAAA AGTATATTGT CCTGGCGCTA AGGGGCCCAA 3780

TTCCTTTTGA TAAGTGGTTG AGTCGGAAGT GCCCACTTCC CGCCCGTTCA TTGCCAGCGT 3840

CACACCGCGC TGATTCGTCG TTAACTGAAC AGAAACAGGC GTTAAAATGA ATTGATAACG 3900

GTGAAAGAAA AATGCTGTTT GACTTGTTtC ACGAATAGTT AAACCATGTA ACTTTTTATC 3960

CTTTAATAGC GCTTTTTCTA ATTTTTTTAG CTCAGTTGGA TTTTTTTGAT AATAAGTGAT 4020

TAAAGGCTTA ATGCTTTGCG GATTCACTTG AAATTCAGAC TCGTTCGTCC TCATTTTTTT 4080

ACTTAATTCT TGGGCATCAC CATTCTCAAT CGCTTCAATA AAATAAGAAA TTTGTTGTTC 4140

CTTAGAAAAA TAATGCGTCC CGAGAAAATA GCTACCCG 4178

(2) INFORMATION FOR SEQ ID NO: 184:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1014 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

GATTTTAATG GGCGGTATCC TTTTAAGTGC TTCCGCGGTT TTACTTACTG CCTGTGGTGG 60

CGCAGAAAAA ACGAAAGCTG AAAATGAAAG CGGTTCAGCA AATAAAGCAG CTGGTGACGT 120

AAAATTATGG GTTGATACTG AATATATGGG CGTCTTTAAA AAAGTTGTCG CCGATTTTGA 180

AAAAGAAAAC CCCGATATTA AAGTTAACTT AACTGCTGGA AATTCGGCAG ATGCTAAAAA 240

AGACATTGCA AAAGATCCTA AAGCTGCCGC AGaTGTTTTC ATGATGCCAC ATGATCAAAT 300

TGGTCAAATG GCAGAAGCTG GTCTGATTTA CCCTAACACA AAATATGAAA AAGAAGTCAA 360

AGAAAACAAT ATCGACTCGG CAGTCGCAGG AGTCACTTGG AAAGACAAAG TCTATGCCTA 420

CCCTTATGCC GTAGAGTCAC AAGTTCTTTA TTATAACmAA GATACGTATT CTCCAGAGGA 480

AATTAAAACT TGGAAGTCTT TAACTGGAAA AGGcAAAATC GGTACCAATT TTGGCGAAGA 540

TGGCGCAAAT TATATCTTCG GTCCTCTATT TATGAGTAAC GGCGATTACT TATATGGTGA 600

AAATGGCGAA GATCCGAAAG GTACCAACTT TAACAACCAA CAAGGTATAG AAGTATTACA 660

ATGGAtTGCC GATCAAAAAA ATAATCCTGG TGTGATCCAA TCAAATACAG AAGCATTGTC 720

TAATCTAGGT TCTGGCAAAA CGGATGCCTT CCTTTCTGGT CCTTGGTCTA AAAACGATGT 780

AGAAAAAGCT CTTGGCGACA AAATGGGGGC CGCTGCCTAC CCAACCATTG ATTTTGGCAA 840

TGGAGAAAAA CAAATGAAAG CTTTCTTGGG CGTTCGTAGT TTTGCTGTCA ACCAACAAAC 900

ACAAGCACCA TTGGCCGCAA TGACATTAGC AAACTATTTA ACAAGTGAAA AAGCACAAAT 960

GACTTACTTC mAAGAAATCg GTTTTGTCCC tCCAACCAAA AATTACCAAn CAAC 1014

(2) INFORMATION FOR SEQ ID NO: 185:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3614 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

CGTTCCATGG CCGTTGTTAA TTTATCCATA TCGGACGCAT AACTAATTCG GACGTATCCT 60

TCTGCTTCTG GGCCAAAAGA AATACCTGGA ATCAACGCTA CTGCTTCTTT TTGCGCTAAA 120

TCAATACAAA ATTTCATAGA ATCTTGTTCA TAACCAGTAG GAATCTTAGC AAAAATATAA 180

AAGGCCCCAT TTGGACGCGC AACCTCAAAC CCCAATTGAG TCATCTGTTC ATAAACAAAA 240

TCGCGCCGTT TCCGATACTC TTCTTTCATA ATTTCAGCAT CATGCATTCC TTCGATTAGG 300

GCACGAACAG CCGCTTTTTG AGACACAGTT GCAGCCGCAG TAACTAAATA CTGATGAACT 360

TTAATGATCT CCGCTGTTAA AATCGCTGGT GCAAAAATAA AGCCAATCCG CCAGCCTGTC 420

ATGGcATGTG aTTTCGACAA ACCGkTAATT AAAATCGTTT GGTCACGTAA CATTTCAGCA 480

ATCGAAACAT GACTCTCATC ATAAGTTAAC TCACTATAAA TTTCATCACT AATCACAAAA 540

ATGTCGTATT TTTTTAGAAC ATCAGCAATT GCTTGCAACT CTTGGCGGTT ATAGGTGACA 600

CCAGTTGGGT TGCTAGGATA ATTTAAAATA ATTGCTTTGA CTTGTTCCCC ATGCTCGGCC 660

ATTGCCGCTT CAATCATTTC AGGGGTCAAG ACAAAATCGT TGTCTCGTGT ATCCAAATAT 720

ACCGGTTCCG CCCCCGCCAA TGTAATAATT GGTTCATACC CTGGGTAAAT AGGTGCTGGC 780

AATAGCACTT TGTCACCTTG TTCTAAAATT GCTAATAACG AAGCTGAAAT CGCTTCTGTk 840

GCTCCAACTG TTACTAAAAT TTCTGTCGCA GGCTGGtAAG TCAAATGGtA CTTTtCTTTC 900

ATAAAATAAG CGGCTGCTTC ACGGaCATCT AATAAACCAG ACATCCCTGT AtAATGAGAA 960

AAATTCGCAT CAATTGCTTC TTTGGCTGCT TTTTTGACAT GCTCAGGTGT GTTAAAATCC 1020

GGCTCTCCTA GCGTTAATTT TAAAATACCT TCAATTCCTG AGACTTGTTC GTCAAATTGA 1080

CGAATTAAAG AAACAGCAAT TTTATTTACT TGACGATTAA AACGATtGGT TAAATCCATT 1140

TAGCTTCACT CCTTCAGTTG ACGTTATAGT AGCACAAATT GACCGTAAAA TGAATGATTT 1200

TCTCATTGTT TCGCTAGATT TTTATTATAA ATATCTCAAA AAAACGACAG AAAAGAAAAT 1260

AGGAGAACTT TGTCCTTTTC AACAAAGTTC TCCTATTTAT TTATTTTTCG AAACGCCATG 1320

TCCGAATTTC AAACGGACGT AAtGGTGCAT CAACTTCTAT CGCTGGTGTC AACGGTTCTT 1380

CTAATACATT TGTTAAACAA GTCGCTGGTT GTGTCTCTTT GTGAAGTTCA CAAGGAATAT 1440

TTTGCGCTGT TAGATTAAAG CCACGAACGA CTAAATGATT TGTTTGCTCT TGACGTTTAA 1500

CGGCTGTCCA ACCAATTGGT CCTTCTGCCG TTGTCAAGTA GACATGGTTC GGTGCTAGTT 1560

TACCTTCATG ACGCGCAGTA GTAGCTGCAG TAAAAGGAAT TTGTGCACTA TAGGCTTGTT 1620

GATACGTAGC AAAGCGTTCT TCTGGTTTTC CATGCAATGA CAAACCATAT TTAAAGGTAT 1680

ATTCTCCTTG GCATTGTGCT TCTGGTGTCG CAAAATAGCC CCAATCACCT AGTTCGCCAA 1740

CACAACGAAT TAATGTGACC GCAATGGTTC CATCTGCTAA GACTTCATAC TCATTTAAGC 1800

CTTCATTGAA GATAGTCACG CCTTTGTTTT GGTCATGAAC ATTCACAAAC GCTTGTTGGT 1860

GCTGCGGGTT TGTTGGATTT TCCCATGTGT CGCTTACTTG ATTTGGTCGT GTCACGACTT 1920

CATAAATACT ATCCGCTTCG TGTGTTTCTG TGACCATGCC AGTTGGGAAA AGCACGCGTA 1980

AACGATGATC TTTGATTTGA TTATTAATGG TTGTTTCAAA GAACAATTGG TTACTTTCTT 2040

kGATCATTGT AACCATTGTC GTTATTTCGA ACGGGCGAAg CGTCGTTGAA CGTTGTGCAA 2100

GGCGTTCTCT AAACTCACGA ACAGCCTTTT GTTCTTCTTC TAATCGTTCA TCTGCCGCAA 2160

CTGGTAACAT GACGGTTTGT TTAATTTGAT AACTTGCTTT TACAGGGGAA TTTTCTTTAT 2220

TCGTGATTTC AGCCGTTACA TTTTCGGTTG TAATTCCTTG GTCTTCGGTC GGTTTGAAAA 2280

AGATATATTC ATTCCCAATG TCGCCAGTAT CTTCAAATGT TAACAATTTT GAAAATGTTT 2340

CTCCTGTTTT TTTGTCAGCA ATGGTTAAGC TACCATTGTT TTCAATTTTC ACTTGAATAA 2400

ATTCGTTTTC CATTTCATTT GTTGCTGGTT GAGCCAAAAG ACTTCCTTCA AAAGCTTTCG 2460

TTTCTCCTTG AATCAATGCA AAGCTATCCC AAGAGAATGG CGCCATTTCT TTGACAGATA 2520

ATTCGACCGT TAAGTATTTG GCCATGTACG GTTGACGGAA AGCAtCTTTA GGTAAATCGT 2580

AACCAAAACG AACATCTTCT TTGACAATAC GGCTTGCAAC GGCTTTTCCA GTTgGATCAA 2640

TCACTTGATA AtCTGGGGTG GCTTTTtGTT TTAATTCATG ATAAAGATCT TCAGGTTTcC 2700

TGTATAAAAT GGCAATCGTt CAATCTCCAC TTCCACAGTG ACAACCTCCG TTTTCTGATA 2760

GCCACTAGTA TTAAATAAAA CAAACGGGTG CGCTTTTTCT GGGAAAACAC TTGTATCAAT 2820

TGCTTCTGTT AATTGTCGTG TTGCTTCATC TGCTAAGAAA TGCCCGACAT CATTGGCATT 2880

TTCAAAGCGA GTCATCATGC CTCGGTGAAC CTCGTCGACA GAACAGCCAC AAATACTATC 2940

GTGCGGATGG TTTTGAAGCA ATGTTTTCCA AGCATAATCA AATTGGTCAT GTGGGTAGTC 3000

GCCAGTTACT TCGTAAGCCA TTGCTGCTAA AGGTTCTGCG ATATTTTCTA ATTGGCGTTG 3060

GACTTTTGTG TTCCACTGTT TTAAATAGAC GCGAGCAGAA GAAGTATTGG CTAATGTGTA 3120

CCAGCCATCC GTTTCTTGAC TCGTTAATTC ACCTGTTACT GTTCCTAAAT CTTCTGGTAA 3180

CTCCTCTTGA ACAGCTTTCA AATAGTCATC GAAATTGGAA TGAATAAATT CATATTCTGG 3240

GAATAGTTCA TTGGCTAAAG CGATGGCCTT AGTAATATCT CTTTGTACAG GTTGGTGGTC 3300

GACCCCATTC ATCATCAATA AATGGTTTGT GGAAGCGTAA CGTTCCACAT CAGCCAATTT 3360

TTGTTTCCAA AAAGCGATTG CCGCTTCTTT TTCTGAAGGA ATTTCGTTCC CATTACTATA 3420

CCAGTTAGCA AAAAGCAAGC CGAAAATTTT CGTTTGATCA GGACCTTCCC ACCACATTTC 3480

aGAATATTGT GAGCTATAGT CGCTTTCTAA GACTTGGTTA TCAAAGCCAA TCGGTTTCAC 3540

CCCGCGTCCA AAGGCAGCCG CAGGaGATTT GCTAATTGCA TCATTTGAGG CGTTTGACCC 3600

ATATTCCCAA ATGn 3614

(2) INFORMATION FOR SEQ ID NO: 186:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2199 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

CTTACATTCC ACTaAAACgT CAAACCgTAT CCAAAGAAGC GATTTATATT GCCATTGGTA 60

TACGAGAAGA CGGCACTAAA GAAGTACTGA GTTATGCGAT TGCTCCAACT GAATCAACAT 120

ACGTTTGGAA TGAGCTGCTA CAGGATATTA ACTCCAGAGG AGTTCAAGAA GTCTTGCTTT 180

TTATTACGGA CGGCTTAAAA GGCATGAAAG ATACTATCCA TCAAATTTAT CCTAAAGCAA 240

AATATCAGCA TTGTTGTATC CATGTATCTC GTAATATCGC TCATAAAGTA CGTGTCAAAG 300

ACCGAAAAGA AATCTGTGAT GACTTTAAGG CTGTTTATCA AGCTAACTCA AAAGAAGAAG 360

CGAATACCTT CTTATCCGGC ATGATTGAGA AATGGAAGAA AAACTATCCT AAAGTGACGC 420

AGTCAcTCAT AGAAAACCAA GACTTATTAA CTTTTTATGA TTTTCCACCT AGCATTCGTA 480

GAACCATTTA CTCAACCAAT CTAATCGAGT CTTTCAATAA GCAAATTAAA AGATACAGCC 540

GTAGAAAAGA GCAGTTTCAA AATGAAGAAT CACTAGAACG CTTTCTAGTC AGCATTTTTG 600

ATACATACAA TCAAAAATTT CTAAACAGAA GCCATAAAGG TTTTCAACAG GTAACCGATA 660

CATTAGTTTC AATGTTTACT GAGTAACTAA TTATtTTGCA GGAGGACAAT TTATTTACAC 720

AAAATTATTG ACGCTCCCAA GAATGTTTTT AAATAATATa AATAATAGTG TTTTTCAAAG 780

TTTTGTTAAT TTttCATTCG GAATCGCAAT TGAAGAAATC ACGTGGTATA CTTTTAAGAA 840

GArACAATGA CACGAAAGTA GCCATTGGAA AAAGGAGATT ATTTCGTGGG AAAAAAACAT 900

ACATCCTCTT ATGAAGTTGC CTACTATGAC GGCGATTTTA CTGGTGCAAT GAAGATACCG 960

GCTTTGCTAG CGGTTGTCAT TAAAGTATCA GAAGAACAAA CAGAATTACT AGGAAGAGAT 1020

GCGGCTTATG TCGCTCAATT TGGTCTAgGC TGGGTCATCA CCAATTATGA AATAGAGATT 1080

CACCGCTTAC CAAAAGTTGG TGAAAAAGTA GCAATTACAA CGCAAGCTAT GAGTTATAAC 1140

AAATATTTTT GTTATCGTAA CTTTTGGGTT CATGATGAAG AAGGCAAAGA ATGTGTCTTT 1200

GTGAAATCAA CCTTTGTCTT AATGGATCAA AAAAATCGGA AAATCAGTAG TGTGTTACCA 1260

GAAATTATTG CACCTTTTGA TTCTGAAAAA ATCACTAAAA TTTATCGCCA TGAAAAAATT 1320

GAAAAGGTCA CAGAGGGGAA CTTTTTACCT TATCGTGTGC GTTTTTTTGA TATTGATGGC 1380

AATCAACACG TTAATAATGC GATTTACTTT AATTGGTTAT TGGATGTGTT AGGGTATGAC 1440

TTTTTAACAA CGCATCAACC GAAAAAAATT CTTGTGAAAT TTGATAAAGA AGTAGAATAT 1500

GGCCAAGAAG TCGAAAGTCA TTACGAAATT GTCGAACAAG AAAATCAGTT GAAAACACGC 1560

CATGAGATTC GGATTGATGG TCAAACCTAT TGTGAAGCGA ATATCGATTG GACCAACTAA 1620

GTTGAGATGA CGGTTTAGAA AAATAAAAAG TTACGCAGAG GAGGAAGGAC ATGATTCGCT 1680

TAGGATTAAC ATCTTTTAGT GAACACGACT ATTTAACTGG AAAAAAACGG TCCACGCTTT 1740

ATGAATATGC CAGTCATTTA CCGCTAGTTG AAATGGACAC GGCGTATTAC GGGATTCCTC 1800

CCAAAGAACG AGTAGCTGAA TGGGTCAAAG CAGTTCCCGA AAACTTTCGC TTTGTCATGA 1860

AAGTTTATAG TGGAATCAGT TGTCAAGGTG AGTGGCAGAC CTATTATGCA AGTGAAGAAG 1920

AAATGATCAC CGCCTTTTTA GAAAGTATGG CCCCATTAAT TGAAAGTAAA AAATTATTTG 1980

CCTTTCTGGT ACAATTTTCa GGAACATTTG GtTGTACrAA AGAAAATGTC GCTTACTgCA 2040

AAAAATTCGT CACTGGTTAA AGACTACCAT CGCCATTGAA TACAAATAAT CCAGGTATCA 2100

ACCAAACTTG TAAACAATTT GCAATTATAA AAAATCATTC ATGTATGTGA TAACGAAATC 2160

TCATCAnCCT TACTATTACA TCTACCnGTC CCGTCnnGA 2199

(2) INFORMATION FOR SEQ ID NO: 187:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2594 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

TnCTCAGGGT TTTTCTTCCC TTCTTGTGAA TGCGGGATTG TACCAATTGT GCATCAATTT 60

GTAAAAAAAG ATGTACCAAC ACATACCGCT TTTGCTTTTA TGTTAACGGC ACCAATCATT 120

AATCCAATTG TGCTATTCTC AACATACGTA GCTTTTAGTG GCGCAACAAA ATTTGTGATT 180

TGGCGAGTTT TAGGTAGTAT GATTGTCGCG TGGGTTATCG GCATTTGGTT AGCTTATTTT 240

CGAAAAGAAT CAATTTTAAC AGAAAAAGAG TTGGCGCAGA TTAAAGAAGA GCAGGTTGCT 300

CATACGCATG AACACGTTGA TGAAGTTCGT TCATTTTGGA AAAACGGCTG GTCAGCGTTA 360

ACCCATAGTA TTGATGAATT TTTTGATACA GGTCGCTATT tAATTTTTGG TtCTTTGTTG 420

GCAGCAATCA TGCAAACGTA TATCCCaACT GGTGCACTtA TGCAATTAGG CCATTCAAAG 480

ATTTtAGCGA TTCTAGTCAT GCTAGTCATC GCTGCCACTC TATCACTTTG TTCAGAAGCG 540

GATGCATTTA TTGGTTCTTC TCTTTTGAGT TTATTTGGCA CAGGACCAAT TGTCGCCTTC 600

TTAGTTTTTG GTCCAATGGT AGATATTAAA AATTTATTGA TGATGAAACG TTATnTTAAG 660

ATGTCATCCA TTGTACAATT TGTTTTGATA GTGGCGCTkG TCGTAmCwAT TTATGCGGCA 720

GTAGTTTAGA AAGGAGGCCT ACAATCGATG ATACGATTTA TTATTTTAAT TGGATATATG 780

GGGTTGATGA TGTATCTTCA AATTTCAGGT GAACTGAACC AATATATCAA TATTCATTAT 840

AATTACTTAG CTGTTTTATC AATGGTTTTA GCTTTTATTA TGGCGATTGT CCAATTGATT 900

TTGTGGAACC AAGCAGATCC ATCAATACAA AAAAAACATG AAGACATGCA CCAACATGGG 960

GGACATGAAC ATCATGGCCA TCAGCATAAT TTGGAAAAAC CCAGCCAAAG AGGGCTTGCC 1020

TATTTACTAC TTTCTTTACC GTTGATTGTT GGCTTATTAT TTCCGACAGT TAGTTTAGAC 1080

ACAACAATTG TTGAAGCAAA AGGATTTAAC TTTCCAGTAA GTAAAGAATC AGTAGGGGAT 1140

CCTGATATGC AAACGCAATA TTTGAAACCT GATACAAGTA TGTATTTCAA TAAGACGGAT 1200

TATGACAAGC AGATGGCAAA AGCCATGAAA CAATATGATG GCCAATCTGT GATTTCAATT 1260

ACGGATGAAA ACTATTTGGA AATTATGGgA ACTAATCTAT AATTATCCAr GTCAATTTGC 1320

aGGGAAAAGG aTTTCTTATA AAGGCTTTGT CTATAATTCC-AAAACGAGAA GAAAGTGTTG 1380

ACCAGTTkGT CtTCCGTTTT GGGaTTATyC ATkGTGtCGC TGATTCtGGt GtTTTTGGCy 1440

TGttGGtTCA TTTyCCTGaA CATACTCAAT TTCAAAACAm TGATTGGGTA ACGATTACCG 1500

GCACGtTGAG TTGTCCTATT ATCCACCGTT TAAACGACAA ATACCAACTG TGCAAGTGGA 1560

GAAAGTCAAA GCAGATcAAG CACCAAAAAA TCAATATGTC TATCGTTCCT TTTAAAAAAA 1620

GTTGGTTGCT TCCTTATCAA ATGGAAGCAA CCAACTTTTT TGCTGTCTAT TTTTTAATTT 1680

TGTTGCGTTT ACGACGAATA AAACCCAAAG ATTCTCTTTT TTCTCGAAAG CTGCGGATAT 1740

TCTGATTGAA AACCTCAGAA GCGACTAAGC CTAACGCAAT TGCCCCAGCA ATCATTACAG 1800

CTTGcACGGT AAAATAGGCA CCTTTTTCAT AATTGCCAAT TACGATATTC CGAACAGCTT 1860

GATAAGCTAA GCCACCAGGT ACCAATGGAA CAATACCAGG rATATTAAAA ATCGTGACAG 1920

GcATTTTCAA TATTTTTGAG AATAGGTAGC TTACTGCCGC AACACCGACA GCACCGATGA 1980

GGGAACCTAA TGCTGCATTT CCACTAAAAA CTTCAACGAT TACCCAATAA AGCATCCAGC 2040

TCACACTACC AGTTAGACCA CAAGCCACCA GTGATCGTCG AGGAACATTA ATAATAATCG 2100

CAAACGCCGC AGAAGCTAAA AAACTAAAGA TAAATTGTAC CAAAACGATT CCTAGCAAGT 2160

AAGTCACTCC TTTTAATAGA AAAGTTGAAA GACAAAAGCA ATGCCAAAAC CAATCATTCC 2220

TGAAGTCATT AATGCTTCAG TCCCACGAGA AACACCCGAT AAGTAGTGTC CAGCTAGCAA 2280

ATCGCGAAGA GCATTGGTAA TTTGCACACC AGGCACTAAA GGCATGACAC AACCAATGAT 2340

GATTAAATCT TGTGATTGTC CGAGTCCCAA TTTGACTGAA AGAATAGCGG CTGCACCAAT 2400

GAACAACGAA GCTAAAAATT CAGATAAAAA TTTAATCTTG AGAAACTTTA AACTrAAATA 2460

ATAAAGACTA TAGCCGCAAC CACCAATCAG AGAAGTCAAG AATAAATCGG ACCAGACACC 2520

ACCAAAAAGA ATCATAATTG TAGCGCTGAC AACGGCCGCA CTAACAATGC GTGCGCTAAT 2580

TGGAAAATGT CGGT 2594

(2) INFORMATION FOR SEQ ID NO: 188:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4176 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

AATCCTTTCG CAAGGAGGAC TTACTCAGCA AAATCCGTTC CAATCGAATG TACCAAGGCT 60

TTAACGTTTC AGCAGCCAAT AAAAATGAAG GATGGGGATC AGACCGATGA CAAGAAAAGC 120

CAAAACGGCA GAAGAGATTA AACATATTGC AGAAGAGGAA AACGTACGTT TTTTACGACT 180

AATGTTTACC GACATTATGG GAACAATTAA AAATGTTGAA GTTCCTGTGA GTCAGCTAGA 240

TAAAGTATTA GATAATAAAA TGATGTTTGA TGGGTCATCC ATTGAAGGAT TCGTTCGTAT 300

CGAAGAAAGT GACATGTATT TATACCCTGA CCTATCTACA TGGATGATTT TTCCGTGGGa 360

AAGCGACCAT GGAAAAGTtG CGCGCTTAAT TTGTGATATT TACAATCCTG ATGGTACCCC 420

TTTTgCCGGA GATCCCCGCG GTAATTTAAA ACGCGCCTTA GCAGATATGA AAGAATTAGG 480

CTTTACTTCT TTTAACTTAG GACCAGAACC AGAATTTTTC CTATTTAAAT TAGATGAGAA 540

TGGTGAAATT ACCACTGACT TGAATGATAA AGGTGGCTAT TTTGATTTTG CACCAACCGA 600

TTTAGGTGAA AATTGTCGCC GTGATATTGT TTTAGAGTTA GAAAGTTTAG GCTTTGAAGT 660

GGAAGCTTCA CATCATGAAG TAGCACCTGG TCAACATGAA ATTGACTTTA AATATGCTGA 720

TGTAGTTGAT GCTTGTGACA ATATTCAAAC GTTCAAATTA GTTGTGAAAA CTATTGCGCG 780

TAAACATGGC TTGCATGCAA CATTTATGCC AAAACCTTTG TTTGGAATTA ATGGTTCTGG 840

TATGCACTGC AATATGTCAT TATTTAATGA AGAAGGTAAT GTTTTTTATG ACGAGTCAGG 900

AGAGATGGGc TTAAGCCAAA CTGCGTATCA TTTCTTAGGT GGTTTATTAA AACATGCCCG 960

TGCCTATACT GCGGTATGTA ACCCAACAGT GAACTCTTAT AAACGCTTGG TACCAGGTTA 1020

TGAAGCACCT GTTTATGTAG CTTGGAGCGG CAGAAATCGT TCACCACTCa TTCGAGTTCC 1080

TGAATCTCGT GGkTTATCGA CTCGTTTAGA ATTACGTTCT GTCGATCCTT CAGCGAATCC 1140

GTATTTAACA ATGGCTGTCT TATTACAAGC AGGTTTGGAT GGTATTCGTA ATGAACTTAC 1200

ACCACCGCCA GCAGTTGATC GTAACATTTA TGTGATGAAT GAGGAAGAAC GTCAACATGC 1260

ACAAATTGaA GATTTGcCAT CAACCTTACA CAATGcCATC AAAGAATTAC GTAAAGATCA 1320

AGTAATGATT GATGCATTAG GTCGTCATAT TTTTGCCAAT TTTGTAGAAG CGAAACGAAT 1380

GGAATGGGCA GCCTTCCGTC AAACCGTTTC TGAATGGGAA AGAGAACAAT ATTTGGAGTT 1440

GTACTAAAAG CCGTTACTTT TAATAATCTT AATGTTTAGA AAGAGGAAGA TAAGTCGAAA 1500

CGCTTATCTT CCTCTTTCGT TCTTTAGTTT GTGAATGAAT AGATAATTTC TATTTCACAA 1560

AATAAAGTAA TTTTAATTCA GTGAAATAGT TTGTTTATAG GGGAAAAAGG GGGTTTTTCT 1620

CAATTTCTCT TTTATTTTTA AGCAAAAAAA TTGACAAAAT CTTTCAAAAT TTGTGAAATT 1680

CCGCTATAAT AAACAAGaAT ATGCGTGCAA CTATTCTATT TTTTAGATAT AAAGGATGAG 1740

AAAAACATGC GGAAAATGAA AACAATGGAT GGGAATGCAG CAGCaGCTTA TATTTCCtAT 1800

GCTTTTAcAG AATTAGCAGC CaTTTATCCA ATAAcACCAA GTTCAACGAT GGCTGAACTT 1860

GTTGATCAGT GGTCGGCCGA AGtAAGAAAA ATATCTTTGG TCAACCTGTG AAAGTAGTCG 1920

AAATGCAATC GGAAGCTGGT GCTGCCGGCG TGGTCCATGG CTCTTTAAAA AcAGGCGCGC 1980

TTACCACGAC CTATACGGCT TCACAAGGGT TACTTTTAAT GATTCCTAAT ATGTACAAAA 2040

TAGCTGGAGA ACTATTACCA TCTGTTTTTC ATGTGGCGAG CCGAGCTTTA ACTACGAATG 2100

CTTTAAATAT TTTTGGTGAT CAAGGAGACG TAATGGCCGC ACGCCAAACA GGCTTTGCTA 2160

TGTTGTCAGA AAGCAGTGTG CAAGAAGTGA TGGATTTGGC ACCAGTGGCT CACTTGGCTT 2220

CGATTGAAGC GAGTGTTCCA TTTATGAACT TTTTTGATGG CTTTAGAACA AGTCATGAAA 2280

TTCAAAAAGT GGCAGTTCTT GACTATGAAG AATTAGCACC CTTAGTAAAT CAAGAGAAAT 2340

TAGCCGAATT CCATCGTCGG AGCATGAATC CGAATCACCC TTCCGTTAGT GGAATGAACC 2400

AAAATCCAGA TATTCACTTT CAACAACGAG AAACAATAAA TCCTTATTAT GAAAAATTAC 2460

CTGGTATTGT TCAAAAATAT ATGACAGAAA TTAACCGGTT ACGGGGAACC AATTATGATT 2520

TAGTGACGTA TTATGGCGCA GAAGATGCAG AAGAAGTGAT CGTTACAATG GGTTCAGTTG 2580

CGCAAACAAT TGAACAGACT GTTGATTATT TACAACAACA AGGAAGAAAA GTTGGCTTTC 2640

TTAATGTCCA TCTCTATCGA CCATTTCCAA TTGAAACATT TCTCGAAAAA GTACCTCAAT 2700

CAGTTAAAGC AATTGCTGTC TTGGATCGGA CGAAAGAGCC CGGTGCTGGT GGTGAACCGT 2760

TATTATTGGA TGTCCAAAGT GCGATGTATG AAGCGGACAT TCGCCCAACA ATTATTGGTG 2820

GGCGCTATGG GTTAGGTTCT AAAGATGTAC TGCCAAATCA AATTGTTGCG GTTTTTGATG 2880

AATTAATGAA AGAACGTTCT GCCATGAAGA AGCGCTTTAC AATTGGGATT GATGATGATT 2940

TAACTTATAC TTCATTAGAA GTTGGCAAAC CACTTGATTT GACGAATCCG AAAACCTATC 3000

AAGCTAAATT TTGGGGCTTT GGTTCTGATG GGACAGTAGG GGCGAACAAG TCAGCGATTA 3060

AGATTATCGG TGATCATACA GACAAGTATG CTCAAGGCTT TTTCTATTAT GATTCAAAAA 3120

AATCAGGCGG GTTGACCGTT TCTCATTTAC GTTTTGGCGA GACACCCATT CGTTCAACGT 3180

ACCTCATTGA ACATTCTGAT TTTGTGGCCT GTCATACAGC CGCTTATTTG CATACCTATG 3240

ATTTAGTAAA AGGGTTAAAA AAGGGCGGTA CTTTCCTCTT AAATACCATT TGGAATGATG 3300

AACAACTAGC ACGATTTTTA CCAAATCAAT TAAAACGGTA TCTAGCTGAA AATGAGATTC 3360

AGTTTTATAC AATCAATGCA GTGAAATTAG CTAGTGAAGT CGGACTAGGT GGGCGAATTA 3420

ATACCGCGAT GGAAACAGCT TTCTTCAAAT TAGCGGAAAT TATGCCTTTT GAGCAAGTGT 3480

TGCCAATTTT AAAAGAAGAA GCACTGAAAA GTTATGGTCA TAAGTCCATG AAAGTTGTCG 3540

AAAAAAATAT TCAAGCAATC GATAAGACCG TTGAATTGTT GCATCAGGTA CCTGTTCCAG 3600

CAGAGTGGAG AACGCTTGAA GTTCAACCGC GAAAACGTTC AGAAAATGTT AGTGATTTTG 3660

TTCATGAAAT TGTTGAACCG ATTAATCGTC AAGAAGGGAA TGCGTTATCT GTTGCAACAC 3720

TAGCGAAAAA TGGGATGACA GATGGGCGGA TGCCTTTGGG AACTGCGGCA GTTGAAAAAC 3780

GTGGAGTTGC ATTAGAAGTA CCTGAGTGGA TTAGCGATCG ATGTACTATG TGTAATGAGT 3840

GTGCCTTTGT ATGTCCTCAT GCGGCGATAC GACCATTCTT AGCCGATGAA GAAGAAATGA 3900

CTGAAGCGCC AGAAGGATTC ATTGTGCGTG ATTTACGTGG TGCCGACGGC TTGAAGTATC 3960

GTATCCAAGT CTCAGTCAAA GATTGTACTG GCTGTGGTCT TTGTGTTGAA GCCTGTCCAG 4020

CTAAAGGGAA AGCTTTAGTA ATGAAACCTT ATGAAGAAGA AAAAGAGCAG GCTATGAACT 4080

GGGCGTTTGC TATGACGTTA CGTCAAAAGG AAAACCCAGC AAAACCGAAT ACTGTTTTAG 4140

TTCGCAATTT AATAAACCTT ATTAGAGTTT TCTGGT 4176

(2) INFORMATION FOR SEQ ID NO: 189:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3528 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

CTATATGATA TATCATTTAA AACAAAGGAG CnACTTTCAT GAATATTCAA GCGATTGATA 60

CGCGACACGG AACTGCCAAC CAACATAGTT TTTCAAATGG AAATTGTCTA CCCTATACTG 120

GCGTGCCATT TGGGATGAAC TTTTACGCCC CGCAAACAAC TGACCAAAAA GGCAGCTGGT 180

GGTTTCATCC AGAAGATCGA ACATTTCAAG GCTATCGAGT AACGCATCAA CCAAGTCCTT 240

GGATGGGTGA TTTTAGTCAT TTATTAATGA CCCCTGTCAG TGGTTCATTG TCGGAACTTT 300

CTTTATTCCA TGCCCAAAGT TCTTATCGTC CAGAAGAAAG CCTTTTTTCT CCTGTCGAAA 360

TTAACTTAAC TCAATTACGC TATCAAATAA CTTCTCAATT AATCCCAAGC ATGTATGGTG 420

GTATCTTAAC CATTGATTAT CAACAAAAAG ATAATCACTT ACTACTAACT CTGCCTGGAC 480

GTTATCAAGT AAAACAATTA GATGACCATC AAGTGGCTGT AAAAGTCATC AACTATTCTG 540

GTTGTGAAGA TCCAGATTTT AGTTTTTATT TTGTGCTTCA TTTTGAACAG CCTTTAACCA 600
AATGGTTTGC CCCTTCTTCA GGTGAAGATG GCAAAATCCT TCTTTCTTTT GGAAATATTG 660
CTCAACAAGT TGTTCATTTT AGTAGTTCAT TTATTTCTGA AAAACAAGCG CAGTTAAATT 720
TAGCGCGAGA AATCTCTTTG CGTTCTACTG AGATGCTGCA ACAAGGAATC GCAGATTGGC 780
ATAATTATTT CGATCGCCTT AAAGTAACGC ATGAAAACCC TGAACATACT AAAACGTTTT 840
ATCATACACT GTATCGAACA TTCCTATTTC CGCAAACCTT TTATGAATTG GACGAAAATC 900
AGCAGCCGAT TCATTATGAT ACGTTCTCGC AAACGGTTCG TCCTGGTGTA CTATATACCA 960
ATAATGGCTT TTGGGATACT TATAAAACGG TCTACCCATT GTTCTCACTA ATTGCTCAAG 1020
AAAAATACGA AGAGATGCTA GAAGGCTTTT TAAATAGTTA CAACGAAACA GGTTTCTTGC 1080
CAAAATGGTT ATCTCCAGAT GAGCGTGGTT TAATGCCCGG CACCTTGATT GACGCTGTCA 1140
TTGCTGATGC CGCCGTGAAA AAGATTCGTC CTGACTTGAT GCCTCAGTTC CTAGAAGCGA 1200
TGAAAAAAGG CGCTACACAA CAAAGCGAAC GAGAAAATTA TGGACGTCAA GGAACTCTGG 1260
ATTATCTAAA ATATGGCTAT GTTCCTTCAA CCTATCATGA ATCTGTCAAT CATACCTTAG 1320
ATTATGCGTA CAGCGATTTC TGTATTtCGC AAGTTGCAAA AACATTGAaT GaTTCAGAGA 1380
CCGCAACGTT TTATCGTCAA CAAGCCTTGA ATTATCAACA ACTTTTTAAT CCTGAAACTG 1440
GCTTTATGCA AGCAAaGGAT ACTGAAGGCA ATTTCCGTCC TGACTTTTTA GATATTCGTT 1500
GGGGCAAAGA TTACGCCGAA GGAAGTGCTT GGCAATCAAG TTTTGCGGTT TATCAAGATT 1560
TTGCAGGTTT GATTAAGCTA TATGGTAGCG AGTTAGCTTT CGAAAAAAAA TTAATTCAGC 1620
TATGCAATCA AGCGCCTAAC TTTAACGTTG AAGGCTATGG TTTTGAAATT CATGAAATGA 1680
GTGAAATGGC GGCAATTGAT TTTGGTCAAT TAGCCATCTC AAATCAACCA AGTTTCCATT 1740
ATCCGTTTCT ATTCAGCTAC ATCGGCAAAC CAGAGATGGC TCAACCACTT TTAAAACAAC 1800
TAATGCAAAC ATTTGATGCT TCACCAACTG GCTATCCAGG TGATGAAGAC AATGGCAGCA 1860
TGTCTGCTTG GTATATTTTT AATAGCTTAG GCTTTTATCC CGTCACACCA GGAACTGGTG 1920
AATATGTAAT CGGGATGCCT TTAGTCCAGA CGGCGGAAGT AAAACTATCA AATGGTAAAC 1980
AGTTAACCAT TCAGACATCA CCAAACAAAG TCCAACAACA ATTTATTCAT GAAATTCAGT 2040
TAAACCAAGA AAAACATACT GCGCCTTATT TTACTCATCA AGAACTATTA AATGGCGGGA 2100
CTTTAGATTA TCAATTAGGG ATCGTCCCGA ATCCACAAAA CACAGCAGAA CGACCGTTTT 2160
CATTAAGTAC GGAAAAATAA TAAAAAGAAA GATGTTCATA AAAATTGAAC ATCTTTCTTT 2220
TTATTCTTCG ACACCTGCGT AATTTGATTG GTATTGCTCT CGAAATTCTT TTGGCGTAAT 2280
TCCATTTAAT TTTTTAAACA TTTTTGAAAA ATAGTTGGTA TTGTTAAATC CTACTTTTTC 2340
TGAAATTTCG TTGATTGTAT CATCCGTGTA CAATAATAAT TTCTGGGCAT GCTTGGTACG 2400
CACTTGaTTC AAATATTGAG aAAAACTCCG CTTGGTTTCT TTTTTAAAAA TTTGGCCTAA 2460
ATAACTAGGG TTCAAATGAA GCTCATCCGC CACCGTTTTT AAGTtTAATT CTTCTAAATA 2520

GCCTTGTTCA ATTCTTTCTA CAGCTTGGCT AACACTTTCT GAATATTTTT TTATTyTAGG 2580

TTTATTTTTC ACTTCTGCCA AGACTTGCTG CATTAACGCA TACAGTGTAT TAAGATTGTT 2640

GCTGGCTCGA ATTTCTTCCA CGATTTTTTC ATAGTTTTCT TCTAACACAG TTGGATACTG 2700

ACGGGAAATA TCTGAAAAAA GTAAGAAGAC TACATAGCGT ACATATTCTG GCGTAGCGTG 2760

TTGTTGACTT AACTGATTGA AAATAGCATC TAATTCCCGT TGAATGGTCT GCATATCCCC 2820

AATCATCAAG GCTTGGTTAA ACGCTAGGAA AGTAAAATCT TCTGCCGAAC GATTCaTTTG 2880

TTCTTTTTGA TTATTTGGCA GAAATTCTTG ATAAAACGTT TCAATTGCTT GCAACTGTTT 2940

TACTTTCTCA TAACTTTCAT AAACCGTTTC CCATTCATTG ACAGATTCCC CTACAAAAAT 3000

CTGACCTTGA CCAGCTAGAA TCGCTTCAAT TTCACGAATA AATAAGAACG CTTCTTGTCT 3060

GGCCCCTTTA TAGATAAGTA ACAATTGGTG GTACGTACTC ATCCAACTTT CAATGAATAA 3120

CAGTTGGCCT TTTTTCATAA ATAGTTCCTT GATTGTTTCT AAAAATGACA TCTCAGCTTC 3180

AACCAATAAA ACGGTAAAAT CACCTGGTTT TATAGTCGCA AATTGCTGCA TTAACTCTTC 3240

GTATTCCCCT TCGCTGATTT CATCATTAAG CCAACGCATC AAGCCATTTT CTAGATATAG 3300

ACGCACGCGG GCATGGGATT TTTTTTGTTG GTCCAATTCT ATTTTAATTC TCTGGATATT 3360

GGCTAATAGT TCTTGTTTAT TGACTGGTTT CAAAAGATAA TCTTGGACCT GCAAAGCCAT 3420

GCCTTTTTTG ACGTAATCAA ACTCTTGATA ACcTGATAAA ATTAAACTAA TAAATTGAwG 3480

TCCtTTTkTC TTtGCTtGCT CaATTAATTC gATCCCGCTC ATTcTGGa 3528

(2) INFORMATION FOR SEQ ID NO: 190:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11410 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

TGGATCCTAC nAAAGGCCTT TACnTACCCG nTTTCCCAAA TTCCCTTAAC CCATATCnTG 60

ATAGGCTCTA GAGGAAGTAT ATCCCCCAGC CGAGAAAaGt AGtGTTAGTA GAAAGATTCT 120

CGCTAATATA TTTTAAAATT TGATATACAG AACTTACCTC TTGTTCTGGT GTCTTTCCAG 180

TAGTTAAAAT AAGTCTGCCA TAGCTTTCAG ATACTTGCTT AATTTCATCT TCTAGATTTG 240

GATATTCTAC AAAAACATCT TGAATATACT CATAAATAAC AGGAGTGCTA TCTATATATT 300

CAACAAACAT TTTAATGATA GAATTAACTT CACTGTAATG AGCATTAAGA ACTCTAGAAG 360

CTAATTGTCT AAATTTCCTA CTTGCCTTTT TTAAATCTTT TTTATTGATA TTAATCTCCA 420

TCTTGCATCT CCCCACAGAA CGGACAATAT CCGCCTTCCA TTTTTAAGCT TGGACTAATT 480

TTTGCTTCTT TATGACAGCA GTCATAAATT TGAAGTTCCA AGTTATCTAG TTTTGAAATA 540

ATCGGATCAA CAGGATCTTT CTTAATTTTT GAACCTGATT TGAATTTTAC GCTTCCATTC 600

TTTGATGATT TACTCAAATT CTTACCAAAA TCATTAAGTA AATCAGATAC ATGATTATTT 660
ACAATTTTCA TTGCTAATTC TATTACTTCA TCAGTAAACA AACTATCTGG AGTTAGACCA 720
CACTTAGGAC ACCAAACATG CAATTGAGAT TCATCTTCCA AGTCCTTAGG CGTTAGCTTA 780
AAATACTCAC CACACAAAGT GCATTGTAAG AAAATAAATC CATCTGAATC TGATGGGATT 840
GAAATATTAA AAGTAGTGTC AGCTGTCATA GATACATCTC CTTTTGGGAA TACTTAATTT 900
ACCTAATTAC TATAGTAATT TTATCACAAA AAGAAGGCGA AAATCTTTTT TATGTTATAT 960
TTCTTTAATA TAAATTAGAT ATTTATATTA TTTTGCACAA TGGTTTTGGT AGATTCCAAA 1020
AAATCATACC CCTTCGGAAC TAAAGGCGTG TAAAACTCAC TAACAACACC CCCACCTGTA 1080
TCTACATAGC CTCGCCCCTT GaTTTGTTTC ATAAAAAAAA TTTTATCAAC CTCACCAAAC 1140
ATCATAGAAT ACCCCAGTTC ACTCATACGA CCCAAAGCCA CACGGAAGTT GAATTGATCC 1200
CGTATTCCGT CTCCTAAATA TTTCGCATCC GGTCTTTGAC AAGCCAGAAT CAGAAAGAAA 1260
CCGGACTGCC GACCTAACAT GACGATTTGT TTGAGCTTAT TCAAAATTAC CGCGCTTTCC 1320
TTAGTCGTTA ACATCTCCAT ATAAGCCACA TATTCATCAA AAACTAAAAA ATTTGGTGAA 1380
AGACCAAGAT ACGCATAATT CTCTCCTGTT TTGTAGTTTG GCATTTCTTT CATCAACCTA 1440
CTACGAaCCA TCATACGTTC ATAGAAATCT TCTACACAAG CGGAAATTTG TTCTTTTTGG 1500
GaATATACGT GAGTCATTAT TGTGCCTAAA TCTGCTAGAT CCGCATTTTT GGGaTCAAGG 1560
ATAAATAGTT CTGCATCAGA TTTCAGTAAA GCCTCGATGA TGGTGAGTAA AAAGTAAGTC 1620
TTCCCACCAC CTGTACCACC TGCGATTAAC ATATGAGGCA AGAAATCATA CGCCCATACT 1680
TGATTTTTCA TCAATCGCAA AGCTCCATTT TCTACCACTG CTTCTTCTAT CCCAATTCGA 1740
TTGGCAATCA TATCGTAAAG CAAGGTATAA TCCACATAGG AGTCTTTGAG TTCCTTTTCT 1800
ACTAACTCAC AATACAGTCC GCTTTCTAAC TTCTTTTCTA GCTTTAATAG TTGATCTTGa 1860
TATTTTCCCA GTGAAATTgA ACGCGAATAG ArAGTAAGCC ATCCTTCATc GATAATAAAT 1920
TTgGGGrAGT AACTGATACT TTCTTTGGTC CGACTACCAT cAAATCTTTG AAAAAcCTTC 1980
ACTTTGAGTT TGTTTCACTT CATACCAATG GTTTTCTAGA ATCATCCGAG CTAGCTTTTG 2040
TCGATGCCAT AAACTGCGCC AATGATCCTG AAAAAAGTGA ATATAAAGCA ATATCATACC 2100
TCCTGAAAGA GCAAAACTTC CTAGTAAACA GATTAAATCT TTCGTGGACC AATCAATTTC 2160
ACTAGTGGAG AATGATAGCT TAGAAATCTC AAATTGTCTA AGCTCTTGAA AATAAAAGAC 2220
GCTCATACAA CCAATGAACA AAGCAAGTAG CCAGCCAAGA ACAAAGTGAT AAATTAAATG 2280
CTGATTTCTT GCGCGTATGC GATGTCCTTT GTACATCTGA TGGTTTCCAC CTCCAATCAA 2340
AAATTGCCCA ATAGGATTTA TCGGGCAATT TCAAAAAGTT ATGCTAGTCA TTTACTACTA 2400
CGGGAAGCTG TTTCGTTTCA ACCAATTGTT TTAAAAAGCT TTCAATGACA TCCGGCATCT 2460
CCACACCGAT TTCTTCTAAA ATAGACTCAG CTTTTCGACA TAAGTTTTCA TTTATTTCCA 2520
CAATCATAAA GAAGGATTCC TCTTCCCCTA ACCAATCCAG AACATTGAAC TCTTCTTCAC 2580

```
TATCGGTATT TCCCCCAAGT AGCCCCTGAT GCATAATCAG AAACATTTTC TCATCATCGG    2640
TCACCCCTTT AGTGAGTTGC TTTTCGGTTA GCTCTCCTTT TCCATAGCGA TCTAGACAAT    2700
CAAAAATATG TTCAACATCA ATCTCCATAC TCTTCACCTA TTTCTTTCCT TCATTTTGAT    2760
GGGTATTTCC TTGATTCGAT TGCCGTTTCA AGACTAAATC ATCCGCCTTA ATAAACCAAT    2820
TGACGGTAGT TCCTTCTCGA AAAGCATAAT TTGCGACCGT GTCCACGATT GGATTGACCA    2880
ACTCTACTTC TGTATCGTAG GCAAACTCTC GTAATGGCAC ATTTGCTGGA ATACTGACTT    2940
GAATCATTCG ACCTTGTGCT GAGGATTTCA AATTATAGGT TCGTTCTTTG ACCTCATTAC    3000
TAACCGTTCC ATCCTCATTT TGTTTCCGAA CTTCTCGGCG TAATGCAGAA AAACGCAACT    3060
TGCCAAACGT TAGTCCTGAA TCTACTACAA TACCTTCTGT CAATCTCATA TTTTTATCCT    3120
CCTATGCTTT CACTAAATCA TCGACATACA ACACATAATC GGTAAAACCG CGGTTTTCAA    3180
TCGCATACCC TTTGGCAACT AAATGGGGAT TCACAGcTTT CAGCGGTTGA TCTTGGTCAA    3240
AaTCTTTTTC GCCGGCTTCC GCTGCTACCA CCACTTCAAT ATCATCTGCT CGTTGTTTAC    3300
TGGAATATAG ATGATAGCTA CGACCAATCA CCACAGTATT tCGGCCATAT CCTTCTGTTA    3360
AAATTtCTCC TTCACCAGCA TAACCTAACT TTCCAAAGGT TTCGGCCATA TCCGGCACGA    3420
CAAATTTCAA TTCCATATTC TTTTCCTCTT TTCATTGTTA GTTTTGGATA CAAAAATAGC    3480
GCGACCACTT TTTTAGGAGG CGGTCGCACA AGTCGGTTCC TACTATTCAC TTGTTTCATC    3540
TCATCACCTT AATTGCGGCG ATTCCAGAAA TAATACCCAG CAGTCGAAAA AATCAAGATA    3600
AAGCCAACTA ACAGTAGAAC GTTGGAATTT TTCTCCCCAG TTTGCGGGAA TGTCTTCACT    3660
GTGGGACTAG ATTCTTGAGA ATTACTTGGA ACAGCTGGTG TTTCCGGTTG TTTTGGCGTA    3720
CTCGGTATGG TTGGCACTTC TTTTGGTGTT AAGGTTTGAG ATTTTTCTTT CAAATCTTCA    3780
TTGTGTTTTC CATTCACGTT TCCATCTTTT TCGTAATTGA TTTCCGTAAA AGTAAACTTA    3840
GTTCCTTCTG GATACTTTCC GGTATCTACT TTTTCCGCAA GTACGGTTTT GGTAAATTCT    3900
TTATCATTCA CTTCATGCTC AATTTTGCCA GATTTCCAAA TTTCTTTGTT CGTACCATCT    3960
GGTAGTAAAG CATACAGAAT TGTTTCGAAA GCTTCTTTTG AGCCATCCAG TACATCATGG    4020
GTAACCGACA CATCATCAAA CATATCCATC ACGTCACCAT GAGTAAAAGT TTGCGAACCA    4080
TCTTCTAGGT GGGCTTTTGT TTGGATGGAA ACATGGCGTT CAATCGTACA ATTGACGGTT    4140
TGTGCTTGAT TGTTCAAGCT AGCATCCTTC GCTACTGGCT CATTGCCTGC TTCAAAGGCT    4200
ACCTTGTTTT CATACACATA ATTGAACAAG ACGATGCTTT TATCCAAAAC TTGTTCAGCG    4260
GTTAATTTAT GCAGAATTTT CCAAGTTCCA GTTTTCTCTT TGTTCGCCAA CGTTGCGGTT    4320
GTTTCGGCAA TCACCACCGG TTTGGCTTTT TCGTCTTTTT CTTGGGCAGC TTTTGTGGCT    4380
TCCACATCAA TGGCTTGAGC TACGACATAC CAGTCTTCTT TGATTTCATG CAAGTTATAT    4440
CTCAATTTAT CAACTAGCTC GGTAAAATCA AGGGTATTCA ATTTTTATT TCCGTCTTTC     4500
CAAGTCGCAA GAGAAGTCAA ACTATCTTCT TTCTCGGGCA AATCATAAAG CATCAAACGG    4560
```

```
TTCAGACTAA CAGAAAACTC GTTGTTAGTT AGTTTCTCGT AAGGAACGGT CACCATCTTA      4620
ATTGGTTGTT TTTGTCCTTC TTCGGTAATG GTAAAGACAT ACTCACTCTT CGCATAGTCG      4680
TCTTTGTTTT CCTTAAATGT AGAACGGATT TCTAGTGGTG TAATCTTTTG AAATCCTTCT      4740
GGAGCCTCTA TTTCTTCAAG TAAATAATCC CCATAAGGAA GATTTTCAAA CTTACCATAG      4800
CCATCAAAAC CTAATTGCTC GTTACAAGCT GTGGTCGCTT TATCTTCAGC ACCTGTGATT      4860
TCmTTGGTCC CTTCCaATGG CGACACTTTA AAAGATAAGT CGTTAAATCC aGTTTCGGCA      4920
GTGCCATCAG CCGATCCAGC AAATTTAAAG AAATCAAAGC CAAAGCGAAT AACTTGTTCT      4980
TTTGCCGTAA CATCTCGAGT AATTACGGCA TTTTTTTCGT TATTATCAAC TTTTTTGATG      5040
GATACAGGAT ACTtCGTTTC ATCCAAAGTA TATCCTTcAG GTGCTTTGGT TTCTTGCCAG      5100
AAATACTCGT TAATTGCTAG GTGTTTAACG GCAACTTGGT TCTTTTCATC TAAAGCCAAA      5160
GTCACTGTTT CATCAGAAGC TTTCGTTCCC TTCACTAATT CTGTTTTAAA AGCTTCACTC      5220
CATTTAACAG CTTGACCATC TTTTGCAGTA AAGAGAGTAT ATTCAGCTCC TTTAAACTCA      5280
GCTTTCCCTT GACTCTCATT ACCGGTATCT TTGTCTTCTT TTGTCAAAGT GGTTTCCCCA      5340
GTAATTTCTT GGTTTTGCCC TTTTACGTTA CTGGTAACAA GAGCCACGGT TTGATTGGCA      5400
TATTTTAACT CGACTTTTGT TGGTTTGAAG GTATTCACGA AACCATTACT AGATTTAGTT      5460
TCTGTCACGT AATAGGTTCC CAGTTCCAAA GCATTAGCAA GCTCTTTTGG TGTTTCCGCA      5520
CGACCTTTTT CATCCGTTGT TATTTCTTGG ACAATTTCAC CAGCTGGGCT GTCTTTACGA      5580
ATGGCAAATG TATTTCCAGC TAGAGAATAA TTGTCATTCC AAAGATCaGT ACCTGTTTCT      5640
ACCCCAGTCT TCTCTAAAAG AATTTGACCT TTTTGTCGCA TATTTTTCGA AGTCATGGAA      5700
ATGGTCTCGC CCGCTTTAAT GGTGGCAGCC ATGGGTGTGG TATCAATCAT ATAAGGATCT      5760
GGCACCGATT TTTCAGTAAT AGTTACCTTT GTACCATGGG GAATTCCATC CAAAATAGAA      5820
ATCCCATCTT TATCTGTTGT CACATCTTTT GAAGGTAAAG CTTTCCCAAA ATCTAAATGG      5880
AAAACCGTTT CTGGTACAAT ATCACCTGAT TCTTTATCGA TTTTTTTGAT CTTTAAAGAA      5940
CCCTTAGTTT CCACATTAAT TTTAATAGCG TAGGTATTGG GCTTATCAAG CGCACCAGCC      6000
ATCACAGTTT GAAGTCCTGC TTTTTTATAA GCGACTGGAG TTCCAGTACC AGCTGATTTT      6060
TtCAATGTTA ATGTTCCTGA TTTGGAATTA GAGTTTGGAG TAAGCACTAA TTGATTCCCA      6120
ATTACACGGT AATCTATATT CGCCGTATTT TGGACGACTT TATCAAACTC AGATAAATTT      6180
AATTCATTTT TATCTATTAA AGTTGTCGAT TGACCTAAAA TTGTTTTTAC AGTGGTATTA      6240
TGAAAACTTG GTTTTTTTTG ATACTCCTCA ATTGCCTTAT TAATTTTTCC TTCAATAGAT      6300
TTTATATCAA CTGAAGCACC ACCTAATCTT TTTATGGAAT GGAGTTTATA ACCGTTCACT      6360
TCTTCCCAAA TCATCTTTTG TGCAACCATA TTTGTATCAA TATCTGTTCC AGCCTTTTCC      6420
CAAAGAACCG ATACTAGTTT CGCTTTATCA GACATTGATG GCAAAGGGTT TTTCTGATAA      6480
CCGTGCGTCA CTTCTGTTGG AATACTAACA CCTGGTTCAA TACAAAATAC TGTTTGTTTT      6540
```

GCACCATTGT ATTCTGCATA TAATGACGTA GTACGTTCTG AGAGCGTGCT GCCATCGCTA 6600

AAAGTTCCGT CAACTTCATA CAGTTTACTT ACGTCATAAT GAATTGTTAC CTCTTGAGGG 6660

TGAGTAATTT CATCGGCAAA AGCAATCGTA GGACCAATCG TTCCGCCAAT TGTTTGACTT 6720

AATAGAGCCA AAGTCGCAAA TAACGACCAA TTTTTAAATG TTGTTTTTTT CATATAATTA 6780

ATCCCCTTTC GTTTTTGAGT AATAAAAAAG AGACATTCAC CTCTGaATGT CTCAAATGTT 6840

TAAAACTATT TTTTTCAAAG ATTATCAGAT ACCATATATC AAAAACTCCA AAGAGACTTT 6900

TCCTATATGG GCATACTGAA TAATGCGAGC CGTATTCGGA ATATCTTTTC CGtCTTCGTA 6960

TTTATCCaTC TTACATTTAT TCATAAAGCC CGCCTTTCTT TTTGAACATA AAAAAAGACC 7020

CTCAATTTTT GACGGTCTCA TCGTAAAAAT TTACGTTATT CATTCACCCA AGCTTCCACT 7080

AAATCCaCAG TATCACTTTG TGGACGTAAA GTTTTATCAC TTATAAAATC AACATTCACA 7140

ATCATTGGTA AGGACAAAGC TGTCCCAGTG ATAATACAGA TACCTTGTGA TAAACCTGGA 7200

AGCATTGATT TAGAAACTCT ATCTAGAGTT GATAATGAAT TATCAATTAT TTGAAGATCT 7260

CGCTCATTCA CTAGCTTGTG TAGAAAGAAA TTATGAACTT GTGAGAGAAG TGTGGGTGAA 7320

ATATCCGCTG GTCTTTGACT TGATAAAGTA AGGAAAAAAC CAAATTTCCG ACCTTCTTTA 7380

ATGATTTCCT CAAAAGTAGA AAGTCTATAA TCCTTCCATG ACGTATTGTC GTTTCTTGAT 7440

TGATCAGACA AAATATTATG TGCTTCATCG ATAATAAGAT GGAAACTAAC CTTATCTTTA 7500

CTCGTCTTCT GCTCATCGAA AAACATTTTT GAAGTTAACA TCGCTACAAT TTGTTTAGCT 7560

TCTTGATTCA AATCACGTAA AGAAATTATC TCAAGAAATT TGACATTGTC GATCTGTCCC 7620

ACTATCTCTA TATACTTCTG GAAATCTTCA ATACGGCTAT CGATTCGCTT TAGAAGCGGG 7680

TCTATATGAT CCGATATGAC ATTTCGATAT AGTAAGTCTT TGACTAATTG TAATTTACAT 7740

CTAAGTTCAA ACTCTTTGAA TATATTTAAT TTTTCCGTTA GAATCACATT TTCAACTTGT 7800

TCAAGATTAA CTGCCTGCAT ATGTTTAGAT TCCCAATTTC CATCAAAAAA AGTACCGTTT 7860

ACAATAAATT TTGCAGGGTC ATCCtTACTG TATATCTGTG CCTGTTTGAT TGGCTCCAAA 7920

GCTTCACTAG AATCATAAAA TTCACCCAGG ATACTTACAA ATCTATTTTT TAAATCCATA 7980

TTGGGAAAAG TCGTAAATAT CTCTCTAAGT ATAGAACTTA TCCATCTTTG CAGTGACTCT 8040

TCACCAAAAC CAAACTTATT TATCCCTTTA AGAACACGTT CCAAAAAAGG TTTTTGAGTT 8100

kGTkGAGTCG CTGrAAAAAG rATwGaCmAG aTCTCAGAAC TAAAAAAGAT ATCCCTTTTA 8160

ATCTGTAACC TCTTATTTGA AACGGTTCTA GTGGATAATT CATATATATT TTTTTCGCTC 8220

TTAGGTACAC CAAATGAGTT ATCATGCACA TACTCACCGT TAAAGTCTAT GACAACAAAT 8280

CTACTTTTCT CTCTTAGCAA TGGGAACGAT TGTTTAAATA GCTCAAAATA CAATTTATGA 8340

AGTGTATTTG ACTTGCCACT ACCAGTATTT CCAAAAATAC CAATATGTGA CGCAAAAAAA 8400

GAACTTATTG GTAAGTTAAC CTCAATTTGC TCATTTAAGC TTTTCCCGAT TTTAATAGTT 8460

TGATCTCCAT TAAAATGATC GAAGCTATTT ACATAGATAG TTTGTGTCTC TTCAGTTGTC 8520

GGTATTGAAC AAATATTTCC AATCATTGGA ATATATGTTG TACCAGTAGT GAAACAGCCT 8580

TCTGAAATAT ATCCAATAAT TTGGATATCA AGAATTCTTT TTATGGTATT CTTACTGAAT 8640

CTGTTATCCA ACTGATAGTT CTGCTTATTA TTTTGAGTAT CCCAAATTGA TTCAGAATTG 8700

ATTCTGCCTA TTATTTTAAT AAAGTTCTGA CGAATAACTA CAAAACTATT TACACTTAAG 8760

TTTTTTATCA CTTCACCATT ATTAATATAT GTCGCTTGGT TCATATCATC AAAAGTGATA 8820

GCTTTACACC TGATTCCGTT CACTTCAACA ACTAAAGCAA CCTCTATATT AGAATTATTC 8880

GTCATCATCT ACTACCTCCG TATATTGACT ATCAATATAC TTTTCAGAAA GATTTCCTTC 8940

GATAATCCTG TTAAAAGCTT CAAAATCTAT AACCGCATCT TTATTATCGT AAACTGATAC 9000

TTCCTCAGAA TAAGTGACAA TAGTACTATC ATTTGATCTA ATCkGATTTG TTTCTAACGG 9060

TAGGGGCGTC TCTTCTTCGT TTTCAGCTTG TTCCATAATA AAATCTGAAG GTGTTATAAA 9120

CACAATATTG GATGGAACTG CATTGGCTCC AAAACTAAAT TGATTTATTA TTTCACCTTT 9180

AGTACCATCT TTATAACAGA AAATGAATAC CAATAAATTT GGATTATTTA AAGATCTTTG 9240

GACAATATCT GTTATATGCT CATCTAAAAA TGAAAAACCA AATGCTATTA AAACCGATTG 9300

TTCCTTCTCA AGCTCGTAAG ACAAGAATCT AAGCATATTA AAATACTGCT CTTCAATTAA 9360

GGTAGACTGA AATTTCCTTT TTGTGGGTAA AACAATCTGT ATTTCACTAA TTTTCTCACT 9420

TATGCCTTGT AATTGATTAT AATACTTTTT ACCAAAATAA TTGATTACTT CAGTAGGACT 9480

ATCAATAAAT CCATCAACAG GTTTGCCTAG TATGACATCT TTCAATATCT CAATGGTAAA 9540

TGACCGTTGA AAATTTGTAT CATGAACTAT ATCGTTTTGG AGTTCTGACC AAACTTCTTC 9600

TGCTTTTTCA GCAGTTTTAA TAGTGTCCAA AAGGCTGTTA TCAACAAGCA CTCTATCTCT 9660

CCCGATAAAG TTTTGCTCAT TCAGCCaATT CACAGATCCA TGACATTTCA CCAAATTAAT 9720

TGTTGGTAAT TCGTTGGCAT AATTATCAAA GGCCCCGGCA TACAACATTG TCTTATTAAA 9780

ATTTTCAGTG CTGAGATATC TTTTCATATA ACCATTCGTC CCGTCATTAA AGAAAATTTT 9840

GGGATTTGAT TCTATATTTT TTTCgACAGC TATTTCaAGA AACAAATCAT AATTCGTAGT 9900

AAAGATATTA GTTCTTCTTG GGGAAACACG ACTATTTCTC AGTTTAGGA ACTCCACGCA 9960

ATTAAAAATA AATCTAGAAT ATTGATTCAT CACAGAAGAC CCAGAATCCA AGTGGTTTTC 10020

CGATTCACCT TTTTTTATAT ATTCAACCAA TATTGAATTA TCTTTGATTG CTATATAATA 10080

AAAAGCAAAC ATAACTAGTC TTTTACCTTC TTCAGTTATA TCTTTGCTAG TCATATATTC 10140

CTCAATTTTT CCTAAGCTTG AAAAAAATGG TACCGAAGCA CCTGCTCCAA ATaAGAATGT 10200

CGTTCtTTTC TCTgAAAGAA CAGCTCTAAG AGCTTTTACC TTTGAAATTG TCATAATAGT 10260

TCCCTCCAAA ATTGGAATTC ACTGTTATTT TACCACAATT TAATAACAAA CGATGATTCT 10320

CCTCAAAATT GTTTTTAAGA TATTTTTTTT TAAACTAATT AACGCCATAT ATCAAAAAAT 10380

TCAAAGAGAC TTTTCCTATA TAGGCATACT GAATAATACG ATCCGTATTT GGGATATCTT 10440

TTCCGTTCTC GTATTTGTCC ATTGACCATT TTGTAATAGG CGGTGAAAAA AACTTTCCAA 10500

AGTCTTCTCT TTCTAAGCGA TAATCCCTTC GAATCTTCCG AAGTCTTAGC CCCATTCGGA 10560

AGGAATCTAA CTTTTCAAAC TTGCTAATCC TACTTCCCTT CTTCAACATC TTACTAAAAG 10620

TCTCTCCATA CAAAATATAA GGGACGGCCA CTTTCCCAAG ATAGGCAACA TTCATTAATC 10680

TTTCGATATC AGGAATATTC GTACCCTTTT CCCATCTATT CACAACATTC TTACCCGCTG 10740

GTCGAGTAAA GTTCTTACCA AATTTTTCCA GTGTCTCACC TCGATTTTTC CGAATCTCTT 10800

GGATTCGTTT TCCCATCGCT ATTTTATCGA TGATGGGTAA TTTAAACGAA CCATCCTCTT 10860

GTTTTATATC CATCAAAATT TCTTTACCCT CGTTAGACAA TAACTGAGTG ATTTTCACAA 10920

TTGCATTTTT TGATAATGGT CGACTTTGAT TTAAGAATTT ATTTAGAAGA GATATTTCGA 10980

TTCCTAACTT CTTAGATAAA TCCTCACGAG AAAAAATATT TTTAGTTTGA TATTTTTCTA 11040

AATCAATCAA GACTTGTTGT TTTCTTTGAA TCATATCTTT AATAAATGAT TCAGCGCCGA 11100

GAACGTCACT CAAAATTCTA TCGtTCTTTT TAGATATGCT GCGTAATCCA TTTTCATATT 11160

TTACGATTGT AGCTTGTGAT ATATTTGATA ACCTTGCCAG TTCTTCTTGA GTTAATTGcT 11220

TAGATAACCT AGCTTCTTTT AAATGATTTC CCATTTTGaT TAACTCATCT TTaTTTaTTT 11280

GCATGAAAAA AGCACCTCCy TATGaGAGrT GcTTTcATCA TACGCTGACA AGTAAAAATC 11340

CTCAAGGAAA TTATTCCTTT TTTGTCCTTT TTTGTTAAGT TGTGACAACC GTATGACAGT 11400

TGAAACTTTT 11410

(2) INFORMATION FOR SEQ ID NO: 191:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1316 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

AATTAAATTT AGATTGTTAT CGATCATTCT AAACTCATTG TTTTCTAATG TCAATAAAAT 60

AATGAGATTT AGATTAAAAT AAAATGAATT TCGAATTTTA TGAGCTTTTT TAATAAATCA 120

GTAAATTCCA TGACATGTGA GATACTTCTC ATTTACAAAC ATTCTTAAAA AAGGTATGAT 180

AGTAAAGAGG CAACTTGCAA AGGAGTAGAA GACGAATGAG TGAACAACCT CAACAACGTT 240

CAACCGAAGA AGTTGAAGCA ATTAAAGAAG ATATTTTAAC TGCATTAGAA ACAGTGATTG 300

ACCCTGAATT AGGCATCGAT ATTGTTAATT TAGGACTAAT TTATGAAGTT GAATTTGCAG 360

AAGAAACTGG TGATACTGTC ATCAAAATGA CCTTAACCAC AATGGGTTGT CCGTTAGCAG 420

ATGTTTTAAC aGAACAAATC CATGGTGTTT TAAAAGATAT TCCTGAAGTG AATAATATCG 480

AaGTAAAATT AGTCTGGTAC CCTGCTTGGA CAACGGATAA AATGAGCCGT TACGCACGAA 540

TTGCTTTGGG TATTCGCTAG TTATTTCGAG ACTAATGATA AAAAACATAT GAAAGAACCG 600

ATTGGTCCGT TCTTTCATAT GTTTTTTTGT TTGTTTAATT TACTTATTTG AAAAAAACTA 660

CCTGATGCCA TCAATACTTG GCTTTTTCCA AATTTTTGCT ACTGGATGGA AGTCTAATAA 720

AAACTGTAGC TGCGTTCCAC CGAAGCCTGA CGCTCCAGCA AACATGATAT GAAAAGTGAG 780

CCCTCTATCC ATAATCAGAT ACCCTGTATC CAGCACGCCA TTTCTTTCAT AAAACTGATA 840

ACGTTTTcTC TTTGGGCTTC ATcGCTGCAT TTTTGcCAAT TTCTTCACTC TCCAAAAGAA 900

CCCCCTTAGG aTAACGTGTC TTcAATAACG cTAACACGGg TGAACCTAAa TCGnTTGtTC 960

GGTGGTGCTG ATCAATTGCA AAGAAAAaGA CATAGGnTGC tTCAGGkGCA AtTGGACTAT 1020

TTAACAAAAT CCCCACGTTC TCATCCGCCT CyTGaATAAC mAACmAATGA AaTTTtCCTG 1080

sCTtGAGAAG ACATCAATAA TTGATCAAAC GGCAGTCTCT CTTCTTCTGG AAAAGCTTCT 1140

ATGTAGGTTG CTTCCCAATA GGGCTTGATT TTCTTGATTT AATAATGTGA ATTGCATAAA 1200

AGCTCCTTTC TAAAATTGAT GCTCGTACTA ATAACTTTTG GTACTTACAT AAGACCTAAA 1260

AGTATTAAAC ATGCGACTGG GnTATCCTAT ACGCAATTAT nTTTGGATAC TAAATC 1316

(2) INFORMATION FOR SEQ ID NO: 192:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2001 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

TTATGGnGGA TnAAATAGGT TCTTACTGGT TATGGCTAAT ATATAATAAT TTTACTTTAC 60

TAATTTTCAA ATTTCAACAn AAAACGACCA CCTTATCGGG AAAAGGTAGT CTGCATGAAA 120

AATAAGAATT TTAATATGTA TCAATATTAC AATATCTGTT TTkTGAGTTC ATTAATTTTT 180

ATGTAAAAAG ATGCCGCCTC ATTGGGGAAA GGCGGCAAGA TGTACTAAAA ATGAAAAATA 240

AAAATTAATT GGTAAAGACA TTTACCGCCT TCTTTATTAA AATACAAGTT TTCATTTTAC 300

AAAATAATTG TATTTTATTT AAATTCAAAT ATACAATTAT ATTAAAATAC TCCTATCTTG 360

GACTTGGGAT AAGATAGGAG CnCTCTTCTC GAAAGAAGAT CATATATGAG TGTATTAAAA 420

AGGTAACTAT ATTTTAAATC ATCTTTTTTT ATATTTCAAC TATTTTTATA AAAAATCCCT 480

ACTyCTCATT TCATTGCGAG AGGTAGGGAT TTACTAGTAC mTaATGTAGC TAAAGTTTGA 540

TAAGTCATAC TAATTTAAAC AGATAGTTCT AGGGTAACCA GATGaAACAT TTACCCATAG 600

ATTAATTTCC CAACACTTTC TGTGTACAAC CAAATAAATT CATCTGGGAA TTGAACATGC 660

GAGATTGTTT TCCCATTTAA AGAGTATTGT TCGGTAATGA TCCCTACAGA AGATGGTTGA 720

AAAGCAAAAT AGTCTCCGAC TTTTAATTCT TGATCTGGCA TATTCTCTCC TGTTGCCGGA 780

TCGATCTTAT CAATATAAGC AACATTAATT CCGTTTTCTG TCCAATCAAA GTCCGGCTGC 840

CCCAATACAT CATTTCTTAC TTGCCAAATA CCGTTTACAA ATTGTAAATC GTCTACTCTA 900

TATCCGTAAC GGCGCTTAGG TTTAGGAGGA ACTGGAGGAG CGGGTGTTGG ATTACTTTTG 960

GGATTAGCGT ATTTGTCCCA CATGCTTCTA TCACCAAAGA ATACGTTTAA ATCTAATCGA 1020

CCATTCCAGC CAGGCAATTG ACCATTTGAC GTATATTGAT ACATTGAAAT AAAATTCCAA 1080

TAAGGGACTG GTGGCGGAGT AGGTTGGGAA TATCCTTGTG GATTATTTGC TCCGTAATCA 1140

GCTAACCACA AACCATAATC AGCTTTTGCA ATAGAGCTAA AATTATATGA TTGTAAAACA 1200

TTGGTATAAG TATAAAACAT AGGTTTTACC CCTGTTTTAC CTTGAACATA ATCTAACCAA 1260

CGTTTAGCCC AAGCAACGTC TCCTTTATTA GTACTTTCCC AGTCTAAAAC CAAGACGGCC 1320

TTACCGATAT AACCTTTAAT ATTTTTTAAA AAGAACTCAG CTTCTTGTTC CGCTGTCCCT 1380

TCAAGTCCTA CCTCATTCGC AAAATGATAT ACAGCTACTT TTTTCCCACT GCTAATGGCT 1440

TGTTGGAATG CTCTATCACA GTCAGGATTT ATATATCCTG TTCCTCCTGT AGCTTTAACT 1500

ATAACAAAAT CTGCAGGGAC ACCTTCTTTA CCTACGTTAA TATTTGATTG CCAACTAGAA 1560

ATATCAATTC CGTTCAATGT CATTTTTATC ATCTCCTATA AAATTTTTAA AAGATTGATG 1620

TAAGCCAGTG CTTGCTAATC CACTTAAAGC CCCATATATT GTATTTtCAG GTGTGATTTT 1680

ACCTTGTATC CACCATGAGA TTAATATACC TAAAATGAAT ACAATTAAAG GAATAAACTG 1740

ATTATATTTA TCTTTTAATA TATTAGTAwt TTyAATAACA AACCCGACAG CTAAACACCC 1800

TAGCACCACT AATGGCATCA ATTCAATTTT AGAAAAGTCC ATTATTTGCC CCTCCGTTCT 1860

AAATCTCGTA TTCTTTCGTT ATTTTCTTTT AAATTGTTCC ATAAAGATTT TATTTGTTCA 1920

TTTAATTnTA TTACTTCATC CTCTACTAAG CCTTGTTTTT CnTTAGTTTG ATCAATTAAT 1980

AAATTAAGCT TATnTACTGT G 2001

(2) INFORMATION FOR SEQ ID NO: 193:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4683 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

AATGATTGTA CGCTTCTTTA AAGGTTACTG TTGTGTCTTC TTTTCGATAA AGGATATATT 60

CGTCATCTAG CATATCCATT TGCAGCTCCT GTTTAAATCC ATTTCTGGCA ACAATAGCAG 120

AAACTAGATG AGCAACAGTC AAGCCATCTA ACGGAATATC CACTCCATTC ATTGCTAATG 180

TGTAGACGGT ATCTTCCACT AAACCACAGG CGTTGCAACC ACCACTGGTT AAATCTACTG 240

TACTTGTTAT AATTGCCTCT TGCATCGTTT CTTCCCCCAT ATTGTTTTTT AAAGTTATTC 300

TTTCATTTGT TATTTACGAA CGTAATTCAC ATCATAATCG TCAACTTCAA TATAGATATA 360

GATTTTTTTA TCAGGCATTT TACTTCGAAT GTCTCGTTCA ATCGCATTCA CAACATCAGG 420

TGCATCTTCT TCCATTGGTT CTACCACATC AATTTTTGCC GCTACTAAAA TTTCAGTTGG 480

ACCTAAATGA ACGGTCTTTA CATCAATTAG TTTTTCAACA TCCGCCCGAT CAAATGCGGT 540

TTTAATTTTT GCTAAATCTG TAGCGGTCAC ACTTTCTCCT ATAATCAAAC TATAAAATTC 600

TTTCGCTAAG AAAATGGCTG CACACATTAA TAGTAAACCA ATCAGTAAAC CACTGATAGC 660

ATCAAaTGCA GCAATCCCAG TTACCATTGT TAAAAGGGTT CCCACCAAAG CTAAACACAA 720

ACCAATAACC GCACAAAAGT CTTCCGTAAA AATAATTAAA ATTTCACTGT GACGACTTTC 780

TCGCAAAAAT TTTAAGAGCG ATAACTTCTC TGTATTCAAT TCTTTGATTT CTTTCATCGC 840

TACACGCAAC GAGCTACCTT CCACAAGCAA ACCAAAAATT AAAATACCAA TGACAATAGC 900

CACATTTTCT ACTTCATGAG AGGGATGCAG TAACTTCTGA ATAGCCTCCA TTACGCCTAA 960

CGCGCCCCCA CCAAAAAAGA GCATCGTTGC AACAATTGTA CTAAAAAAGT ATTTTGCACG 1020

GCCTTCTCCA AATTGATGTA ACGCACTTTG TCCACGAGAA GCGCGCTTAT TCCCAAAAAG 1080

TAATAACACC TGATTACTAC AGTCCACGAC ACTGTGGATA CTTTCATTCA TCATAGCGGT 1140

ACTACCACTT AAAGCAAAAC CAACAAATTT ACTAATCGCT ACTAAAATAT TCGCAATCAG 1200

TGCGGCAATA ACTGCAAACA TGCCACTCGA TTTTTTTGTC TCCATAATTA CCTCCTATAT 1260

GTAACGTAAA ACTAAGATAA ACCACCCGTT TATTCATCCA TCATTTTACC ACTTCTGAGG 1320

TATTTGTAAA AAAGAGTTCT CTTAAAATTC AAAAAGAAAT GTACTGACTG TTTTAGTTGG 1380

CCCTCATTTA TCGGTGTCCA TCCCACTCTT GATAAAATTC CTCCAGATAA AGTTCCATAA 1440

ATTGATGACG CTTTTCCGCC AATAAGCGAC CATAATTTGT ATTCATTTGT TCCTTTAACT 1500

TTAAAAGTTT CTCATAAAAA TGCATAATAG CAGTATCTTC CCCATTCCGG TATTCTTCTT 1560

GGGTTAACTG TTCACGAGGT TGTTTGTGAG GATCGTGAAT TAAACGGCCT TTATTTCCCG 1620

AATAAGCCAT CACCCGTGCA ATTCCAATCG CACCGATGGC ATCTAATCGA TCCGCATCCT 1680

GGACCACTTT TCCTTCTAAT GTAGTGACTG TTTTCCCAGT ATTTCCTCCT TTAAAGGACA 1740

TATTAGCAAT AATGGAAAGG ATGTGATTCG TTTGCTCTGT TGATACGTTA TTTGCTTCCA 1800

ACCATTGTTG CACCTTTAGC AAgCCAGCCT CTTCACTTTC GTTAAATTTT TCATCGCCAA 1860

TATCATGTAA CAATGCTGTT AATATACATA AAAATAAATT TGCTTGTTCT TTCTCTGCAA 1920

TTGTTTTTGC TAATTGTGTC ACCCGATAAA TATGCCACCA GTCATGGCCA CTTGCTTCGT 1980

TGGCTAATTC TTGTTGAACA TAACGCTCTG CTGCTTGAAT AATTGTTTGT TCTTCAAAAC 2040

TAAGAGATTC ATTTTCAAAC ATCCTAACAC TCCTTTATTC TGGATTTAAA ATTTCCACTT 2100

GGATAATCCC GCTAGCTTTC AGAAACATAT TGGTGACAAT CGGACCTACA AAAGAAAATC 2160

CTCTTTTTTT TAAGTCTTTT GCTACGTTAT CTGATAAAAG CTGTGTGCGT GGGACTTGGT 2220

ATGCTTCTTC ATAAATAGAT AAAACGGGGA CGTTTTTTAC AAATTGCCAT AGATAATTGG 2280

CAAAACTGCC AAATTCTTCT TGTACTTTTA AAATTGCTTG CGCATTTTTA ATGGTTGCTT 2340

CAATTTTGCG CCGATTGCGA ATCATATCTG GATCAGCTAC TATTCGATCG ACTTCGTCAG 2400

GCAACAGCGC GGCAACTTTT TCCGGCTTCA TTTGGCAAAA ATTTTTCATG AATGCTACTT 2460

TTTTCCCTGC TGCTGCTTTC CAACTTAAAC CAACTTGAAA AACACCCACC GTCAACAATA 2520

AAAACAGCAA ATCATCCTCT CTTGTAGGTT TTCCCCATTG TTCTGTATAC CATTGATAGT 2580

TGTTCATCGC CGCCTCCTTT TAAGACTGTT CAAATGTTAA TAATGCTTGT TTCATTGCTA 2640

```
AACCACCACG GTAACCGGTT AATTGTCCCT TGCTGCCGAT GACTCGATGG CAAGGAATGA   2700
CCATTAATAA AGGATTTTTC CCAATCGCTG TGCCAACTGC TCGAACCGCA CGTGGACGCC   2760
CAATTTTTTC TGAAATTTCC TGATAGGTCA TTGTCACACC ATAAGGGATT TCTTGCAACG   2820
CTTGCCAAAC TTCTTTTTGA AAGGGCGTTC CTTCGATGTC TATGGGGCAA GCAAACGTTG   2880
TCGAGTGTTG GTTAAGATAA CTTATTAGTG CTTCTTTATA AGGCGCAGTT TTCTTTTCAG   2940
AAGCTTCTAA TTGTGCTTGA GGGTACCAGT TTTTCAGCTC TTCTATCGAT TGATTGGCTG   3000
AACCTACAAA AGCTAGTCCA GCATCTGTCA CACCAAGATA ATAGTGATCG TTCTGCCATT   3060
CTTGTTGCGC ATAATAAATT TTTTTTGCCA TATTTCTACA CTTCTTTCTC GTATTTTCTT   3120
ATCTTCATTA TAATTCTTTT GGCTTTCAAA ATCTTGCGCT CCCATCACTT TTTACTTGTT   3180
TTTATTAGGG ATTCTTTTAC AATGAAGAAA ATGGGAGGTG TGCTTACAAT GAAAATAACG   3240
TATCGACTTA CTCAGAAACG TTGGGAAGCC ATTCAAAATA ATAATACGCA ATTTGATGGT   3300
GATTTTTTTT ATGGTGTGAC CACAACTAAA ATTTTTTGTC GTCCTTCGTG CCCTTCCCGT   3360
GTGCCCAATA AGAACCACGT TCTGATTTTC AAAACGGCTC AGGAAGCTTT AACACTTGGT   3420
TTTCGTCCAT GTAAAAGATG TCAACCAACC GGGCAAGTCG TTCCTAACGA GGAATGGGTT   3480
GCTCAAATTA AGCAATTTAT TGATCGGCAT TCAGCGGAAC CATTAAGTTT AGACTATTTA   3540
GCGCAAgcCT GCCACGGTAG TCCCTTTCAT TTGCAGCGAA CCTTTAAGGA ACAGACAAAA   3600
CAAACCCCGT CAGCTTATTT AACAAGTGTC CGAATTAGGC AAGCCCAAGA ATTGCTTCTC   3660
CATTCACCCC TTACGATTCA ACAAATCGCG AAAAAAGTTG GCTTCCAAAC CGCAGCATAT   3720
TTTGCCACCA CTTTTAAAAA AGAAACAGGC GTAACCCCCT CTGTTTATCG CAGTAAAAAT   3780
GAGCCGAAAA AATAACATTT CGGCTCATCG GATTATTTAT TGGGGCGTGA TTGTAGCCAT   3840
GCTACAATCA ATTGACTGAT TTCTGTTTGC TGTTGCTCAT TAGAAATTGT CGCTTTGCTA   3900
TCTCCTTTTT GGACACCGTA ATTACCAAAC CCTGCATGAT TCCCACCTTC GATAGATACA   3960
TACTGCGTAT CTTTTGGCAA AAACTGTTTG GCTTTTTGGT AGCTAGTCCA ATTTAAAACA   4020
CCATCGTTTG TGGCTGTTAT TGACAGAACC GGCAGATTTT TCTTGTTTAG TGCACCTTTT   4080
TCGTCAGGAT AGCTAGCTAA ATAAATCATG CCTTTGATTC GTTGCTCATT TTCCGCTACA   4140
AAACGGCTAG CCATCACACC ACCCAGAGAA TGTCCAGCCA AAACGAAGTT TTGTTGTGGA   4200
TGATCCTTTA TAATTTGTTC TGCAGCATTT GGTGCAAGAA CCGCTAAATT CAAAGGAAAA   4260
TGAACAACAT AGACATCATA ACCTTTCGTA GCAACTTCTT GCGCCCATTC ACTATAACTA   4320
GCGGGCGCAA CCAATGCACC TGGATAAAAA ATCAGTGCTG TATCATTTTT TTGCCCGTTT   4380
GTATAATAAT CATAGTTTTT TTCTTGCGTG GCACTTTGGC TTATTTTTAA GGCTGAATCG   4440
GTTGCTTGAT ACGTACTATT TTTAATAAAA ATAGTCAGAC CGATGACTAA TACCGCTAAA   4500
AAAAGCCCGA TTCCTAGGAC AATTTTTTTC CAACGTGATT TkGTtCTATT TTTCATCTCA   4560
TTTCCTCCCA ATGAAAAAAT TCTTCTTTCT GATTCTACCT TTAATAAAAA AAGrACACAA   4620
```

CAATaGGaGT yGGTGAAGCa ACCCGTCTGT TGCTCCACCA ACTCCTAACT AGACCGTTTC 4680

TCG 4683

(2) INFORMATION FOR SEQ ID NO: 194:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9235 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

CGCATTGAAA TTTGTTGCTC TCTTGATCGT GATnGGTnGC CTTAATAGTG ACGGCTTTTA 60

CTGTCGGGCT TGGTTTGACA TCAATGCTAG AAACACTAGG AACAACCTTT GTGAATAATC 120

GAGGCATGGC TATCTTTATT ATTATyATGT TAGCAACAGG AACGTTAGAA AGAAATGGCT 180

TGAAAGAGTC AGCAGCTACG TTGATTAAAC GATTCAAAAA GGTATCTGCT GGTGTGATTA 240

TTGACATTTA TGGTGTCTTT CGGATTATCT TCGCCGCTTT TAACGTAAGT TTCGGTGGGG 300

TAGCAGGTTT TGTGCGTCCG ATTATTTTGC CAATGGCGTT AGGAACAATT GAATCAAAAA 360

ATTTACCAAT GGTGCCAGAG TATGAAGAAG AACTAAAAGG GATGGCTTCA GCCATGGAAA 420

ATATCTGTTG GTTCTTTGGT CAAGTTTTGT TTGTTGGTGG GGCTGGGGCA TTGCTTGTTC 480

AATCAACATT GAAAGATTTA GGCTATGAAG TGACATTAGG TAAACTGGCT TTAGTGGAAG 540

TACCAGTCGC CTTAGTTGCC TTAATTTCGG CAAGTATTTA CTTTACTTTA AAAGAAAAAA 600

GATTACGTAA AAAATATTAT GGACAGGAGG GACTGAAATG AGTTTTTTCA CAGATCCAAG 660

TATTGATTTG AGTGCAAAGT TATTGGAAGT CCTTTACATT TTTATGGGGA TTATCTTGTT 720

ATATACGAGT TTTCGTAGTT TTCGTGATAA AACCAATCAA CATGCGTATG GGACTAGTTT 780

ATTCTGGGGA ATTTTAGGAA TCTTAATCGG TTTTGGTCGG TTTATTCCAC CAGTTGCCTG 840

TGGCGTACTT ATTTTTGTGA TGACAATTCC GGCAATTCTT CAAAAAGTGA CAAAAGGAGA 900

AAGTCGCTTA CCTTCGAAAG AATACATGGA AAAAATGTCT CAAAAATTAG GAATGAAGAT 960

TTTTATTCCC GCACTAAGTA TCGGGGTTTT TGCAATTCTT tTTGCCACAT TGACTACTTT 1020

AGGAGCTTTA GTCGGTGTCG GTGTTGGTGT CTTTGTAGCG ATTTTCATCA TGATGTTCTT 1080

TTCTAAAAAT AATACGCCGA AAGTATTTTT TGATGATGCG GCGGAAATGT TAGAAACAGT 1140

CGGTCCTCTT AGTACGTTGC CAATGCTGTT AGCTTGTTTA GGTGCTGTTT TTACTAGCGC 1200

TGGTGTTGGT GAGATTATTT CAAAAGGGGT TAGTGCCGTA GTACCAGCAG GAAATGTCAA 1260

TGTAGGGATT ATTATTTATG CTGTGGGAAT GGtTATTTTT ACGGCTATCA TGGGCAATGC 1320

CTTTGCAGCA ATAACCGTGA TGACAGTCGG AATTGGTGTG CCGTTTGTCT TTTCTTTAGG 1380

TGCAGATCCC GCATTAGTTG GGATGGTTGC CTTAACATGT GGTTTTTGTG GTACACTGTT 1440

AACACCGATG GCTGCGAACT TTAATATCGT ACCTGTAGCC ATGTTGGAAA TGAAAGATAA 1500

GTACGGCGTC ATTAAAAATC AATTGCCGAT TGCACTTTTC ATGTTAATTT TCCAAATTGT 1560

TTATATGATT CTATTCAAAT AAGAAAGGAT GATCGAAATG AAAGTTGTCG TTACTGGCTT 1620

TGATCCATTT GGAGGAGAAG CAATCAATCC TGCCTTTGAA GCAGTCAAAA AATTACCTGC 1680

TGAAATTGCA GGAGCTGAAA TTATTAAAGT TGAGGTGCCA ACTGTTTTTG GCACTAGTGG 1740

TGAAAAAGTG GCTGAAGCAA TTGAAACCCA TCAACCAGAT ATGGTGATTT GTGTTGGACA 1800

AGCAGGCGGC CGCGAAACAG TCACTGTTGA AAAAGTGGCG ATTAATTTAG CGGAAcGCGT 1860

ATCCCGGATA ATGCAGGACA ACAACCAAGC GATGTGCCGT TGGTTGAAGA TGGAGCAACA 1920

GCCTATTTTA CAAATTTACC GATTAAAGCG ATGGTAAAAA ATTGTCATGA CCATCAATTA 1980

CCTGCmGCGA TTTCTTATAC AGCAGGAACC TTTGTATGTA ATGATATTAT GTATCACTTA 2040

CTACACTTAA TTAACACTAA ATATCCAACT ATTCGTGGCG GATTTATTCA TGTTCCGTTC 2100

CTACCAGAAC AAACAATTGA TAAACCAACG TTTGCTTCTA TGTCATTAGA AGCAATCACA 2160

GATTCTTTGT TCTATATGAT TGAAGCAGCA GTAAAAACGC AAGAAGATAT TCAACTACAA 2220

GGTGGCACAA CCCATTaATT AACAGGaGCG TGGAACAAAA ATCACTTTGG ATTTTTGTTC 2280

CACGCTCAAA AACTGATAAA CGGCGGGAAC AGAAGCAACT CcTTCGGAAA TAAGCCGAAA 2340

TTCTCCAAAA ATTAAAGAAC AATTTTCGGA AATTCCTTCT TATTTCTCGG AGTTAAACGC 2400

TGCTGTCCCG ACCTCTTTCT ATTTTTACAA AAAGAAAAAG TAATTTTTGT TGGCGTGTGA 2460

AAAAAATTTT TTATACTATT TTCGTTGACA AGTCGGTTTC TTCTTGGTAT ATTGAGATTC 2520

ATAAATTATT AGCCCaGTAG GAAATAGATG TTTTGTTCAG AAAGTCGATG GTTGCTGCGA 2580

ATCGATCAAG TCTATTTGTG AATTACACTA ATAAGTAGTA ACAATTACAT AATACATAAG 2640

CGAGAAGCAA GTAAGCTTCT AATTAAGGTG GTACCGCGGA GAGATTCGTC CTTATTCTTT 2700

AAGGATGAAT CTCTCTTTTT ATGTAGCTTT AGGAGGCGTT TTCAATGAGT TGTGGAGGGG 2760

TTATTCGTTA TCATAAGATC CTTTAATGTC AGAGTATTCA GAAAAAAGAT TAGAGGAGAG 2820

TAACGAAATG AAAGAAAAAC AGAAAAAGAA GACCGTGAGT GTTCGAGAAG CAAGTATCGT 2880

ATTATTAATT ATTGTGGCAG CTATTGCAAC AGGAGTAATT GGCTTAAAGA TATCGCCAAA 2940

TATTACTATT TTGTTTGTGA TTGCATTGAT TTTAGGCTAC GCAATGTTTC GTAAAGTACC 3000

ATTTGATCAA ATGCATAAAG GGATTGTTGA TGGCTTGAAG CCAGGAATTA TTCCAATTTT 3060

TATTTTCATT TTAGTGGGGG CATTGATTGC CGTTTGGATT CAAGCTGGAA TTATTCCAAC 3120

GATTATGGTT TTTGGTTTTA AAATGATTAG TGTTAAATGG TTTGTGCCGT CAGTTTTTAT 3180

TGTCTGTGCG AtTGTCGGCA GTGCAGTAGG TAGTGCTTTT ACAGTGATGT CAACAATTGG 3240

AATTGCCTTT TTCGGTATTG GCAGTACCTT AGGGTTAAAT CCTGCTTTAG TTGTGGGTGC 3300

GATTGTTTCT GGTTCAGTTT TTGGTGATAA AATGTCACCT TTATCAGAAT CAACAAACTT 3360

AGCCGCAGCG ATTGTTGAAG CGGATTTGTT TAAACATATA AAAAACTTGA TGTGGTCGAC 3420

TGTACCAGCA TTTATCGTCT CTTTAATCTT GTTTATGATT TTAGGACAAA CGTACGCGAA 3480

TACAAGTCTG ACAGAAGTAG TTCAAGTTAT TCAAGTATTG GAAGACCACT TTACCATTTC 3540

AATTTGGTCA TTGTTACCAT TAGCCTTAAT GTTAATCTGT GCCTGGCGCA AAATTCCAGC 3600

GATTATCACG ATTTTGTTAA ACATTATTGT GGCAGTAATT ATGATTATTA TTCAAAATCC 3660

GAAAGTATCG TTGCAGGCAT TAGGAAACAC GTTGGAAAAT GGGTTTGTTT CTCAAACGGG 3720

GAACGCACAA ATTGATCAAT TATTAAGTCG TGGCGGGATT ATGAGTATGA TGCCGACTGT 3780

GGCACTTATT ATTCTAACGT TGTCTCTAGG TGGCTTATTA ATGGAATTAG GATTGATTTC 3840

TGCGGTCATG GAAGTTGTTT CGCAGAAAAT GAGCAGTACA CCGAAATTGA TTTTATCAAC 3900

GTTATTAACA GGAATTGGCG TAAATATTTT TATTGGTGAA CAATTCCTAT CGGTTATTTT 3960

ACCAGGGAAT GCCTTTAAAG AAGTTTACAA AAAAGAAGGC TTGGATCCGA CTGTTTTAGG 4020

CAGAACCTTG GAGGATGGTG GGACGGTTAT TAACTACCTC ATTCCTTGGG GGATTGCAGG 4080

TAGCTTCGTA GCGGGGACCT TTGGCGTACC AACGTTGACT TACTTACCAT TTGTTTTCTT 4140

TAGCTTATTA TCGCCTGTCT TCTCAATGGT CAGTGCATTT ACTGGTTTAG GAATTAAGCG 4200

GCTATCAGCA GAACCTGTTA CCGAAAAATA GATTGATTAA AATAGAACAG TTGATGGGGA 4260

ATACTGTAGT TTCTCCCTGT CAATTGTTCT ATTTTTCTTG TAAAAAAGGT TGCATTTTTA 4320

GAAAGAAACA GTTATAGTTG ACTAGTCAAT TGTTGACGAG TCATCTTTTT AAGGAGGAAA 4380

CTATGTCGTT ATCATTACGG GAAGTCAGCC AATTATTATG TCAGTTAAAG GTATTAGATC 4440

AAAAAATTAC CAAAGTTTTT GAAGAGCAAG TTGGCTTGAG TTTGACTCGT TATGAGTTGC 4500

TGATGATTTT AAAAGAACGT CAGCCGTGTC TTCAAACGGA GATTCAAGAA CACTTAAAAA 4560

TCGATAGTGG TGCGGTCACT CGCCATTTGA AAATTTTAGA AGAGAAACAG TATGTGACAC 4620

GACAGCGAAA TCCAGAAAAT AATCGGGAAG TGCTAGTGCA TTTAACAGAA AAAGCACAAC 4680

AAGAACTACA GCAATGTACA GCCAAGCAAC AAACGGTAAC AGAAATTATT CCCAGTACAT 4740

TTACAAAGGA CGACTGTCGC CAGTTAAAAG AACTTTTGAC AAAACTAGAT CAAAGCATTA 4800

CAACGAAAGA GGTATGATAG AATGACAACA TATACAACGA ATGATTTTTC AGAGATTGTT 4860

TTTGGCCGCA AATCGGTTCG GGTGTACGAT GAAACCCATA AAATTTCTCA TGAAGAAATG 4920

TTAACAATGA TTCAAGAAGC GACCACCGCC CCTTCTTCTG TTAACATGCA GCCATGGCGT 4980

TTTGTCGTAG TGGAAAGTGA AGCAGGCAAA GAAAAATTAA AACCCTTAAT TCGCTTTAAT 5040

ACGCGTCAAA ATGAAACATC TTCTGCAATG CTCTTAATTT TTGGTGATTT AAACTGCCAT 5100

GAGCGAGGAG AGGAGATTTA CAATCAAGCC TATGCTAGTG GTAAAATGCC TAAGGAAGTG 5160

CGGGACCAAC AATTAGCAGC CATTATTCCC CATTATGAAA GTTTAAGTAG AGAACAAATG 5220

AATGACATCG TAAAAATTGA TGCGAGTTTA GCAGCTATGC AATTTATGTT AGTGGCCCGA 5280

GCTCACGGAT ATGAAACAAA TCCAATCGGT GGTTTTGAAG CGGAAAAATT AGCCGAAACA 5340

TTTGGTTTAG ATCAAGAGCG CTATGTCCCA GTAATGATTT TATCTGTGGG TAAGGGCATG 5400

GAAACTGGTT ATGAATCTGT TCGTTTAGCA CCAGAAAAAA TCACGACATT TGAATAGGCA 5460

```
ATAAAAAATG ACGGCGTACG CTtGCATTAA CCAACTTTCT TAAAGAAAGT TGGTTAATGC     5520

mACAGCGATG CCGTCATTGT CGTTGGATAA AGTGACTTCC GTTGCGATGC TTTTAAGCTC     5580

GTCAACGGCG TTTTCCATAG CAATCCCCGT GCCGGCCCAT TCCAACATAG TATAGTCGTT     5640

GTGACCATCA CCAAAAGCAA TTACTTCTTC AGCAGCTATA CCTAACTGTA GGGTTAGTTT     5700

TTCTAAAGAA CGCGCTTTAT CAATATTTTT AGCGGTAAAT TCAAAATAGA AAGGAGCGGA     5760

AAAAGCAGCG GTGACGGTTT CTTTGAAAGG CGCGTAGATG GCCTCATGAT ATTTTTGTAA     5820

ATAGGCGGGT TCACCAGCAA TTAAAATTTT GTTCAAGGGA AAGTTTAGTC TGGCGGCTAA     5880

ATCACGCCAC TCACATAATT GAAAGTTGCC ACCCCGTGAT TCATATTCAA TAATATTAAA     5940

ATCACCAGTC TCTAAATGTA AGGTGTTATG AAAAACATCA TTGACATACA TAAATGTTTC     6000

ATCATTAATC ATTGGAATTA CGTCAAATTG TTTCAAATGC TCTAAAATTT GTTGACTCAA     6060

TGACTTTGAA ATTGTCTGAT TAAATAGTTG TTGCTGACTG CCGTGATGGG TGACACAAGC     6120

GCCGTTATAG GAAAGTAAGT GTCCATTATA ATGAGCCATC TCCAATTCAT CTGCTAAGGG     6180

CCGCATACCG TTAGTCGGGC GGCCAGAAGC CAAAATCAGT GATAGGCCTT GTTTTTGTGC     6240

AGCTACTAAT GCTTGTCGCG TTTTTGGTGA AATCTTTTTT TCTGAAGTTA ACAAGGTGCC     6300

ATCAATATCC ATTACAATTG CTTTGATTGC CATAATTGGC CTCCAATCCG AAGAGTTTCT     6360

TTCTTTTAGT ATAAAAGTTT TTTCTATTAA TGTAAACGGA TTCTCTTGTG AGGTGTATAA     6420

AAGAATTATT TTACAAAAAT CTATACTATT TTTTTGTGAA ATTTCATATA ATAGAAGTAT     6480

TCTAATTATC TACATAGAGA GCGAGGGACA AGGAATATGA AGGAAAAAGA AATGCATTCG     6540

CTCTTTTTTA AACATAAGTT TGTGAAAGTA ACTCCCTATT TACGTCGTTT TGGTCATCGT     6600

TTGAGTGGGA TGATTATGCC AAATTTGAGT ATTTTTATTG CGTGGAGCTT ATTGTCTTTG     6660

GTGGCTGGCT ATACGACTGG GAATCTACGG CTAGCTCTTT CTGAAGTCGA AACGATAATG     6720

ATTCGAGTTG TTTTACCGAT TCTAATTGGT TTTACAGGCG GAAAAATGTT CGAGGAACAA     6780

CGTGGCGGCG TTGTTGCTGC TATTGCGACA GTGGGCGTGA TTGTTTCCAC AGATGTTCCA     6840

CAGTTGTTTG GTGCTATGTT TATTGGCCCT TTAGCAGGAT ATACTTTCGC CAAAATTGAA     6900

CAAATTCTCT TACCGAAAGT TAAAGAAGGC TACGAGATGC TGACTAAAAA CTTTTTAGCA     6960

GGAATTGTGG GAGGACTGCT GTGCTGTTTT GGTATTCTGG TTGTAGCTCC GGCTGTTGAA     7020

AGCGCTAGTT TTTGGCTGTA TCAATTTTCT TCTTGGTTAA TTGAAGCCAA TCTTTTACCA     7080

TTGGTTCACG TTTTCTTAGA GCCCTTAAAA GTGTTATTTT TTAATAATGC GATTAACCAT     7140

GGCTTATTAA CGCCTCTAGG TTTAGAAGGT GCTAGTCAAA CAGGTCAGTC CATTTTATTT     7200

CTATTGGAAA CAAACCCTGG ACCAGGCGTG GGCGTTTTGG TTGCTTTTCT GCTGTTTGGG     7260

CCTGTAGGAC AACGAAAAAC AGCAGGAGGT GCCACCATGA TTCAACTGAT TGGGGGCATT     7320

CATGAAATTT ATTTTCCGTT TGTTTTGATG GACCCGCGCT TATTTTTAGC AGTAATTGCT     7380

GGAGGAATGA GTGGTACGCT TGTTTTTCAA ATATTTAATG TGGGTCTAAG TGCTCCAGct     7440
```

TCGCCAGGTT CATTGGTTGC GATTTTAGCC AATGCCCCGA CTGATGCGAG GCTGGCGGTT 7500

TTTAGCGGAA TTTTTGTTAG CTTTCTGTGC TCTTTTGCAA TAGCAAGCTT GTTATTAAAA 7560

CGTCAACGAG GAATTGAACC AGTTTCAATG ATAAAGATGA AGGAGGAAGA CCAAGTGGAA 7620

ACAGTCACAC CTAACTATCA GCAAATTTTA TTTGTTTGTG ATGCAGGAAT GGGCTCAAGT 7680

GCCATGGGGG CTAGTTTGCT AAGCCGACAA TTAAAAGCTG TGAACTTGGA GATGCCTGTG 7740

ACTTACCAGT CCGTTCATCA GATGAAGTGG CAGCCTAAGA CATTAGTGGT CATTCAAGCA 7800

GAATTGAAAC AGTTAGCACA AAAGTACGTC CCAGAAAAGG ATATGGTGAG TGTTCAAAAT 7860

TTTTTAGAAA TTAAATCCTA TTACCCGCAA GTTTTAGCCA AACTGACTGC TTCTTCTCAA 7920

GAGCAATCTT CACTTGGTTC AGAGTCTACT GAAACGAACT CGACAAAACA AATACAGAAG 7980

CTTGTTTTTT TATATGCCGA GAATGTTCGA GGATCGCAAA CAATGGGAAT GGAATTATTG 8040

CGGCAACAAG CGGCGAAACA AGGAGTCGCG ATTGAAGTAT CTAAAGAGCC ACTGGAAACA 8100

GTCTTTTTTA CCAAGGAGAC AACCTACGTA GTGACTCGTG AACTGGCGCA AGCCTATCAT 8160

TTAGATCTAA CGCAACAAAA TTTATACGTA GTTACTAGTT TTTTGAATAA GAAAGAGTAT 8220

CAAGAATGGC TGGAAGGAGG AGCTGATAGA TGTTTTTAAC GACAAGAGAA ACTGTTCTAT 8280

TAACGGAATT AGTGAATAGT CCTACCCCCG TTTCAGTAAA CCGCATGATG AATTTGTTAA 8340

AGGTTAGCCG CCGGACGGTA TATCGTGAAT TAGAAAATTT AGAAACCTCG CTAGCATCGA 8400

TGGGCGCGAC ATTGGAAAAA GTAGCTCGTG GGCGCTTTTC GATTCAAGCA GATGAAGCAG 8460

CGATGACGGA AATTCAGGCT GCTATTTTAG GAGAAGAGAC ACAGGAACTT TCTACTTTGG 8520

CGCGGCAACA TGCGATTTTA TTGACATTAC TGCAAACAAA AGAGCCAGTC TCTATGCATT 8580

ATTTTCTGGA AACGTATTGT ATTAGTAATA CGACTTTTTA TGCCGATATT AAACAATTAG 8640

AAACACGGAT TGCTCGGATT CCGTTGACGA TTGCTCGGAA TCAAGGTTAT GAAGTGACTG 8700

GTTCTGAAAA ATATCGCCGT CTTCTAATGG CGAATATTCT AAGTATGGAA ATTAATGAGT 8760

ACCAATTTTT TCATTTTACC GAATTAACGA CAAACGACCA TTTTTTCTTT CAGTTTATTC 8820

ATGCAGAACA TTTAGCATTT GCACAAAAAA TTGTTCAGCC AGAAGTAGAA ACTTTATTTC 8880

CAGCATTAAG TGATCGCAAA TTACAACATT TGATTTTAAT GTTAACAATT GCGATGGATC 8940

GGGTCACTGC AGGTTTTTAT TTAGCTGATG AGACCTATAC TGGGCAGATG AATAAGGCCT 9000

TTCTAAATTG TTCTAAACGA CTTTTTTCAA AAGTCGCTGC TGAAACGAAA CAACTATACG 9060

CTGTCAGTGA GATTGTTTTT TTTGCTAGCT TGTTAAGTGA TTTTTCTAAT TCnTTTGATG 9120

AAGATTTTTT tGATGAACAT TTTGACACTC AGCTGGcTTA TGAAGTGAAA CAACTGATTG 9180

AAGCTGTAAG TCAAGAAACG GAAGTTnATT TTTTCGAAGA TACCAATTTG TCTAA 9235

(2) INFORMATION FOR SEQ ID NO: 195:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16798 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

CCAGCTCAAA ACGCATTTAA TCACAAAAGA GGTAAACCGG CTGAAAAAAG AGAAGAACGC 60

CcTGCACTAT ATCGGCATTG CCGCAGACGA AGCGAAACGG TGCAAGGATG CCCCAAACCA 120

CCGCTATCCC CTGGTGGAAT GGGGCGTCAC CGAGGCGCAG GCTtTACAGA TCTGCTATGA 180

CCGGGGCTTC GATTTTGGCG GGCTGTATAA AATCTACCAC CGCGCTTCCT GCTGGTGCTG 240

CCCCTTACAG CGGATTGGTG AATTACGCAA GCTGCGTCAG CACCACCCTG AACTGTGGGC 300

GCGGCTGTGG GATATGGATA ACCGGGCGCG GGCGCAGTTT GGCGCAGGCC CCCTTGGACA 360

GTTTAAGGAG CGTTGGAGCA TTGAACAGTT AGACGCACGG TTTTCCCGCG AAGAAAAGGA 420

GGAACGCGCA TGAAACACCT GctCCGCCGC CTGCTGGTTC CGCCTTAGTT GCTCCCaGGC 480

AGCCAATGAA AAAAACGCAA AACAGAAAGA GGTGAGAACA ATCGAATTTC AAGCGAAACA 540

ACTGCTGATG GAAAAGCTGG ATAAAAACCT GCTGGATCAT GCAAAAAGCC TGACAGACGG 600

GAACGGGCCG CAGATTGTGG ATGTGTATGG GCTGCAACAA CTGGCCGAGG TTCACTGTTA 660

CTTGAAAACA ACCCACCAGT TTGACCCTGC TGAAGTGGAA GCACTGCTGC AATTTGCTGA 720

CCCGCTGGAA GCTGCCTGCC ATTGTTGGGA AATGAACACC CATGAATACA GTTTCCCGAT 780

CTGCGACCTG CTGCGGGAAA CCAGGGCGCG GGAACTCTTT GAACTGGCAA AGCCGGAACC 840

CCCGGCAAAA TTGTCGGTGC GGGAGCAGct TCGCGCCGCT GTGAAAGAAG CCCGCCAGAA 900

TCTGGCACCA GAGGAACGGA CAAAAGGCGG CGAAGCCCGG TGAAACAACC TGCCGCTTGC 960

AGGAAAGGAG GTGAACCCCT ATGAACTATG ACGCAAAGCA GCCCCCGGCA AACCTGGCCG 1020

ATTCCATCCA GCAGAAAGAA ACCGTCCAGG CATTATACGA CCTGTTGGAT AAAAATATGG 1080

CCGACTTCGA TGCAAGTCTG GCTGGCATGG ATAAAGCTGA AACCGCTGAC CTGAAAGAAG 1140

TTTCTGATAC GCGGGAAATA TACGACTTCA TTCGATACCA TTATGGCGAA TTTGGACAGC 1200

GTGAAGCAGA ACTTTTGCTC CGCATGGAAA ATCCCCTGAA TTTCATAGTG GAAAACTGGC 1260

CCTGCCATAA TCTTGAAACG CTGGCTATGA ATTATCTGGC GGATAAAACG ATTAAGGAAC 1320

CGAACAAAAA GGAAATGCCG CAGTCCGGCC ATGACACGAA AGATGCTGAA AAACCCTCTA 1380

TTTTGAAACA GATTAAAACC AACCAGGGAC AGATCAAGGC AAACAGACAG GAAACCCGCC 1440

AGCAATCTAT TACGGCAGAG CGACCCAGGG ACGGTGAGGC CCGCTGACGC TTTACGAGAA 1500

AGGAGGTAAC GCACCATGCA GGACGAAGTA CGCGAAAAAT CAGTCGCGCT CACCATCAAG 1560

GTCGGCAGGA CTGGCGGCAG GCTGACCGCT GAACTGCTGA AATGGGCCAT CCGGAAATAC 1620

CTTTCCCAAT CCCAAAACCC GAAAATCCAC CACGGCAAAC AGACGGTGAA ACAGCTTGTG 1680

GGCCAGGGCG CGGGTGTGCA GAATATTGAG ATTACCGGAA AAAATATCAA ATCCTTTGAG 1740

CGCGTGGCCC GGAAATACGG CGTGGATTTC GCTCTGAAAA AAGACCCAGC ACAGGGAAAA 1800

TACCTTGTCT TTTTCAAAGC ACGGGACGCG GATGCACTCA ATGCCGCCTT TTACGAGTAT 1860

GCCGGAAAGA GCCTGAACCg ctCCGCGCAA AAGAAGCCGT CGCTGTTAGG CCAGCTCTCG 1920

CATTTCAAGG AGATCGCAAA GAATCTTACA CGGGACACCG CTAAAAACCG GGAGAAAGGG 1980

GGCGTTGAAC TTTGAAATTT GATGTAAAGA AGATTCTGCT TCCAAACCTG CCTTATCTGT 2040

TCATCTTCTG GTTTTTCAAC CGTGTGGCGG AGGGCTACCG GCTGGCGGAG GGCGCGGATA 2100

TGGTGACAAA GGCAATGGGC GCAGTATCCG GCCTGGGGGC GCTGATTTCC AAAAACCCGC 2160

TGCCCAGCTT CTACCCGCGC GACCTGCTGC TCGGCGCTGT CGTCGCGGTC ATTATCCGGG 2220

CCGTAGTTTA TTTCAAGGCG AAAAACGCCA AGAAATACCG CCACGGAGTT GAGTATGGTT 2280

CGGCTCGTTG GGGAAATTCC GACGATATTA AGCCTTTCAT CAATCCCCAA TTCGACCAGA 2340

ACATTCTCCT GACGCAGACA GAGCGTGTCA TGGTGGGGAG AAACAAAATC CCCAAGTACA 2400

ATATCAACAA GAATGTGCTG GTAATCGGCG GTTCCGGCAG CGGCAAGACC CGCTTCCACA 2460

TCAAGCCGAA CCTCATGCAA ATGAACGCCA GCTATATTAT TACTGATCCG AAAGGAACAG 2520

TCGTTCTCGA ATGTGGGAAG ATGCTCCAGC GCGGCGGATA TGAGATAAAA ATACTGAATA 2580

CGATTGATTT CAAGCAGTCT ATGAAGTATA ACCCGTTCCG CTATATCTAC TGTGAGAACG 2640

ACATCCTGAA GCTGGTCAAC TGCATCATGG AAAATACCAA AGGCGAGGAC AGCAAGGGCG 2700

GCGAAGATTT CTGGGCGAAA GCCGAAGCCC TCTACTATCA GGCGCTGATC GCCTATATCT 2760

GGTACGAAGC GCCGGAGGAA GAAAAGAACA TGACCACCCT GCTGGAAATG CTCAATGCCT 2820

CCGAAGTGCG GGAGGATGAC GAAAATTTCA AAAATGCCGT TGACCTGATG TTTGACGCGC 2880

TGGAAAAACG CGATCCACAA CATTTTGCGG TGCGCCAATA TAAGAAATAC AAAATGGCGG 2940

CTGGCAAGAC GGCAAAATCC ATCCTGATTA GCTGCGGCGC TCGTATGGCT CCCTTTGACA 3000

TCCGGGAGGT GCGGGAACTG ATGGAGGGCG ACGAACTGGA ACTTGAAAAG ATTGGCGACC 3060

GCAAGACGGC GCTGTTCTGT ATCGTATCTG ACACGGATAT GACATTCAAT TTCATTTCTG 3120

CGATGGTTTA TACGCAGATG TTCAACGTCC TGTGCGACAA GGCGCTGGAA AATGGAGGCG 3180

CACTGAAAAC CCATGTTACC TGCCTGCTGG ACGAGTTTGC AAACCAGAAA ATCCCCAACT 3240

TCCAGCACCT GATTTCCGTT ATCCGAAGCC GGGAAATTTC CGCGCATATC GTGGTGCAGA 3300

CGCAGAGCCA GTTAAAGGCT GTCTACAAAG ACCATGCGGA AACTATCATC GGCAACTGCT 3360

CCTGCGTCCT GTTTTTGGGC GGCAAGGAGC GCAgCACCCT GAAAGAGATT TCGGAAACGC 3420

TGGGAAAGGA AACCATCGAC CTTTTCAATA CCTCCGACAC GCGGGGAAGC CAGCGCAgcA 3480

TGGGGGTAAA TTACCAGAAA CTCGGAAAAG AACTGATGTC AACCGATGAA TTGGCGGTGA 3540

TGGACGGCGG CAAGTGCATC TTGCAGGTGC AGGGCGTAAG GCCGTTTTTC AGCGACAAAT 3600

TTGACATTAC CAAACATCCG CAATATAAAT ACCTGCTGGA TTCCAGCAAG AAGAACGCTT 3660

TTGATATTGC AAAATATCTG AGCCGGAGAT TGACCGTCCG ACCGGACGAC GTTTTTGAAA 3720

GCTGTGAACT CGAAGCGCCG CCTGCGTCTG ATGCACCCGC CGATTGACTT TTCAACCGGC 3780

GGTTTTTTTG TTCCTATCCA ACCAATCACA ATAAATTTAT GGAGGTAATT TATGGAATTT 3840

```
TTTGCAAGCG CTATTGAAAC CCTGAAAGTC CTTGTAATCG CGCTGGGCGC TGGCCTTGGC   3900
GTGTGGGGAG TGATCAATCT GCTGGAGGGA TACGGCAATG ATAACCGTGC GACACGTTCG   3960
TAAGTGAAAA GcTTGCGCAC AAAGTCGAAG GGCAATTGTA GCCCGAAACT TTGGAGAACT   4020
GATATTCCGA TGGGTGAAAT GAGGGGGTAA CGTCCCGAAG CGCCACCCTA ATACTCCGAA   4080
TGGCGGGGAG CATGCAACGG CTCTCTCCAA AAGTTTATAG AGCTCGGTGA AGTCGGCAGA   4140
GAGCAACCGT AAGACTGGTT TTCAGACCAG ACACGAAAGC TGTCCAGAGG TGGATGGTGT   4200
TGCCTATATG CCGGGGGTCG GATATCCATG GTGAGAATGT ATGAAAATAC TGACGAAATT   4260
CTGAATGTAC GAGTCTAAAC GTTGATGCTG TCGAAAGGCA GTACACCTTT CATGGTGGTA   4320
TCTGTGGGCG AGTAAAATTG GTTGTTATGA AAGTCCATGC ATCGTTACAG GCGGTGCGGT   4380
TATGCATGAC CAGATACAGG CTTAGAGGAA GAACCTAAAG ATATCCAATG ATAGGAATAT   4440
CGGAACGTGG AAAGCTGGGA ACACGGAGGC GGTTGCAAGC CAGTGATGTG GTGGATGGGA   4500
ATACCTTCGC GAGAGGGCAA TAACCATTCT TTATTCCAGT GAGAGTAGTG GCACAGTACC   4560
TGTGAAGCCG TGAATAAGTA AGCGGTGGAG GGATAGCCAC AAGTCGAATT CAGCAAAGAA   4620
AATGAAAACA CAACAAACAC AGCTTCGAGT ATGACAAAGA AAATCCAAAC CGAAGGgAGA   4680
GGATGCCTGT GTTGACTTCG GGGCAGCAAG AAAGGAAAAG CAAACAGAGA AAAATCCGCA   4740
ACTCAGAGTA TTACGACATG GAAGGTACTT TTGACAGGCT GTATGCAGAA AGTAAGAAAG   4800
ATAAAACCTT CAACCATCTG ATGGAAATCA TCGAGAGTGA AGAAAATATC AAGCTTGCAT   4860
ACAGAACGAT AAAGAAAAAT ACAGGAAGTG ATACTTCGGG AGTGGACAAA AGGACGATAG   4920
CCGACCTTGC AAAGTTAAGC GAGGAAGAAT ACGTCCGTCT CATACGGAAA CAATTCAGTA   4980
ATTATCATCC AGGACCCGTA AGGCGAGTGG AGATACCAAA GCCCAATGGA AAGACCAGAC   5040
CTCTGGGAAT TCCAACTATC GTGGATCGGA TCGTACAGCA ATGTATTTTG CAGGTCATGG   5100
AGCCGATATG TGAAGCAAAA TTCAGTGAGA ACTCCAACGG GTTCCGCCCG AACCGTTCCG   5160
CTGAGACTGC TATTGCACAA TGCATGAGGC TGATACAGGT ACAGCATCTG TACCATGTGG   5220
TTGACCTCGA TATTAAAGGC TTCTTTGACA ATATCAGCCA TACAAAGCTG ATTCGTCAGA   5280
TATGGGCATT GGGAATCCGG GACAAAAAGC TGTTATGCAT TATAAAAGAG ATGCTGAAAG   5340
CACCTGTTGT TCTGCCAAAT GGAGAGAAAA CATATCCCGC ACGGGGTACC CCGCAAGGCG   5400
GTATTTTGTC GCCATTGCTT GCAAACATCG TATTGAATGA ACTGGATTGG TGGATTGCCT   5460
CACAATGGGA GGAGATGCCA ACAAAAACAA AATTCAAAAC AAGGAGCAAT GCGCAGGGAA   5520
CAGAGATTAA GAGCCACGCC TACAGGGCTC TGCGCCGGAG CAGATTAAAA GAGATGCATG   5580
CAGTTCGGTA CGCAGATGAT TTTAAAATTT TCTGCGCAAC ACATGAGGAT GCTGTCAGAG   5640
CATATAAGGC AACAGAGTTA TGGCTAAAGG ACAGGCTGGG GCTGGAAATC AGCCCTGACA   5700
AATCCAAAGT AGTTAATCTC AAAAGACAGT ATTCTGATTT TCTCGGGTTC AAGCTGAAAG   5760
TCCGTAAAAA GGGTAAGAAA TATGTGGTTC GGTCTCATAT GAGCGATAAA GCATACAAAA   5820
```

```
AGGCGCATGA AAAAGTCTCG GAAGAAGTAA AGAAACTGGC TTATTCGTCA GATGATAATG    5880
CTCAGTTCAT GCAGCTTCAG AAATATAACT CTGTAGTTGC CGGACTTCAT GAATACTATT    5940
GTATTTCTAC AGAAGTATCT CATGACTTCA GCAGACTTGC CTTCAGTATC AACAAACAGC    6000
TCAGAAACAG ACTGAAAGGC GACTTATCCA AAAAGGGACA GCTCAGAAAT GGGTTCATTA    6060
AAGAAAAGTA TGGAGCAAGC AGACAGATGA GGTTCCTTCA TGGACGTCCg GTGGTTCCAC    6120
TGGGATATGT CCAGTCAAAG AATGCCCAGC ATAAAAGGAA ATCCATCAAT AAATATACGG    6180
TCAAAGGTCG GGAACAGATT CATAAGAATC TTGCAATCGA CACAGCTACA ATGTTATGGC    6240
TGATGAGAAA TCCTGTAAAG GGCAGAACCA TTGAATATGC GGATAACCGT ATTTCTTTGT    6300
ATGCGGCACA GTATGGTAAA TGTGCAGTGA CAGGAATTCC GATGAATTCC CACGATATTC    6360
ACTGCCACCA TAAAGTACCG GTAAGCAATG GCGGTTCGGA TGAATACGCA AATCTTATTC    6420
TTGTCAGCAA AGCGGTTCAT ATCCTTATCC ATGCTTCTTC AGAACCTACG ATTGAGAAGT    6480
ACCTGAAATC CTTAAATCTT GACAATAAAC AAATAGAAAA GCTTAACAAA CTTCGTTCTA    6540
TGGCAGAAAT GCCACCTATT ATTCTTTGAA TTTAATGCTG TAACGAAGTG TGTAAGTTGT    6600
GTTTAAAAGC CCTGGACGGC TGAATTCGAT GGAGCGCCGT GTGCGGTGAA AGTCGCATGC    6660
ACGGTGTGGA GCGGGGGAAA ATCCGGAGAT AACTTCAAAG GATTACCTAT CGCTATCTGG    6720
CGCGAAATCG CAGGGAATGA AACAACTTAT GGCCGGCGGC GGAGTTGCCC TGATTGGGAT    6780
GCAGCTTATC CCTCTGCTGG CAAACCTGTT CGGTTAAGGC GGGATAGCCT ATGCTGAACA    6840
TCCTCGACAT AATCACAGAC TGGATCAAGG AAATCCTGCG GGACTGCATC ATGGGCAACC    6900
TGGACGGGAT GTTTGACCAG ATTAACAATG AAGTCGGGGA GGTTGCGGCG AATGTGGGGA    6960
CGACCCCGGC GGCATGGAAC GCCGGGGTCT TTTCCATGAT CCGCAATCTG AGCGATACAG    7020
TCGTAGTGCC GGTTGCGGGG ATAATCTTAA CCTTTGTCCT GTGCTACGAA CTGATTTCCA    7080
TGCTGATGGA AAAGAACAAC TTCCACGACT TTGAAACTTT TGCACTGTAT AAATGGATAT    7140
TGAAGGTGTT TATCGCAGTC TATCTGGTCA CACACACCTT TGACATCACA ATGGCGATTT    7200
TTGAACTGGC GCAGAACGTA GTACAGCAGA GCGCCGGTGT GATTACCGGC TCCACCACCA    7260
TAGATTTTGC GGCGGCGCTG GGCGATGTGG CCGCCCAACT GGAAACAATG GAAATCGGGG    7320
AACTTTTCGG CCTGCTGGTG GAAACCATGC TGCTGAAAAT CACCATGCCG GTGCTGTCGT    7380
TCTGCGTCAT GCTGGTTCTG GTGGGGCGCA TGATTGAAAT TTATATCTAC TGCTCGGTGG    7440
GCGCAATCCC CTTTGCGACG ATGGCAAACC GGGATTGGGG GCAGATGGGA AATAACTATC    7500
TGCGTGGTCT TGTGGCTCTT GGGCTGCAAG GCTTTTTCAT TATGATCTGT GTTTCGATTT    7560
ACGCTGTGCT GGTGGGGCAG ATTGGGAGCG CCGACAACAT CCACATAGCC ATCTGGCAGT    7620
GTGCGGGCTA TACCGTTCTG CTGTGCTTCT CCCTGTTCAA AACCGGCGCG GTCAGTAAAT    7680
CAATTTTTAA CGCCCATTAA GGAGGTGGAA CCCATTTGAA ACAATATGAA ATCATTCTGG    7740
CCGACCCTCC GTGGAGATAC TCACAAAAGG GGGGTGCAGG GAGCGGCAGA ACGTCACTAC    7800
```

GCCACCATGA GCATGGAGGA TTTATGCTCA CTGCCTATCA GCCGTTTTGc tGCGCCGGAC 7860

TGCGCGCTTT TTCTTTGGAG TACATTCCCA CAGCTTCCAG AAGCACTGCA TCTGATTCGT 7920

GCATGGGGAT TCACCTATAA AACAGTGGCT TTTGTCTGGT TAAAGCAAAA CCGCAAATCC 7980

CTTAGCTGGT TCTATGGGCT TGGATTCTGG ACGCGCGGCA ATGCGGAAGT GTGCCTGCTG 8040

GCTACCAAAG GGCATCCCAA ACGCCGCCGG GCCGACATCC ATCAATTTAT TATCTCACCG 8100

ATTGAAGCCC ATAGTAAAAA GCCCGATGAA GCACGGGAAA AAATCGTTGC TCTGATGGGC 8160

GACCTGCCCC GCGTCGAACT GTTCGCCCGC CAGAGAGTCC CCGGCTGGGA CGCATGGGGC 8220

AACGAGGTAG AACCCAGCCC CGGACTGGCG GAGGGGCTGG CCCCTGAAAA TCCAACGTAA 8280

AGGAGGGAGC GAATTTATAT GGACACTGAA AAACTGCGAC GGCAGCTTAT TGACCGCATC 8340

ACGGACAATT ACAGCGAATA CCGCGCCAGC CTGCTGTTAA AAGACCGGCA GGAATTGATT 8400

GACGGCGCAG GCAGGATTGC CGATACCGCC ACCGTATTCC AGCACCTGGC TGACCGAAGC 8460

TACAAACCGG AAGAACTGGA ATACCTTTTG AAATTCCAAA ACCCCCTGGA GGTGGTCACG 8520

GAGCATTGGG CAGATTATGA ACTGATGCCG GATGTGCTGG ATGCGCTGAT TGCAGACACG 8580

GTTGACCGGC AGGACGATCT GACCCTGTTC CCGCTGGCAA CGGACGCACC GGCACAGAAG 8640

CCGGAAAGCC TGCGGAAGTT TCTGAATGTA GACCTGGAAA CCGTCCTGCC CCAGATCATG 8700

AAGCAGAAAA CCGCCTTTTA CCAGACAGAC TTAAACTATG CGCTGGATGC CATGAGGGAG 8760

GGCGCAACGT CCGGCGACCC GGCAAAACAG AACTTTGTCC TGATTTTCCG GCAGAGCGGC 8820

GTGGAGTGCT TAAATGAGCG GGATATGTTC ATCGGCGGCA CACGCAGCTT TAACACACTG 8880

CAATTCTACC ACACACAGAC AAGGGAGCCG GTTCTGACCT ATTCGGTGGA ACTGACGGGG 8940

GACGGCAAAG ACGGCcTGCG GGGAAATCTG TACCAGCAGG ACAACCACCG CCTTGCGGAA 9000

tACGCAGGcG TACCGCyTCC CCCTATACAG ATGTAACGGT TACATTTACA GGAGGGAAAG 9060

AAGTACGGCT GCCGGAAAAG GAATTTTACT ATGAAAAAGC CGCAAGCCTG AAATATCTGC 9120

ACGGCAGCAT TAAGGAGGTT CGCAATGAAG CGGAGGACGA AACCGTGGTG CAGGGGGCGG 9180

TCAAACGGGA ACACGACCGC CGGCGGGAAA TGCCCAGGGG CTATTTTGCC GTCCACCTGC 9240

AAGCGCTGGA TGAGCAGCGG ATTCAGGCCG AAGCAGATCG GCTATCGTCC GCGCTCACGG 9300

ACTTAAAACA GCCCAACTCG CCGGACAAGG AGAAGTTTAT GGTGGAAATA TCCGTGTATT 9360

TCCTGCCGCT GGCTTCCCAG CACGATATGA ACCGGCTATT TGAGAAAGTG CGCGACCAGG 9420

TAGGGCAGAC GGTGTATATC ACTCCGCCGG ACGATACCCA CAAGCGGTAT TTTGTCATGC 9480

GGCGGGArGA ACGCGCACAG GAGCAGCCAG AACGCAAGCC CTCTATCAAG GCCCAGCTTT 9540

CTGCAAAGCC CGTCCCCGGC GATAAGCCCG CCACAAAAAC CAAAGACCGG GAGGTGAGAT 9600

AAATGCCCTT TGTACCCGTA CCCAAAGACC TTGCCAAAGT AAAAACAAAA GTGGCGTTCA 9660

ACCTGaCAAA GCGCCAGATC GTCTGCTTTG GCGGGGGCGG GCTGGTGGGA GTCCCCGCCT 9720

ACATCTTTAC ACGCAGCAGC ATCGGCAACG AGCCGGCGGC TCTCCTGATG ATCGGCCTGA 9780

TGCTGCCCTT TTTCTTATTT GGCATCTATG AAAAAGACGG TCAGCCGCTG GAAGCAGTCC 9840

TGCGGCATAT GATACGCGCA AAATTCCTTT GTGCAGGGAC GCGCCCGTAC CAGACCGACA 9900

ATCTGTACGC CGCCCTTTCG CGGCAGACTG AGGAGGTATC GACGATTGAA AACAAAACAG 9960

CAGGCAAAGC AGGCCAAAAA CCTGCAAGCG GCAAAAAAAG AACTGAAAAA AGAAAAGTCT 10020

GACTTAAAGG CCACCCTGCG CCTGACCCCG GACGCAAAAC TCAACAGCCG GGACAAGCGG 10080

CGGCTTGCGG CATCGGTCAG CAAGGCAAAG CGGGACGGCA GGCTTCCCAA GACCGCCCAG 10140

CAGACCATCC CCTACCAGGA AATGTGCCGG GACGGTATCT GTATCGTGGG CGACCGCTAT 10200

TTCACCAAGC AAATCCAGTT TTATGATGTG AACTACCAGC TTGCACAGAA CGAGGACAAG 10260

AACCTGATTT TTGAGAACTA CTGCGATTTC CTCAACTATT TTGACAGCTC CATCAAGGTG 10320

CAGCTTTCTT TCCTCAACCA GCGGGCCGAC ATGGAGGAAT ACCAGAAGTC CATCCATATC 10380

CCGGAGCAGA CGGACGCATT TAACGGTATC CGGCGGGAGT ATTCGGGAAT GCTCAAAGGC 10440

CAGCTTGCCA GGGGGAACAA CGGCCTGACA AAGACCAAGT ACATCACCTT TGGCATTGAA 10500

GCAGACAGCC TGAAAGAAGC AAAGCCCCGC CTGGAACGGA TTGAAGCGGA TGTGCTGGCG 10560

AATTTCAAGG TGCTGGGCGT CCGCGCCCAT TCGCTGGACG GCTATGAACG TCTGGCAATC 10620

CTGCACCGGA TGTTTCACCC GGCGGACAAT CAGAAATTTA AGTTTGCCTG GGATGCCATC 10680

TGGAAAACCG GGCTTTCCAG CAAGGATTTC ATCGCGCCGG ACAGCTTCAC CTTCAAGTCC 10740

GGCCAGTATT TCCAGATTGG CAGGACTTAC GGCGCGGTGT CCTTTCTGCA AATCCTCGCC 10800

CCGGAACTGA CCGACCGGAT GCTGGCGGAT TTCCTCGATC TGGAAAACAG CATGGTGGTG 10860

ACACTGCATA TCCAGTCCAT CGACCAGAGC GCGGCAATCA AGAACATCAA GCGCAAGATT 10920

ACCGACCTCG ACCGCATGAA AATTGAGGAA CAGAAAAAAG CAGTCCGGGC AGGGTACGAC 10980

ATGGATATAA TCCCCTCCGA CCTTGCCACA TTCGGCAGCG AAgcAAAGAC GCTTCTGGAA 11040

GATTTGCAGA GCCGCAACGA GCGGATGTTC CTTGTGACTT TGCTTGTGAT GAACCATGCC 11100

AGGACGCGGC AGAAGCTGGA AAACAACATC TTCCAGGCGG CGGGTATCGC GCAGAAATAC 11160

AACTGTGCGC TGCGCCGTCT GGACTACCAG CAGGAGCAGG GCCTGATGTC CTCCCTGCCC 11220

CTGGGCTTAA ACCAGATCGC CATCCAGCGG GGGCTTACCA CCTCCAGCAC TGCCATTTTC 11280

GTGCCGTTTA CCACCCAGGA GCTTTTTATG AACGGTGAAG CGCTGTACTA CGGCTTAAAC 11340

GCCCTCTCCA ACAACCTGAT CATGGCTGAC CGGAAAAAGc TGAAGAACCC CAACGGCCTG 11400

ATTTTAGGCA CCCCCGGCAG CGGCAAgTCC TTTTCCGCCA AGCGGGAAAT CACCAACGCT 11460

TTCCTGATTA CCCAGGACGA CATCGCCATT ATCGACCCGG AGCAGGAATA TTCGCCCCTT 11520

GTGCAGCGGC TGGGCGGGCA GGTGATTGAC ATTTCCCCCG CCTCCGACCA GTACATCAAC 11580

CCGATGGACC TGACCCTCAA TTACAGCGAG GACGACAACC CCyTGaCyTT AAAAAGCGAC 11640

kTTATCCTCT CCCTGATGGA GCTGATTGTG GGCGGCAAGG CCGGCCTGGA GCCGGTGGAA 11700

AAGACAGTGG TTGACCGCTG CGTCCACATG GTCTACCGGG ACTACCTGCA AGACCCGAAG 11760

CCGGAGAAAA TGCCGGTGCT GGGCGACCTT TACCGCCTTT TGCGGGAACA GCCGGAgCGG 11820

AAgCACAGCG GCTTGCTACC GCGCTGGAAA TCTACGTTAC TGGCAGCTTA AATGTGTTCA 11880

ACCACCAGAC CAACGTGGAA ATCAAGAGCC GGATTGTCTG CTATGTGATT AAAAACCTGG 11940

GCAAGCAGCT TAAAAAATTC GGGATGCAGG TGGTTCAGGA TCAGGTGTGG GGACGGGTCA 12000

GCGAAAACCG CGAAgcCCAT AAATCCACCC GGCTCTACAT TGACGAGATG CACCTGCTTC 12060

TCCGGGAGGA ACAGACCGCC GCCTATACCG TTGAGATTTG GAAACGGTTC AGAAAATGGG 12120

GAGGAATCCC GACCGGAATT ACCCAAAATG TCAAGGATTT GCTGGCAAGC CGGGAAATTG 12180

CGAACATTTT TGAGAACAGC GACTTTATTT ATATGCTCAA CCAGGCGGGC GGCGACCGGC 12240

AGATTTTAGC CAAACAGCTT AATATTTCCC CGCACCAGCT TGGATATGTG ACCAACTCCA 12300

ACGCCGGCGA GGGGCTTTTG TTCTATGGCA ATGTGATTAT CCCATTCGTT GACCATTTCC 12360

CCAAGGACAC GGAACTCTAT TCCATCATGA CAACCCGCCC GGAGGATCTG GCGGAACAGT 12420

AAAGGAGGGT AAGATTTGGA ACACAATCAG GAAATTATGG AGCAGTTTTT GAAGCTGCTT 12480

TCGGAAAACG GGcGGGCGGG ACAGTCAGAG GATCTGTCCC GCCTGCTGTT CTATATGGAC 12540

GGCATGAACC GCCAGCTTGA CGCGGTGCGC CAGGAATTGC AGACAGTCAA GGTACAGCTT 12600

GCACAGGCCC AGGACACGCC GCAGAAAACC GTCTTGCAGG GCATGGTGGA CGGTCTGCAA 12660

AGCAAAGTCC GGCAGGCGCA GGAAAAGCTG GACAGCCTGC GGGAAAAAAT CATGCAGTGC 12720

GCCGCCAACG CGGTAGAGAG TTTCAAGCGC GTGGGCGTGA CTGCGCTGGA TAAGGCGGTT 12780

TCCGCTATGG GCATCCACAA GACGCTGGAA GCGGTACAGC AGAATATCAG CGGTTCCCTG 12840

GACGATGCCA GAAAGAGCAT TAAAAAGGTG GAAACCCTTG GCCACGAACT CCGCAGCGTG 12900

GGCGGGCATC TGAAAAACGC CGGGCGCGCC GTGGCCGGCA GGGAAACCCA GGCCGTGGAC 12960

GGCGGGCAGG AAGGACGCTT CCAGGCTGCT TTCCTTGCCC CCATGCGGAC TGCCTATAAG 13020

ATGCTTTCCA GCATGAACAA TACCACCCTT GCCGCTATTG GCAACGTGGA ACGACTGGAA 13080

ACAGCGGCGG AAGCgTGCGG GAAGCCAAAA AGCCCCCTGT CCGGCAAACG GAAACGCAGG 13140

AAAAGGAGCA GGTAAAGGAG CTGGAACATA AGGAAGAACC CACTGAAACC CAGCGCCGTG 13200

AGAAGCCCTC TATCCGACAG ACGCTGACAG AGAAAAGTGA GCCTATGGAC GTAAGCGATA 13260

TTTTGAGCGA CCCTGCCAAT ATCCAGATTT TTGATACGGT TTTAGAAGAA TCCTGCCGTC 13320

AATGGTGCGA CTGTATTGGC GATGCGCCGG AACGTGCCAG CGGCAAAGGC TTTGCGTCCT 13380

TTTTCTATGA GGTTTTTGAA CAAAAAGAGC AGGAATACAA AGCCGAGTTA AAGGCAGACA 13440

CACGCTCTGA AAAACCATCC ATCCGAAAAG CGCTGAACGA GAAAAAGGCC GAAGCCGCCG 13500

CACGGCCTTC CCCAGAGCCG GAACGTAAAG CGCCGGAGmC GCACGGTGAG CCGGGAACGG 13560

TTTCAGGCAC GGGAGAAAAA GGTTCAGAAG ATGGGCCGGG ACGGTCTGGT GGAACAGAAC 13620

CGCGCCACCG GCGAAGAAAA gCGCGTCAcC AGCGCACGGC AGACACCGCC TTTGACCGCG 13680

CCCGCCCTAC GGAACGGGAG GACATCCGCC GCCCGGCAGC ACGGGGCGCA GACACCGGCA 13740

```
AAAAGCGGAA ACAGCCCCGC ACTGCGCCGG AGATGGCGGC AGAACAGACG GAACCGCTGc    13800
TTACGCCGGA ATATCTGGCA GACAGTACTG CCGAACCGCC GCCCGATGCC CCGGCATCCA    13860
TGCGCGGCGC GGTGGATATG CCGTTACTGG ATGCGCCCAT AGCCGAAGAA CCCACTCCCG    13920
CGCCGGTGAG GAAACAGCCC CAAAAGCGGG GGAAAAAGGG CAAGAAGAAG CAAAGCACGA    13980
AATTCACGGA AGACGCTGCC CGACCGGGTA GTGAGGAAAA AACGGCTATG CGCGGTGTGG    14040
AGGGCGCTCC TGCCCCAAAG CGGCGCACTT CTGAGGGCGG CTCCCGCCTG CAATTTGACA    14100
AGGAAGAAGA ACCGGCAGAC CCCGCAGCCG ACCGCGCAgc AATCAGAAAG CGGCAGACAG    14160
CAAAGCTGGC CGAACACGCC GCCAAACCGG AGGATaAGCT AAAATCCGAT GGCAATCCAA    14220
AGGCCGGGGa AACACGCCTG AAATTTGAGG ACACTcCCAA AGATACGGAC ATTTCCGGCG    14280
ATACTGCGCC CGATATGCCC ACCCGCCAAC AAAAGAAATA TAACAAGGCC GCCTGCCGCA    14340
CGGAACGCGC AGACCGCCGC GTGGAGCAGG CACAAATGAA GCTGCCTGCC AAACGCCGCC    14400
TGCACATGGA TCACCAGCCT GACGAAGCCA GCGGGAAGAT AAAGCGGCGG CTCCGTTTTG    14460
AGGATGAAAT CCTGCCGGAG TACCAAAAGC CCTCCCTGTC CTCCCGTGCC GGAAATGCGG    14520
CAAAAACGGC GGCAGTCTTA AAGCTGCATG GCAAACTCCG TGAATATGAG CGTGACAACG    14580
TGGCTCTGGA ATCTGCCCAC AAAATCGAAC TGGTTGCCGA ACAGGGCGCG GGGCGCGTCC    14640
TGCGCTGGGA ACGGAACCGC CGCCGCTcCA AACCCTACCG CACTTTGCGG AAAGCCCAGC    14700
AGAAAGCGGC GCGGGCGCAC ACTGACCTTG CATGGCAGAC CGCCnTGCGG GATAACCCCG    14760
AACTGCAACG CAAAAACGCG CTGGCAAAAT GGGTGCAGAA GCAGAAAATC AAGCGGAAAT    14820
ATGCCCAGGC CGCCCATGAA GCCAAACAGA GCGCGAAATT TACCCAAAAC GTCCTGACCG    14880
CCACCGgCAA GATTGCCCGC GCCATCGCCC AGTATGCGGC GGCACATAAG GCGGTCTTTC    14940
TGGCGGTTGC aAtGcTGGCG cTGGTGGTGA TGTTCTTCGC CACGGGCCTG ACCTCCTGCA    15000
cCGCCATGCT GTCAGGCTTc CAGTCCTCGT ACATTTCCGC GTCTTACATG GCTAATGAGC    15060
AGGAAATCTG CCAGAGCGAC CTTTATTACA CGGAAAAGGA AACGGATTTG CAGATTGATA    15120
TTGACAAAAC TGAGGAAAAT TACCCCGGCT ATGATGAATA CCGCTACAAC ATCGGTGAAA    15180
TCAGCCATAA CCCTTATGAA CTGTTAGGCT ACCTCTCCAC GGCGTTCAAT GCTTTTACCT    15240
TTGCAGAAGT CCAGCCCGAA ATTGACCGCA TTTTTTCCCG GCAGTACACC CTCACCCGTG    15300
AGGTGATCGT GGAAACCCGG TATGACGATG ACGGCGACCC TTATGACTGG TATGTTTTGC    15360
AGACCACATT GGCCGTGCGC CCGCTGTCCT CTGTGTTACA GAGCAGCCTC ACCCCCGGCG    15420
AACAGACCGA CCGCTACGGC GTGTATATGC AGACtACGGC AACCGGCAGG CGTTTGGAAA    15480
TCCCTTTGGT TTTTCCTGGC TTGGATATGT GAGCAGCGGC TACGGCTGGC GGGTACACCC    15540
TGTAAACGGC GAAAAGAGCC TGCACCGGGG CATTGACCTC GCTGTGGCGC AGGGGACTCC    15600
GATTCTGGCG GCGCAGGACG GGCGTGTGGT ATCGGCCGGC GATGCAGGCA GCTATGGGCT    15660
GTGCGTGGTG ATCGAGGACG ATAAAGGCTA CCAGTCCCGG TATGCCCACT GTTCCAGCTT    15720
```

AAACGTCAGC GCCGGCCAGG AAGTAAAACG CGGGGATGTG ATTGCTGCTG TCGGCAGCAC 15780

CGGCAACAGC ACGGGGCCGC ATCTCCATCT GGAAGTCATG TTAAACGGCG AATACTTGAA 15840

CCCTTATTTC TTCGTGGACA ATGGCGACGA TGGAACCGGC GCAATCCCCg GCACCCCCGG 15900

CGGGCCGGCT aTCCCgGACT ATTCCgGCGA AcCGATGGGC GATGGCAGyT TTGAAaGCCA 15960

TGCTGCGGGA ArCGGAGAAa TACCTGGGCT ATCCCTATGT GTGGGGCGGc TcCCATCCAT 16020

CCACTTCCTT TGACTGTTCC GGTTATGTTT CGTGGGTCAT TAACCAGTCC GGCGTGGGCA 16080

GCGTCGGGCG GCAGACGGCA CAGGGCCTTT ACAACCTGTG TACGCCGGTA TCTGCGGTGA 16140

ACGCACAGCC CGGCGACCTG ATTTTCTTTA CCGGGACATA CAGCTCGCCG GGGCCGGTCA 16200

GCCACGTTGG GATTTATGTG GGAGGCGGGC GGTTTATTCA CTGCGGCGAC CCGATTTCCT 16260

ACGCCAATAC GGGAAGCCCC TATTGGTCTG CCCATCTGTA CGGGTATGGC AGAATACGCT 16320

GAAACACTTT TTTCAGGCAT TTTTGCCAGA AACGGAGGTA TGATTTGAAA ATTGGACTGA 16380

TTGATGTGGA TTCCCACAAC TGGCCGAACC TGTGCCTGAT GAAGCTGTCC GCATACCATA 16440

AGGCGCGGGG CGGCCTGGTT GAGTGGTGGA AyCCCAAAAA CCACTACGAC CTTGTTTATA 16500

AAAGCCGGGT ATTCACAGAT ACCTACTCGA AAGATACGAT TGCCGTTACC AATGCCGATG 16560

CCGTCATTTT AGGCGGCACG GGCTACGACC TTGGGGAAAA CTTACCGGAT GCGGTGGAAC 16620

ACACCCGCCC GGACGCTGCA CTGTATCCGC AGTTTTCTGA TACCGCTTAC GGCTTTTTGA 16680

GCCGTGGCTG TCCGCGAAAC TGCGGATTCT GTATTGTCAG CGGCAAGGAG GGAcGGCGAA 16740

GCCGAAAGGT GGCCGACCTT TCCGAGTTTT GGGACGGGCA GAAAGAAATC AAGCTGCT 16798

(2) INFORMATION FOR SEQ ID NO: 196:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1409 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

ATAAAATCTA GCCGTCAATA TCATAGGAAA TTATTAGAAA ATcCCAATTC TGcACCAAGT 60

aTGTGGTTTA TcAATcAAAA ATATGGTTCA TGGaAAAACT TATTAGTTAG TCTTGGGTGT 120

GATAATGGAG AATATGGGAA ATGGGCAAAA ATATCAGAAA AAGATTTATT GAAAATAGTA 180

GAGTCTTTTA TCACGGTTGA GAAAATAACT TCTCAGCGTA TGTATGAAAA AAGATCAGTA 240

GGAAAAGACG TTCCTTCATT GAGTACATTA AAAAAGAGAT TTGGAGATAT AAGGTATCTA 300

TTTAGAAAGA ATACAGAGaA GTCTTCTTTT ACTGATTTTG aACTAATGAT TGAACTAAGA 360

AaTGAAATAG TCCGATTAAA ATTACAGGAT GATTTATCAA TGACGAAATT TCGAAAACTA 420

GTTCAATCAC CAAAGTTACC ATCAGTTGAT ACAATAATGA AAAGAACAAA TAAAAATTGG 480

GAAGAGTTGA TGACAGAAAT AGGGTTTGAC TATCGAAAAA TCAAAATCAA TAAGCAGAGA 540

AATAACTTAT CTAAAAAGAA GAAAACTAAA TAGTAAAACT CTAGACTAGA CGTTCAAACT 600

AAAAAGTTGA ACGTCTTTTT TATTACCTAA AAACGAAAGG GGATTTTATT TTATGATAAA 660

GCCGATATTT AAAAAATGGC GATTGTTTGC GGCTTTAGCT TTATTAGGTC AAACAATTGT 720

AGGGGCGATT GATCCTATGA TTGTCTTTGC TGATGAAATT GCTCATCCAC AAACAGTGAC 780

GGTTGAATTA GATTTAGCGC ATCAATATGT TGTTGAAGGA ACTTTCAGTG ACGGTCGTCC 840

AATGTCAGAA GTCACTGTTC CACATTATGC TGTGTACAAC GGCGTGAAAC AAGATGTTTT 900

CTGTATTGAG CCAGGAGTCC CAATTGACAA TGAGTTCACG CCGGGTTATG AGAAAAATCC 960

ATTGCCTGAT ATGCCAGAGA AAGCGAAACT AGTTTCCGTT CTATGGAAAA AAGCCGGTAC 1020

AGATGTAGAT ACTCATATTG TGGCACAGAA GATGATCTGG CAAGAAGTTA ATGGGTATAC 1080

CCTTCACTCA ATCAAACGAT TGAATGGTAG TGCCGTAAAT ATCGCAGCCA TTGAGGCAAA 1140

AATCAATCAA GCCATTGCCG ATTATCAGAA GAAACCAAGC TTCCATAATA GTACGGCTAA 1200

AATAGTATTA GGGCAATCGA CTACTATGAC GGATACGAAT AATCTGAATT TGTCAGAGTT 1260

TGATGAAGTG GTGGAAAATA CCGCAAACAT TGATTATCGT GTGAATGGCA ATCAGTTGAT 1320

AATCACCCCA AATGCCAGTT CAAAAGAAAG TGGTGTGctG ACGCTTAAGA AATCTGCTGG 1380

CACCGGAACA CcTGTCsCCT ATaAAAAGG 1409

(2) INFORMATION FOR SEQ ID NO: 197:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9619 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

AAATGGGAAA AGTCGCTCGA GCAATCGCTT CGGGTCAACA AAAGAGTCAA CTTTTCTTAA 60

AGGATATTGA AGAGGATACG AAAAAATTAG TCCGTGAAAT TAAACAAAGC GAGAAAAAAT 120

ATCAAGATTT TTCCATTACG CAAAAAAAAT GTCCAGATTG TGGCTCAAAC TTACGGGAAA 180

AAAATACGAA AGATGGGAAG ATTTACGTTT GTACAAACCA AGAATGTTCT TATCGTCGGC 240

GAAAAGATCC TAAAGTATCC AATCACCGTT GCCCACAATG CCATAAAAAA ATGGTCATCA 300

TAGAAGGAAA AAATGGTCGT TCTTTCAAAT GTAAATTCTG TTCGATTACT GAAAAAATTC 360

CCGATAAAAA AGAACGAAAA CAAAAAATGA CTAAACATGA AGAACGTCGT TTGATGAAAA 420

AATATTCACA ACCAGACGAA CCAGAAGAAA GTGCTTTAGC GCAGGCCTTA AAAGCCGCTA 480

TGAATCAAGA GTAGCGACTT GACGAATGTT GGCTCTAAAT AATCAAGTCA TCCTTGTTGG 540

TCTTTAGTAA ACTTTCAAAT AGAAAGAGAG GGACACGGAT GCTCGCAAAT TTAAATCAAG 600

CAATGGATTA TATTGAAGAA CATTTAACGG AAGAAGTATC CTTCGATGAA TTGGCGAAAA 660

AGACAGGGAT TTCTGTTTAC CATTTTAAAC GAACTTTTTC GTTTATTACG GGGATGAGTC 720

TGGCTGAGTA TATTAAAAAG AGACCTTTGG CGGAAGCAAA TTTAGCATTA CTTGCTGGTG 780

AAAAAGTGAC AGATGTAGCT TTTAAATATG GTTATCAAAG TATTGAAGGA TTCTCAAGAG 840

CTTTTCGCGA TTGGTCAGGA CAAGCCCCTT CTGAAGTAAT GAAAACACAA ATCCAAAAAA 900

CGTTTCCTAA ATTCTCTTyC TATATTGATA TTAAAGGAGG ACAATCAATG GATGTTAAAT 960

TGATTGAAAA ACCaGCGTTT CAAwTAGTCG GAGTTAGTCA GAAGGTTCCC TTGCAGTATC 1020

aAGGGGAAAA TCAAGCAATC ATGGAATTAG CACAGCGCAT TACACCACAA CAACGAGCGG 1080

AAATGCATAC ATTTGATGAC ATATATCCTC ATCAAGTAGT AAATGCTTCT TTTGACTTTC 1140

AAGAGGGACG AACAACAGAA GGTGGAGAAA TGACACACAT GATTGGCTTT GCCACTTCGC 1200

AAGAAAATAC CTATGAAGAT TTGGAGCAAC TAAGCGTACC TGCTCATACA TGGGCGGTTT 1260

TTCCAAATGA AGGTCCTTTC CCACAAACTT TACAAGAAAC CTGGGCAAAG ATATTCTCTG 1320

AATGGTTGCC TTCATCTGGT TACCAAGTCG TTGCAGCACC AGAAATTTCG TTTACGCAAT 1380

ATCAAGGACC AGCAGAAGCT AAGTATAGTG AAATCTGGCT TGCTGTTACA GCTACTAAAT 1440

AAAGAAAACC CACCATTGAA AAATGGTGGG TTTTTCCGCC AAGAAGGAAA AAGTTkGGTA 1500

TAATAAACGT GAAGAAAAAA ATTAGACCTT CTAAACTGAA ATCTAGCTAT GGATAAGTGA 1560

TGCGAATACG GAATCATGGA GAAAAAATAA TTCTCCGAGG TTTTAGAGTC ATGTTGTTTA 1620

GAATGGTACC AAAACAACAA ACCCTTGTTC AGCATAATGA CTATAGTTTT AGAGTCATGT 1680

TGTTTGGAAT GGTACCAAAA CAAACTTTTT TGATTTGGCT TTTTCTCCCC TGTTTTATAG 1740

CCATGTTGTT AAAAAACAAA CTATCACCTC AAGCAATCCG TAATTTTCAT CACAAGATTA 1800

GAATTTTTAT CATCAAAGAC GAGCTGCTTC AATTTTCGAA AACTAATCAA GTGATAAGTC 1860

AAAATGTTGA TAGATTGAGA TTAGTTCTTC CTTTTTTTAG GAAGGGCTTT TTTTATTGAT 1920

AGAAGAACGG GAAGAAATCT TGGTAGCACC TAAAAAAATT GTGAGAAATG CTTCTGTATT 1980

GTTTGGTTTC TGCCTCAGAA AGTTGTATAC TTAGATTGGA ATTATTCTAT ATTACATGCA 2040

TTTTCACACT TTTTGGAAAA GTTCACTTGA ATTTTCTTTT AGTTTCGTAG ATAAAAGGAG 2100

TCATCGCTAA TGGAAATGAA CAATTCAGGT AAGCTTGTTT CTTTGTGTGG AGGCAAATCA 2160

GGGAGGAAAT AATATGTTTG ATATTGTAAC ATTGGCGAGA ATCCAATTTG CGATGACAAC 2220

AGTTTTCCAC TTTTTCTTCG TACCGTTTTC AATAGGATTA GCACTTGTTG TAGCAATTAT 2280

GGAAACAATG TACGTTGTGA AAAAAGAGGA ACGTTATCGT AAAATGGCCA AGTTTTGGGG 2340

AAATATATTC CTATTAAgcT TCGCTGTTGG GGTTGTAACC GGGATTATTC AAGAATTTCA 2400

GTTTGGGATG AACTGGTCTG ACTATTCTCG ATTCGTCGGA GATATTTTTG GGGCACCGTT 2460

GGCGGTAGAA GCGCTCTTAG CTTTCTTCTT AGAGTCAACA TTCTTAGGTT TGTGGATCTT 2520

TACTTGGGAT AAAATGAATC CAAAaGTtCA CGTAACCTTT ATCTGGTTAG TCGTATTCGG 2580

TTCAATGATG TCTGCTTTCT GGATTTTAGT TGCGAATAGT TTTATGCAAC ATCCAGTAGG 2640

TTATGTCATC AATAATGGCC GTGCTGAAAT GGTGGACTTT GTAGCTGTCA TCAGCAATCC 2700

AAAAGTTTGG TATGAATTTT CACATGTTAT CATGGGCGCC TTTACATTGG GCGGCATGGC 2760

```
AGTAGCTGGt ATGGCTGCTT TCCAAyTGTT GAAAAAACGG GATATTTCaT TCCACAAAGC    2820
ATCTATGCGC ATTGGCTTAT GGGTTACTTT GTTCGGTTCA ATCGGCGTTT TATTAGCTGG    2880
CGACTTGCAA ATGAAAGCTC TAATTGAAGG ACAACCAATG AAATTTGCGG CGATGGAAGG    2940
GGCCTATGAA GACTCTGATG ATCCTGCTGC TTGGACATTG ATTGCTTGGG CAGATGAACG    3000
GGAACATAAA CAAGTTTTTG GTATCGATAT TCCGTATATG CTAAGTATTT TATCGTATAA    3060
CAAATTATCA GGTTCGGTTG ATGGAATGGA TACAGTCAAC AAGCGTTTGG TAGAAAAAtA    3120
TGGGGATGAT AAAAATTATT ATCCACCAGT TAACACTTTA TTCTGGAGCT TCCGTATCAT    3180
GGCTGGTTTT GGTGCCTTAA TGTTATTGGT TGCTGCCTTA GGACTATTCT TTACAAGGAA    3240
GAAAAAACCA TCGCTGTACG AGAAGAAATG GATGCTTTGG ATTGTTGCTT TATGTACGTT    3300
CGCACCATTC CTTGCCAATA CAACAGGTTG GTTAGTCACT GAATTAGGTC GTTATCCGTG    3360
GACTGTCTAT GGATTATTTA CGATTGAACA AAGTGTTTCA CCAAATGTAT CGGTTGCTTC    3420
ATTAATTACG TCAAATGTCA TTTATTTCTT ATTATTTGCT GGTTTAGGTA GTGTCATGGT    3480
TTACTTGGTT ATTTTAGAAT TACGTAAAGG ACCAGACTAT GAAGCGAAAA AATTAGCGAA    3540
AGAAAATGAA CCAGCGTTGG ATCCATTTGA TAAGGGGGTA TTTGGCGAAT GAGTACACTA    3600
CAATTATTAT GGTTTGTCCT TATCGGCGTC TTGTTCTCAG GTTTCTTCTT CTTAGAAGGA    3660
TTTGACTTTG GTGTTGGAAT GGCCGTTCAA ACGTTGGCAC ACAATGACGA CGAAAAAGAT    3720
CAAGTCGTTG CAACAATTGG ACCTGTTTGG GACGGAAACG AAgTTTGGCT ATTAACAGCG    3780
GGCGGCGCAA TGTTTGCCTC TTTCCCATAT TGGTATGCTT CATTATTTAG TGGCTACTAT    3840
TTAATTCTAT TTACGATTTT ATTTGGCTTA ATTATTCGCG GAGTTTCCTT TGAATTCCGC    3900
CATAACATGC CGGAAGGTAA ACGTCGTCGG ATGTGGAACT GGACACTATC AATTGGTAGT    3960
TTCCTTGTAC CGTTCTTCTT TGGTATTTTG TTTATCAGCT TAGTCCAAGG GATGCCTTTA    4020
GATGCAAACG GCAATATGCA TGCGCAATTT ACGGATTATT TTAATCTGTT TTCAATTGTT    4080
GGTGGAGTAG CCTTAACATT ACTTTGCTAC TTACACGGCA TGAACTATAT TGCTTTAAAA    4140
ACAGAAGGCC CaATTCGTGA ACGTGCTCGA AATTATGCAG AAATCTTATA CGGCGTTTTA    4200
TATATCGGTT TAGTTGTTTT TGCAGTGTTG ATGTACTTCA AAACGGACTT TTATGAAAAG    4260
AACTTTGCGG TCACTTTAAT TTTAACGTTA GCAATTGTGG TCTTAACGGT CATTGCTAAT    4320
GTTGGGGTGT TTAAACGCAA AGAGATGTTA GCTTTTCTAG CGAGTGGTTT AACATTAGTT    4380
GTACTAGTTG CTCTCTTGTT CAGTGGCTTG TTCCCACGTG TTATGATTGG TAGTGAAGGT    4440
TTTGATTTGT TGATTAAAGA TGCAACAAGC ACACCATATA CACTAAAAAT TATGACGTGG    4500
ATTTCTTTAA GCATTTTACC ATTCGTGTTA GCTTATACTG CTTGGTCCTA TTATATTTTC    4560
AGAAAACGTA TTTCACAGAC AGCCGTACCA GAGGGGTATT AAATCTATGA TTGATAAAGA    4620
CATCTTACAG ATGCCTAAAA TCAAAAAAAT TCTTGTCTTG CTTGCAGGTT TTTCTTTCCT    4680
GCAAGCAGTC TTTATAATTG GGCAAGCTTT TTTTCTTTCA AAAGCCATTG TCGGCCTTTG    4740
```

GAGTGGCGGG CATCTTAACC AACAACTACA GAGTATTTTA CTCTTTTTTA TTTTTTATCT 4800

TGGTCGACAT GTTATTACGT ATTTTCGTGA GAGAATGCTC GATACATTTA GTTATGAACG 4860

CTCGAAAGAA TTACGCGAAC AATTATTAAC CAAAATTTTC CGTTTGGGGC CAAATGTTGT 4920

TCAAAAAAAT GGAACTGGAA ATATGGTTAC CATGGCTTTA GAAGGCATCA GTCAAACCGA 4980

AAATTATTTA CAATTGATTT TGACAAAAAT GATGAATATG ATGATTATTC CTTGGATTAT 5040

TTTGGCCTTT GTTTTTACAC AAGATATTCG TTCAGGCATA ACATTGCTGA TTGTTTTTCC 5100

GATTATTATT ATTTTTATGA TTGTGTTAGG CTATGCAGCA CAAAGTAAAG CGGACAAGCA 5160

ATATGCGGCG TTTCAAATGC TTTCCAATCA TTTTATTGAT TCCTTGCGAG GAATTGACAC 5220

GCTGAAACTG TTTGGGTTGA GCAAAAAATA TGCTGGGAGT ATTTACCATA CAAGTGAACG 5280

ATTCAGAAAA GCGACCATGA GTACATTGAA AATTGCTATT TTATCAACCT TTGCGTTAGA 5340

CTTTTTTACG ACATTATCGG TAGCTGTTGT GGCAGTTTTC CTTGGATTAA GTTTATTAAA 5400

CGGCACGATT TTATTATTTC CTGCCTTGGT TACTTTAATT TTAGCACCTG AGTTCTTTTT 5460

ACCGATTCGT GATTTTTCAA GTGATTATCA CGCAACCTTA GATGGAAAAA ACTCTTTTCA 5520

AGCGATCCAA GCGGTATTAG CATTGCCAGA AGCAGAACAG ACCGATGTGC TCACGTTAGA 5580

TGATTGGCAT CAAGAAAGTC AGTTGTCAGT CACAGCGTTA AATTTTTCCT ATGAAGATGC 5640

AACGCAAGAA GCGATTCAAT CGTTGCAATT CTCTGTTGAT GGTCGCAAAA AGATTGGAAT 5700

TATCGGGGCG AGTGGTTCTG GTAAGTCAAC GTTAATTAAT TTATTAAGCG GCTTTTTATT 5760

ACCAACAGAA GAAACGAGTC AATTGGCGAT AAATGGCCAA ACGATTCCAC ATTTTTTACA 5820

AAAAGACTGG CAAAAACAAA TCCTTTATAT TCCGCAAGCA CCATATATTT TCCAAGATAC 5880

TTTGGCAAAC AATATTCGCT TTTATACACC TGAAGCAACA GATGAGGCGA TTCAGCAAGC 5940

GATTCAGTTA GTTGGCTTAG ACGAGCTGGT TAAAGACCTT CCCGAAGGAA TTCATACCTT 6000

GATTGGGGAA AGTGGTCGAA TGTTAAGTGG GGGACAAGCG CAACGAGTAG CGTTGGCCCG 6060

CGCCTTTGTC GATCAAAAAC GCCACGTCTT GCTGTTTGAT GAGCCGACAG CGCATTTGGA 6120

TATTGAAACA GAAGTTGAAA TGAAAGAGCG CATGTTGCCA TTAATGAACA ATCATCTTGT 6180

CTTTTTTGCG ACACATCGTC TGCATTGGAT GGAAGAAATG GATTACATTT TAGTGATGGA 6240

AAAAGGTCAA TTAGTGGAAC AAGGAACGCT AGCAGAGTTG ATTGAAAAAG ATGGCTACTA 6300

TGTTCAATTA ATGAAGCAAA TGAGAGGAGG ACGCCAATGA TGGAAAAACA GCCGACAACG 6360

AAAGACGTTT GGAAAAACGA TCAATGGGTT CGTCCTTTCT TAAAaCGCTA TAAAAAAACC 6420

TTGTATTTTG CCTTACTTTT AGGCTTCTTG ACGTTCTTTA GCGCAGGGGC ATTAATGTTT 6480

ACGTCTGGTT ACTTAATCAG TCGTGCGGCT TCCTTACCAG AGAATATTTT ATTAATTTAT 6540

ATTCCCATTG TTTTAACAAG AGCATTCGGG ATTGGACGTC CTGTTTTTCG TTATGTGGAG 6600

CGTCTAACGA GTCATAACTG GGTGTTAAAA ATGACCTCTG ATCTGCGCTT GAAACTTTAT 6660

AATGTACTAG AAAAAGATGC GATTTTCTTT AAGACAAAAT ATCGCACTGG GGATATCTTA 6720

GGTTTGCTTT CGGAGGATAT TAATCATATC CAAAACTTGT ATTTACGAAC GATTTTCCCA 6780

ACCGTTATTG CTTGGATTTT GTATATCTTT TTAGTGATTG CATTAGGTTT CTTTTCTTGG 6840

TGGTTTGCTT TATGCATGTT ACTGATGTTA GGAGTGGTTG TTTTTCTATT ACCACTTGTT 6900

TCAGTTCTAG TGAATGGTGC GCGTCAAGAA AAACATAAAT ATGCCAAAAA TGAGTTATAT 6960

CAAACACTAA CCGACAATAT TCTAGGTGTG TCTGATTGGG TATTTAGTCA ACGAGGCTCT 7020

GAATTCGTTG CTCGTTATGA AACAGATGAA GCAAACGTTC GTGCATTAGA TGAAAAAATG 7080

AAGCAATTTA ACCGTGGCCG AGACTTTGTT TTACAACTTC TGTTTGGTGT GATTGCGATT 7140

GCTGTCTTAG CTTGGACAAG TGTGCGGTTT CCAGGTAATC ATGGGGGCGC AgCcAATTGG 7200

ATTGGTGCCT TTGTTTTAAC CGTTTTTCCG TTAATTGATG CCTTTGCTCC GCTACCTGCT 7260

GCAGCACAGG AAACAACGAT TTATAAGGAT TCCATTCGTC GTTTTAATGA GCTTCCTGAA 7320

GGGGAAGATG ACTCAACAGA AGCGCCTGTT CAGCCAAATG GCACCGGTTT ATCGATAGAA 7380

CATTTGTCTT TtGCATATGA AAATCAAGCG AAAAAGGTAT TAAATGACTT ATCGTTAACA 7440

ATTCCTGAAA AACAAAmATT AGCGATTTTa GGGCGCAGTG GTTCCGGAAA GAGTACGTTG 7500

GCTTCGCTCA TTCGTGGTGA TTTACGGCCT ACTTCGGGAG AAATCCTTTT AGGAAATATC 7560

CCAACTGAGG CATTTGGTGA GACGATGACT GAATATATTG GGGTTATGCA TCAAGCACCG 7620

TACCTATTTC GAACAACCAT TTTAAATAAT ATTCGGATTG GTAGAGAAGA AGCAAGCGAA 7680

CAGaAGTTTG GGCGGTTTTA GAAAAAGTTG GCTTAAAAGA AATGGTTGCC CAATTACCAG 7740

AAGGTTTACA AACAATGGTG GATGAAGCAG GTTTGCGTTT TTCTGGTGGC GAACGACATC 7800

GCTTGGCGCT AGCGCGGATT CTTTTGCAAG ATACACCGAT TGTGTTATTA GACgAmCCAA 7860

CTACTGGTTT AGATCCAATT ACAGAACAGC AATTGCTGGA AACCTTTTTT GAGGCATTAA 7920

AAGATAAAAC CGTCATCTGG ATTACCCACC ATTTGCAAGG CGTAACTTTG ATGGATCAAG 7980

TGATCTTTAT TGAAGATGGC CAATTAGAGA TGAGTGGTAC GCCAGAAGAA TTACTTGCAA 8040

CCAATGCGCA TTATCAGAAA CTTTACCGAA TTGATCGTGG CATTAGTTCT TTCGAGGAAT 8100

AAAAAAACGA CTGGTTGTTG ATTAAAGCAA ACAATCAGTC GTTTTTCTAT TTCAATTAAT 8160

GGCAATTCTT CTATAAAACG TTTAAAATGA ACGATGTAAT AAGAGCTCAG TTAAGGAAAG 8220

TAAGGTTTCT TTTGCTGTGC CATCAgGCAA TTGTTGAATA TCAGTGATTG CTTTGTTGGy 8280

TACTTTTTTA GCAAAAGCTT GTGCTTCTGT AACGCCATGA TaGTCATTGA CTAGCGCCGC 8340

AACTTCAGCT GCCTCTTCTA AGGTAATGGC GCGTCCTTTA TCCAAATAAG GGCTGAATAC 8400

GTCAGGTGCT GCTTGATAAG CGAAGAGAAG AGGTAACGTA TAAACACCTT GTGACAAGTC 8460

TTCTAAAGCG GGCTTTTTCA AGGTTTCCGT GTCGGCAGTA TAATCTAAAA TaTCGTCATA 8520

GaCTTGAAAA GCAATGcCAA TATGACGGCC AATTCGTTTA GCTAAACGTT GGACTTCTTT 8580

TGAACTATGA CCGAAATAGG CCCCTTCAAG ACAGCTTAGT GAAAAAAGTT CGGCTGTTTT 8640

GcCGTTGACA CTTCTTAAAT AAGCAGGAAT CGACATGCGG CGGTCAAATC GATGACTCAT 8700

TTGATCGAGT TCACCAAGTA ATAACCGTTT CATTCCTTGT GCATTGATTT TCATAAATTC 8760

AGAGCCATTC ATGGCATCTG CAATCAAAGT AAAGAATTCA GTGAATAGTA AATCACCTGT 8820

GTAAACAGCC ACGTCTTTAC CATATTGTGA TTGAATAGTC ACAGCGCCTC GGCGCAATGG 8880

AGAATCGTCA ATAATGTCAT CGTGAATCAA CGTAGCCATA TGTAGGATTT CCAAAGATGC 8940

AGCTATTTTT AATAGTTGTT GTTTCTCTTG TTTTTCTTCA TCACCCAGTT GAGCAAAAAG 9000

AAAGAAAAAA GCGGGGCGTA ATAATTTGCC GCCAGAACGA GCCAAAGTTA CTAAGGCTTC 9060

TTCAATGTCT TCGTTGCGTG TATGCAACGT TTGTTCAATT AATTCACATG TTTCCGTTAA 9120

AGAACGTTCA ACGTCAGGaA ACGACTTCCA AAATGTGTTC ATAAATATAC TACGATCCTT 9180

TCAGTGGTTC AATACTATTT GATTGTATAC GAATTGGCGA CAGAAATACA GTAATTAAGT 9240

TTTAAAAGTA ACAAAGTTTG TAAAAAAGTA GCGGAGGGAT AGTTTGAATC GAGAAATATT 9300

TTTTGAATTA GTAGAATTAA AAGCAAAAAC AGCTAGTGTT TTGCCATTTT TGTTAgGAAT 9360

TkGTTTTAGT TGGTACCAAT ATGGGCGGCT TCATGTTGGT TATGTTTTAA TTTTTTTCAT 9420

TGCGATGTTT ATTTTTAATA TGGCGGTCGA TATTnTAGAT AATTATAATG ATTACCATCA 9480

TGCCACAGAA GTTCATGATT ATAAGGGAAA AAACGGAATA TTATCGGTCG AGAAAATTTA 9540

TCTTTGCCAT TAATACGCAA AATGATnTTT TGGGATGATC CCGGTTTCTG CTTTAATGGG 9600

GATTGGGGTT GCCATTnAA 9619

(2) INFORMATION FOR SEQ ID NO: 198:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1521 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

GCTCGCTTCC AATAGAATAT TTAAAAAAAT TTAAAAAGTC TACATAATTC CAAACAAGAT 60

TACCCTTTCA TTTTACACGT GATAAAATAT TTTTGCAAGA ATTGTCCTCA TTCTGAAGCA 120

AAAAAATTCT CTTCTCTCAA ACTAATGTAT CCGCTATCAC CAAGAATAAT ATGATCCAAC 180

AATTGAATCC CCATCATTTC GCCACATTCT TCCATTCGTT TGGTAAATTG GATATCTTGT 240

GGAGACGGTG TCGGATTGCC AGATGGATGA TTATGAGCTA ATAGTATGCG AGCAGAAGAA 300

TATTTAACCG CTTCCCGAAA AATTTCGCGT GGATGAGCAA TACTTTGATT CAGTGATCCT 360

TTGAAAATCG TTTTCTGTTG GATAATGTCA TTCTTGGTAT TCAAGTAAAT ACAAATTAAA 420

TGTTCTTGTT GCAGCCCCTT CATCTCTTGG AGTAATCGTT GAGCGACTTG TTGACTGGAT 480

GTAACTTTGC CAAATTTGAT TTGACTACTT TGaTAAATTC GACACCCTAA TTCAATGGCC 540

GCACGTAATT CAATGGCTTT CACTTGGCCA ATGCCTTTAA TAGCCATTAA CTCGTTCAAC 600

GTGGCTTGCT GCAAATAAGA AAGTTGCTTA AACTGTTTTA ACAGCGTTGC GGCAACAGTC 660

ATCACGTCTG CTTCTTTTGA GCCAGTACGT AACAGAATCG CCAACAACTC TTGATTCGAT 720

AATGCTTTTT CTCCTATCGC CAAGAGCCGT TCTCGTGGTA AACAATCACT CGGCATCTCG 780

CGAATAAACA GATCCGATAC TTGCATAAAA AAACCTCCTT TGCTAGTCAT TAATAAACTA 840

CGCAAAGGAG TGTTGTTTTT ATTATTTTAA ATAATGTGGC GCTTGATGTA AACTCTCCAA 900

GACATGTAAT GTTTTCTCTT GGATAACTTC AGTTGGTTGC AATAAATCAG GTACCATGAT 960

CACTGGAATG CCCGCACTAT GAGCCGCACT TACCCCATGA AATGAGTCTT CGAAAATCAA 1020

GGTTTTCGGG GCTTCTGTTC CTAAAAGTTG ACGAGCTTTT TGAAAAATTT CTGGGTCCGG 1080

TTTTGCTCGT TTAACATCCT CTGCTGAAAC AATACCTACA AAACGATCCT GAATTCCAGC 1140

ATGGCTTAAT AACATTTCAA TTGCTGGACG GACGTTACTT GAAGCAACTA ATCGAGGAAT 1200

TTTTTGATCA TCTAAAAAGT CGAGGAATTC CACCACCCCT GGTTTTAAAG GGACGTTACC 1260

AGATCGAAAC TCTTGCAAGG TATCATCATA AGAACGGCGG ATAAATTCTT CCACTGTGTC 1320

ATGACCATAG GATGCATAAA TACGTCGGTA ATTCTCTTGT ACTTCTTCAT CGGAAATCCC 1380

AACATAGTCT AAATAGACTT CTTTACTATA TGGTAACCCC ATTGCATCTG CTACTTTTTG 1440

AGTAGATGTA TAGTAAATCA ACTCCGTATC AAAAAGCAAC CCATCCATAT CAAAAATTAC 1500

GCCAnCAATT nCTnCATCAT T 1521

(2) INFORMATION FOR SEQ ID NO: 199:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5738 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

TCACCTAACT TATTACAAAG AAATAATAGT CTTTTAAGTA ACTCTTTGTC AAGACATAAG 60

CAGAAGTCAA AACAGTTCTA TTTACAAATG AATGCGACGA ACCATTTCTT CTGTTATTTG 120

TTCTACAATA TCAATGGTTG TTTTTGGCGA ACGTTGATTT CTGGGGATAT CGATTGAGTA 180

AATTTTGGCA ATACGACCAG GATTTGCATC CATTAATACA ACCCGATCAG AAAGTTGAAC 240

TGCTTGTTGA ACATCATGCG TGATCATCAC AATTGATTTA GTCGCAATTT GATTGTTTAG 300

CCAAATTTCT AATAAATCGT TAGATAATTG GCTGGCTGTA ATGATATCTA AAGCTGAAAA 360

AGGTTCGTCT AAGAGCAATA AGCTAGGTTC GATGGCTAAA GCACGAGCAA AGCCTACACG 420

CTGTTTCATT CCACCAGATA ACTCTCTAGG ATAAGCTTTT TCAAAACCAG TTAAACCAAT 480

CATTTCAATG AGATTACTGA CTCTTTTATT TATTTGTTCA CGGGAAATTT CTTTACGGTT 540

TTCTAAGCCA AAGCTGATAT TTTTTTCAAC TGTCAACCAA GGAAAAAGGG CAAAATGTTG 600

AAAAACCATG CCGATATTGG CATCAGGCTC ATTAATGATT TTATCTTCAA AAGTAATTTC 660

ACCTCTAGTT GTACGAAGAA GTCCAGCAAT ACTTCTGAGC AACGTTGATT TTCCTGTTCC 720

TGATTGTCCG ATAATACTTA AAAATTCATT CGTTTTAACT GTTAAATCAA TATTTTCAAG 780

GACTTTTAAA TGTTCTGAAT GTCCTTCAAT TTCAAATTCT TTATATACTG ATTTTATGCT 840

TAATAATTCT TTATTTTCCA TCGTCTTATT AGATAGGTCA AGGGAAATCA CCTATCGTTC 900

TCCTCTCTAG TTTATCTTAA AAATTTAGTC GAGTTGAGTC GTCCAGCTAT TAATAGTGAA 960

ATTTTTGTTT TGCTGCTCGA TATAAAGGTT GCCAAACAAA GGCGATACAT AAACCAACAA 1020

CGAAACACAT CGCTAAGGTT CCTAAGGCAG ATTTATCTTT TACACCATCG TTTGCCGCAA 1080

TATATTGTCC TAGGCCAGTA ACATCTATTG ATTTTTTGCC CCATTGAATG CTTTCTGCTG 1140

CGATATCAGC GTTCCAAGCA GCACCAGCAG CGTTAATAAT TGCTGTAATG AGGTAAGGTA 1200

AAATAGAAGG GATCAAAAAT TTTGTCCACC ATTGCCATTT AGATAATTTA AAAAGTTTGG 1260

TCACGTCACG AATATCATCA GGGATGGCTA AGTAGCCAGC AATTACATTA AAGAAAAAAT 1320

ACCATTGGCA ACCAACCATG ATGAGTGGTA AAATCCACCA TTCTAATTTG TTGaAGCCCA 1380

TACTAATAAT CAACACAATG AATGGAGTAT AAACATTTGA CGGAATGGAG CCTAAAATTT 1440

GTCCAATCGG TTGAACGATA CTTAAGCGCC GATTATCATT CGCAACCCAG ATTGCTAAGG 1500

GAGTAAAGAT AATGCCAGCA ACGACCATAG CCAAGGTAAC CCGTAGCGTA GTATAAAGGG 1560

TTAATCTAGG AATCATGCTG AAATCTTGAT GGGGCAAGAA ATGATAGAGC TTAATAGAAC 1620

AATAAACCCC AATAATGGTT AAGCTGACAT ACCATAATGT TTTGAAGAAA TGGCGTAATT 1680

TTGCTAGTAA TCGATAGAGA TAGTCTAGCT TGAAAAAGTA CCAAACTTTT GGTAAATCGT 1740

AAATCCAAAA ATGACTGAAT TTCACAAAAG CTTTTGAAAA GTATTTCGTT AAAGTTGCTT 1800

TTTTTAATAA ATGATAAATG ATGGATTTAG GTGGCGTTGT ACTTTCGACA TCTGTTTCAT 1860

ATTTAAAATA ATAAGTGGCT CGAACTAAAG GTTGAAAAAC TAAAAAGTTT AAAATCACCA 1920

CATTGATTAA TAAGGCAAAA ACAGCCCATA AACAGGATTT AAAGTCGGCT CGATTTAATG 1980

CTACTTGGAT ATAAGAACCA ATTCCTGGTA AATAAAGGTT GGTTTCTTTA GGAAAGGCAA 2040

CAGAAGCTTG CTCGCTGGCC ACCAAGGCAA ACCAAGCCGC AGTTTGAGAA ATAATAACGT 2100

TCCAAAGTAA GCCAGGAATT GAGTAAACAA ATTCAAGCCG CCAAAATTTT TCCCAAGCAT 2160

TGTACTTGAA CTGATCGGTC ACATTGACTA GATCTGAAGG AATAACTTCC ATTATTTGAT 2220

AAAGAGTCAA CATGATGTTC CAAGCTTGAC CTGTGAAGAT GGCAAAAATA GCAACCGCTT 2280

CTGCGCCCAT CACATTTCCT GGAAATAAAC CAAGTAACCA AGCAGTTGTA AAGGTCAAAA 2340

AACCTAGCAA TGGAACAGAT TCAAGGAAAT TAACTAATGG TAAAATGACT CTTCGTGCCG 2400

TTTTATATTT TACGGCTAAA ATACCAAAAA CAAAGGAAAA CAGGAAGGAC CAAGCCATAC 2460

CCACAATCAT TCGAAACATG GTTCGAATTG AAAAGTAAAA AAGATTTGAT GGAGATAAGA 2520

AATCATCTGT TTTACTTGGT GTTGCGAGGA ATTCTTCTGA GGGAAGGGTG TGGAGAGGGC 2580

GATTCATTTT TAGCCAAGAG AAGACGATGA TAAAAATGAA AATGCCTAAA ATGATGAGCA 2640

AGGTAAGATT TCCGATACTT TTAGCAAGTA GGGTATTTTT CTTAGATGAT GCAGTGACAG 2700

CTTTCAATTT AAAAACCTCC AATTAATCAT ATTAAAGAGG TCGCCGAATA AAAGAAGAAG 2760

ATTATAGGGT AATCACAATA ATTGAGCTTT GGCTTTCTGA GACGTAAGAA AAAAATTCTA 2820

GTTATCTTAG GAAAGCCTTA ATTCGACAAT CCCTGCTTTA CTCGTTACTG TCTCTCGACA 2880

TTTCCGAGCG ATAACCTGTG TTCTTCAGAG TAACCTCACC TAATATGATT CTATTTGATT 2940

TACATGCTTA TCATACTCGT GACTAAAAAT AAAACAATTT CTAGCTGGTT ATACCTGAAT 3000

TACTGTTTGT TGTTATTCTA TTCaATGATT CTTAAAAAGA TTGCCCaCAT TCTGTTTTTT 3060

TAGGGAGAAA TTATCAGATG AGTATGTTAT GTAACGCAAG GTGAATAGTA TAGGATGTTG 3120

TATTTTTAGT TTGTCTCGTT TACTGTTAAA GATGTAAGTA GTAACTATAA GTGAATGGCT 3180

AAAAAAATTA AAAGTAATCT AGCACCTAAA TTaAACAACA AAGAAATTTG TTCAGAAGAA 3240

AGGAAATTAT GATGATTATT TGGTTAAACG GTGCGTTTGG AGCCGGAAAA ACGACCATTG 3300

CACATGAGTT ACAACAAAAA CTTCCCAATG CAATCATTTA TGATCCAGAA ATTATTGGTA 3360

GTGCATTGAT GGAGTTAGTT CCCGAAGAAA TGAAAGAAAA TGATTTTCAG GAGTATCAGG 3420

AATGGCGTTG CTGGAATGCT CATTTATTAA AACGAATGTC CAAAGAGTCT GGACGGCCGA 3480

TTATTGTTCC AATGACCCTT TATAAAAATG AATATGAAGA AGAGCTGATT GGATATCTGA 3540

GACGAGCAGG CATTGACGTT TATCATTTTT TGTTAGCGGT CGAGAAAGAA GAAATTTTAC 3600

AGAGATTATT GAAAAGAAAT GATGGAACTT TTGAATGGGG GAAAAATAAG TTGCCCGAAG 3660

TACTTGaAGG TTTTCGTCAG ATTCAGTTTA CCGAGATTTT TAGAAATCAT TCAGCTGACA 3720

CGACAGAAAT TGTAGCAACC ATTCTTAATA GAATAACCGA ATAAAAATAA AATAAGGGAC 3780

CTAAGCCGAA ATGACTTAGG TCTCTTATTT TCTACTTCTT AAGATAATCA ACTACCAGCT 3840

ACTCTGATTT CTTCAAACTT TAAGACCGCA CATTTACGAC AATCTATTTT TACACTGTCA 3900

AAATTTTCTT TCAAGTCTTG CATTTTGTTA TTTTTAGCAG TATATTTTAG GTGTACCTAA 3960

AATATATTAG AGGTTAACCT TCTAACAGGA AGGGCCTTTT AAAATTATTG CAAAACGAAA 4020

GAAGGAATTG AATTGAAAAA TTCAGAAGAA CATGAACCAA AGCAAAGACA TCATTTGATT 4080

GAATATGCAA ATGGTCCtTC GCTTGAAGAA ATTAATGGCA CCATCGACGT GCCTAAAAAT 4140

ATGAGTTTTT GGAAAACGTT ATTTGCTTAC TCAGGTCCAG GAGCATTGGT AGCAGTGGGG 4200

TATATGGATC CAGGAAACTG GTCTACTTCA ATTACTGGGG GACAAAATTT TCAATATTTA 4260

TTGATGTCGA TTATTTTAAT TTCCAGTTTG ATTGCAATGT TGCTCCAATA TATGGCTGCT 4320

AAATTAGGCA TTGTTTCACA GATGGATTTA GCACAAGCGA TTCGTGCTAG AACTAGTAAG 4380

ACGTTAGGTA TTGTATTATG GATTTTAACA GAGTTAGCAA TTATGGCTAC AGATATCGCT 4440

GAAGTTATCG GGGGCGCAAT TGCCTTATAT TTATTATTTC ATATTCCTTT AGGTCTGGCT 4500

GTCTTCATtA CGGTATTTGA TGTTTTACTT TTGTTGTTAT TGACAAAAAT TGGTTTTAGA 4560

AAAATTGAAG CTATTGTTGT TGCTTTAATT GTTGTTATTT TTGTGATTTT TGCTTATCAA 4620

GTGGCATTGT CAAATCCAGT ATGGGGAGAT GTAATTAAAG GGCTGGTTCC TAGCGCAGAA 4680

GCTTTTTCTA CATCACATGC GGTGAATGGG CAAACGCCAC TGACAGGTGC ATTGGGGATT 4740

ATTGGTGCAA CAGTGATGCC TCATAATTTG TATTTACATT CTTCTGTTGT GCAAAGTCGT 4800

AAAATTGATC GTAAAGATAA AACCGATATT CAACGCGCTT TACGCTTTTC AACtTGGGAT 4860

TCTAATATTC AATTAACGAT GGCATTTTTT GTTAACTCTT TATTATTGAT TATGGGTGTG 4920

GCCGTTTTTA AATCCGGCAG TGTCAAAGAT CCTTCATTCT TTGGTTTGTT TGATGCTTTA 4980

TCAAATCCAG CGGTTATGAG TAATTCTATT TTAGCGCATA TTGCAGGTTC TGGAATTTTA 5040

TCAATTTTAT TTGCGGTAGC CTTATTGGCA TCAGGACAAA ATTCAACAAT TACAGGAACG 5100

TTAACTGGTC AAATCATTAT GGAAGGGTTT ATTCATATGC GCGTACCAAT TTGGTTACGT 5160

CGGATGGTCA CACGTTTGTT ATCTGTTATT CCTGTCTTAA TCTGTGTTTT AATGACCAGC 5220

GGAAAAAGTA CAGTGGAGGA GCATATTGCG ATTAATAATT TAATGAACAA TTCACAAGTT 5280

TTTCTAGCAT TTGCTTTGCC ATTTTCGATG TTGCCTTTAC TGATGTTTAC AGATAGCCGT 5340

GTTGAAATGG GTGAACATTT TAAAAACTCG TGGTTAATTA AATTGTTAGG CTGGGTTTCT 5400

GTCATTGGCT TAATTTACCT CAATATGAAA GGTTTACCTG ATCAAATTGA AGGATTCTTC 5460

GGTGATAATC CGACAGCGAG TCAAATTACG TTAGCCGATA ATATTGCCTA TGTCATCATA 5520

GCACTCGTCA TCCTCTTGTT AGTTTGGACG GTTGTGgAAT TATATAAAGG CGATAAACGA 5580

TATGCACAGC AGCTTGcAGC TATGGaGCAA CAAGTAGAGG AGGTTAAATA ATGAAATTTT 5640

CAAAAATTaT GGnTtGGTGT TGaAGAGTCG CCAGATGcGT TaAAAGCATT TCCATTATGG 5700

CTATCCCAAA AGCAAAAGAA GAACAAGCTG GATTGGGG 5738

(2) INFORMATION FOR SEQ ID NO: 200:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13865 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

AAGCCATCAG CTGTTTGnTT ACATGGTGAA TTTTGAGACA AGTTAATTGT AAAAGTCCCT 60

AAATTATCGT TGTTCTTCGC AGCACTAGAT TGTCCTTGGG CATCTGTTTG TACTTCAATA 120

TACGAATTAT CTTTTGAAAA AGATACCATA TCGCCATCTT GAAAGCCGAA TTGGACGACA 180

TTGTCAATAA TAGTAGATAC TTCTTTGGCA TCATAAGTTG TCATGCTTTG CATTAGTTAG 240

TtCCCCCtTT AGACTTCAAT TGTTCCAGTA ACATCAACAG TGTGAATTGC TCCAGAACGT 300

TTGTATTTAA ATGATAATCC TTTATAGTTT CGTGCTGCAA TATCATCTGG ATTTAAATCT 360

TGACGTCCCA AAGCAGTCAC ACTATAATTT CCAACACCAG TTTCATCTAC AATATCTACA 420

ATTCCGTTAT TAAATGCAGT TTCCAAAACA TTTGCAACAG TTGTATCCAA TAAAGCAATC 480

CCATTAGAAT CAAAAGTTAA TTTATCTGTC GTTGATAACA AGCGCTGAAC ATTCGTTTCA 540

ATGTTTGATT TAACCCAGTG ATCCCCATGC AGGGCATCGA TAAATTCGCC ACTCATTGTT 600

```
TTTCCTTCTG ATGTTTGGGG GATTCCCGCT TTTGATACAT AAGCAATAGC ATTAGCTTTC    660
TCAATCGCTT GTAGTTGAGA GGTAGTTAAT GTGTTAGGTG TAATCCCTAC TAAATTATGT    720
CGGAATTTCC AAGTAACACT TCCGACTGTT AAGCTTGCAG TATTACCAAT TAATGCAGCA    780
TCTAAAAATT CTTCTAACGG ATGAACCAAG CCGATAGTTA GATTGTTCCC TGTAAAAACT    840
GTTAATTCGT CAACTGTAGC TGTCTGAATT ACTAAAAACT TAAATTCGTT TTCTTCAATT    900
AAATTTGATA GTGCCAAAGC ATCTGCCTCA ACAAAGTTGG CTAATAATGC AAAATGCCAA    960
TCATTATAAA AGTAACTTGT AGCTGCGGCG ATAATTCCAC TTCCAGCAAT TAAGCTATCT   1020
GGCTGTTCTG GACTAGTAGG GGTATCTGCT ACAAAAGTTA CTACTGgCAA TAGTTTGTGG   1080
TTTATTTTCC TGTTTCCAAA TTATTTCTGC TTTTTTATAA ACAGTAGTTG TTGTAGCAAA   1140
ATCTTTTGCT AATGTTTCTA AACTTGTATA TTCTTTATAG TTTTGAGTAG ATCCTTGAAC   1200
AAAAATTGCT GGATTTCCTA ATCCAACAAT GGGTTGTGGA TGCATTATGT CAATTTTTAC   1260
ATTAACATCT GTAATTTTTT CAATCATTTG TTTTCCTCCT GTATGTCTAT ATTTTCTATA   1320
ATCACAGCAT TGTCAACGTA GCTGTCTTGT ACTCTAAAAC GAGCATCAAA GCCAGCCAGA   1380
CGTTCATAAT CAATACTAAT GAAATTATCT CGTTTTTGAG ATTGTGTTGT TGAAACTAAG   1440
GCTATTTTAG AGTTTTGAAG GTCTACTTTT ACAGAAAAAC CGTTTAAATG CTTTCTTAAC   1500
TGCTCTGACA AATTAAGTGC CTGAATACTA GAATCTGTAT GACATTTAAT AGACACTACT   1560
AATTCAAATA CTTCATTATC AGTGACATCT ATAGTCACTG GAATATATGG CGAGATGACC   1620
TCATAAGAAA AAAACGGTTT ATCTGGCTGT GGACCAGCTG TGCTACTTTC AATAAGTTGA   1680
AGTCCAGTTG ATTTCTTAAC AATTGCTATT AGCTCATCAG CTAAAAGCTC ATAACTAAAA   1740
GCCCCATTAA TTTGCGGCAT TATTGGTCAC CGCCTTTAAT CTGTAAAGGT TCACATCTGC   1800
ATAGTGTTCC GCATATGGCT CTTCGCTTTC TACTTTGTAC TCTTTTCCCA TGTCAATAAT   1860
TTTAGAACCT AAAGGAATCG ATCCAACGTA AGCAAGTTGC CTGTCGCTAG ATGTGAGTGT   1920
TCCACCACTT TGATAAATCG TCCGATTGTC GTATGGGATA ATCGCCCCAG AAACGGTTTT   1980
GGGTTCTTCG TTAGTAGGAA ACCATTCGCC GTGAATGTAT TTACCAGCAC CTTCAACTGA   2040
AGGTAACACT AACTGAAAGT CAACCGCAAA AGTTTCTATA AGACTAGAAA AGTTCATTCT   2100
TTGCATTACC CTTTCACCAC CTTATAAGTC ACTGACTGTC TTAATTTTCC TGTGTCAATT   2160
AATGGATTAC TTGACTTCTT ATTATTGACT GTTATTGGAG AATTTGGCGG ATCAGATAAA   2220
TTTCTGATAG TTCGTTGAAT ATCCCTTTGC ATACGAGCAC CAACTTTATT TGTCAATTCT   2280
TCTGCTGTCA TTTTTCCAGC AATAAGTTTT TTCACAAGGT TTAAGGCGTA ATCCGACCAA   2340
GCCTTTGCTT GCTCGTCAAA TGTAGATCTA ATAAAGGAAC GTTCTGGAAT AGTTATTTGC   2400
TCCGCTAACA TGTATGCGAA TTCTAACTGA TCTTTACCTT TATTTCTTAC TAAAAACTTG   2460
TGATTTTCTT TAGTTTGCAT AAAAAACAAA TCAAATTGAC GTGGGCTTTT ACCTCTATAC   2520
TTTTTCATAA GTGGTATAAC AAGAAAACGA CCTTTGGGAC GAATAGTCAC ACCAAATTCA   2580
```

TGAACTTGGG CCAACATCGC AATAAAAGAA TCATCTTCCC CAAATATTCC TATCTGCAAA 2640

GAATATTTAT TAAGCTGATT TAGTTCATCA ATCAGCTTAA TAATCCTGTT ATTCTCAGTA 2700

ACTTTCATCA AATCACAACC AaaCTCAATG AACCTTTTTT TGCGTATTCG TTGTAAAGAC 2760

GTAAATACTC TTGACCATAA ACGGTTCTTT TAAGGTCGGT AAAAGTTGAG TGAAAACCCG 2820

AATACTCTTT TTTTAGCGAG CCTACTTGCT CGGATTTAGT ATTTTGGTTG TTCAAAACGG 2880

CTAGATGGCA AGCGAGATAG CGACACGCTT TCTCCTTAAC CTCTTCTTTA AATGGCAATG 2940

CATCTACTTC TAGCCAAGCA TCATCAATAA ACAACTTAAT AGAATCATTG TTAACCCCTA 3000

CTAGTTCTGC AGCTGTTAAC CTAACATTTT CAACTGTGCT TTTTGGCATA GTACCACTTC 3060

CTATTCTGGA TTAACTACGC TGTCTTCATC AGGGTTTTTA ATATCATCAA TTTGATTTTC 3120

GATTGCTTTG ACTACAGTTG TGCGATTTTT GTTTGCTTGC TCTTCTTCCA ACCATTTTTC 3180

CAACAATTCT AAATCGAACG TATCAGCAAT GGATTCTACT GCTTTATTGG CACTCAATTC 3240

AGTAAAACCA ACTGCTTTTT TTGATTTTCC TTTTGTTATA TGGTCCAAAA TCTCAATTTC 3300

TCCTAATCTT TCCAACGATT TATTTAATGG TAATTCCATA CCTTTAATAA ATTCTTCAGC 3360

ATCTGAATTA TCTAAATCAT TTACACCTGG AATTAATCGA ACATTTCCAA TATGTCTAAT 3420

ATATGAACCT TTATTGTGTw CTAACATAAT CTTTACCTCC TTATTTTTAA ATACCATCCA 3480

CACGAACAAT TGCATATGGC GTTCTAATTA ACGCACCACC ACAACGTTCT ACAAATGGTA 3540

CTTTCCAATT TGGAAAAGAC CATTCAACCT CTAAACGCAC GATGTCTTCT GGAAGTAAAA 3600

TTTCACAAGT TGATGGTTTT GTATCCATGA TAATAAATGA ATCAGAATTA TCAGTACCTA 3660

CACCTTTTAA ATCATAAACT TGCTCAATAG ATGAAAACCA GCCATTTTCC TGAACAaCTT 3720

TCATAATTGA TCGGGCATCA TATTCACCAT AACGACGATT CAATTCTTCA TATTGTTCTG 3780

GAGCAACCAT TAATTTTAGA CTAGATCCTT TAAATCCTGG AATAATAGTA ATTTTAGCTC 3840

GAGATGTACG TAATTGTTCA ACAATCTCTT CACTAGTCAT TTCTTTCCAC TTCTTAGGTG 3900

AATTAATAAC TTGAATACCT TCAGCATTCG CAACACCTTT ATGGTTTACT TTAGGATCTC 3960

CCACAAAAAT AAAGCTATTT TCTTTTTCAG CAATAGTACG TCGTGCCACT TCTGCTTTTG 4020

TTGCATCAAT TGAAGTTCCC ATCATTTGAG CTTGACGAAT TTCTTGTCGA CTATAACGAA 4080

TACCAGCAGC AATTGTAAAA ATTGGTGaTT GATAACGtTC aTaTCAAtGk CAACTAAAGG 4140

AAGATCATCT GCACCATTTG CGATAATTTT TGCAGCGCCA CTTCTAGTCA TAACATTATA 4200

TGCGTATGTT TCCGCTCCTG GATTGATAtC TGTTTTAACA TTAAACATGG TTCGCGCCAC 4260

AAGTTCCTCC TGTGGTGCTT GATAAATGAC CTTATCAATT TCTTGTAGGT CACGTGCTTC 4320

TAAAGTTGCT GTTACATCAT TTCCCATTTA GTTGTCCTCC TAGTTTTTTA AGGTAAGTTA 4380

ATTTGTAAAA CTGCTAAATT ACCAGCTGAT GCTGTTGTTT TAAATGTACC AATCACGGTA 4440

TCTGATGGTG TCGTAGCTGG ATCAGAACTA ATAGTTGCTT TTCCAAAATT TCCAGTTGAT 4500

AAAGCTTTTG CATTTTCACC TGCTAAAACA TCTTCGTCAA CTTTCACCCA AATAGCCCCT 4560

TTACGTAAGA TCGGTACCAT TTCGTGTTCT TTATACTTAC CAACTTTTTC CACATCTCCG 4620

TAAGGGATTT CTTCTACATA ATTTTTAGCA TATGAAATAC CATAAAATTG CCCATCTTTA 4680

TATGTGGTAG CAACGTCTTC TGTTACTTGA ACGCCGACAC CGAAACyTAA GCCACCGACA 4740

CCAACGACAA GACTATCTGC TTGTACACCT TGATAATTAG CTAATTTCCC AATACCAAGT 4800

TCAGGCTTCA TATACTTTTC TGGATAAGGA ATAGTCATAG TTATTTATCC TCCTTTTTGT 4860

TCATGTTCAA ACGGTTATTT TTTTGTTTTT CAATATCTTC CGCCGCTTCT TTATCTTTAT 4920

CTTTTGCATC GTTGAACGCT GCTGGATTTG TAAATCCTTT TTTCTCCACA TTTGCAACTG 4980

CTGAATCAAA GTATGCTGCA ATATAATCGT CAGATTTTCC ATCACCTTTA AAGTCAGGTG 5040

ATGTTTTGGC AATTACTGAT TCTTTAATTT CACGATCAGA TTTGCCAGTG AAATCAACTG 5100

AATCACCAAG AAACTTTTGT GCTTTATTAA TTAGATCAAC ACGATCTTGA ACACGCTTAT 5160

CCAAAGCATC TGCAGTTACT TCTTTTTTCT TTGCTtCGTC AATTyCCTTA TTCAACTTAT 5220

CAATTTGAGC TTTTAAAGCA TCTCGCTCAC CTTCTAATTT CTCAACGTTT GCTTTACGTT 5280

GTTCGGCTGC ATCTAATTTT GCTTCTAAAG ATTCAAATTT TGCTTTTACA ATTGAATCCA 5340

CTTCAAATTC CTTTGAATCA ATAATTAATT TAGACATTCC ATTTCCTCCT GTTTGTTTAT 5400

CTTTAGTATC AATCATGAAA GCGACTGAAT CTCCCCGAAT AGCAACTTCG GGACCAGCTC 5460

TCCCCTCATC TACTATAGCA ATATGATTAA TTTGCATATT CCTTTGCACA GAATCATATT 5520

GCATGCCATT ATACACGCCA GTTTCTTTTG AAACATCTGC TTGAAAACCG ATGCTCAGTT 5580

CACGTTTTCC ATCGTTAATT TTCTTTATkG TCTCTGCATC TGTAATtGTG AACGAAACCA 5640

ATAGCTTATT ATCTAAAACG TGAGCATCAT TATGAGTCAT GCCTTTTGAA TATTTATTGT 5700

AATTAGCTGC TGTTACTGGT TCGGTCGGAT GATCATCAGT CATTGGCTTA GCATTTGCTG 5760

AAAGTACAGT TGCCTTAGAA AATAATTCAT CAGGCAATTT TGCCTCCATT GATAATCCAC 5820

CATCAGTTCG GCGGTAAGGG AAAACACCTG GTCGTGTAAT TGGACAAGCT GTGATTGTCA 5880

AATAACCTTC ATCAGTTTCT TTAAAATCCT TAATAAAGGC TTTGTCATAT CTAATTACCA 5940

TGTCTTCACC CCATTATTCG ACTAATAAGC GAATGcCTTC GGTTACTTGA TTACTCGATT 6000

CTAACTCATT AAGATCCAGT AATTTTCGTA GGCTCATATG ATAGCGTGTC GCAATTTCAC 6060

TAGCAGTTTC TCCCTCAGAA ACAGTATGAA TCGTTTTCTG CGTAGTTTTT GACAACTCTT 6120

TTTTGTTTTC TTCTTTACTA GAAGCTTTTT yCTCTTTAGC CATTTTGATC ACCTCCTTCa 6180

GkTACATCaA ACCTGATATT tCTCCACTtT TTATAAGCAT CGAAATATAG TTCATTTTTA 6240

TCACCGTTGA AGGTAAyTCG TAATACATAC CATCAAATAG ATTGGTACTT AATAATGCTT 6300

TATTATTTTG CAGGGTTTTC GCTAACCAGA CCACGTAAAT ATCAGACTCG CTTATTTTTT 6360

TTTGATCTGT GTAATCCAAT TGACTGTTAG CAAAATCAAC TACATGCTTT TTACATAAAT 6420

CAATAAATTT TTGACTATCC ATTAATGTCC CTCCTGTTTT TTGAGTATAA AAAATAGCCC 6480

CAACCTATAA GGTCGGTGCT ACTCTTCTTT TTCTTTTAAA GATTTTTTCA ATTGCTGCAT 6540

AAAATCTTTA TCTTTCTTTT TCGATGTATC TATTTCCTTT TTTGTCATTT CATCTGGGTA 6600

ACCTAGACTT TTTGATTTGG GAACGTCTTC CCATCTACGT TTGTTTTTCT CATTTTTATT 6660

CATGGTACTC CTCCAGTACT ATAAATGGTT TTCCATCCaA TAAATATCTG TCTTTAACAA 6720

TAAACTTGGc ATCACGATTA AACAGAACTT CTTTTTCCGA CTTGTTGTAA GAACCTAAAT 6780

CTTTTCCAGT TGAGCTCATT ATAACAAATC TTAACGAGTC TTGcTCTGAG TAAATGTCTT 6840

TTGTAGTCGA AATGTATTCA GGAAATTsAA CAACATCATT TAAATTGTAG TTGTTAGCAA 6900

ACCTCACTAG GTCATCACTG CTATCAAAAA ACATGGAACG TGTTACTTCA CCATCATAGT 6960

TGGTCATTTT GCCTAACGCT TTGTCTAAAT TGTCTATCAA TTTCAAGTCG CTTTCATCTA 7020

ACTGATAACC ATTTCTGAGC TTATCGTTTA ATTTGTAAGC TTCAGAGCTA ACATATGTCT 7080

TGACCGCATG CTCTTCTTGA TTAGAAAGGC CAAGATCAAT GCTTTCTTCA AGCAATTCTT 7140

CATCGTCGAC TGGTTCAGCA ACACAGCGAC AATTATATTC TTCGCCTGGT AATAATGGAT 7200

TATCAGCATA AAAGTAAATC TTCCCGTTAC GTTCTCGATG AGAATCCCTA ACTCGTTCAT 7260

CTCCACTATC ACTCCATCTA AAAGCTTGAA ATCCCGCTCG TTTTTGGCGT TCACTATTCA 7320

TCTGACCTAA AATAGTACCT GTCTGATCAC GAGCAATAAA CGCTGCTTTG TCTGATGACA 7380

TACCAGCTTG ATGAACAAGC TCTTCTCTTA TCTCGTTTGA AGATTTGCCT TCTGTGATTC 7440

CGCGATAAAT AACCTGTTCA AACTTTTTTG AGTAATCATC ACGAATATTA GTGACATAAC 7500

TGATGTTTTC CGCTATTTTA GCTTGAACAT AGCTGTCTAA CCATTTTTCA GTCTGTAGTG 7560

GGTTTATTCC TCTAGCGCTA AGTTGAGAGT TCACATTAGA CTTGTTAAAT GCGTTGACAC 7620

TATTAATATG CTTACGAACT ATTTTTTGTG CGGTTCTGTT TTGAAGAATA CCTAAAAAGT 7680

ATGTTTGCGC ATTTCTGATT AGTTTCGATG CGGCATCGAA TAATCCATCA TTAATTAAGT 7740

TGGAaTCATT AACAATATTT TTATCATCTA TCATCGGTGC TAGATACTTA TCAAATTCAT 7800

ATAATGAAAC TTTTTCTATT TCTTTTACGG CCTTCTGAAT GTTTTTAGCA TAGCTTTCTT 7860

CTAAACGTAA CGGATATCTC GTTTTTGGAT TATTTTTCAT TATCTCTGTC CTGTTTGTAA 7920

TTTTCATAAA CTACAGCAGC CATCTTATCT AACTCATCCC TGGACAAGCT ATCCGCGTTA 7980

AATTTAGAGG TTTCTGTTAT GCCAAAACGA CCAAAGCGAG CTTCGTGAAC ATCATCAGGA 8040

TCAGATACAC CTGCCTCAAT GTAAATCTTA TCTGTTTCAG CAGTGAGTTT TCTAATCTCT 8100

GCATCTGTCT TGCTATCAAC GTTCCATAAC GGATTGAATT CAACGGACCA TTCGATTGAA 8160

TCAGGGTCAA GTCGACCACC ACATTCGTCC TCTGCCCACA TTAGACACCG CATGAGATAT 8220

TCAAGGTGTG GCCTTAGCTG ATTTTCTTGT ATAGCAGTTA TACGAGAATA GTAGTTCATG 8280

ACATCATATT GTGCTCCAGT AACAGTTCCG CCCTCTTGCC CTTTTAAAAC AGTCTTAGGC 8340

ATTCGAGCAG CGCCTGCTAA ATAATCCCAA ACAAAGTCTA AAAGTTCACC GATACCTGCT 8400

ACCGAAGAAC TCTCTTTmCC AAGACTTTCy TCCtTGTCtA TGATAGCAAT TGCTTCTGTT 8460

CTGAATTGAT AATCCAyCTT AGTTTCAATT TCTAGTTTGT CTTGTGGTGT AAGACTTCTA 8520

```
ACATCTGCTG ACTTGTATAC TTTAAAAATA AAATCATACA GAATTTGTCC GACTGACCAA    8580
ACAGATGTGT CAGCTACAGT TAAGATATCA TAAAGATTCT CTAATAAAGA TGATCCTTCT    8640
AATTCATCTT CAAATCTCAA ATTTTGTTGA TGTAAAACTC GTGACCTATG AATTGTTGTA    8700
GCTCTGCTGT AAGTCGTATT ATTTAGAAGG TGTATGCGAC TGTTGTTTGA TCTATTATTG    8760
ACTTCAAAAG ATTCAATCTG TCCATAGTTT GGACTAAAAA CATCTTCATC AATTATTCTA    8820
TTACTAATCT TTTTGCCTGA AAATGCATTA ATATACGGAA CACTTTTAAT ATTCTCAAAG    8880
TCTAATGGCT CACTTAGAGA GTACTCTCTT TTCTCTATTG TTCCAATACT TACGAATCCA    8940
TCTCCATATA ATCGTTCATA CGTGAATAAT TGTTGTAATC GTTCCTTAGC TTTCAGTTGT    9000
CTCAGCTTAC TTTCGTACAG AGCTTTTACT TTATCATCTT GCATTTTGAG ACTCCACCCA    9060
TTTCTAGTTA AGTCTTCTGC TGGAATGTCT ACGATATTTT TCGCCATCGA GTTAGACGAG    9120
TAAAGCGACT CTAATTGTGA GCATGACAAT CTCTTGCTCG TTCCTGGTCT TTGTCTAGAT    9180
AAGTTGTCTC TAGCGTGGCC TTTGCCATTT CCAAGCATAA AATCATTTCG ATATGCTTTA    9240
CCATCCAGCT TTAATAACTT AGCTTCGTTT GCTATATTCC CCATTTATTC ACCGCCTTTC    9300
AATTTTTTAA TTGTCGAGAT ACCTTTCTCT TAAACTTACT TTGCCAATCA TTTTGTTAAG    9360
AATTTGAGTC ATGCTGTCAA CGTCGTCATC ATGTTCAGCA TTTGGGAATG CTACTAGTTC    9420
GTTAATGAAA TCATCTACCC AAGGACAAAC TAATGGGTGC GGAAGATAGA CATTACCAGC    9480
TTCCCATATA GGAGCAACAG CGTTTGcCCG TACCTCTTTG CCGCCTTCTG GATTTACTGC    9540
AATGATACCG CTCAATTCAT TTTTAAGCAT TTCAATTACC GCCGTCCCAT TAGCTTTATC    9600
TTCaATGTAA ATACCGCGAG CTTTTGGCCA CTTAGCAGCC ATGGCATCAA TTGCTTTCAT    9660
CGTCTCAACT ATGCCCATAC GCTCATGATG ACGGTCTAAT AAATAAAAAT CGGCTCTCTT    9720
CTTGCCCCAT ACCTGACCAG ATACATAGTC AGATGTTTCG GTGTCTTTGA AAGTACAGTC    9780
CCAAGACTGA GCTTGTCTAT CAAAAAGACG TGGTAAAATA ACCACGTCAT CGCCTAAACT    9840
TAATTCAATT TTTTTGGATA ATGTTGGCAC ATAGAACTTA ATCCATGAAC GTTTGAAAAT    9900
GTTACCACCT GCTGGCGTCG GTCGTTGTTG GTACAAAGCA GCCCAACCAC GAGAACCAGT    9960
AACTGCTTTT GTTTTAGCAG CCCACTCCTC ATCTTTCCCA ATTTCAGGTG cTAACGCTTC   10020
TCCTGGTTTT CTTCCTAGTA AGTCGTCATC TTCGGCTATA GCTGGTATTT TAATTTCTTC   10080
CCAAGGCAGT GTTTGTTCTT TTAACAGACG ACCTGCTAAA TCATCCTCAT GCCATCTAGT   10140
CATGATTACA ATGACACTGG CATCCGCCGT TAAACGAGAA TAGAAAGTGT CTTGCCACTC   10200
ATTATATATT TTATCGCGAA TTGTTTTAGA TTCAGCTTCT GCTCTATTTT TTATCGGATC   10260
ATCGATAATT AAAAGCCTTG CTCCACGTCC TGTTGCACCA CCTAAAATAG AAGTACTGTA   10320
AAGCGAGCCT AGATGACCTT CAACTCCCCA TTCACTAACA GATGAAGTTT CAGAACTTAT   10380
TTGTAAGCCG AATAATTGGT CCGAAAACAA GCGGAACTTT TCGCGGTTCT TTCTACCAAA   10440
CTTTTGGAAT AATTCTTTTG AATAAGAAAC AACCATCACC AGACTATCTG GATGCCTCAT   10500
```

TAAATAATAA GCTGGAAATG TCTCTGTAAT TACAGTAGAT TTACCGTGTT GGGGAGGTAT 10560

TTCAATAATA TAATACTTTT GCTCCCCATC AATTATTTTT TGAAGTCGAT CTGTAATGTA 10620

TTTCTGATGT CTTAATAAAT CCCATTGTTT ACCATGTGAC AAATAGAAGA AATCTCCATA 10680

ATTTCGTCTA GCTAGTTCTT CCAATGCAGC GTTAGCTAAT GCATCAAGTT GCTCTTGGTC 10740

CATTTGCTAA TCGCCTCAAT TCTTCTTCCG ACAAATTCGC AAGCGGATTA ACATCCACTT 10800

TTCCGCTATG CTGTATCTGG TCAATTGCTT TAAAGCCACC TCGGTCTAAA ATGTCTTGGA 10860

AAATAGATTT CTTTAGACTC TGTAATTTTT CCCATTCCTT GGCATCAAGA TAACGAGATT 10920

GATGCTTTGA CAATTCATTT AGTAACACAA CTCTTTCAAG ATTCAGTTCA AAATATTCTT 10980

TTTTTATGGA GTCAATCTGC TCTAAAAGAC TTCTTTTGTG GnTTCGCATT TCTTTCTTTC 11040

GACCATCAAT AGCATCTGCT GCCTTTTGTA CTTTTTCTAA GTCGTTATTA GCCTTAAGCA 11100

TTTCGAGCTC TAAATCGCTA AGTTCAGACT TTATTTTATC AATTGCTATC TCAGCTTCGT 11160

TGTGATCTTG TATTTTCCCT TCAGTTTGCA TTGCTATATC TAACAGCATT GAAAAACTAC 11220

GCAATCCCTC GTCCTTCATA CGATCGCGTA GTCTTTTCAT TTCCAATTGA ATTTTTTCTT 11280

TAATGTAAGC ATTTGTTAAT AGCTTGTGTC CTTGTTGTCT AGCTGTCTTT TGAGAATAAC 11340

CCACTTTAAT TGCTGATTGT GTTGCATTAA AAGTTTGCAG ATAACAGTCA ACAAACAAGT 11400

CATATCTCTC TTTTGTTTTA TTTGATGGCT CTTTTTCACT TCTTGCCATT GACCGCACCT 11460

CCCCTAACTT CTAAGTATTT TATCAGCTTC AATAAGAGTT TTTAGATCAC TAACTGAAGT 11520

TAGTTTGATT TCGCCATTTT TAAGATTTTT TAACCATTGG GCCATCGTGG CTCTTATGAT 11580

TTTTCGGTAT TGTTCATAAT CTTCGGCATC TTCAAACTCT TTCTCTAATT CCAAATCTAA 11640

AATATCAATT TTTTCATGTT CTTCTTTTGT TCCCATTGTT AAAACACCCC GCATTTGATA 11700

AAATGCTAAA AGACACAGAG GGTGTCGAAA ATCCACGCGT GGGAATTCTC TGTGTCTTCG 11760

GGGTGTTTTT TGCGTCCTGA GGAATTAGTC GAGTGTTAGC TGCACTCGGC TTTTTTAATT 11820

TATTCTACAA GTTCTGCAAT GATTCCTGTT GATTTTTCAA AGCGTTCTAT GATTACATCA 11880

CAGAAGAGTG GATCTAATTC AAGCGTATAA CAGATACGAT TTAACTGTTC ACAAGTCATT 11940

AATGTACTTC CTGAACCACC AAATAGATCT AAAACAATAT CTTGTCTTTT AGAGCTATTT 12000

CTAACTGGTA TCGCAATAAG TGATAACGGc TTTTGTGTAG GATGATAGTA CGTTGCTACA 12060

TCATCTCTTG GTACTTTCCA AATAGTAGCT GGTAAATCTT CTAATAGATC ATCTTGCCAA 12120

ATTGTAGTTT GCTTTCTGTC TCCGTACCAC GAAGGTGCCT GTTTCTTCTT GTGGGCATAA 12180

AAAACTGGTT CATGTTGCCA TCTGTACTGG CTCCAACCAA ACGTAGCATT ATTTTTTACC 12240

CATATACATT GTGAACGGAC AACTATACCA GCAGCATTCA TACTATTTTC AAATTCACGT 12300

TGATAAGATG ATCCGtGAAA CACATAAATC GCTGAGTCAT CTCTCAtTGC GTTAGAATAG 12360

TTTTGAAATA CTGACATCAA GAAnTGGTCG AACTCTTCAT CACTCcATGT CATCGTTCAT 12420

AATTTTTTtC CCGACCAGAT TCGTTTAATt CTTTATTGTC AGATTTTACT GCTACATTAT 12480

AAGGTGGATC AGTCACAACT AGATTTGCTT TTTyTCCTkG TAATAACTTT TCaACATCCG 12540

AAAGCTTTGT CGCGTCGCCA CACAATAGAT AATGATTACC AAGTTTCCAT AATTGGCCCA 12600

ATTTAGTTTT AGCTTCATGA TGATTTTCTA TGAACTCATT AACTTGAAAA TCATCTTCAA 12660

TAATCGGTTT CTCAATATCT TCTTCATAAT TGaAAGAaGC GAGCArGCTA TCTACTTCCT 12720

CAGTATCAAA mCCAGTCAGA TTAACAGCTT CATCATcTAA TTCATTTAAT AAAATAGAGA 12780

GTTTTTCTTC ATCCCcACCG ACCAGAAATC TTATTaAGAG CCACGTTGAG CGCTTTTTCT 12840

TTATCAAGAG GTAAATCTAC CACGGATACC TCTATTTCAT TGAACAAGCC CAGTTCTTTG 12900

GCAACAGCGA CACGCTGATG TCCACCAACT AGATTGCCTG TTTGAATATT GTAAATAGGC 12960

GGATCTACAA ATCCAAATTC TAGGATAGAT TGTTTTAACT TCTCATACTC TTGCATTCCT 13020

GGTTTTAAAT CAACTCTTGG ATTGTAGTCA GCTGCTTTCA GATCAGATAA TTTCATTTTT 13080

TCTATTTGCA TGTTTTTTCT CCTTTAAACA AATAAAAAAG CCTAATTAGG CTTTTTTAGA 13140

TAATAGTCAA TAAGTATTCT TGTTCAAGAC AGTTTTTTAT TTCAATCGCA GCTATTTCTG 13200

CTTCTTTATA TCTTTTTAAT AATTCTTGAT ACTCATCTAT AGCCAGCTCC GTATATGCTT 13260

TAGGGTTAAT AGATAAAATA TAAGTATTTT CTTGATTCGG CATATTAACA AAATTAATCG 13320

ATACCGAATC ATTATTTTCT AAAATTACGT TAATTTCTTG TTCGTTAATC ACTACTGTAT 13380

AATTCTCATT TCCAAATTGT TTCATCTTAT TCCCTCCAAA AATGTAATAA CCCCTCATCT 13440

AGGTGTGTCT ATATAACCAT ACTATTAGCT ACTTGTACAG TCATCATAGA ACATTTATCA 13500

ACTTAATTAT GGTATATGTG ATTTGAAATT ATTCTGAATG TGATAATCCT TAAAAAAGCC 13560

ATATCCAATG TAAATTAAAT CTAGCTTTCT AGACATTTTT TTACTAACAG GTCGTATACA 13620

TTCAACAATT GsAAAGTCAC GATTTAATTT TACTTCTCGA GTTATTTTTG GkGGGTAATT 13680

TTTACTTAAA CTTATGTwCC AGTAATTCAT AATGTaTCCC CTATTTAAAT TTGTGTGGAA 13740

CGCCTAGCCG AATCCACACk GAATAACAAT ACTAGGTTGC TCGCCTTTTC CTGTTTCCGC 13800

AGGCAGGCAC TTTGAAAGGA AGGAGTGAAA ATCGTGTGTC ACAGTAGTAG AAACATTTAT 13860

TGACA 13865

(2) INFORMATION FOR SEQ ID NO: 201:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5154 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

GGTATCTAAC GCCCAACCTG TAATGGTTAA AATTATTGGT GGTTTGGTCG CGGATnGATA 60

CTATCCAAGG ATCCACnTTA ATGTCTTCaT TCAGnGCaTT CtCCtTAaGA TGTTTCGGAT 120

TTCGGTAGGG ATCcAACTAA AagCTTCCGC GGCCCaTTTT TCCGcAAGTc TTTATAAGAT 180

CCTTTTTTAT CATGGTTTTG CATTTTTTCG TCTGCTTTCA AAATTGCATA GGACACATAT 240

ACATTTCGTT TACCAGCGCT AGCAAATTGT AACGATAAAT CGACACCCAT CCATTCTTTT 300

CCTAAGCTTT CGTTAATTCC TTCTAAATTT AGGTAAAGTT GTTTGTTGAA TAGTAAGCAG 360

TCTTCAGTTG CTGCTTGAAT ATTTCTTGGT AAAGCAATTC TGTAATAGTA ACCTAAGCTT 420

TTACCAGGGG TGCCTTTTTC AGGATAAAGT AATCGACCTT TGTCTGTATC AACTGATACT 480

CCCACTGTTT CCACGCGATA GCGTGmtCaA CTACGCGGCC AGTCACTAAA CCAGAGGTCT 540

CTTGTTGACC AATGTTCATT AACTCTTTTA ACCAATCATT TTTCGTTGGT ACCAATCCAG 600

CATTTAATAA AGCCACATAT TCTCCCTTAG CAGCTTGAAT CATGCCATGA CGAGTTTCTG 660

TTGAAACATA ACGAAGACGG TCATTTTGGA TATTTATTTG ATTCTCGAAG CGTGCAGGCA 720

ATAGAATTTC ATAATTTGTA TAGGCAGTTT TTTCTAATAA TTGACGTAAG TAGCTACTCA 780

TATTCTCTGT ATCATTCGTA ATAATCAATG ATACCATCGG CACATGATCA TAAAGATAGT 840

TAATTTTATA GGAACCATAA AATTCGGCAA TCTCTACTTG TGCTTTTAAG cCACGGCGTT 900

CCGTCGCCGC TTGAACCGCT TTTTGACCAG CCACATAGGC GTAACCTTTG CTTTCAGGGT 960

TTAGAGCAGT CGATGATTCA ATGGCACGCC AATGATACAT CATACCAGGA ATGTGCTyGA 1020

TTTTCGTTGC TTGTTCCGTT GCTCGCAATA CGAAATCATA GTCTTGAGCG CCGTTATATG 1080

CTGAGTTCAA TCCGCCTACT TTTTCTAATA AATCACGTTT CACAACAACA AAGTGTGTAA 1140

TGTAATTATG ATTTAAAATC AGTTCTGGAT TCCAGTCCGA TTTGTAAAAC GCGTTGAAGC 1200

GACGACCATT TTCTGTAATT TTATCTTCAT CAGTATAAAG AAAATCAATC GTTGGGTCTG 1260

TATTTAAAGC TTTGACCACT TCATATAATG CTTGTGGCGC TAATTCATCA TCGTTATCCA 1320

TGAAACCAAT AAAATCGCCA GTAGCAATCG ACAAAGCTGA GTTAGTTGCT TCTGAAATAT 1380

GTCCGTTTTC TTCGCGATAA ATAACTTTGA TTCGTTGATC TAGTTCTTTA TATTTTTCTA 1440

GCATTGGCTT AATATGTTCG CTTGGCGATG CATCATCTGC TAAACATAAC TCCCAATTTT 1500

CATAGTACTG ATTTTGTAAA GAAGAGACGC ACGCTGCAAG CCATTTTTCT TCGACATTAT 1560

AAACAGGCAC AGCGACAGAA ATTTTAGGCT GAAATTGGAA CGTTTTAATT TCCGCTAAAA 1620

CTGCCTCCAC GTCAACAGGG GCTTCTGCTT TTTCTTGTCG TTTAAAGTAC CACTTCACTG 1680

AGCCCATCAA ACCACGTGTC TGAATCGAAC GCCATGCTCG ATCAAATTTC GTGCCTGGTT 1740

GAACCATGTT TGGACGCTCC ACCACGTATT GAACCTTCTG GCCTGATTTG GTGTCAATGT 1800

TTAACGTGTA GGCTGGTAAA TCTTCTGGAT CTTTAATCGT AATTACAAAA CCATAATCTT 1860

GATCGCCTTC TAATTCAAAC ATATCAATAA CATCACGACG GTATTCTGTT TCCACCGTAT 1920

AAGTGACTGA TTCACTAGCT TCTGCAATGC TGAATCTTAG CGGATAATTT GTTTGTTTGT 1980

CATACGCCCA CCCTCTGATA ACAGAAACGT TATTTTCTAC ATCGCGAGCA ATATCATCGA 2040

TGGTTAAATT AATCATTTTA TCAAATTTAG TTAACGTGAT TCGTTTATAT TTTCCAGTAG 2100

GCATAGGGGC AGTCTCCGTC TGTTTATTCA TTTCTTCTAA AAATTGATTT CTTTCATCTA 2160

ACAGTGCATG AATAAAGGTA TTTTCACGTT CTGAATAGTT GCCTTCAAAG AAATAATCAA 2220

TGTCTACTGT CAGGTCTTTG CCTGCCATTT GTGCATCGGT TACTTCGATT ACAGGTACTT 2280

GCGTATCTAA GAAATAATAG ATATTCTCTT GTACCAGAAA AGCATTCGTG GCTGCGACTG 2340

GGACTGTCTC GCCGTCTATC GTCATGGATA ATTTAACAAT TGCCCCCGTT AAACTTGGGA 2400

ATAATTTTAT CAGTTGGACG TCTTCTGGTA TCTTCAGTTG GAACTGTTTT GTTTCAGGTT 2460

TTTGTAAATC AACTGGTTGA TCCAACTCAG TCTCTTTTTC TTCGCCGCAA TTAATTAAAA 2520

TATGAACGTT TTTTTCATTG TATGAATTTA CTGGTGTCAC TCGCTCTGGC TGTTCTACTG 2580

GTTCAGCAGT CAAGGCAAAA AAGTACTGAT ACACTTCGCC AAAGGGACGC GCTTTAAAAG 2640

cTGCTTGCAC TTCTACAGGT AACGCTTCAT ACGTCGTAGG AATTTCgTTG TAGCCAACTT 2700

GATTAAATGT AAAATCTTCT TTGGCAATGT ATAAACCAAC TTCTGCAAAA ACTTTTTCAA 2760

ATCCTTCTTG TAAATAGAAT GATTTATGCG TAGCATCCAA CAGCCCAGTT TCATTCCACG 2820

TTAAACGATT ATTAAATAAA TCAATTAAGA CCGAATTATG CGCCAAATTA GGGAAAGTGA 2880

TTAAGATTTG ACCGCCTGGT TTTAAAAACG GTTTGACTTT TGCTAAGGCG CTTTGAGGAT 2940

TCATTAAATG TTCCAACACA TCTGCAAAAA CGATATAATC AAACGTCTTA CCAGCAAAGT 3000

ATTCGACCCA ATCATTTTCA TCAATATTGC CATAAAAGGC ATCTGTTGAA AATTGACTGA 3060

CATGATCATA AAGTTCTTTG TCTAATTCAA CAATTGAGAC CTGGcACTTT TTTTCTTCCA 3120

TCAAATAGCT AGTCATTCGC CCATTACCTG GaCCAAATTC CaAGACATCA CTATTTTCTT 3180

TGATTTGTGC TACGATTTTC CCAACACTGG TACTCTCGTC TGTTTCCATT TCAAAATCAT 3240

ACTTCATCTT TTGTCGTATC CTTCCATTTA TGATCTAAGA TGACAATTCC TTCTCCGGCA 3300

TAACGACCGG TAATAAACAT CGTCACGTAT TTTGTTTTGT AGACTAATGG GACTGTTGCA 3360

TGTTCCTCAA AGAAAGCAAT ATCAAAATAA TATTCACCTG TTAATAATGA CATTTTTTCA 3420

TATTCTAAAT AAAAAACATT CCGTCCTTTT TTCCAAGGAA TTTGCACGCC ATCTAACAAC 3480

GTATTTAAGC CACAGACATA ATGATTGTCC ACTGTTCGAA TAGCGACCCC ACAAACAGGC 3540

TTCTTAATTT GGTCATTTTT GACAGTATAT TCAATTTTTA CAATCACTTT TTCATCTTGC 3600

GTTACCATTT CTAATGGTTC ATAATCCGCA TTTAAAAGTT CCGCATGATC CACTTCAATA 3660

ATATCTGGTG TTGGGGTTTC ATCTTTCATG GGTCtTCGCG GTCAACCGTT TTAATTGATT 3720

TCTTTTTCAA AAAGTTTTCG TAGTTGGTCG TCACGTCCAT CGTATCTCCA TATTCAACGA 3780

CTTTCCCGTG TTGCAACCAA AAAGTTCTGT CACAAAAACG ACGAATGGAG TTTACATCGT 3840

GAGAAACAAA TAAAATTGTT TTACCAGAAT TTTTAATTTC CGTAAATTTT TCCATACATT 3900

TTAATTGAAA CTCTAAATCC CCGACAGCCA ACGCTTCATC GACAATTAAA ATATCGGGAT 3960

CCACGTTAAT GGcTACCGCA AAAGCCAAGC GGACAAACAT ACCACTAGAG TACGTTTTAA 4020

CTGGTTGATA CAAATGATCG CCGATGTCAG CAAAATCAAT GACATCTTGG ACACGCTCTT 4080

CCATCTCTTC TCGGGAAAAA CCCAAGACCA TTCCATTTAA GAAAATATTT TCATATCCTG 4140

AATATTCAGG GTTGAAACCT GAACCTAATT CAAGCmACGC TGAAATTTTC CCATTAATCA 4200

CCATGGACCC GGAGGTTGGT GTTAGTACTC CTGTAATAAT TTTTAACATG GTTGATTTCC 4260

CAGaGCCATT TTcGCCGACA AAmCCAaTCA TCTCGCCtTT TTTAATTTGc ATCGTGrCmT 4320

CGTCTAACGC ATAAAACAAA TCGTGGTAGG ATTTTTTGGT TGGACTAAGC GCTTCTTTAA 4380

ACCGATCGGA CGGCTTTTTA TACATGTTGT ATGTTTTGGT AATATGTTGG ATATCGATTG 4440

CATATTCTGA CATGTCGTTC TCTCCTTATA GTACATCTGA AAAATGTGGT TTTAAGCGTC 4500

TAAACACGGT TGAGCCAATT AATAGTAAAA CAATAGTAAC TCCCCAGAAA TATAAGCTAT 4560

ATTGCCAATG TTCCCAGAAC CATGCACCGC CTAAATAAGA TTcGCGGTAT CCTTGAACAA 4620

TATAATATAA TGGATTGATT TTTAAGATTG TCGCAATCAT TCCAGTTGGT TGCCATAAAA 4680

TTGGCAATGT CCACATAACT GTTTGCATAA TGATATTGAT AAATTGCATA ATATCAGGTA 4740

AAAATGGTTG TGTTGAAGCA GTAATCCACG TAATTGCTGT TAAAAAGACT AATAAACAAG 4800

CTAAGTAGTA AATTAATTGC AAACTGTACA GACTTGGGTA ATAACCGTTA GCTGTAGTAA 4860

TTACAAAACC AATTAGAATG AAAAACAAGT GGGTGTACAA GTTCGATAAA ATTTTAGATG 4920

TTGGTAAAAT TTGTACATTA AACACGACTT TTTTAACTAA ATAACTATAT TCACGAAACA 4980

CATTCGTGGC TGATAATAAC GAATCTGAAA AGAAGAACCA AGGTACCATC CCTGTTACTA 5040

AATAGACAAT GAATGGCATA TTGTCGCCAG CTCTAGAACG AAGGCCTACA GAAAAGACAA 5100

ACCAATACGT CAAAACAGTT ACTAATGGCG TTAAAAACGC CCAGACAATT CCTA 5154

(2) INFORMATION FOR SEQ ID NO: 202:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10996 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

ACTGCCGCTA AACAAGAATT TAAAGTAGTA GTTCAACAAG AAATGCCTTC TGCAGATTTA 60

TCTTTAGGAA CAGATACAAT TAGTTTTACG GCACTGAATA ATGCGTATGA AGGAATTTAT 120

CGTTTAGATG ATAAAAGCAA GCCGCAACCA GCAGGGGCTA AAGAAAAAGT GCAAGTCAGT 180

GATGATGGTC TGACGTACAC AGTGAAACTA AGAGAAGAAG CAAAATGGTC CAATGGTGAT 240

CCTGTCACAG CTGCCGATTA TGTTTTTAGT TGGCAACGTA CAGCAGATCC TCAGACTGGC 300

GCTGAGTATG CTTATTTTCT AGAAATGATT GAAAACGGTG CAGACATTGT AGCTGGTAAG 360

AAACCAGTCT CTGAATTAGG CATTAAAGCA AACGGAGATT ACGAGCTAGA AATTAAACTG 420

GCAAAACCAA CACCATATTT TGATTATTTA CTGGCCTTTC CACTATTTTT CCCGCAACAT 480

CAAGCAACCG TAGAGAAATA TGGCAAAGAT TACGCTGCTT CCAGTGAAAA AGCAGTCTAT 540

AATGGTCCGT TTGTCTTAGC TAACTTTGAA GGGGCTGGCT CAGATACCAA TTGGACTTTG 600

GAAAAAAATG AAAACTATTG GGATAAAGAC AATGTCAAAT TAGATAAAAT CAATTTTGAT 660

```
GTTGTCAAGG AGGCGCCAAC CGCCTTGAAC TTATTCCAAG ATGGCCAGGC AGACGATGTT     720
ATTTTAAGCG GCGAATTGGC CCAGCAAATG GCGAAGGATC CAGAATTGGT TATCGAAAAA     780
GAAGCACGAA CCAGCTATTT GGAATTTAAT CAGCGAGATA AAAATTCACC GTATAATAAT     840
GTGAATTTAC GCAAAGCAAT TTCAGCAGCG ATTGATCGCA ACGCATTAGT AGATAAAATT     900
TTAGGCGACG GCTCAGTAGT CGCTACCGGC TTAATTCCAG AAGGGATGTC GTATTCGCCG     960
ACAGATGACA CAGATTTTGC TGATGAAAAT AAAAAAATTG TTGAATATAG CCCAGAAAAA    1020
GCAAAAGAAT ATTGGGCAAA AGCGAAGAAA GAATTGGGCA TAACAACACT CAAAATGGAT    1080
ATTGTTGCCG ATGATGTCGA TTCTACGAAA AAGTTAGCGG AATATATTCA AGGTACGTTA    1140
AAGGATACTT TAGAGGGCAT TGATGTAACA GTGAGCCCAG TGCCGTTTTC AGTTCGAATT    1200
GATCGAGGTA GTCGCGGCGA TTTTGAGACA ATCTTAGGTG GTTGGGCTGC GGATTACGCA    1260
GATCCCAGTA GTTTCTTAGA TCTTTTTGTT ACTGGTAATA ACTATAATCG CGGGCGTTTT    1320
TCAAGTAAAG CCtACGATGA GTTGATTGAG GCTTCAGCTA CAAGAGATGC TTCAGATCCT    1380
GAAAAACGTT GGGAAGACAT GGTTAAGGCT GAGAAGTTAT TGATTGGTGA AGAGACCGCT    1440
TTAGCACCAC TTTATCAAAA GGCGACTGCC CATCTACGCA GTAAAGAAGT GAAAGGCGTC    1500
GTTGCCCATG GTGCTGGCGC ACAATATGAT TATAAGTGGA CATATGTGAC AGAAGAGTAG    1560
TTAAAAAGAG ACACAGAATA GTCAAATCGA CTGTTCTGTG TCTCTTTTAA TCAAAATATT    1620
TCTGCTGTTC TTCTTTTAAT TTTTCTGGGG TTACATCTTC ACCAGTAGGA TCATAAATAG    1680
TGACGCCCAT TAATGTTTTA TTGAATTGTT CACGAATGGT CCGCAAATAT TCTTGTCGCA    1740
AATCGACACG AAGAGCTTGT TCTTCAGATG TTAAACCTTG TTCACGCTCT TTTTGTGCAA    1800
GCGTATTAAT TCTTTTTAAA ACCTCTAACA TTTCTCATTT CCTTTCTATA ATAATTCTTT    1860
TGCTAGGCGT TCAAACGATT GAAAGCGCGT TTGTTTGTCA TAACCTGGGG TGATTAGCAT    1920
TGCTTCTGTC AACCCTAGTt CtTCtTGTTG GTGGAAAAAC TCCGCAACTT CTTGGCTACT    1980
GCCAATGACA AAGCGTTGTT TATTTTGTTC TAAAACCAGT TTGTCCATTT CACTGTAAGG    2040
GAAAGCTAGG GCTTCTTCTG GCGATAAAAA TTTATTTGGA AGCCCATGTT CGGTGGCAAA    2100
GATATAATGC TCGAAACTAG CCGCATGGTA TTCTGCTTCT TCTTTCGTGT CGGCGGCAAA    2160
AGTGAAATAC GCAGCGCTCA CTTGCGGCTC TTGTTGTTCA GTTACTGGTC GGAAATGCTT    2220
GTAATATGCG GACAATACTT CTTTGGGCAA ATGTCCATGA ATAAACTGAG CAAAGGAAAA    2280
ATTAATGCCA TATTCACCAG CTGAACGTGC GCTTTCTTTG CTAGCACCTA GTAACCAAGG    2340
TTGTGGTAAT AGAACTTCCC GCGGAACAGC TGAAATTTTA GAATATAAAT GATCTTCAGG    2400
AAATTCCTCT CTTAAAAAGA GTAAAATTTC AGGTAATTTA TCAAATTGGT CTAGCTGTGG    2460
TGGCTGACCA TTATTTAACG CCAACGTAGT TAAAGGGCTG CCGCCAGGCG AGCGGCCCAT    2520
GCCTAAATCG ATTCGATCAG GATATAAAGC TGCTAACGTT TTAAACGTTT CTGCGACTTT    2580
AAATGGCGCA TAATGCATGG TCATGATGCC ACCCGCGCCA ACACGAATTT TTTCGGTGGC    2640
```

ACTAGCAATA TGTGCAACTA AAATTTCTGG CGTTGCACTG GCAAAGGCGT CTGAGCCGTG 2700

ATGTTCGGCA TACCAAATTC TTGTGTAACC AAGTTTTTCC GCTAATTGAG CGAGCTCTTT 2760

GGCTTCTTGT AATGTGTCGC TTGCTGATTG ATTTTTAGGA ACCGTTAATT GATCTAAGAT 2820

ACTTAATTTC ATGTATGTAG TCTCCTTCTT GGTTATTTTT TTGAAGCTAC CGCTATTATC 2880

ACGCAAGCAT GTCAAAAAAA ACAGTACGCA CTTTTTAATT CCCTAGTGAG CGAAAAGTAA 2940

CAAATAGATG TAGCCAAGAT CGATGATAAA AAATTCTCCG CTATGGCTTT TTGATGCTTT 3000

GTGTTAAAAT GGAACATATG TACGCAATCA GCAATTGGAA AGGTGACAAC ATGTACGAAT 3060

ATATTATTGG AAAAGTAACG TTTGTTAGCC CATACTATAT TGTCGTTGAA ACAAATGGCA 3120

TTGGCTATCA GATTTCAGTC GATAACCCGT ATCGCTATTC AGGAAAAATG GATACGGATA 3180

TTAAACTATA TCTTCACCAA GTCGTGCGGG AAGATGCACA ACTTTTATTT GGTTTTGGCA 3240

GTTTAGAAGA AAAACAATTG TTCTTAAAAT TAATCAGTGT CTCTGGAATT GGACCTAAAA 3300

GTGGCTTAGC TATCATGGCT TCTGTGGCTG ATCATGGTGG TTTAATCAAT GCAATTGAAG 3360

GGGAAGATGT TACATATTTA ACGAAtTCCC AGGTGTTGGC AAGAAAACAG CCCAACAAAT 3420

GATTTTAGAT TTAAAAGGAA AATTAGGCGA ATTAGAGTCT TCCGAAGCAG CGGTTGCGGC 3480

CATGACCGCA ACCGAAGCGG TGACAACAAG CAATCAAGCA TTGGCAGAAG CGTTAGAAGC 3540

ATTAAGTGCC CTAGGCTACA GTGACCGAGA GATTAAGCGT ATCACTAAGC AATTGGAAGC 3600

GTTAGGTGAG ACAACGACAG ATGTTTATTT GAGCAATGCG TTAAAATTCA TGATGAAACG 3660

CTAAGGAGGA ATAACGATGA CAGAAGAAGA ACGGTTGCTT TCAGCTGCAA GTCAAGAAGC 3720

AGAAGCTTCA ATTGAAAAAT CGTTACGTCC ACAATTTTTA GCACAATATA TTGGTCAAGA 3780

TAAAGTAAAA CAAGAACTAA CGATTTATAT TGAAGCGGCG AAAAATCGGA ATGAAGCGTT 3840

GGATCATACC CTTCTTTATG GTCCGCCAGG TTTAGGGAAA ACGACAATGG CGATGGTCAT 3900

CGCCAATGAA ATGAATGTGA ACATTCGTAC CACAAGTGGA CCAGCCATCG AACGGGCGGG 3960

AGATTTAGTC GCTATTTTGA ACGAATTAGA GCCTGGTGAT GTGTTATTTA TCGATGAAAT 4020

TCATCGTTTG CCGCGAGTTG TGGAAGAAAT GCTGTATTCC GCAATGGAAG ATTTTTACAT 4080

TGATATCATG GTCGGTCAAG GAACCACGGC ACATCCCGTT CATTTTCCGT TACCGCCTTT 4140

TACTTTAGTG GGCGCTACCA CACGGGCAGG GATGCTTTCA GCGCCATTAC GAGATCGTTT 4200

TGGGATTATT TCTCACATGG AGTATTATCA AGAGCAGGAT TTAAAAGAAA TCGTTCTCCG 4260

TTCCGCGGaT ATTTTTCAAA CGGAGATTTT TGAGGAAGGT GCCTTTGAAA TTrCGCGCCG 4320

CTCAAGArGA ACCCCCCGAA TTGCTAACCG TTTGTTAAAA CGTGTCCGTG ATTTTGCCCA 4380

AGTCCAATCA GATGGCAAAA TCGACCGAGC GATTGCAGAC AAAGCGTTAA CCTTACTGCA 4440

AGTGGATCAT CAAGGCTTGG ATTATGTCGA TCAAAAATTG TTGAAAACAA TGATTGATTT 4500

ATACGGCGGC GGTCCGGTTG GTTTAAGTAC CTTGTCTGTA AATATCGGTG AAGAAACCGA 4560

AACGGTGGAA GATATGTATG AACCCTATCT TATTCAAAAA GGTTTCATTA AACGAACACC 4620

ACGAGGACGA ATTGCAACGC CATTTGCGTA TGCACATTTT GGGTATGATT ATTTAGAGGG 4680

TCGTAAAAAT TAAAGGGTTA AATAAGAGAG AAGGGCCTTC AACAGGATAG AGTATGTTGA 4740

AGGTCTTATT TTTCATTTTT TCAAAGTTTA TGTAACGGAA TAAATATTCT TTTTTTGATA 4800

GACCACTATA AGTATATTAA CGAATAATTT TTAAAAGTTT CATCTCTTGA TAATATAAAT 4860

CTTTGTAAGA TAAATCAGAC AGTTTGCGAG TAACTGTTTT TGATATCGGT TTGTTTTCAA 4920

AACCATTTAA AGCGGACAAA ATATATAACT TAGTTGTGCA GGTATTTTTA AAATGATTTC 4980

GGATGTAGCC AAGAAGAGCG ATTGCTGAGG GTGTTTCAAT ATTTTGATTA TCTGACGGTA 5040

GAGAACACAT AGAGCAAACC TGATAATTCA ATGTTTTTAA ATAGTTGCTG TATGAAAATT 5100

TACCTGGTGG ATCACCTTCA AAATCATTTA TAAAAAAGAA GAACGCTTTA TTTAATAGAT 5160

TGCCAGTGGG CTTGCTCAGT CCATTTAAAG TTATACTGCC TGTAGATAAA GGTTGTTCGC 5220

TAGCTAAGTA GATGTATTCT GTCAATTTTA TCACCTCTTT CTTATCGATA TGAATTATCG 5280

TATCATGATA TAATCAAATC CGCTATATTT TAAGGTGAAA TGCTAAATGG AAACAAAAGA 5340

AAGCAAGCCT TAACCGCTTG CTTTCTTTTG TTTCACGTGA AACGTTTGTT CTTATTTAGA 5400

AGAGCGGCAT ATTTCAGCTG TTGTAGTTAT ACTGTGTTTG TAATCGCTTT TAAATTGGAG 5460

GGATCAATGT TGAATAAAGA AATCAAAGGG GCATTAATTG TTTCTTGCCA AGCGTTGGCA 5520

GACGAACCGT TGCATAGTTC TTTTATCATG TCCAAGATGG CTTTGGCCGC TAAACAAGGA 5580

GGGGCTAAAG GTATTCGAGC AAATTCTGTG GTAGACATTC GAGCAATCCG GAAAGAGGTT 5640

GATTTACCAA TCATTGGGAT TATTAAACGA GATTATCCTG ACGCAGAAGT ATATATTACT 5700

GCAACTATGC GTGAGGTAGA TGAATTAATG GCTGTTCAGC CAGAAATTAT CGCTCTCGAT 5760

GCAACCCATT CGAAACGACC AAATGGCGAA ACGTTAGCCA CATTTTTCAA AAGAATAAAA 5820

GAAAAGTATC CTAACCAAGA ATGGATGGCG GATTGTTCCA CGCTGGAAGA GATGATTGAA 5880

GCAGCGGAGT TAGGTTTTGA TTATATTGGC ACAACGTTAG TCGGATATAC GCCGCAAAGT 5940

GAAAATCAGC GGATTGAACA AAATGATTTT GAACTGTTAA GAGAAGCAAA GAAGCATTTT 6000

ACTACACCAA TCATTGCAGA AGGAATGATT GATACGCCAG CAAAAGTCCA ACGTGTCCTG 6060

GAGCTAGGTA CTTATAGCGT GGTGGTGGGA TCTGTTATTA CACGACCGCA AAAAATTACT 6120

GAAAAATTTG TGGAAGCTAC GAAAATAAAA CAGAATCAAG CATAAACTAA ATACACTATC 6180

GATACACTCG TTATTTTTTG ATCAATACAC TTCGATACGG TCGCTTTTTT TCTAGAGAAA 6240

GTTGAATCTT TCAATAATAA AAAGGGATAC ACTCCATTTG GCATAGTCCT TGCTGATAAT 6300

AAATCAGTGT ATAAAGCGCT ATCATTTTAT AGGAGGGGTT TTATGAAGGG TTTATCAAAA 6360

AAGAAACGGG TGTCTACTTG GTTAGCGTTA GGAATCACCG TAGTCAGCTG TTTTGCGTTA 6420

AGCAGGGAAG TGCAAGCAAG TGTTGAAAGA ACAAAAGTTG ATGAATTTGC AAATGTTTTA 6480

GATGTGAGTG CATCACCAAC CGAACGGACG AATGGCGTAT ACGATACCAA TTATTTTAAT 6540

AATTTTTCTG ATTTAGGTGC ATGGCATGGC TACTATTTAC CTGAAAAAAG CAATAAAGAG 6600

CTACTGGGTG GTTTTGCGGG GCCATTGATT ATTGCGGAAG AATATCCAGT AAACTTGGCG 6660

GCAAGTTTAA ACAAATTAAC GGTCAAAAAT AAAAAAACGG GAGAAACCTA TGATTTAAGC 6720

CAAAGCAACC GCATGGACCT GTCTTATTAT CCTGGGCGCC TAGAGCAAAC CTATGAATTA 6780

GACGATTTAA CGATTCATTT AGCTTTAATT TTTGTCAGCA ATCGAACGGC GCTTATCCAA 6840

ACGACACTTG AAAACACTGG TGAAGAGCCC TTGTCACTTG GAGCAAGCTG GACAGGTGCG 6900

GTCTTTGACA AAATTCAAGA GGGAACGGAA ACCTTAGATA TTGGCACTCG TTTAACTGCT 6960

AAAGACAATG ACATTCAAGT GAATTTTGGT GAAGTCAGAG AAACGTGGAA TTATTTTGCT 7020

ACGAAAGACA CAAAATATAC GATTCATCAT GCGGATAAAG TTTCAACAAA AATTGATAAT 7080

CGGAATTATA CAGCAACCGC TGAACCAATT GAATTGAAGC CTAAACAAAC GTACAACACC 7140

TATACGACAG AAAGCTATAC TTTTACAAAA GAAGAAGAGG CAAAGGAACA ACAACAAGCA 7200

CCCGAATATA CCAAAAATGC GGCGCGCTAT TTCAAAGAGA ACAAGCAAAG ATGGCAAGGA 7260

TATCTAGATA AAACGTTTGA TCAAAAGAAA ACAGCAGAAT TTCCTGAATA TCAAAATGCG 7320

CTAGTCAAAT CGATTGAAAC GATTAATACC AATTGGCGAA GTGCGGCAGG TGCCTTTAAG 7380

CATGACGGGA TTGTTCCGTC CATGTCTTAT AAATGGTTTA TTGGTATGTG GCTTGGGAT 7440

TCGTGGAAAG CGGATGTAGC AACGGCTGAT TTTAATCCTG AGTTAGCTAA AAATAATATG 7500

CGGGCCTTGT TTGATTATCA AATTCAAAAA GATGATACCG TACGTCCACA AGATGCAGGA 7560

GCGATCATTG ATGCTGTCTT TTACAATCAA GACAGTGCGC GTGGTGGTGA AGGTGGCAAC 7620

TGGAATGAAC GAAATTCTAA ACCACCATTG GCTGCATGGG CAGTTTGGCA TATTTATCAA 7680

GAAACCAAAG ATAAGGAATT TTTAAAAGAA ATGTATCCCA AACTTGTGGC TTATCATAAT 7740

TGGTGGTATA CCAACAGAGA CCACAATAAA AATGGGATaG CAGAATATGG AAGCATGGTC 7800

AGTGATGCTC ACTGGCAAAA AGACGACAAG GATCAAATCA TTAAAGATAA AAATGGCCAC 7860

CTAAAGTGGA TGATGATGCT GTTATTGAAG CAGCCGCGTG GGAAAGTGGC ATGGATAACG 7920

CTACACGGTT TGACAAAGAA GGTGTGGGCA AAGGCGACGT TGGAGTTAAA GTTTTTGAAA 7980

ACAAAAATAA AGGAAAAGTA GTGGGGTATT CGATTAATCA AGAATCAGTG GATTTGAATG 8040

CGTATCTCTA CGCTGAAAAG GGCTATTTAG CTTCGATAGC AGAAGAACTA GGCAAAAAAG 8100

AGGACTATAA GAACTATCAA AAAGAAGCGA AGAAACTAAA AAAATATATT CAAGAAAATA 8160

TGTTTGATGA AAAAACAGGC TTCTTCTATG ACTTACAAAT AAATGAAGAC GGCTCGAAAA 8220

CAAAATTGTT AGTCAATCGT GGCAAAGGAA CAGAAGGCTG GTTGCCGCTC TGGGCAAAAG 8280

TAGCGACCAA AGAACAGGCG GCGGCAGTTA AGAAAAACAT GATGAATCAA GAAATGTTCA 8340

ATACCTTTAT GCCATTCCCA ACGGCATCCA AAGACAATGA AAAATTTGCT GCAACGAAAT 8400

ATTGGCGGGG ACCAGTTTGG CTTGATCAAG CCTTGTTTGG TGTAGAAGCT CTTCAAAACT 8460

ATGACTATAC AAAAGAAGCG AAAGAAATGA CACAAAAACT CTTCTTACAT GCCGAAGGAT 8520

TGATGGGAGA AGGCCCTATT CATGAAAATT ATGATCCACT GACAGGGAAG GGCTTAAGTA 8580

CGAAAAACTT TAGTTGGTCA GCAGCAGCGT ACTATTTATT GTATAAAAAC ACCTTATTAA 8640

GTAATAACCC CACAACGCAA ACTGCTTTTG AGATAAAATA GAAAAACAGC TCTTGAGAAA 8700

ATATCTCTCA AGAGCTGTTT TTTTCTTCAC TTGCTACAAT CTGTTTCCCT CTGATTTAAT 8760

TAAATGTAAT TTATTTGTTT AAATAAATAT ATTTCATATA ATTAAATAAA ATATCTGAAA 8820

ATAAACGATA TGTACTTTTA TTAAAAATGT TATTTAATAA AAATGAAATA GATATTCAAT 8880

GAAATTGATT GTACAATGGT TAATGAAGAG TGGGTTTAAC GGAATAGTGT AATAAGGAAT 8940

AAGCTGATAT CAAAGCCAAA CATATGCAAA ACCTCCTTCT CTGTTAGCTA TATTGACTTC 9000

TATTATAGAA CATTTCTTTT GGTTTTTTCA AGGGAAGTTA AGTATATAAG TTTTTATGAT 9060

TATCAAAATA TTACAATATT ATAATTACGG AAATAATAAA AATGAATATT TTgAAATATT 9120

TATTAATGTT TTCGCTGATT TTGTTTAATT TGTACAGGTA TTGTTGGATC TTTAAGTTAA 9180

TAAAAAAATA ATATTTAATA TTATTTTTAT TATmAAAGGG GGATATTATG GATGTGCTAG 9240

GTGCTCTAAA ATATTTTAGA AAAAAGAAAA AAATTGCTCA GAAAGATGTT TTGCCAAAAA 9300

AGGGGAAACA AACCTATCGA AGAATCGAAA TTGGTGAAGC CAAGCTAAGT TACGAGGATT 9360

TAATGACGGC ATTACAGTCA TTGGGGATTA CTTTTAATGA ATTTTTTTTT ATGGTATCTG 9420

ATACGAGAAT CATGGCTTCT TCAGAAGTCA AACATCAAAT TGAATGTTGT CAAATGGGGT 9480

TGAACAATAC AAGTGAAAAA AAGAATCTGA TTCATTATTT TTATCAATTA GAGCGCAATC 9540

CTCATAAAAA TGCGTTGGAA ATGTCTATTT ACACAGATAT TAAATTAACG TTTAGTAATG 9600

ATTGGGAGGA AATTCCAGAA TTTGATGAAA CCGATCGAAT GGTAATTCTC GCGCTTATCT 9660

CTAGCAAAAG CTACTATACA TACTACGATT ATCAAATGGT GACTAATCCT GCCGCCTTGT 9720

TTTCTGAAAA TGAAGTCCTT CAGATTCTTG AGCAGATGTT TCCTGTCAAA GATGCTGAAT 9780

TACGTGATAC CCAAACACTA AATGTTGCCT ATGGTTTTTA CCTAAATATC ATTACTGGGG 9840

AACTTTATAA GAAAAATTAT GCGAAAGCTC GGGAATATTT AGCGCTAGTG AGTGTTACCA 9900

CGATTCCTGC TGAAATCTAC TATATTCATT TTAACTTACG TTACTTAAAA AACCTTACGT 9960

ATTATCTGTA CACTGGCAAA ATGAGATATT ATAAAGAAGT AATCGCCGTT ATTGATATGA 10020

TTGAAAGCTT TGGGGATGCA CGTTTGGCGG AAGGCATGAA AAAAGAAATG CTTCAGCTGA 10080

CGGCTGGTCG GACATTTAAT TTAGAAAAAG GTCAGTTTCC ATTGAATATT GTCACTGAAA 10140

AATAAAAAGC CCCCCcCCCC CGCAAGGTCA AACTTGCGGG GGGTGCTGAT GTTAAACAAT 10200

AATGTTATCT CTAATCATAT CTAAATCTTT AATCAAAATT TTGCGGTTAC GTAGTTCAAT 10260

TGCCCCTAAC TCTTTTAGTT GCTGCATCAT CCGGTTAACG CTACTAGCAG AAGTTATCCC 10320

ACAAAAATGC GCAATTTCTT CATTGGTCAC CACGAAATCA ATTAAAATAC CTTCCTCTGT 10380

TTCTACGCCA AATGAATCGT ACATTTCATA GATTTGTGTA CAAACGGCCC CAAATTTGCC 10440

ATTCATCAAC ATTTGCTGCA TCTTTTTCAT GGAATGCATT AAACGGTCCC GATAGTATTC 10500

TTTGACATAC ATTTGCAAGT CTTTACTATT ATTGATATCC CGCCAAAATT GAACACGGTC 10560

GATTTGGTAA AGTTCTGCTT GTGGAGATTC GATGCGAATA TTAAAAGGTG CATCAATAAA 10620

TTCTGAATAT TCATCACGTA GCAATGAAAC AATTTCCAAA TTGTTGATAT AACGTAAATT 10680

AAATTCACGA CCGTCTTTGG AAATAACGCT AGTTTTGATG ATGCCACTTT TTAAAATATA 10740

GGCATAACGA TCTTGCAGTC CTTCATAAGT TAAATAATTT TTCTTCTTTT TTACAACGAC 10800

TCGAAAGGAA TGTTCATCTA AGTAGTCTTG TAAAACATTT CTCATTGGCG GTAACTCCTC 10860

AATTATCTCT AAATATTTTT GTTTTTGCCT GTTGGGCTTC TTTCTTATTT AAGCGAATTG 10920

GaCATCATAT GTCAATAACa AATGTTGAtA AAAAAgGTCa ATTAAACgCT ATTTTGAAAA 10980

ATAAAATCAG CnTTTA 10996

(2) INFORMATION FOR SEQ ID NO: 203:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 957 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

CGGCCCAnTT TTTTAAATGC nTnCATTCTT TTAAGGCTTT CCAATCCCAT CATTATCATT 60

TGnATCTGTG ACCATCGTGG CCACTTTTTG AATGGAgGAC TCGCATTGCC CaTTGCGACG 120

CCAATCCCTG CATGGCTTAA CATTTCTCGG TCATTTTCGC CGTCCCCAAA GGCAATGATA 180

TTTTTTGCAT CAATACCCAC ACGTTTAGCT AAATACGTTA ACGTAGCTGA TTTTGAGCCA 240

TCTTTAGGCA CAACATCTAC GCTATGTTGA TGCCAGCGAA TAAAGCGAAA TTCTGGGTAT 300

TCTTCTTCAA ACGTACACTG ATCTTCTGCG TCATAAAAAG CCAGCGCTTG GTAAATATCA 360

TTTTGACGAT AAAAGTGTTC ATCATAAGCA GGTGAATGAA AATGAAAAGA GTTCATCGCA 420

ATTTCCATTT TTTCAGCACG GTGCTGGkTG CTCTTCTTAA TATCATCTAA GCCCACACAA 480

GCAAATCCTA AGCCGCGTTG ATCACTGTCT TGTGCCAAAC GTTCCAACGC TTCAGCGTGC 540

AATAAATTTT GATAATATTG CTCATGATCC ACAAAAGCCG CGGCGCCATT ACATAATACG 600

TAATTGGTAA AATCTAAATC CCAAATCACC TCTTGCGCCA TATAACGACT TCTGCCAGTT 660

GCCAGCGTAA CTAAGTGTCC TTGCTCCCGC AATTTTGCTA GCGCTTCTCT CGTGCTATCC 720

AATGCTTTCT TCTCCGAATT TAACAACGTC CCATCAATAT CAAATGCAAA AAGTTTTCTT 780

TCCATCATTT CTCCtTATTC GCTCGCTTTC GCTAAATATT TACTGCCTTC TCGTTGTGCT 840

AAAGGACTCA ACACTAATTC TTTCATCAAT TCCTCGACCC ATTGCGGATT CGTTTTACTA 900

TATTGGGCGT AATGGACCAG CCAATCGCTT TTTGAATGAA GACTCCTCTG nAGTCnA 957

(2) INFORMATION FOR SEQ ID NO: 204:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6835 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

AGTCATTnGT AATCTTTTAA ACTATCGATT AGTCGGGATT GTGTTAGGTA GTTTACTTGG 60
TTTTGCCGGC GCTTATGTTT TCACAACTGC GCTACGTTTC TATAATCGGG CCTACTCAAT 120
TGATGAAACC TTTGAAAATC AACAAGAAGA AAGAGGCGTT TTATGAAAAT AACTTATCCC 180
TTGCCACCAA ATCTACCAGA ACAACTTCCT CTATTAACGA ACTGTCAATT AGAAGATGAG 240
GCAATCTTAG AAAATCATTT GTATCAGCaG ATTGATTTAC CGAAtCAAGA AGTTCGTAAC 300
TTAGTCTTTC GGGATGCTGT TTTCGATCAT CTTTCATTAG CAAATGGCCA GTTTGCTTCT 360
TTTGATTGTA GTAATGTGCG TTTTGAAGCT TGCGATTTTT CTAACGTCGA ATGGCTGTCT 420
GGCAGTTTTC ATCGAGTGAC TTTCCTTCGT TGCAATTTGA CAGGAACGAA TTTTGCAGAT 480
AGCTATTTAA AAGATTGTCT ATTTGAGGAT TGCAAAGCCG ATTATGCTTC TTTTCGTTTT 540
GCAAATTTTA ATCTTGTTCA CTTTAATCAG ACTCGTTTAG TCGAAAGTGA ATTTTTCGAG 600
GTCACCTGGA AGAAACTCCT TCTAGAAGCT TGCGATTTAA CCGAAAGTAA TTGGCTAAAT 660
ACTTCTTTAA AAGGATTGGA TTTTTCCCAA AACACCTTTG AACGCTTGAC TTTCTCACCA 720
AACTATTTAT CAGGTCTTAA AGTCACCCCT GAACAAGCCA TTTATTTGGC AAGTGCCCTT 780
GGTTTGGTGA TTACCTAAGA AAATTTTAAA AAAATAGCCG AATAACTGGT TCTGTCGGTT 840
CTTTCTTATT AAATTGTTAT TAATTTTTTG GAAAAACCAA CGGAACCATT TTACTTGTGC 900
CTTATTTTTA GGkAAACTAT CyTAGGTTAA GCAATCTCTT TTTTTGGAAC AAGAGAAACT 960
TTTGGAATCA TTTCCTATAA ATCATGAAGA GATTATGACA AACGTTCTTC TAATCCAAGA 1020
ACGTCTTTTG TCCTATACTC GATTCTACTC AAAAGGAATG TCAGTGAATG AATATAAAAA 1080
AACAAGGCCT CGGTCAAGCG ACGGGAAAAA TCATTTTAAT GGGAGAACAC GCCGTTGTTT 1140
ACGGCGAACC AGCAATCGCC TTTCCTTTTC AAGCGACAGA AATCACAGCC GTCTTTACCC 1200
TAGCAAAAAC TATGCAGATT GATTGTGCAT ATTTTACAGG ATTGCTTGAA GACGTGCCCC 1260
AAGAGCTAGC AAATATCAAG GAAGTTGTTC AGCAAACCTT ACACTTTTTA AAGGAAGATA 1320
CGTTTAAAGG CACTTTGACC TTAACAAGTA CGATTCCCGC TGAACGAGGA ATGGGCTCAA 1380
GCGCAsAACC GCTGTGGCCA TCGTTCGAAG CCTTTTTGAT TATTTTGATT ACGCTTATAC 1440
ATATCAAGAA TTGTTTGAGC TTGTTTCCTT AAGTGAGAAA ATTGCTCATG GCAATCCTAG 1500
TGGTATCGAT GCCGCAGCAA CAAGCGGCGC TGATCCCTTA TTTTTTACTA GAGGATTTCC 1560
GCCCACACAT TTCTCGATGA ATTTATCTAA TGCCTACTTA GTAGTAGCTG ATACGGGAAT 1620
TAAAGGTCAA ACACGTGAAG CCGTGAAAGA CATTGCGCAg cTAGCTCAAA ATAATCCCAC 1680
CGCAATCGCT GAAACAATGA AACAATTAGG TTCTTTTACT AAAGAAGCAA AGCAAGCGAT 1740
TTTACAAGAT GATAAACAAA AATTAGGTCA GCTAATGACG TTAGCGCAAG AGCAACTCCA 1800
GCAATTATCC GTCAGCAACG ATATGCTGGA TCGACTAGTG GCTCTCTCTT TAGAACATGG 1860

CGCTCTAGGA GCAAAATTAA CCGGCGGCGG TCGCGGTGGC TGTATGATTG CCTTAACAGA 1920

TAATAAAAAG ACCGCACAAA CCATTGCACA GACTTTAGAA GAAAATGGAG CTGTTGCTAC 1980

ATGGATTCAA TCATTAGAGG TGAAAAAGTA ATGCTTTCAG GAAAAGCACG AGCGCATACA 2040

AATATTGCTC TGATTAAATA TTGGGGAAAA GCCAATGAAG AATACATTTT ACCAATGAAT 2100

AGTAGTTTAT CATTAACATT AGATGCCTTT TACACAGAAA CAACTGTGAT ATTTGATGCC 2160

CATTATTCAG AAGATGTATT TATTTTAGAT GGTATCTTGC AAAACGAAAA ACAAACAAAA 2220

AAAGTCAAAG AATTTTTGAA CCTTGTTCGT CAACAAGCCG ATTGTACTTG GTTtGCAAAA 2280

GTGGAAAGTC AAAATTTTGT GCCTACTGCA GCTGGTTTGG CTTCTTCAGC GAGTGGTCTA 2340

GCTGCTTTAG CAGGGGCCTG TAACGTAGCC TTAGGATTAA ATCTTTCAGC AAAAGACTTA 2400

TCACGTTTAG CGCGACGTGG TTCAGGTTCT GCTTGTCGCA GCATTTTTGG TGGTTTTGCT 2460

CAATGGAACA AAGGCCACTC TGATGAAACG TCGTTTGCTG AAAATATTCC AGCCAATAAT 2520

TGGGAAAACG AATTGGCCaT GCTCTTTATC TTAATTAATG ATGGCGAAAA AGATGTTTCC 2580

AGCCGTGATG GAATGAAACG AACAGTAGAA ACTTCTAGCT TTTATCAAGG TTGGTTGGAC 2640

AATGTGGAAA AAGATTTATC CCAAGTTCAT GAAGCAATTA AAACAAAAGA CTTCCCTCGT 2700

TTAGGAGAAA TCATTGAAGC CAATGGGTTA AGGATGCATG GAACAACCTT AGGTGCTGTC 2760

CCTCCATTTA CTTACTGGTC CCCAGGCAGT TTACAAGCGA TGGCTTTAGT TCGCCAAGCA 2820

CGGGCCAAAG GAATTCCTTG TTACTTTACA ATGGATGCAG GTCCGAATGT CAAGGTCTTA 2880

GTCGAAAAGA AAAACTTAGA AGCATTAAAA ACATTTTTAA GTGAACATTT TTCAAAAGAG 2940

CAGTTAGTCC CAGCTTTTGC TGGTCCCGGA ATTGAATTGT TTGAAACGAA AGGAATGGAT 3000

AAATAAGCAT GATTGAAGTT ACTACGCCAG GAAAGTTATT TATTGCAGGA GAATATGCCG 3060

TTGTTGAACC TGGCCACCCT GCCATTATCG TTGCTGTGGA TCAATTCGTA ACTGTAACTG 3120

TCGAAGAAAC AACAGATGAA GGCAGTATTC AATCTGCACA ATACAGCTCT TTACCTATTC 3180

GTTGGACACG CCGAAATGGT GAGCTTGTAT TAGATATTCG CGAAAATCCT TTTCATTATG 3240

TTTTAGCGGC GATTCATCTA ACTGAAAAAT ATGCGCAAGA GCAAAACAAA GAATTGTCAT 3300

TTTATCATTT AAAAGTGACG AGTGAATTAG ATAGTTCAAA TGGACGAAAA TATGGTCTTG 3360

GTTCAAGCGG TGCGGTAACC GTTGGAACTG TCAAAGCCTT GAATATTTTT TATGACTTAG 3420

GTTTGGAAAA TGAGGAAATT TTCAAATTAT CAGCATTAGC TCACTTAGCC GTTCAAGGAA 3480

ATGGTTCTTG CGGAGATATC GCCGcCAGCT GTTACGGGGG CTGGATTGCC TTTTCAACCT 3540

TCGATCATGA TTGGGtCAAT CAAAAAGTAG CCACTGAAAC ATTAACTGAT TTGTTAGCAA 3600

TGGACTGGCC TGAATTAATG ATTTTTCCGT TAAAAGTACC GAAACAaCTA CGTTTACTAA 3660

TTGGTTGGAC AGGTAGTCCT GCGTCCACTT CAGACTTAGT TGATCGAGTC CATCAATCAA 3720

AAGAAGAAAA ACAAGCGGCT TATGAGCAGT TCTTAATGAA AAGTCGGCTT TGTGTCGAAA 3780

CAATGATTAA TGGCTTTAAC ACAGGAAAAA TTTCTGTTAT TCAAAAACAA ATTACTAAAA 3840

ATCGCCAATT GCTCGCCGAA TTATCTTCAC TGaCTGGTGT GGtAATCGAA ACAGAAGCCT 3900

TGaAAAATCT TTGTGaTTTG GcTGaATCTT ATACAGGAGC TGCGAAATCT TCTGGCGCTG 3960

GCGGGGGCGA TTGTGGGATT GTAATTTTCC GCCAAAAATC TGGGATTTTA CCATTAATGA 4020

CTGCTTGGGA AAAAGACGGA ATTACCCCAC TGCCACTTCA CGTCTATACC TATGGTCAAA 4080

AGGAGTGTAA GGAGAAGCAT GAATCGAAAA GATGAACATC TATCATTAGC TAAAGCGTTC 4140

CACAAAGAAA AAAGTAATGA CTTTGATCGT GTGCGTTTTG TTCACCAATC GTTTGCTGAA 4200

TCCGCTGTTA ACGAAGTGGA TATTTCCACT TCGTTTCTTT CTTTTCAGCT TCCCCAACCT 4260

TTTTATGTCA ATGCAATGAC AGGTGGTAGT CAGCGTGCAA AAGAAATTAA TCAGCAATTA 4320

GGCATTATTg CCAAAGAAAC TGGCCTTTTA GTTGCGACAG GATCTGTCTC GGCAGCGTTA 4380

AAAGATGCTA GTTTAGCGGA TACGTATCAA ATTATGCGAA AAGAAAACCC AGATGGACTC 4440

ATTTTTGCCA ATATTGGTGC AGGCTTGGGT GTGGAAGAAG CAAAGCGAGC GCTTGATTTA 4500

TTTCAAGCGA ATGCCTTACA AATCCATGTA AATGTGCCCC AAGAATTGGT CATGCCTGAA 4560

GGAGATCGTG ATTTCACTAA TTGGCTAACC AAGATTGAAG CTATCGTACA GGCCGTAGAA 4620

GTGCCTGTCA TTGTCAAAGA GGTTGGCTTT GGCATGAGCC ArGAAACCTT AGAAAAACTT 4680

ACCTCTATCG GCGTTCAAGC AGCGGATGTG AGCGGCCAAG GCGGAACGAG TTTTACACAA 4740

ATTGaAAATG CCCGGCGGAA GAAACGAGAA CTTTCTTTCT TaGATGaTTG GGGGcAATCa 4800

ACGGTCATCT CTCTTCTGGA ATCACAAAAT TGGCAAAAGA AACTAACTAT TCTCGGCTCT 4860

GGCGGTGTGC GTAACTCTCT TGATATTGTC AAAGGACTCG CTTTAGGTGC CaAAAGCATG 4920

GGAGTTGCTG GGACTATCTT AGCTTCCCTT ATGAGTAAAA ATGGTTTAGA AAATACCTTA 4980

GCCCTTGTAC AGCAATGGCA AGAAGAAGTG AAAATGCTTT ATACTCTTTT AGGAAAAAAG 5040

ACGACAGAAG AATTGACGAG TACCGCACTT GTCCTCGATC CAGTTTTAGT TAATTGGTGT 5100

CATAACCGTG GTATCGACAG CACTGTTTTC GCAAAACGTT AAAGAGACCA GAACATTAAT 5160

GTTCTGGTCy CyTTATTTGA GTTCTTTTAT CCTACATAAG CATCTAAGAT AATTTGTTTT 5220

AGTTCACTAA TTAATGGTTC TTTCGGATTT GCAGTTGTAC ATTGGTCTTC ATATGCAAGT 5280

TCTGCCATAC GATCAACCGT TGTGTCTAAT GTTTCTTGTG AGACACCTTG GGCTTTCAAG 5340

TTCATGTCAA TTCCGACTGA TTTACCTAAA TCTGCTACTG CTGTTGCTAA AGCTTCTACT 5400

AGTTCTGCGG TAGTATTTCC TTTTAAGCCT AAGAATTTTG CAATGTCTGC ATAATCAGTA 5460

TCGGCGCGGA AGTAATCATA TTTAGGGAAC ATTGCGTGTT TTGATGGATC TTTCGCATTG 5520

TAACGAATAA TATGTGGTAA TAAAATCGCA TTTGTACGAC CATGTGGAAT GCCATATTCC 5580

CCACCAATTT TATGAGCAAT AGAGTGACAA ATACCTAAGA ATGCATTGGC AAATGCCATC 5640

CCTGCCATGG TTGATGCATT ATGCATTTTT TCGCGAGATT CCATATCTGG ACGTTTCACT 5700

GAATTTTCTA AATGATCGAA AACTAACTTG ATTGCTTGTA AACTTAAACC ACGTGTGTAA 5760

TCTGAAGCCA TCACGGAAAC ATAAGATTCA ATAGCATGTG TTAAAACATC CATACCTGTA 5820

TCTGCAGTGA CTGAAGCTGG CACAGACATT ACGAACTGTG GATCAACAAT CGCTACATCT 5880

GGTGTTAACG CATAGTCTGC TAACGGATAT TTCaCGTGTG TTTCACTATC AGTAATAACG 5940

GCAAATGGTG TAACTTCTGA ACCAGTTCCT GAAGTAGTTG GGATACAAAC AAATTGTGTT 6000

TTTTCTGGTT TGGCAATTTT ATATGTCCGT TTGCGGATAT CTAAGAATTT TTGTTTTGCA 6060

CCAAAGAACT CTGTATCTGG GTGTTCATAG AACATCCACA TTCCTTTCGC CGCATCCATT 6120

GCAGAACCGC CACCCAAGGC AATCACTGTA TCTGGTTGGA AATCAACCAT CATTTTGGTA 6180

CCTGCATAAA CAGTATTTGT TGATGGGTTT GGTTCAACAT CAGAAAATAC TTCGATTTTC 6240

ACATCATTTT TACGTTTTTG TAGCTCTTTA CGTACGATAT CTGCATAGCC AAATTGAACC 6300

ATACCTGGGT CACAAACTAA CATGACACGT TCAACATTTT CCATTTTTTG TAGATATgTA 6360

AAGAATTTTT TTCAAAGAAA ATTTTTGGAG GTAATTTAAA CCATtGCATA TTATTTCTCC 6420

GTTTCGCTAC CGTTTTAATA TTAATTAAGT TAATGGcTGa TACGTTTTTA GAAACAGAGT 6480

TTTTCCCgTA TGgAACCaCA ACCAAGTGTT AAGGATGGAA TCATTTCGTT ATAAATGTTA 6540

CCGATACCGC CTTCTGCTGA TGGAGTATTC ACTAAAATAC GACAAGCTTT CATACGTAAA 6600

CCAAAAGCAA CTTGTAACTC TTCATCTTCT GTATGAATTA CTGCTGTATG TCCTAAACCA 6660

CCTAAGTTTA ACATCGCTTC ACATAGTTCA AAAGCATGCT CTGCATTGTT TGATTTCATC 6720

ATTGCTAAAA CTGGCGATAA TTTTTCTCTT GATAATGGAT ATTCTGGACC TGCACCTTCT 6780

AATTCAGCGA CTAAAATTTT TGTGCCTTTT GGTACACTAA TTCCTGCTAA TTCAG 6835

(2) INFORMATION FOR SEQ ID NO: 205:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3977 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

AGTACCAAGC ACTTGACCGC TTTTTTCCTT AGGTGTGCCT TGTAGCAACT nACGCAGTCG 60

CATTACTGAT nTAACCAGAA AAAACACCTT GTAATAAGCG TAAGCCAATC AATTGATAAA 120

CACTTGTGAc TGCGCCCATT AAACTAATGA CGATAGCCAT CCCTAATGAA GCGCGTAATA 180

ACATTAACTT GCGTCCTTTT CGGTCGGCTA AACGGCCCCA CCAAGGTGAA ACAATGGTTG 240

TGACTAAAAA AGTGGAAGAG AACGTGATTC CGCTCCAAAA ATTCAATTGT TGGTGGGTAA 300

AAGTACCTAA TGTGTTGATA TAAAGAGGCA TAAATGGCAT CACTAAGCTA AAGCCAATAC 360

CAGCCATAAA TGTGCCAAAC CATAAAACGA GCAGATTTTT TTCCCAAGGT TGTCTTGGAC 420

GGAAAATTTT ACTTTTAATT AATTCATACA TTCTATTTCC TCACTTTCTA GTAAAGAAAT 480

TCAAATTTGG CTTTGGCGAG TAGTTGATCA TTCAATGCTT GATGCAATTC ATTCGCAAAA 540

AAGCCTTCTT TTAAATTTAG TGGATTTTTT GTAGTGTACA CATATAACGT ATAGGTGTGC 600

ACACCAGAGC GAGGGCTTAA CCCCAGATAA CGTTCGGTTA ACTCAGGTAA CGTTTCAGTG 660

```
GCAAGCGCAG GAATCATCTG TTTCCCCAAA GGAGAAGCAA GACTATTTTT TCCTTGAATG      720
GTAGGCTTGA GTGCTGTTCG ACTAAAGTCT TCAGGAATCT CATTGGTTAC AGGAACATTG      780
GCAGCTAACC AATGAATAAA AGGGAAACCG ATTAGAGGAA TCGTGTCATA ATCAATTAAT      840
TCCCAGGATA AATAACGGGC TTCTTTAGGA ATGGCCGCTA AAGAAATAGG AAAAGATAAA      900
ATAGGACGTT CGTGTAATAA AAATTCTGCA CTGGTTCGTT TGCCATAACG ATCGGCCAAA      960
TGAGAACGGT TGGTAATATG CATGCCAAAC AGCCTTCTTT CAAAAATGAT GTTTACAAAT     1020
TATTATTATA AACTATAATT TATAGTTTAA AAAGAGTCAA TAGATATGAT AAACTAATAC     1080
TTGTCCAAAA TTTAAGAATT GGAGGCGTCA GAAAGATGGC AAAAAAACGG AATTTAACGA     1140
AAGCCAAAAT CCTTGAAAGT GCCCGCAATC TAGCGGAAGA ATTGGGCGTT CAACAATTAA     1200
CCTTTCAAAA TTTAGCCGTT GATTTAGGCA TCAAGTATCC TTCACTCTAT AATCACTTTA     1260
AAAATATTGC TGAAGTAAAA AATGCTCTAG TGGACTTGTT AATTCAAGAA TTAAATGATG     1320
CATTGCGAAG AGCCTTAGTG GGTAAGAGTG GCGCAGAAGC AATTCGGaTA TATGCkGAAA     1380
CGTACCAACA ATTTGCyTTT GAAAATTCTG CAGTCTATGA ATTACTTATC AGTGTGCCGA     1440
AAACGCAAAA CCAACAATTA ATTGAAGGCA TCCATGAAAC CAATCAAATT ATCTTACAAC     1500
TTTTAGCTTT TTATCCATTT AATAATGAAG AACGACTGCA TAAAAGTCGA GAATTACGTA     1560
GCTTAATCCA TGGTTATATT ACCTTACGTT TTTtAGGTTA TTTTCAACGC GAAGAAGCCA     1620
CGCCGGAAGa AAGTTATCGT CGCATGATTG aAGaTTTTAT TGTTTCTTTA gAAGTGGcAG     1680
GaTAAAATAA ATTTCAAAAA TTTCTCAGTC CGAAACCAAA TCCTTGcAAC TTTTTcaGAA     1740
AGTGaCTATG ATACCTACAG TAGCaAAAAA AgGAGCGGTT CAATGTATCm AGtAATTACA     1800
ATGtATGGcG aCmATGAACC TTGGTGGTTT TTTGATGATT GGCAAGAGGA CATTGTTCAA     1860
GAAAAAACCT TCGATAACTT GACCGATGCA GAGACATATT ATGTGGAAGT TTGGCAAAAA     1920
CTTTCTCAGA ATTATACATA CGTCAACACA AAGCCTAATT ATTTAACGGC TTTCTGGAAT     1980
GATGGGGAAG AACGCTGGTG TGAAGAATGC GATGAAGATT TACAACAATA CTTCGGTTTA     2040
GCATTACTAG AAAATCATCA TGCGGTTACT TTTGAAAATA AAGCTGAGTT TACCACAATA     2100
TCCAATCATT TTGGCAAAAC AAAATTCTGC AAACGGAAAG TTATTTGAAA CGCATGAGTT     2160
TAAAGAAAAA CCTCAGGGAA TTTTCTTTAA ACTCTTTTTC TATTCTCGGA ATTTAGAGAA     2220
AATCGAACCA ATATTTGGTA TAATACGAAA AGATAGCACT CAAAAAGGAG CGAAACTAGT     2280
GGTTGCAAAA CAATTAATTA TTGCTGAAAA ACCTAGCGTA gCAAAAGATT TAAGCCGCGT     2340
TTTAGGAGCA AATCAAAAAA ATAAGAATTA TTATGAAGGA CCCAACGTGA TTGTCACATG     2400
GGCTTTAGGC CATCTCTTAG GACTAAAAAT GCCAGAAGAT TTAAACAAAG AATGGCAAAC     2460
GTGGCAAATG GAAACTTTAC CGATGATTCC TAAGAATTTA GGGATTAAGC CGTTACCGAA     2520
AACGGGACAT CAATTAAAAG CGATTAAACA ATTAGCCAAT CGTAAAGATG TTTCTGAAGC     2580
CGTGATTGCC ACTGATGCTG GTCGTGAAGG TGAATTGGTG GCCCGCTGGA TTTTAGAATA     2640
```

TGTCCGTTTT AACAAACCAG TCAAACGCTT ATGGATTTCA TCTCAAACAG ACAAAGCCAT 2700

TAAAGACGGG TTTAAGAAAA TTCGCCCAGC GAAAGACTAC GATAATTTGT ATTACTCTGC 2760

CTTGGCGCGC GCAAAAGCTG ACTGGTTAGT GGGGCTAAAC GTCACACGGG CCTTGACAGT 2820

GAAATATCAA GATAACTTAT CCGCAGGTCG TGTTCAAACA CCAACTTTAG CAATGGTACG 2880

CCAACAAGAA AAAACGATTG AACAATTTAA GCCTCAAACT TATTTTACTA TCTCGTTATC 2940

TGTTGAATCA GAAAAAGCAA AGATGACCCA AAAAAATCCA TACGCTTTAA AAGAACGGCA 3000

GGAAGCAGAA CAATTAGTCA AaGAACTCTC AAAACAAAAG GGCTTAGTTA CCGATATCCA 3060

AGAAAAAGTT AAAaCAGAAA ACGCGCCGTT ACCGTATGAT TTAACCGAAA TCCAGCGTGA 3120

AgCCAACCAG CGATTCCAAT TTTCTGCAAA GAaACGTTGT CTCTTGTACA AAGTTTATAC 3180

GAAACCCATA AAATTGTTTC CTATCCTCGA ACTGACAGCA AATATCTAAC GACGGATATG 3240

AAAGGAACGA TGAAAGAGCG CCTTCAAGCT ATTGCTGATT TTTCACCAGA AGTGAAAGGC 3300

TACCTTAAAA ATGGGGCTGT TGTTAAACAA CAAAAAGTTT TCCAAAATGC AAAAGTCACG 3360

GATCACCATG GGTTGATTCC AACTGAACAA CGTCCTCGTT ATGAAAAGCT AAGTAATGAT 3420

GAACAAAAAA TCTATCAAAT GATTGTTCAA CGATTCTTAG GTTTGTTTGC TGAGCCAAAT 3480

CAAACCAAAC AAACAAAAGT TACAGTCGCT TTTGGAAAAG AAACATTTGT CTTTCATCAA 3540

AATAAAGTAG TTGTTGCTGG CTGGAAAACA ACAGCCGAAC AGCCGCTATC CACTGTCCAA 3600

TGGCAAAAAG GCATGACCGT TGCGCCAAAC TTTACCATTA ACAAAGAATT AACGTCGCCG 3660

CCAAAACCAT TAACCGAAGG AACGTTGTTA GGAAAAATGG AGAAACACAG TTTAGGCACA 3720

CCGGCAACGC GAGCTGAAAT CATTGAAAAA CTGATTAAAT CAGAATTGAT GGaACGAACC 3780

AATAGCGGTT TAAGTGTCTC TGCCAAAGGa AAGCmACTAT TAGACTTGGT GaATCCATCA 3840

CTTGTGaCAC CTGGAACTtA ACGGGAAAAA TGGGGaAAAG TcGcTcGaaG CaATcGCTTT 3900

CGGGGTTCCA ACAAAAGAGT TCAACTTTTT CnTTTAAAGG GATATTTGGA AGAGGGnTAC 3960

GGAAAAATTT AGTTCCG 3977

(2) INFORMATION FOR SEQ ID NO: 206:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5530 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

nnTATTTTTC AATGTATCCG CGTAATTCTT CTTCAGTGAC ACTTTCGTCA AAGATATACT 60

CACCAATGAT GATCGTTGGT ACAAATTGAA TGTGCGCCGC ATTGGCTTCA GCAATCACTG 120

CGGAAACAAG CGTTGCATCT TTTTGTTCTT TTAAGCCCAG ATTTTTTTCA GCATACGTCG 180

CTACTTCTTC TAAGGTTAAG TTTCCCCATT CATCTTGCGT ACAAACATTT TATGCAATGC 240

TGACAGCGCT TGTTCAGGCG CTGAATAGTC AATGTAGTGA TGCATCACAT TGCCGCGTTG 300

TAAACTTTCT TTTTCTTTAT CAAACAACTT AATGATACGT TCAACTTTAC CACTTTTGAC 360

AGATTGTGCT AACAGTTCCT CAGACTCTTC AAACCATTTT CTGCAATAAG GACAGCGAAC 420

ATTGATAAAT TCTATCATTT TGACAGGGGC ATTACTCTCA CCAATGTGAA GCCCTGTTTC 480

TGTGTTAACT TTTGTTGCAT CAATTACTGA AATATCCATA AAAAACCTCA TCCTTTCAAA 540

CCTTTATTTC TCTTTTAGTA TACCGTTTTT CAATAAAAAA ATCATAAGAC ACGAATTCAA 600

ATTTTCTGAA TTCGTGCCTC TGTTCACTTT TTTAACGATT AATCGAACGA GCCATTCTAG 660

CAACAGTATG tCTTTTCCCA CGTATCGCAT CAGAAAGCTC TTTAATAwTT TCAATACCAA 720

CTAAGCCCAT ATtGGCATCG CCAaTATGTT GAATATCTAC GCCTAAAATT TTATTTCTAA 780

TGCCGATTGC TTGGgATAAC TTCAGGTTGA GCACTTTCTT GTGAAGTTCC AATGGCACTC 840

aTAACTAATT TTCCTTTTGA ATGTGCAAAA TCCACAACTT CTTtGACTTG TTGATCATCA 900

ATGCCCCAAA CCGTGCCGAC TGCTGGGACT AAGACAATAT CAGCGCCCGC TTCAATAAAT 960

TGTTCAGCTG ATTTTAAACT GACCACTGGT TCATCTACGC CTGaACTGTG CATTTTTCCG 1020

GCAATAATGA TGCCGTCAAA ATGTTTTTTG GCTAAGGCCA CGTTCTTAGC AATCAAGTCG 1080

TTGGTCACGC CAGTTCCTGG ATTACCTGTC AATAAGATAA AATCCAAACC CAGTTCATTG 1140

GCTTTTTTGA AAGTAGCTGC TGATGCTTTT CTACCTGGTT CAATCGAAAC TCTCGTTGAA 1200

GCTAAGTCTA AATTTTCATC AACAGGTTCG ACATTCATGC CaATCGGTAA ACCGACTAGT 1260

TCTTTCAAAC GATGAACAAC CTCATCATTA GGGACACATG TTTCCGCAGT ATATTGCCCT 1320

TCATAAGCGG CTACTACAAT GGGTTGATTA ACATCGAAAA CATTCAATAA AATCATGTCT 1380

GCGCCATATG CTTTTGACAT TTCCGCACAA GTGATATCAC CTGCTACGGA TGGATGAACT 1440

ACCACATTTT CCGATAATAT TACGCGCCCT TCGCTCGCTT TAATACTCTG AAACAGCTCT 1500

TCTTTTGACA TCCGTTTGAT TTCTGATGCA TTTGCACTAA TTAATCGTkG AACCATTTTG 1560

ACCCCTCCAA TAGTTTTTAA TGTTTTAGTT AAAAACAGCA ATAAGTAATC GTTTGCATTT 1620

TATTACATGC AACATTCGTG cCAACGTCGT TTCTTTAAAc CACAGGTCAA CACTACATTT 1680

CAGtGCTACA CACTTTTTGc ACACTTTAAT TTTTTTTACA CGCTTAAAGA AAAAGTGAAG 1740

GATAGCACTT CAAACGGAAA TGTTATTCTT CACTTCTGGA GCAGAAAAGT TACGCACTTT 1800

TTTCTCGTAA TTTTTCTGCA AATTCATTAA TTTTTCGAAT ACGATGATTA ATACCTGACT 1860

TAGAAATAGC TCCTGAAGGA ATCATTTCCC CTAGTTCTTT GAGACTAACT TCCGGATACT 1920

CTAACCGTAA TTCTGCGATT TCCTGAAGTT TCTCCGGTAA TGACGTCAGC CCTACTGTAC 1980

TTTCAATAAA TTGAATATTT TCAATTTGTT TTGAAGCCGC ATCAATGGTT TTATTTAAAT 2040

TCGCTGTTTC GCAATTCACT AAACGATTTA CTGAATTCCG CATATCACGG ACAATCCGGA 2100

CATCTTCAAA TTTCAACATA GAGTTCGTGG CACCAATTAA TGTTAAAAAG TCGGCAATTT 2160

TTTCAGCACC TTTTAAATAA CAAATATACC CATTTCGGCG TCCCAATGTT CGAGCGTTTA 2220

AGTCATAGTA ATTCAACATC TTACAAATAT CATTATTGTG TTCTTCATAA ATTGAAAAAA 2280

TTTCTAAATG ATAGCGACTT GTTTCCGGAT TATTAACTGA GCCAGAAGCC ATAAAGGCCC 2340

CGCGTAAATA GGAACGCATT TTTTGCGCGT TACCCATGAT TTCATTTGAC ACATTGCCAT 2400

TAAAAACGAC CCCATCCATA ATATCCAAAT CAGCTAGAAT TTTTTGTGTA TCTTGTTTTA 2460

GACGCACAAT GTAAACATTA TTTTTCTTTA GTTTCATTTT TTTACGAACT AATAATTCTG 2520

CTTGTGCATG GTAATGATCT TTCAGCAATG AATACATTCG ACGAGCAATT GCAGCATTTT 2580

CCGTTTGTAC ATTCAAGACA AACTGTTGAT TAACTAAGCT AAGCGAGCCG TTCATCCGAA 2640

TTAATGCTGC TAATTCTGcT TTCGCGTGTT CTCGATGGAC TTCTAAACCA GTTAACTCTT 2700

TTTTGACATC CGAGGCAAAA GACATCTAAG CTCCTCCTCT CTTGATTCTT TTGATTAATT 2760

AATATTTCGT TCCAAATACG ATACGGAAAA GTTCTTCAAC AACTTTTTCC CCATCATGAA 2820

AAACTCCGCC ATCTCTTAAT TCTAAAAAGT CTGTCGAAAT GACTCGGCAA CCTTCTTCAC 2880

GTAATCCTTG AAAATCATGC TTTACTTGAA CTAAATATTC ATCATAAATT TCAGGATCCA 2940

TATAATTTTC TGGTACTTTT TCTGTGTTAA CCAGCACCGT ATCAACAAAC TGCGCTTGTA 3000

GATGCTCGTT TAAGACACGA ACATGATCTG CATCCGTAAA ATGTTCCGTT TCCCCTTTTT 3060

GGGTCATGAT GTTACAAATA TAGACTACTT CGGCTGCCGT TTGTTTGATG GCTTCACCAA 3120

TTTCTGTAAT AACTAAATTA GGCAAAATAC TTGTAAAAAG ACTCCCTGGC CCTAACACGA 3180

CCATGTCTGC TTCTTCAATT GCTTTGACGA CTTTACGCGC AgcTCTTGGC TGCTCTTCTC 3240

CGTGTGTATT TGTTACAAAC ACATGGTCAA TCGTCTTGCG ATCCAGTGCA ATTTTTGATT 3300

CTCCCACCGC CACTGAGCCA TCTTTAAATA CAGCATGTAA CGTTAATGGC CGTTCTGATG 3360

ATGGATAAAT CCGTCCATCA ACATGCATCA TTTTTGAAAG CAATTGAATC GCTTCATACG 3420

TACTTCCTCT CATTTCTGAG ACAGCAGCAA TGATTAAATT ACCGATTGCA TGATTGGCAA 3480

AATGACTATC TGATTTATCA AAACGATATT GAAAAATATC TTCATATAAT TGTGGCATAT 3540

CTGACAATGC CACTAAAACA TTTCTTAAAT CTCCCGGAGG TGTCATGTTA TTAATAGAGG 3600

AACGAAGCTC GCCACTACTA CCGCCATCAT CAGCGACCGT TACGACGGCA GTAATATCAA 3660

CACTTTGATT TCTTAAACTT TTCAAAATTA CTGGCAAGCC TGTTCCACCA CCAACAACCA 3720

CGATCTTCGG TTTACGAATT CGATAGGTTT TCATTAAGAG CGATTAACCG TTTCTTTTCT 3780

TTTTAAGCGG TCGCGGTGGG TAATATTTAC ATGGTAATCT TCCTCTTTTA ATTCAGCGCC 3840

AACACGTTCT GTCAATGCAA CAGAGCGATG TTGTCCGCCA GTACAGCCAA TTGCAATGGT 3900

GACGCTACTT TTACCTTCTT TTTTATATCC AGGTAAGACT GTTCGCAGTA AATCGATAAA 3960

TTTTGTATAA AATTCATCTG TTTCTGGGAA GCCCATCACA TAATCATAAA CAGGTTGATC 4020

CATTCCAGTT AACGGTCGTA GTTCATCAAT ATAATGGGGG TTAGGTAAAA AGCGAACATC 4080

CATAACGATA TCTGCATCAA TCGGCAAGCC ATATTTAAAG CCGAAGGACA CCATTTCCAC 4140

TCGAAATTCG TGTGTCTCTC TCGTTGCTAG TTCTTTATTC AAACGTTCCC GCAATTGACG 4200

AGGGGACAAG TCCGTAGTGT CAATAACCAA CTGAGCATCC GCCTTGATTT CTTCCAGCAT 4260

TGCACGTTCT TTACGAATAC CTTCTGTCAC TAAACCATCC ATCGCCATTG GATGCGCACG 4320

CCGTGTCTCT TTATAACGAG AAACCAACTC TTCATCGGTG GCATCTAAAA ATAAAATCGT 4380

TGTATCAATA AAATTAGTAT TTTCAAGTTC GACTAACATA TCCTGAATTT CTCTAAAGAA 4440

GGTCCGTGAA CGTAAATCAA TCACTAACGC AATCTTGGTA ACTTTTCCAG ATTCTTTAAT 4500

TAATTCCCAA AATTTTGGGA TTAAGCTTGG TGGTAAATTG TCGATACAGA AATAGCCCAT 4560

GTCTTCAAAG CTTTGAACAG CGACTGTTTT TCCAGCGCCA CTCATTCCTG TAATAATAAC 4620

CAGTTGTAAA TTTTCTGGCA TATTCAACAC TCCCTTCACT TGCATCACTA TATTATAACA 4680

CCCCGGGCGT ATCAAAGAAA GCTTTTCAAT CATTCTTATA AAGATTTCTA TTCTTTCATA 4740

GAAAAACAGC ATTGACATTT TTGTCAATGC TGTTTTAATT GCTGAATTAT ACTATATAAC 4800

TCTGTGTTAT TTTAATACAA ACTCAATTTC TGTTTTATCA GTAAAATTAG CTTCAAACTC 4860

AAATTGTTTT ACTTGTTCAT TCACGGCTAA CTTTTTCGTA ATGTGTTCTT TAATTTGTTC 4920

ATTTGTAGCT GTTTTGATTG ACAATCCATC TAAAAGAGTC ATGATTTTGT CTTGTGCTGC 4980

CTTACCTTCT AATTCTAAAT CAGCAGATTC ATCTTCATAC GTTGCTTCAA TTGTTCAGTT 5040

CGATTTTATT TCATAGGTGA ACTCAATTTC ATTTTCGTCG TATTCAATTT CTACGTCAAT 5100

CTCTTGAAGA CTGGCAACAG TCGGTGTCGG TGTGGTTAGT TCATTGTCTG CTGTTTCATG 5160

GTCCCCACCA ACTTCAGGAG TTACAGTTTC CCACTTAAAG GTACCAATTT TAGCTGTTGC 5220

TAGATAGGGA TTATTAATGG TTATGACTTT CAATTGAAGG TCTGAATGGT TGGCGATATC 5280

TGTTAAAATC CCCAGAGCCT GATATGTATT TTGTGTAAAA CGTGTCGCAA TATTTTGCGC 5340

ATCAATTGTG GAATAGCCTA ATGTTTCCCA GGTGATTTGT TGCTTTAATT CCTGATGAAC 5400

GGCTGTACTG TCGCCGGCTC CAATTGAATG CCCGTTGcTt CTtTTAAGnC TTCCTGGnAA 5460

AACTCATTAC TTTTTCATGG GCTGATTGTG AAATTAAnGT CGCTGCTCCA TGGGGTACnA 5520

ATGACAAAAG 5530

(2) INFORMATION FOR SEQ ID NO: 207:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2571 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

TTGTTAGGGG ATGAGATGTA TCATGGAAAG ATGGATCAAG GAATTGAGCG GCAAGCGTTG 60

CATTGCATGG AACTAATTTT TACTCATCCT TTTACCAAAG AAACGGTTCA TCTGATTTCC 120

CCACTnGCGG AGGATATGAA AAGTGTAGAT GATACGTTAT AATAATAAAG AGGCAGGAGG 180

GGAAAGTGTA AGATGAATGA AGGCCAAGAA ATGGAAGAGC AATTTGCCCT TCTTTTAGAA 240

ACGCTAAAAA ATCAACAAAT GAACGAATTT CGTGAGCTAT TTTTAGCGTT ACACATTTAT 300

GAACAAGGAC AATTTTACCA ATCGTTAGAT GAAAAAGACC GACAACATTT ATACAATTAC 360

CTATCTCCGA AAGAACTAGC GGATATGTTT GATGTTATCG AAGAAGATAA TGAGAATATG 420

AAAGATTATT TAGCTGAAAT GCGCCCaAGT TATGCAGCCG ATATGTTAGC TGAGATGTAT 480

ACCGATAATG CGGTGGATTT ACTAAATATG TTAGATAAGA GTCAAAAGGC GAAATATCTT 540

AGTTTGTTAA GTAGTGAAGA AGCAGGCGAA ATTAAAGAGT TACTTCATTA CGAAGATGAA 600

ACTGCTGGGG CGATTATGAC CACCGAATTT GTTTCGATTG TGGCGAATCA AACAGTGCGT 660

TCTGCGATGT ACGTCTTAAA GAATCAAGCG GACATGGCGG AAACAATCTA TTACGTTTAC 720

GTAGTCGACC AAGAAAATCA TTTAGTCGGC GTTATTTCGT TGCGGGATTT AATTGTTAAT 780

GATGATGATA CCTTGATTGC TGATATTTTG AATGAACGGG TTATTTCAGT TCATGTTGGC 840

GATGACCAAG AAGATGTCGC GCAAACAATT CGTGACTATG ACTTCTTAGC TGTGCCAGTT 900

ACAGATTATG ATGATCATTT GTTGGGGATT GTAACAGTTG ATGATATTAT TGATGTTATC 960

GATGATGAAG CCGCTAGCGA CTACTCTGGT TTAGCCGGGG TCGACGTTGA AGAAGTTAGT 1020

GAGAACCCTT TAAAAGCCGC TTCTAAACGA TTACCTTGGT TAATTACGCT GCTGTTTTTA 1080

GGGATGTCTA CGGCCTCGTT AATCAGTAAT TATGAGTCGT TAGTTAGCGA AgCGAGTATT 1140

TTAGCCGTTT TTATTTCTTT AATTACCGGA ACAGCCGGCA ATGCCGGGAC ACAATCATTG 1200

GCGGTCGCGG TACGCCGTTT GGCTATGAAA GATGAAAAGG ATAGTAATTT TGGTCGCTTA 1260

ATCTTAAGTG AAGTGCTCAC GGGACTAGTG ACGGGGGCTG TGACAGGTTT AACCATTATG 1320

ATTGTGGTTG GCGTGTGGCA ACATAATCTT CCGCtTGGAT TTGTGATTGG GATGGCAATG 1380

CTCTGTGCAA TTACTGTGGC GAATTTAGCA GGGAGTTTAA TCCCAATGTT AATGGATAAA 1440

TTAGGGTTTG ATCCCGCAGT TGCCAGTGGC CCGTTTATTA CTACCTTAAG TGATTTAACC 1500

AGTGTATTAA TTTATTTCAA TATTGCCAGC ATGTTTATGC GTTATTTTGT TTAAAAAATT 1560

GCACTTATTT ATCATGAGAC TGAGGTAAAA GGATTCGTTG GTTTGCTTGA AATCAAAAAA 1620

GTGAGATAGC CGTTAATACG ACGGCTrCCT CACTTTTTTA GATACAAGAA ATTAAACAAT 1680

TTTTGTGCCA TGGACTTCTG TGACkTTTAA AgTGTCCATC ACAGTTTGAA CGTTTTCTGT 1740

GGAAATACCG CCACCTGGkA AAATAAGAAT CCGkTGaTCT GCGTAGACaA TTAAwTCTTT 1800

TAAAkGATCA AAGTwGtCTy CGATAGGGGk ACCAGCyGGc CCGCCATGAG TTAAAATACG 1860

sGkCACACCA CGCTCAGCTA ACCaGTCGAT CGCCTTAAAT TGaTTTTCTT TCGaCAAGGC 1920

aTCAAAaGCC aTATGGAAAG TGATTTGCaA GCCTTCaGCC gTTtCAATGA ATAAGTcCAA 1980

AgCTTCTTCa TCTAACCaAC CGTCCTCAGT TAAAaCAGCC cAATAACcAA TGCCCGTCGG 2040

TTCCCTAATT TCTTTGCCTC GATTAAATCG GTATGCATAA TTTTTAATTC AATATCATTA 2100

TAGACAAAAT TGCCGCCACG AGGACGAATG ATTGTCAtCA CTGGTACAGA ATGTtCGCCa 2160

GCGtATGCTA AGaCTTCTTC AATAACACCA GTACTTGGCG TCGTTCCACC CACGGCTAAG 2220

TTATCGCATA ATTCGATGCG ATTCGCACCT TTTTGAATGG CTTGGGGAAT CTTAGTGAAA 2280

TTTTCGGCAC AGAATTCTTT AATCATTTCT TATCGCTCCT TGCTTTTTCT TCACAGGCAT 2340

TGTACCATAA TTCTGCTTAC TACCATAGCC CTGAAAAAAA GAGTGCAACT TTTTATAAGA 2400

GAAAATGTTC ATTTAATGAA CGTAAGGAGT TTTAAATGTT AAAGAAAATT GAACGAGCAC 2460

AACAATTATT GAAAAAAGAA GCAGCCGTTT TTCGCTGTCC AACTTGTCAT GAACCAATGC 2520

ATTTAGAAGG GGTTGGTCTG ATTTGTCAAC AGAGGCATCA ATTTGATTTA T 2571

(2) INFORMATION FOR SEQ ID NO: 208:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2988 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

AAAGTTGTTG ACAAATAACA AACGAGATGn TAAGATATAG AAGTTGTCAC nACAGACAAC 60

AACAACCAAG CAAAACAACT TAAAAAAACT TTTTyAAAAA AgTTCTTGAC ATTCAAACGr 120

AAGTTTGTTm TGATATAAAA GTTGCTGCGA GGTAACGTAG ACCTTTGAAA ACTGAACAAA 180

GAATAGACGA ACCAAATGTG TAGGGCGTCT TGATTCAATT CAAGACAACA AACATTTTTA 240

ACAAGCAAGC AATATGCTAG CAAACAAATT GAGCTTAACA ATCGTAAGAT TGTTCTAACT 300

TTTATTATGA GAGTTTGATC CTGGCTCAGG ACGAACGCTG GCGGCGTGCC TAATACATGC 360

AAGTCGAACG CTTCTTTCCT CCCGAGTGCT TGCACTCAAT TGGAAAGAGG AGTGGCGGGG 420

CGATGCCAAA AAGCAAACGC TTGTTAAAGT AACACTGAAA GGGCAATTAA AAGAAGTCAC 480

GTTTCCAGAA ACAATTAAGG TCTTTAATAA TtGTAAAACT GGGACATTGT TTGGCGATTG 540

GGCTGACGTA AAACCTTTTT TAGAAGCAAA CAAAGAAAAA ATTGAGGATT ATGTTGTAGA 600

AAACGATGCA CGTAACTCAG CAATTCCTTT TCTTGATCTG AAAGACATTA ATGCTCGGAT 660

TGAACCAGGT GCTTTGATTC GCGAAAAAGT TGAAATTGGC GATCAAGCGG TGATTATGAT 720

GGGGGCTATT CTAAATATTG GCGCCGTCGT TGGTGCAGGG ACTATGATTG ATATGGGCGC 780

TGTTTTGGGC GGTCGTGCAA CTGTAGGGAA GCATTGTCAT ATTGGCGCAG GAACGGTTTT 840

AGCTGGTGTC ATTGAACCAC CAAGTGCCGC ACCTGTGGTT ATTGAAAACG AGGTTGTCAT 900

TGGAGCGAAT GCGGTCGTAC TAGAAGGGGT TCGTGTTGGT GAAGGTGCCG TTGTGGCTGC 960

TGGGGCGGTA GTTGTGGAAG ACGTTCCAGC GCATACCGTT GTTGCAGGCG TGCCAGCAAA 1020

AGTAATTAAA CAAATTGATG ATAAAACCAA GAGTAAAACA GAGATTTTAG AAGAATTACG 1080

TAAACTCTAA AAATAAATTA GTCAGTCGCG ATGAAGAAAC TTCTTCATGG CGGCTTTCGA 1140

TTACCCAAAA AATGAAGGAG TGGCTTTTGT GGAGCAAGAA GAATTAATTG CGATTAGACG 1200

GCAACTACAT CAAATACCtG AAATCGGTTT AGaAGAAAAG GAAACACAAG CCTTTTTACT 1260

TAATGAAATT GATAAAATGA AGCAACCTTA TCTGCAAGTT CGAACATGGC AAACGGGGAT 1320

TTTAGTCTTC ATAGAAGGCA AAAATCCTCA GAAAACCATC GGCTGGCGTG CTGATATTGA 1380

TGGTTTACCT ATTCAAGAGG AAGTGGTTTC TGCTTTTCAA TCAAAACGAC CAGGGTTTAT 1440

GCATGCCTGT GGGCACGATT TTCACATGAC AATCGGTTTA GGGGTTTTGA AAGAATTAAG 1500

TCAACAGCAA CCAGACAATA ACTTTTTATT TTTATTTCAA CCTGCTGAAG AAAATGAAGC 1560

AGGGGGGATG TTGATGTATG AAGACCATGC GTTTGGTGAG TGGTTACCTG ATGAATTTTA 1620

TGCGCTTCAT GTTAATCCTG ATTTGCCAGT CGGAACTATT AGTACCAGAG TAGGGACATT 1680

GTTTGCTGCC ACTTGTGAAG TGAACATTAC CTTAAAAGGA AAAGGTGGAC ATGCGGCTTT 1740

TCCGCATCAA GCGAATGATA TGGTCTTAGC GGCCACAAAT TTAATTCAGC AAGCGCAAAC 1800

GATCGTTAGT CGAAATGTTG ATCCTGTGGT TGGGGCAGTC GTTACCTTTG GAACTTTTCA 1860

TGCAGGGACT GCCTGcAATG TGATTGCCGA AGAAGCAACT TTATCTGGTA CAATTCGCAC 1920

ATTGACAGCA GAAACGAATG AACAAACGCA GCGAAGAATT CGCGAAATTA GTGAAGGTAT 1980

TGCGCAATCT TTTCAATGTG AAGTAACTGT TCATCTTGAT CAGAAAGGGT ATTTACCTGT 2040

AGTGAATGAG CCAGCATGTA CCACGAATTT TATTGAATAC ATGAGCAAAC AAGCAACAGT 2100

TCAATTTCAA CAAGCACCGG TTGCAATGAC AGGTGAAGAT TTTGGCTATC TATTGTCTAA 2160

AGTCCCAGGC ACGATGTTTT GGTTAGGCGT GGCAAGTCCC TACAGCCTTC ATTCTGCTAA 2220

GTTTGAACCG AACGAAGAGG CGTTATTGTT TGGAGTGGAG GCAGTTAGTG GCTTTTTAAA 2280

ATCATTGGAC AATTAGAAAA ACAGCCACCA GATAACTAAT AAATATCTGG TGGCTGTTTT 2340

TTATTTTGCA ATAATGGGCG TATTGGGAAT CGTAAAACCG GCCTCTGTTA ATGCTTTGAT 2400

ATATTGGGCT AGAAATTCTT CTTTGACCGC AAATTGTTTG CCGTTTAGTA CATACATGGT 2460

CGTTCGCACC GCAAAGTTGC CATTGCCTAA ATCGACCATa CCAAAAATAT CTGGCTCAGT 2520

TTGAATACTT TCGATATACT TTTCTTTTAA TGTTTCATTG ACTTCTGTAA TCACTTCACA 2580

AATTTTTTCA TAACCTTCTT CAGGATTGAT GCGGACATCA ATTAACACTT GCATATTGGC 2640

ACGAGACAGG TTACTGATTG TGGTGATGTT TCGGTTTGGA ATAAAATGAA CAGTTCCATC 2700

AACCGCTTTA ATTTGAGTGG TTCGTAATCC GACAGAAGTA ACAGTTCCTT CAATATTTAA 2760

CGCCAATAAG CGAATATAAT CGCCTACATC CATTTGTTGT TCCAAAATAA TGAAAAATCC 2820

GGTAATTAAG TCATTCATGA ATCCTTGAGC CCCTAAACCA ATGGCAACAC CGGCAATTCC 2880

AGCTCCGGCT AATAATGAGC CAACAGGCAC ACCGACAATG GTCAAAAGAG AATAGATGAA 2940

GAnAAAGAAC AGCGTATATG GAAATGCATT AATGGnTTAn AnTGGGTA 2988

(2) INFORMATION FOR SEQ ID NO: 209:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1510 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

```
CnTATAGACA TATTCTCnCC AACTTATTTT TTGGTGGGAG TTCCCGGTTA TTCnTCCATA      60
TTGTTTTCnG GTAATTCGTT GGTAACTAAC AAATGTTTCC ATATTGGTCA TCTGTCAATC     120
CAAGTAAGCC GTATTCAGCT TTAATATCTG CATATTTAGG ATAATCTAAA ATAACTTTGC     180
CGATTAGTGG TATATAAGAC AATGCCCAAG AAATAACCGC AAAGTTCAAC ATTACATTCC     240
CAACAAAAGA AGTTGTTCTA TTATTTGTTA GGCCATTTAA ATCATCAGGA TAATGAACTA     300
ATTTGCCATT GCGTAAACCC CAGTTCAAAT AATCTTTTGG CTCTTCATTT ACTTCAATTT     360
CAATATAGTT TTTCTTAGCA ATAAACGAAT AGTCCGTAAG CATTTGATTT TCGTCCATTG     420
TAATCCATAT TTTTATCTTA TCCATGTTGC ACTCCCTCCA AATTCTACAT CATCATTTTG     480
AAGTCCATAA ACTCGAATCA TAATTCGACC AGATGTTTCT ATTTGTATTG CTGCATATTT     540
ATTTGTTTGA ATCGTATTGC CTATATTTAA GGTTAACGGT ATTTTAATAA TCATATTTTG     600
AGGTCTATAG CTAGCTTGAA TATAGGTATC CACACCATGC CAACCATTGG CTTTtCCTTT     660
TTTAAAAATT CCTTCAACAT TACATGTAAT AATTGTATTT GATCTTTGAA TGTGAATAGT     720
AATTCCATTG TAAGAAATAT GTTGACCAAC AGCATTAAAT GCCGTTTTGT CAATCTTTTT     780
ATTTAAATCA TTTGTATACT GTGTTTTAGT AACTAAATTG CTTATATCAG GCGTATCACC     840
TTTAGGTCCT TGCTCTCCTT TTGCGCCTGG AACACCATCG TTTCCTTTTT CACCTTTTGG     900
ACCTTGTGGA CCAGTATCCC CCTTGTCACC TTTAGGTCCT TTTAAACCAA TCGGACCTTG     960
CGAGCCTGTT TCTCCTTTTT CTCCTTGTGG TCCAATGTCT CCTTGTAAAC CTTCTGGGCC    1020
TCTAGGGCCT ATATCCCCCT TATCTCCTTT TGGACCTTGT GGACCAATAT CGCCTTTGTC    1080
TCCTTTTCGA CCTTTAAAAT CTGTAAAACT TTTTTTATCA AAGGATAAAA CATGAGCACT    1140
ATTGTTTGGC AATATATACG CAACACATTC TTCAGCAGCT TCTATAGCTG GTAGTTCATC    1200
TACACGTTCA AATGAACAAA TACTTACATC TGATTCACAT GGACTAAATC CTATATTGCC    1260
ACAATCAGAA CAAAAGCTTC TAACCCTCAT CATCAATAGC CCCTTTCATC ATATCGCTAA    1320
AGCAAGAAGC ACATTCGGTA ATGGGTACTC CTGTAAATTG AGTGATTTTT TTGATGAAAT    1380
AAGCATTTTT TTCAATACAT ATTTGATAAA GTAAACGATC GTTATCACTG GcTTGCCAAG    1440
CTTCAAAAGC TGTATACATT GCCATTGCAG CGTGTTtAAC TATACACCAT TGTCCCTTAT    1500
CACCTTTACC                                                           1510
```

(2) INFORMATION FOR SEQ ID NO: 210:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7287 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

```
ATTGAAGTTG CTGCAAATGC yTCGkTGATT TCATACAGAT CAATTTCTTC CGTAGTAAGT      60
```

```
TGATTGCGCG CTAACAGTTT TTGAATGGCT TTaATCGGCG AAATTCCCAT ATAGGCTGGA    120
TCAATACCGA CTTCCACACT GTCTCGAATA ATAGCTAAAT AAGGAAGACC GTGTGCTTCG    180
GCATATTCTT GTGAAGCAAT AATCAAAGCA GAAGCCCCAT CATTAATGGT TGATGCATTC    240
CCTGCTGTTA CAGTACCGTC TTCTTTAAAA ACTGTTTTAA GCGTTCCTAG CTTCTCAACG    300
CTCGAATTAG GGCGAATCCC TTCATCTTTC TCCACAAGCG TTCCTGATAC TTCTAATGGG    360
GCTATTTCGT CAGCGAATAT CCCTTCTGCT TGTGCTTGAG CTGCTTTTAA TTGTGAATGT    420
ACAGAAAATT GATCTTGCTC TTCTCTAGTT ACATGATACT TTTCGGCCAC ATTTTCAGCA    480
GTTAAGCCCA TTGCCTGACC ACTAAAGGCA TCCGTTAATC CATCATACAT CATACTAGAA    540
AAAGGCGCAT CGTAGCTTTC TGTTTCGTAA TTAAAACGTT GTAATTTAGG TGCTTGGGAC    600
ATATTCTCAA TCCCGCCAGC AATTAAAACT TCCgCTTCTC CTAATTGAAT CAATTGTTTC    660
GCCAAAATAA CGGCCTTCAT TCCTGATCCG CAGACCTCAT TAACCGTCAT TGCGGGAATT    720
TCATGAGACA AACCGCTGTT TATTGCTATT TGTCGTGCGG GATTTTGGCC ATTTCCAGCT    780
TGTAAAACAT TTCCAAAGAT TACTTGATCA ATTTCTTCAG AAATAGTGGA ATGTCTTTTT    840
AAAAGTTGTG TTGTAACATG TGTTCCTAAG TCTACGGCAC TTACTTGACT TAAGCTGCCT    900
TTATATTTTC CAATTGGTGT TCGTAATGCA TCAATAATAA CTACTGTTTT CAACTAAATG    960
CACCTCCATC TTGaAATATA TCATAAAATC CCAGAATCAC AAGTGCCAAT TACTGAATGT   1020
TAACAGACTT ATAAGATATT TTTTAATAAT CTATGAAGTT AATACAAAAA TTCCCCATTT   1080
TGGTTCATTT CGTGCTATTT TATATAATGA ATCTATTTAT CGTAAAGGAG TTAAAGAAAT   1140
GACAATTGGG ATTGATAAAA TTAGTTTTTT TGTGCCCCCT TATTATATTG ATATGACGGC   1200
ACTGGCTGaA GCCAGAAATG TAGACCCTGG AAAATTTCAT ATTGGTATTG GGCAAGACCA   1260
AATGGCGGTG AACCCAATCA GCCAAGATAT TGTGACATTT GCAGCCAATG CCGCAGAAGC   1320
GATCTTGACC AAAGAAGATA AAGAGGCCAT TGATATGGTG ATTGTCGGGA CTGAGTCCAG   1380
TATCGATGAG TCAAAAGCGG CCGCAGTTGT CTTACATCGT TTAATGGGGA TTCAACCTTT   1440
CGCTCGCTCT TTCGAAATCA AGGAAGCTTG TTACGGAGCA ACAGCAGGCT TACAGTTAGC   1500
TAAGAATCAC GTAGCCTTAC ATCCAGATAA AAAAGTCTTG GTCGTAGCGG CAGATATTGC   1560
AAAATATGGC TTAAATTCTG GCGGTGAGCC TACACAAGGA GCTGGGGCGG TTGCAATGTT   1620
AGTTGCTAGT GAACCGCGCA TTTTGGCTTT AAAAGAGGAT AATGTGATGC TGACGCAAGA   1680
TATCTATGAC TTTTGGCGTC CAACAGGCCA CCCGTATCCT ATGGTCGATG GTCCTTTGTC   1740
AAACGAAACC TACATCCAAT CTTTTGCCCA AGTCTGGGAT GAACATAAAA AACGAACCGG   1800
TCTTGATTTT GCAGATTATG ATGCTTTAGC GTTCCATATT CCTTACACAA AAATGGGCAA   1860
AAAAGCCTTA TTAGCAAAAA TCTCCGACCA AACTGAAGCA GAACAGGAmC GAATTTTAGC   1920
CCGTTATGAA GAAAGTATCG TCTATAgTCG TCGCGTAGGa AACTTGTATA CgGGTTCaCT   1980
TTATCTGGGA CTCaTTTCCC tTTTAGAAAA TGCAACGACT TTAACCGCAG GCAATCAAAT   2040
```

TGGTTTATTC AGTTATGGTT CTGGTGctGT CGCTGAATTT TTCACTGGTG AATTAGTAGC 2100

TGGTTATCAA AATCATTTAC AAAAAGAAAC TCATTTAGCA CTGCTGGATA ATCGGACAGA 2160

ACTTTCTATC GCTGAATATG AAGCCATGTT TGCAGAAACT TTAGACACAG ACATTGATCA 2220

AACGTTAGAA GATGAATTAA AATATAGTAT TTCTGCTATT AATAATACCG TTCGTTCTTA 2280

TCGAAACTAA GTATTTAAAT CCCCCGATTC AATCAACAAA AGAGCTAGAC ATTCAAGCAA 2340

AAACTTGAAT GTCTAGCTCT TTTTATGAAA GCCAAACTTC TTTGACCCCG TCTGTCACGG 2400

CGATACCATC TGTCAAAGCT AAACGTTGTG AATAAGAAAC GTAGCCATAA TAAACATCAA 2460

CACCGCCGCC ATTGTTCCAA GTAACAGTGA TTTTAACATT TGGAGTAATC GATTTTGTAG 2520

CTGGTTGGGT CGTCGGGCGA TAACTAGTCG TATCATAACT TAAAATAGTG CCCGAAGCTC 2580

CACTTATTTC GGCTTTTCCT GTGTAAATTT CGCCATTCGT ATTCGTCTGC GTGACCGCTC 2640

CATTCGCTTC AATGATTAAA CCCGCTTGCT TTTGATCTCC TCGTGGATAA TAAAAAAGAT 2700

ACGTGGTTGC CCATTTAGCA AATTCATCTG GAATTGGTTT AAGTGATCCC GTTGGTGCCG 2760

GACTCGACTT CGTTTGCTTT TCAACCGCGA CTTCTTCTTG TCTTTTTAAG CCTTCACTAT 2820

TTTGATTATT TAGTTTCTCA ACCGTTTCTT TAACGATCGA ATAATGAGGT TTCCCTAAAT 2880

CGCTAACAGG GATCGTAGCT AATTTTTCTT CTGATTCTTT ATATTGCCCC TCTGCCGTTA 2940

ACTTTTTAGC TTCTGTAATA ATGGCATCAT ACTTGGTTTG GTCTACGGAA GAGGATGTGC 3000

TTTGTCCCGT AGCTTCGACT TGTTTTGTGA CACTGGTTTT TTCTGTGGTC ACGTCTGATG 3060

AATTTCCACA AGCACTTACA AACATTGAGC TAACGATTAA TAAACATCCC ACAATCACAT 3120

GCTTTTTCAT TTTTAAATCC CCCATTTGTC TATCTTTGTT ATTTTTTGTT ACTTAAATTA 3180

TCCCATAGAT GAACAAAAGA AACCATTATT TATAAAAAAG CTTAGCAAGC GCTTCAAGAG 3240

AGGCACTTGC TAAACTTTTT TCTCACTGTT ATTGATTGAT TACTTTGACT AGTTCTTCAA 3300

ATAATAATGC ACTAGATGCT GCTCCTGGAT CAATATGACC AATTGCTCGT TCCCCTAAAT 3360

AAGAGGCACG TCCTTTTTTC GCCACCCAAT CTTTTGTTGC TTCAGCAAAT TGTTCAATTT 3420

TTTCTTCATT AAATTGTTCC GTTCCAATTA CTTCAGCAAC AGGTGCCCAG ACATCAACCA 3480

TCGTTTTATC TCCTGGTTCA GCTTTGCCAC GCATCTGAAT CCCAGCTAGA CCTTGCTGAA 3540

CAATTTCTTT TAATTGCTCT GGCGAACTGA TGGTTTCCAA ATCTTTTGTC GCTTTAGTCA 3600

TATTCATAAA CGCAGTACCA TATAAAGGAC CAGAAGCTCC GCCAACTTTT GAAATCAAAG 3660

CCATGGATAG GACTTTAAAT GTCTCCGTGA TTGTTGTTGG TACTTTTTTG TCAAACGCAA 3720

TTTTGTATTC AGCCATACCA CGAGCCATGT TGTTACCGTG ATCCCCATCA CCAATCGGCG 3780

TGTCTAATTC ACTTAGATAG TCTTTTTTTT CTTCAATAAC TTTTGAATAA TTGTCTAACC 3840

ATGTTTGAAT GTCTTTTACT GTTAATTCCA TCTTTTATCA CCCTTACCAA GAAATTGTTG 3900

TTACATTGCT TTCTAGCGCT GTTTGCCAGT CTTTTGTCGC CAAGTCAATC ATCGTTAAAG 3960

ATAATCCTTG CATGTCAATA GACGTCATAT AGTTACCCAC TTTATGGAAG GTCACTTGAA 4020

CACCTTTATC TTCTAACAAT GTTAGCACGT CGTTCATAAA TACAAATTGT TCCATTAATG 4080

GCGTGCCACC CATGCCATTG ACAAGTACCC CTACTGTTTG TGGTTGTTGC TCATAGCTAC 4140

TTAATGTTTT TTCAACTAAT TCTTTCGCTA ACTCTTTCGA TGGTTGGATT TTTTCACGAC 4200

GATAGCCTGG TTCTCCGTGA ATCCCTACGC CAAATTCAAT TTCATCTTCT GGTAAAGTAA 4260

AACCTGGTTT GCCAACTTCA GGCACAGTGG CTGCTTTTAA AGCCACTCCA ATTGTTTTCG 4320

TGGCTGATAC AATCTTTTCT CCTAAAGAAA CGAGCTCTTC TAAAGAAGCA CCTTGGCGGG 4380

CATGATGTCC TAAAATTTTA TGGACAAGAA CAGTCCCGGC AACGCCGCGT TTTCCAGCGG 4440

TATAGGTGCT ATCTTCGACA GCAATATCAT CATCGACAAC GACCATTTCT ACATGAATGT 4500

CTTCCATATC AGCCATATCT TTGGCCATTT CAAAGTTCAA AATATCGCCT GTATAATTTT 4560

TTACGATTAA CAGAACACCG GCCCCTTGGT CTGCCGCTTG GATCCCTGTT AAAATTTGAT 4620

CTGGTGTTGG TGAAGTAAAG ACATCACCTA AAACTGCGGC ATTTAACATT CCATCACCAA 4680

CAAACCCAGC ATGTGCAGGC TCATGACCAC TTCCGCCACC AGATACTAAC CCAACTTGTT 4740

TGTTTGAGTC GTTGCGAGCA ATCACGCGAG AATCTTCGAC TTGATGAACA AGGTTTCCAT 4800

AGCTTTTCAC AATTCCTTGG AGCATTTCTT CGACAATTTT ACCAGGTTCA TTAATAATTT 4860

TTTTCATTTA GTTCCACCCT TTCTTACCAG CATTTTGAGC TTCTGCTAAA ATTTGCGTCA 4920

AATCATCTGA GACACCTGCA GTAATTGctG ACGCAAAGGC ACCTTCAACC AATGGTGCAT 4980

CGACTAAATG ATAATGTTTC TGTTGGTCTT CTTCTAGCAT ATCAAAAGCT AATTCTGAAC 5040

TTAAAACGGC ACTGCCTAAA TCGGCAAAAA TTAAAAATTC TCGGTCAGAG TCCGCTTCGT 5100

TAATCGTGTC AATGATTTTC ATTGGATCTG AACCTAATGA ACCATCTGAT GTACCACCAA 5160

GTGAATGAAT TGTAATTTCT tCACTGGCGT TCATCTGTTC AATCATCTCT TTAATACCAT 5220

CTGTAATCAT TTTGCTATGT GAAACGAGTA AAATATCTGC TTTCATGTGG ACACTTCCTT 5280

AATCTAATTT ATGTCTTGTT TTATATTCTT GTCCGACTTT ATCTGCAACA ATGATTGCAT 5340

TCGCTACTTC TTCAACGACA ATTGGGAATG GCATTGAATG AATTGATTCT TCTGGAATGC 5400

AAGCAATTTT GGCTACTTCC AACGCTTCTT CATAAGTGAT TTCATCGACa CCAATGTCTG 5460

AAAGCGTTAA TGGTAAACCT ACAGAAGCGC TAAAGTCTAA AACAGCTTCT AATTCGTCTT 5520

TTGCTGCATC TTCTAATACA AGTTGTGCAA TCGTTGCAAA AGCTACTTTT TCACCATGTA 5580

AATAACTGTG CGCACCAGGT AAGACTGTCA TACCGTTATG AATCGCATGC GCTGCAGCTA 5640

AACCGCCACT TTCAAAACCT AAGCCTGAAA GTAAGATATT TGTTTCGATA ATCTTATTCA 5700

ATGATTCAGT AACCACATTT GCATCACAAG CTAATTTCGC TTTTAACCCA TCTTCCTGGA 5760

GGTTTTCCCA GCAAAGTTTT GCCATGGCCA AAGCTGCATA TGTGCCGCTA GCAGCAGGCG 5820

TTAAGCCTTC TCTTGAACCC ATTGGCAAAC TAGCATTAAC CCGAGAATAT GAATTGTGTG 5880

TTGCGCGTGC TTCAAAATAT GTGGATAAGG CATCCCCCAT TCCAGCAACT AAAAATCTAG 5940

TTGGCGCTTT GGCAACCACT TTTGTATCAA CCATAACGAC GCTTGGACTT TGTTTAAAGT 6000

AAGCATAATC ATCAAATGAG CCATCTTCAT GATATAGTAC AGCTGAATGC GAAGTTGGTG 6060

CATCTTGAGC CGCAATCGTT GGTACAATAA TTAAGTTTTC ACCTTCTGCT ACGACTTTTG 6120

AAGCATCGAT TGCTTTACCG CCACCTAAAC CAATGATACA GTCTGCTTGT TCTTTTTGTC 6180

CAACGTTTTG CAAGCGCTTT GTTTCTTCAC GCGTTACTTC TCCACTAAAG CCACCCTCTA 6240

CGAAGGTAAT GTCAAACTTT TTCGCTGTTT CTTCCAATkG TGrACGwACA CGATCAACGT 6300

CAGCAGGATT TGCaATTAGT AACGCTTTTT TTCCaAAAGT TGAAACGAAG TAACCTAAGT 6360

TTAGTAATTC ATCTTCCCCT TGAACATATT TCGTTGGTAA AATCAATGCT TTTCTCATGT 6420

TGATAGCCTC CTATATTTTG TTnACATCCA TAGTATAAAA CGCTTTCATC TAAAAAGCAT 6480

TGAACTTTTG GTAtCGTTTy CTTGCAAAAA ATGATAGAAT TATTTTGTGA AATAATAAAA 6540

TATTATGATT TTTAAACTTT AGATTGAAGG ATTGCCTTAT GACTGTAGAA TTAAAAAATA 6600

TTTTAGATTT AGAAAAGTGG GAAAAATTGC AAGAATCATT GGCCTTAAGC ACTCGTTTAG 6660

CCATTATTTT AGTCGATTAT AAAGGGCGAC CTGTGACGAA ACATAGTCAA GTTCAACCCT 6720

TTTGCCAACT GGTACGTCAC TCCCCTGAGC TCTCGAAACT TTGTGAAAAA TGTGATGCTC 6780

GTGGTGGCTT GGAAGCTGTT CGCACTGGTC AGCCTTTTAT CTACCGTTGT CACTTTAACA 6840

TCGTTGATAT GGCCATCCCC ATTATTGTTG ATGACCAGTA TGTAGGAGCC ATTATGGCTG 6900

GTGAAGTTTT GCTCGAAGAT CACCAAGAAG AGTTAGAACA GGTTTTAACC ATGAATGATG 6960

CCTTCATTCA AGAATTTTTA GTAACACACC AAGAGCTGTA TCAACAATAC CCTGTTTTAG 7020

AACTAGCAGA CTTAGAAAAG AGCGCGACAT TAATCGCTGA TCTAAGCCAA TATATTATTT 7080

CTGAAGCCAT CAAAAAAGAT TATTTAATCA ATACGTACAA ACAATCCTTA CAGATTAGTT 7140

CGCGAAAAGA AGAGCTGGAG CCATTGAATA TGATCCAAAA AGATATTCmA CAAmCTTTAT 7200

TAACTGCTCA ATTAACTCCT TCAAGkGCCT tCCAAmCTAA AAwTAAGCAa TaCCAcCGGG 7260

TTTAGATGnC CCTTGGGGAA ACCAATn 7287

(2) INFORMATION FOR SEQ ID NO: 211:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3756 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

GTATGCCAAT CGTTTTAATT ACTTGTTGGT CACCGCTAAC GATGCCCACA TCAATCAACT 60

TGCCAACTAA TTTAGGGACG CCAAGGGTTC CTGCGATTTG CAAACAGATG GTTAAAGTTG 120

TCGCAAGAAC AATCCATTTA TTTTCTTTAA TAAACTCTTT CATTAATTTC ATGAAATTTC 180

CTCCTTAAAC AATACCGTTT TAGCTGTTAC AAAGAGTCAT TGTACACGTc AAGTAACTTG 240

AATGTCAACA GTAGAAATGC TAAACTTTTT GATGGAGGGA ACAGTATGGA AAAAAATTAC 300

ACAtAAAAGa TGTTTCTGAG TTGTTAAATA TCCCTAAGTC CACGATTCGT TATTGGGACG 360

AGCAAGGGTT GATTAGTACA ACACGCAATG AAGAAAACGG ATATCGAACC TTTGATTTAG 420

AAGATTTGTT CAAAATATAT GATATTGATT TTTATCGAAA AATGGGGATT CCTATCAAAG 480

ATATGTTGAA TTTATATAAA AAGACACCTA CCGATCTCTA TGCCACCATG GAAGCAACAG 540

AAGAACAACT GGGGCAAGAA ATAAAAAGGC TTCAACATAA ATACAAAGAA ATCAAGAAAA 600

GAAAATTACA TTTACATGCG TTAATTGATT TAGAAACATT GGTTTTTTCA GATGAACCAA 660

TTCCTTTTGA TAAAATTGTT GTTGTTGATA TTGAGGATCC AGAAGAGATG AAAGTTTACA 720

TGGAACACAT TTCAACTTTA GGTATTTACG GGGATTTAAG TGAATCCAAT GAACTGACGT 780

ATGGGATTTG TTTGGCAAAT CATGTGAATA AAAAAAGTTT AGATAAAGTG ATTTGGCTCA 840

AAGAATCAAG TAGCCAATAC AAAGGCTTTA TTTTAAAAGT CAATACACAA GATACCCGTT 900

TGAATAATAT CGAAGAAGTA AAAAAGCGTT TAGCAAAACA AGGATTTGAA ACAGGCATCG 960

TTGTCGGAGA GCATATCTTG ACACAACTAG AAGAGGAAAA CCAATATATT GAATATTATT 1020

ATGGTTGGAT TGAAATTTTa GGGGAAAaGT AAAACAATTA AaAAAAGGAA CTAGCAATAG 1080

CAAATGTAGC TAAGCTAGTT TCTTTTTTTG GTAAATGTAA TGAAAATTAC ATGAAAAAAG 1140

TTTTAGTTTT TAAAACAATT TAGAACGAAT GTTATTAAAG GGAAAACAAT AAAAAAATTA 1200

CTCTTAAAGC TAGAACTTAC CTATGTTTGT AATTTTTTAG GGAAAGTCAA TATATTTGCT 1260

TCATATAAAA CGTGTTATTT AATTCACGAT TTTTTGTTGA AATATCGGGT TTTTTCTTTT 1320

TTTAGCTAAT ATAATGTGTT ACTTTTTGAT TATAAGGAGA AACGGAAAAT CGTACGTTAA 1380

AAAAAAGTTG TTGTAGCTAA ATAAAGAAAG AATTGAAGGT TGAAAAGTAT TTTATGGAGG 1440

TGGATGTTGA GAAACAATTA GTAATTGTAA AAGTTGTTTG AAGATGATTT GTTTAAGTAA 1500

ATGAAAAGAA CGAAGGAGAA TGTAAATGAA TAAAGCAGTT AAAAATTTTG TGAGTTACCT 1560

TATGATTACC ATGCTGTTTA TTTTAAATTT GTTACCAATG ATGAACGCAT TTGCTCAAGA 1620

AGTAACAAGT GATGCTGAGA AGACTGTTGA AAAAGACGGG CTCAAAGTAA TAGGGAAAAT 1680

AGAAGATACA TCGTCTCAAG AAGATATTAA AACGGTCACT TACGAAtAAC AAATACGCGA 1740

GATGTACCAA TCAAAGATCT CATTTTAAAA CAAAAAAATA CCAATGATAG TCCAATCAAA 1800

TTTGTTTTAG ATACTTTAAG TGAAGAGCGA GGACCTACGT CTTTAGAAGA ACAAGCAAAG 1860

GTTGAAACAA ATGAGAAAGA TCAAACAACC GATATTAAAC TGCTTAACCT TCAACCAAAT 1920

TCGACAAGAA AGATTACCAT AAATGGGCAA ATAACTACAA AAGCGTCGAA CAAGTTATTG 1980

GTCAGCGTAC TAATAGAAGA TAATGAAAAA GGAACACTAG TAATTGATTT ACCTAGTAAA 2040

GATATACTAG CAGATAAAGA GTCGGTTTCG AAGGAAAAAC AGGAAACTTC AGAAACAAAA 2100

GTAGAAAATC AGGCAAACGA AACAGCCTCT TCTACAAATG AAATGACCGC AACTACTAGT 2160

AACGAAACAA AGCCTGAAGC AGGAAAGGCG ATAGAAAGTA TTCAAGAGAC AGCACTAACG 2220

CAGGCTACCG AAAGTCCTGA GCAACCTCCA TTAAAGGCGC AACCTACTGG TCCGTTAGTG 2280

CCACCAACAC CTGGTCGGGG GTTTAATACA CCGATTTATC AAAGCGTTCA TAAAGGAGAA 2340

CTTTTTTCAA CGGGAAATAC CAACTTAAAA ATTGCGAATG AAAATACGGC GGCTGCACAA 2400

ACATTTTTAA ACACACGAGG AGCAAGTAGT GGTTATGCAA TTAACAATTT TCCTTTAGAA 2460

TTTGCAGATG TTGACAATGA TCCAAATACG TATAACTCTA GTCGGGCTTA TATTGATTTA 2520

AATGGTGCAA AAGAGATTGC TTGGGCGGGC TTATTTTGGA GTGCATCTAG ATATAAAGGC 2580

CCTGCTTACG GAACAAATCT TTCTGATGAA GAAATTAGTG CACCAGTTCA ATTTACTACA 2640

CCAAATGGAA CCGTACAGCG TGTTTCGCCC CAAAGGTACC ATCGTATCGA TCAAGATGCA 2700

ACAAACCCAG GACAACGTTT CGGGTACAAT AACACTGGAT TTTCTAATTA TGCAGATGTA 2760

ACTTCAATTT TACAAGGGGA TAAAAGTGCG ACAGGGAGTT ATACGTTGGC AGATATTCCT 2820

ATGACAAGTA GTTTAAATGG TCAATATCAA TATTATAACT TTAGTGGTTG GAGTTTGTTT 2880

GTTGTTACAA AGGATCAGGC AAGTAAGTCA AGAGCTTTTA GTATTTACTA TGGAGCACGT 2940

GGTAATGCTG CTGGAACCAA TAATGAATTT ACTATGAGCA ACTTTTTAAC AGCAAAACAA 3000

GGAAATCTTG ATCCAATTGT GACCTGGTTT ACTGTTCAAG GAGATAAATA CTGGACTGGA 3060

GACAACGCAC AAATTAAAAA TAGCGCAGGA ACTTGGGTAA ATATTTCGAA CACGCTCAAT 3120

CCAGTTAACA ATGCTATGAA CGCAACTGTG ACAGATAACG ATGAACATAT GGTAGACAAG 3180

TATCCAGGGA AATTTGCGCC GGATCATCCT AATTTTTTAG ATATTGATAT CGATCGTATG 3240

GCAATTCCCG AAGGTGTTTT AAATGCTGGA CAAAACCAAA TTAATTTTAG AACAACAAGT 3300

AGTGGTGACG ATTATTCTAC GAATGCGATT GGTTTTGCTG TAAACGCAGA AACGCCTGAA 3360

TTTGAAATTA AGAAAGAAAT TGTGGAACCA AAAGAAACTT ACAAAGTTGG TGAAACAATC 3420

ACGTATCGTG TCTCACTAAA AAATACAAAA GCTGATTCTG AAGCGATAAA TTCAGTCTCG 3480

AAGGATGCAC TAGATGGTCG CTTAAATTAT TTGCCAGGCT CTTTAAAAAT TATTAGTGGA 3540

CCAAATTCTG GTGAAAAAAC GGATGCTTCA GGCGATGATC AAGCTGAATA TGATGAAACC 3600

AATAAACAAA TTATAGTCCG TGTCGGCAAT GGTGCAACTG CTACACAAGG AGGTAGCTAC 3660

AAAGCTGATA CAGCTGAAAC AATTTACGAA TTTAAAGCGC GGATAAATGA ACGAGCAAAA 3720

GCAAATGAAT TAGTGCCTAA TAGCGCACCG TTGAGG 3756

(2) INFORMATION FOR SEQ ID NO: 212:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1046 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

ATAGAAAAGC ATTATGtGTC CCCaGCGGTT TcCaTGAGGC CaCAAAAGGT TCCATACCGA 60

CTTTAATaTG TTGAGGATCA GATAACCATT TAGCAAAAAA CTCATGCCAA ATTCGTTCTC 120

TTCCTGTAAA TAAATCAACA GTATCCGATT CCGCACATAA GTAAAATAAC AATGGTCCCA 180

ATGCAGCTAC TCCGACAAAA CCAAATAGCC AAACTTTCGA ACGTTGCAAG AAACGTTTTG 240

GTAAAAGATT ATCCACCAGA ATAAAACCAA CTAAAATGGC AAAAGAGGTT TTTGATTGAC 300

TGACCCATGT ACCTAAAACA CCACCAATGT ATACAGGTAA AAGCAGCAGG TTAAAAACCT 360

TATTTCTATA TGCTTTGATC AAACTACTGG CCATCATTGT TGAaAACATC AACGTTGCAC 420

CAATCGTATT GGTATTAATC CAAATACTTT CTAGTTTATT TGATTCTTTA AAAATTTCGT 480

CTGCTGGAAT CAACTGAGGC AACTCCACGG ACATCCGATA CAAAGTCGCA CTCAAAGCAA 540

CTAAAGAGAC AATTAATAAA AGCGTTTGAT CCAAACGATC AAAGACACAA ACTTTAAAAC 600

AAAGAACCAA CAAAATAATT AAAGGAATCA TTGTATCTGT TTGCATCGTG CTATGACGGA 660

TTGATTCCGT AAAGTGAAaC AAGAGCCAGG TGGCTAGCGC CAACGCTACG GTCAGCCAGT 720

CAATAAATTT TAATTTCGTA CAATTTAATA GCAGTACAAG TATTACTGCT ATTAATAATA 780

AAATTGAACT ATTGGCATAC ATATATTTTG TGACTGGCGC ATCAATTCCT ACGACCCACA 840

TCGTATACGA GCCGATCATT GTAACTAATA AAATAAGTCG ACCGATTCTA TTCAATACTT 900

TTCCCATATT CmAACCTCAT TCCATGAATG AATCACTAGT CGTTTATTCG ATAAATATAT 960

GCCCCGTTCT GATCAGGTCC ATATAATTTT TCAAAATGAA TCGTTGGTGT GTCTAACGGT 1020

TCTAATGGTT TATAAACGAC CGCATA 1046

(2) INFORMATION FOR SEQ ID NO: 213:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5233 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

TAAGTCGGAT TATTGTCGCC TATAATGACC AAAGAGAGCC GGTGACGGTT AAAGATTTAA 60

ACGTGGCAGG TGCGCTGACT TTATTATTAA AAGACGCCTT AAAACCTAAT TTAGTTCAAA 120

CGATTGAAGG AACTCCAGCT TTTGTTCACG GTGGTCCTTT TGCCAATATC GCTCATGGAT 180

GTAACAGTAT TTTAGCGACA AAAACGGCTT TGCATCTAGG GGATTATGTC GTAACAGAAG 240

CAGGTTTTGG TGGCGACTTA GGTGGCGAAA AATTTTTAGA TATTAAAGTA CCTTCTTTAG 300

GAAAAGCACC CGATGCCGTT GTGATTGTGG CAACGATTCG CGCGTTGAAA ATGCATGGCG 360

GTTTGGCGAA AGATCAGTTA GCAGAAGAAA ATTTACCAGC GTTACAAAAA GGCTTTGCCA 420

ACTTAGAAAA ACATATTCAA AACATGCAAC GTTATGATGT ACCAGTAGTT GTTGCAATTA 480

ATGAGTTTAC ACAGGATACA GAGAAGGAAA TTCAACTGTT GAAAGAAGCC TGCCAAGCGT 540

TAGGCGTGCC CGTCGAGTTA ACAAGTGTTT GGGCGCAAGG TGGCGCAGGT GGCACGAATC 600

TAGCGAAAAC AGTAGTAGCA GAGATTGAAG CCGATACAAA ACAGTTTCAA CCTTTATATA 660

ATCCTCGCCA AACTATTGAA GAAAAAATTC AAGCCATTGT ACAAACGATT TACGGCGGGG 720

AACAGGCGAT TTTTTCTCCG AAAGCGCAAA AACAAATCGC AGACTTTACA AAAAATGGTT 780

GGGATCAATT ACCAATCTGC ATGGCGAAGA CACAGTATTC ACTTTCTGAC GATCCTcAAA 840

AATTAGGACG GCCAGAAGGA TTTACGATTA CAATTCGTGA ACTAGTTCCA AAAATCGGTG 900

CAGGGTTTAT CGTGGCGTTA ACGGGTGACA TTTtAACGAT GCCTGGCTTG CCGAAGgTTC 960

CTGCTGCGTT AAAtATGGgA tGTGGACGAA ACGGGTCAAG CGAGCGGTCT CTTTTAAAGC 1020

CGCTCTTTCC TGAGACCTAT TTTTAGTATA TACTTGAAAG TAGCTCAATA AAAAATAATG 1080

ACGGAAATTT ATTGCGTAGT TTATTATAAG GAGCAACGAC TATGGGGGAA CGAGCAGATA 1140

CATTTTTAGC ACACTTTAAT CGCATTGAAA AGTGGATGAG AGAGAGACTG AACAATCCTG 1200

TCAACATGGG ATTCAGTGAA ATGACAAGAC GTCTTTCCCG AAAAGAAAAA GGACAGATTG 1260

CTCAATTTGA AGAGGATTTG CTTCAAATGG CTCAGTTAAG AAATGCCATC GTTCATGAGC 1320

GAATTGCTGA TGATTTTGTA ATTGCAGAGC CAAACGAATG GGCAGTTCAA CGAATTGAAA 1380

CGATTGAACA AGCTTTAACA AAACCAGAAA AAGTAGTGCC TAAATTTGCG AAACGTGTTA 1440

CAGCTTTTGA AATCACCACT TCTTTAGAAA CGTTATTAAC GATTATTGCA AAAAAACAAT 1500

ATTCTCAATT TCCGATTTAT GAAAAAGGTG TATTTAAAGG ATTAATCACG GTTCGAGGGA 1560

TTGGTGTATG GTTTGCCATT GAGAGCACCA AAGGCGAAGT TCACATTGCC AACCGAACGG 1620

TGCGAGAATT GTTGGCATCA AATTATAAAC GCAGCAACTA TCAATTTGTT TCAATTGAAG 1680

CAACCGTTTT TGAAGTTGAA CAAATGTTTC GGGAGCAGCC ACGCTTAGAA GCTGTGCTTA 1740

TTAGTAAAAG TGGCCATCCT AATGGTGAGT TAGTAGGCAT TGTTCGACCA AGAGATTTAG 1800

CAAGTATTCA TAGAGAAAAG GAATGACGCG ATTGTTAGTA GTATATTTTT TAAyCaGTGC 1860

CTTATTAGTA GGGCTTGACC AATGGAGCAA ATATTTGACG GTCCAAAACA TTTCTTTGGG 1920

tGAAACAAAA GAATTCATTC CTGGATTTTT GTCGCTGACC CATTTAAGAA ATACCGGAGC 1980

TGCTTGGAGC TTATTAGAAG GAAAGATGAT ATTCTTTTAT GTCATCACAG TAATTGTTAG 2040

TGTCGTGATT ATTTATTTAT TAATTAAGAA TTATAAAAAA AGTATCTGGT ATTCTGTAGG 2100

TTTATCTTTT GTCTTAGCAG GTGCCATTGG GAATTTTATT GATCGTGTAC GTTTAGGCTA 2160

CGTAGTGGAC ATGCTCCAAA CGGATTTTAT GAATTTTCCT ATTTTTAATG TTGCAGATAG 2220

TACGTTAGTG GTCGGCGTTA TTTGTATATT TATTTATTTG ATTTTAGATG AAAAAGCAGC 2280

AAAGGAAGGC AAAAATGGAA CAAATTAACG TAACGATTCA AACAGAAAAA GGCCGGATTG 2340

ATAAAGTATT AAGTGAACAT TTACCCAATC ATTCCCGCTC GCAAATTCAA CAATGGTTAA 2400

AAGAAAAACA CGTAACAGTT AATGGTGAAA CGGTCCGCGC AAATTACAAA GTTACAGCAG 2460

GCGATGAAAT TGTTGTAACA ATTCCTGAAC CAGAAGTTCT AGATATGGTA GCAGAGGATA 2520

TTCCTTTGAC CATTGTTTAT GAAGACAATG ATGTGATTGT TGTCAACAAA CCTCAAGGCA 2580

TGGTCGTTCA TCCTTCGGCA GGGCATCCCA AGGGTACATT AGTGAACGCG TTACTTCATC 2640

ATGCGGATCA CTTATCGGCG ATTAATGATG TGGTGCGACC GGGAATTGTT CATCGAATTG 2700

ATAAAGATAC TTCGGGGCTA TTGATGGTGG CTAAAAATGA TCAAGCCCAT GAATCATTAG 2760

CGAAACAGCT AAAAGAAAAA ACATCGTTAC GTAAATACGT TGCCTTGGTC CACGGTGTTA 2820

TTCCTCACGA TAAGGGAGAG ATAAATGCGC CTATTGGTCG TTCCAAGGTT GACCGCAAAA 2880

TGCAAGCGGT CATTGAAGGC GGTAAAGAAG CAGTCACGCA TTTTACTGTT TTAGAGCGTT 2940

TTGATGCATT TACCTTAGTT GAATTACAAC TAGAAACAGG CCGGACACAT CAAATCCGTG 3000

TTCACATGAA ATATATCGGT TATCCATTAG CTGGAGATCC GTTATATGGA CCGAAAAAAA 3060

CATTGCCTGG CAATGGACAA TTTTTACATG CGAAGTTATT AGGCTTTACA CATCCTACTA 3120

CAGGTGAACA ATTAGTCTTT GAAGCACCTT TACCAGAAGT TTTTGAAAAA ACTTTACACC 3180

AATTAAGAAA TAATCATTGA TTTTTTTCTG AAGAAGAGGT AAACTCTTTT TTAAGAAGAA 3240

AAATAACTAA ATAACTGTAA CCTTTAAGTT TAGTCCCGTG AGGCTAAGAA GGAATCATGC 3300

AAAAGAACAC TAGGTGCGCT TAGTTAGTAA GTTGTGTGGG CTACATGGTA CCCTTGTACC 3360

TTTTTAAACT GTAACCAACA TCCTCCTATC CTCATACGGG TAGGAGGATT TTTTGTTGGA 3420

AAAAATAACA GGACAATTTC AATAAAAACC CACTTATTTT TCAGCTGCTC AATTGAAGGA 3480

AGAATCGCTG ACCCCAAGAC AGCGAGGACG TAGATAAAAA AGAAATGAGG AGGAAATTAT 3540

TATGCCAAAG AAAGAAGTTG TTGACGCAgT CACAATGAAA CGAGCGCTTA CTCGAATCTC 3600

GTATGAAATT ATTGAAAGAA ACAAAGGAAT TCAAGACATT GTTTTAGTAG GAATTAAAAC 3660

GCGTGGGATA TACATTGCAC AACGTCTAGC CGAACGCTTA AAACAATTAG AAGATATTGA 3720

CGTGCCAGTT GGTGAATTAG ATATTACTTT GTATCGTGAT GATGTGAAAG ATATGGAGGA 3780

ACCTGaATTA CATTCTTCTG ACGTACCAGT CTCAATTGAA GGTAAAGAAG TTATTTTAGT 3840

GGATGATGTC TTGTACACGG GCCGGACCAT TCGTGCTGCG ATGGACGCTG TTATGGATTT 3900

AGGTCGTCCG AGAAAAATCT CTTTAGCAGT CTTAGTAGAT CGTGGACATC GGGAACTACC 3960

GATTCGTGCC GATTATGTAG GAAAAAATAT TCCAACATCA AAAACTGAGG AAATTATTGT 4020

GGAAATGGAA GAGCGTGACG GCGCAGATCG TATTATGATC AGTAAAGGAA ACGAATAAAA 4080

AGGAGTGAAC AACGTTGTCA GAAAAAGAAT TTCGAAATGA GGATGCAGTC TTAGATATTA 4140

AAGACAGACC GCAAGCTTTT CACTGGGTTG GCCTAAGTTT ACAACATTTA TTTACGATGT 4200

TTGGTGCAAC GGTTTTAGTA CCAATCTTAG TGGGAATTGA CCCAGGGATT GCTTTAGTCA 4260

GTTCAGGATT AGGAACAATG GTCTATCTGA TTACTACAAA AGGAAAAATC CCCGCTTATC 4320

TTGGCaGTAG CTTTGCTTTT aTTGCAGCGA TGCAGATGTT AATGAAGAGT GATGGTTATC 4380

CAGCCATTGC CCAAGGTGCT ATGACAACCG GCTTGGTATA CCTAATCGTT TCATTAATCA 4440

TTAAAAAAAT TGGTTCCGAC TGGTTGGATA AAATTTTACC GCCTATCGTT GTAGGACCTG 4500

TCGTGATGGT TATCGGTTTA GGCTTGGCGG CTAATGCCGC AAATAACGCA ATGTATAACA 4560

ACAACGTTTA TGATTTTAAA TATATTGCCG TAGCCTTAAT TACTTTAGGC TTAACCATTT 4620

TTTACAACAT GTTTTTCAAA GGCTTCCTAG GATTAATTCC AATTCTTTTG GGAATTGTCA 4680

GTGGGTACCT AGTCGCTTTA GCTTTCGGAA TTATTGATTT AACCCCTATT AAAGAAGCTG 4740

CGTGGTTTGC TTTACCGAAT TTTGAAGTAC CTTTTGTTCA ATATGAACCG AAATTGTATT 4800

TAAATGCCAT TACAACAATG GCACCAATTG CTTTTGTAAC AATGACAGAA CATATTGGTC 4860

ATTTAATGGT CTTGaATAAA TTAACCAAGC GCAACTTTTT CCAAGACCCA GGCTTATCTA 4920

AAACTTTAAT GGGAGATGGC TTAGCGCAAA TTGTAGCAGG TTTTGTTGGT GGACCACCAG 4980

TTACAAGTTA TGGTGAAAAT ATCGGCGTGC TAGCAATTAC AAGAGTACAC AGTGTCTTCG 5040

TGATTGGTGG AGCCGCGGTG TTCGCTGTAG CTTTAGGCTT TGTTGGAAAA TTAAGTGCCT 5100

TGaTTCTAAG TATCCCAGGA CCAGTTaTTT nCAGGAATTA GTTTCGTCTT GkTTCGGTGT 5160

GaTTGCCGCA ATgGTTTGaA AATCtTAATC GATAATAAGA nTAATTTGGA TAAGAAAAAG 5220

AATTnTTGAT TGC 5233

(2) INFORMATION FOR SEQ ID NO: 214:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4473 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

TnATAGTATA TTTTTTTAGT CAAAAAATCA ATAGTTTTTG CTAAAAAAAA GTAAAAAAAG 60

CTTTTTCTTG CtTTTTTTTA CTAAAAAGAA ATATACTATA TGCTATAATA GTACTTGAGA 120

AGGAGGCATT TAGAATGGCA ACTATCACAG ATATCGCAAA AGTCGCGAAT GTTTCAATTT 180

CAACTGTATC GCGAGTTCTG AACTATGATC CAAACTTATC CGTTACAGAG GAAACCAAGC 240

GAAAAATTTT TGAAGCAGCA GAAACCCTGA ATTATACGAA ATATAAAACT AAAAATAAAA 300

ACAAACAGCA GGAGCAACTG ACACCTAACC TTTCCGTACA AACAGCTCCG CAAGAGCCTT 360

CGATTGCCGT CGTTCAGTGG CGTTCAGATG ACGAAGAACT AACGGATATT TACTATATGT 420

CGATTCGTTT AGGCGCTGAA AAGCGGGCCG AAGAGTTAGG TTACAATATT CTAAAAGTTT 480

CTCAATTGGA ACAACATAAC CTTCAAGGGA TTGATGGCAT TTTGGCAATT GGCAAATTCA 540

CTCAAAAGAC GTTACAAGAG TTGCAGCAAC TGCATCCCAA TCTTTGTGTA ATTGGCAGTA 600

ATTTCCCATT GGAAGAATTT GATTCTGTAA ATACTGATTT TTATCAAGCA ACCGAGATTG 660

CCTTAACCCA TCTTTTGGAA TTGGGCCACG AAAAAATTGC CTTTATTGGT GCCGAAGAAA 720

GCGAAAATAT GTATGGTTTC CGTCGCTATA AAACACCGAC AACCAACGCT TATCTTGACA 780

TTATGCAGCA TTATCACTTG TTTAACGAAG ATTATTTTAT CTTAAAAGAA AATGGCATGT 840

TGGATGTCAA AACGGGGGAA CAATTGACAG AAGAAGCGTT AGAAAAATGG GGAAATGATT 900

TACCTACTGC GATTCTTGCT TATAATGACG CCTTTGCAAT CGGTGTGATA CATACGTTAG 960

CCGCACACGG TATAAAgGTT CCtGAAGAAA TTAGTGTCAT GGGGATCAAT GATATCTCTA 1020

TTTCTCAATA TGTCTCACCG CCGCTATCCT CTGTCCATGC CTTTACTGAA GAAATGGGCG 1080

AGACAGGCAT TAATTTATTG CACAAACGAA TTCAAATGCC TAGCATTCCT CGTAGAGTCC 1140

TGTTAAACAC AGAATTGGTT GTCCGCCAAT CGACAACTAG CCCACGCCAG AAATAAAAAA 1200

TAAGGACCGC TCCTATAAAG GAACCGTCCT TATTTTTTAT TTGTTTTGAC CAAATAATTC 1260

AATTTGTGTT AATGCTGGAA ATGGTGATGG TTCATCTGCT TGAATTAACT CTGTTAGTAT 1320

CACCCACTCC GCTGCCCGTT TCGCAAAGGT AAAGCTCTGG CGTTGCTCAG TCTTAACAGT 1380

TTTAAAAACT TCATGTGTAC CGTCGGAAAA TTTTAAAGTC ACTTGTTGCC AATAATTATC 1440

ATGGGGAAAG TCCGCTCGTA ATGTTAAGAT CACTTTATCC AACAAGACTT CACGCCCAAA 1500

TTCAATTTTT AATGCGGCAT CTAACTGCTG ATTAATCCCC CATGATTGAT AAGGATACGG 1560

CCCATGATAA AGATTAGCGA AGGTACCATC AATCGCATTA CAAGCAAAAA AAGTGGCATC 1620

ATTTCGCGTT TCCACGTTGG CAGATGCATG AGGATACGCA CCAGTAAAAT GTTGTAAATC 1680

ATGTGGATTT AACGCCCAAT TACGATAACT CTTAATTTCT TCTTTAGTGG CAATACGCAC 1740

ACTTATATAA GGTTTCTTGG TTTCAAAACG ATACGCTGCT CGCGCTTCGA TAGCGTTTTC 1800

ATTTTTTATG ATTGGAAACG TCCATTTTTG TTCTGGAACA TAAATCAAAC TGCTGTCAAG 1860

CGTAgcATcT AACTTCACCA TTAAATAAGT ATTGGGCGTA TCTAAACAAA CTTGAATCAC 1920

ATCGCCGTCT TGTAATTGAC AATCCCGAGC AGCTAAACAA ACAAAATCTT TTCCTGTAAT 1980

AGTTAACTCT TTCAAAGrAT CTGAATCATC TATCTTAGCT TTTCTAACTT CTTGtTCTTG 2040

aTTTAAAATC GATAATGrAA TACACGTCAT ATGCTTCCCT CCTAGTTTTA CCTnTATTAT 2100

ATATTATTTT AGTAAATTAA TATATTTTTT ACTAAAATAA CGACTAATTA AAGCATAAtG 2160

TTTAGATCAA AAAAtAAAAA GGCTTTTCTT AAAAGTCTTC TTTTGTTCAA ACTGATTATC 2220

TTTAAGATTT AAAAAGTGGG ACATAAGTTG AGCCGACCCC CAAAAGTTAG ATTTCTAGGT 2280

CTAACTTTTT GGAGTCGGCT CAAGTCGAAA TGACTAATGT CCCACTTTCA ATAAAAAAAG 2340

AATAAATTTA TGATTTTTTC TGATCATCTT CATCCATTTT CGATTCCGAC AACTGGCCTC 2400

TAAAGCCCGG TATATCAACA TCGAAAATTA AAAAATCGTC TAAAAACAGC TTGTGACAGC 2460

AATTTGACAG TAGTAAATCG CAAAAAAAAA TTACACTTCA AATTTGGTAA ATTCATCTGC 2520

TAAAGTTTGG TTTGTATTAG GATACAGATG ACCATACTTA TCCAAAGTCG TCTTAACAGA 2580

GGAATGCCCC AAACGCTTTG AAACAGCATA GATATTGTTG TTATGCTCAA TCAAAAAAGC 2640

TACATGAGAA TGCCGTAATC CATGTAAGTG AATTGGTTTG ACACCAGCTA TTTCAGCTAA 2700

TTTTTTTACT CGACCTAAAA AGGTCTTTGG GGACGGCGGT AATCCATCAA AGGTAAAAAT 2760

AAAACCTAAG TCATCCCCGA TCGTTTTTTG GGTCTCTTTC CATTCCCTCA AGTCTACTAT 2820

CGTTTGCTTA TCTAGAGTAA TCGTTCGGAT ACTTGAAGCA TTTTTTGTAT CTGAGAAATC 2880

ATAATCTGTT CTTGTTCGAA TATAGAGTGA TTTATTCACG CAGCAAGTAC CTTTATGAAA 2940

ATCAATATCT GGCCATTGTA GCGCAAAAAG TTCTTCACTT CTTAGTCCAG TTACGAACAA 3000

AAAGCGCATC ATTCGTTTAT AAAATGCTTC TTGGAAATCA CAGCGATATG AACAATTATA 3060

TACTTTCTTG AATTCATCAA CAGTCCAAAA TGCCACAACC GATCGTTGAG TTCTGATTTT 3120

CCCAATAGTA TCTACTGGAT TCTCATCAAT GACTTCCAAA ACCATCGCAC GTTTAAAAAT 3180

AATCCGTAAC TTATTAAAAA TCTGCTTAAT ATAATTATTC GATAGCGGAC GTATTTTcCC 3240

TTCTTTATTC GTGACAAAGC ATTCTTGAAC AAGAAATTGT TGAAmCTCTT GAATATGAAT 3300

CGGTCTAATT TTyTCCATTT GCATTTTCCC GAAATAAGtA ACGCGCGTTT TAAAGTCTTA 3360

TCAGTTTTTA CATAAGTTTT CTCAATCGTC CCAAGCTTAT ACCAAGGGAA AAAATACTTC 3420

TTATAAAATT CTCGAAAArT AaTTCCTTGT AAATAAATCT CTTGaGAGTC AGAGTAATTC 3480

TTTTTTAATT CTTCCATCGC ATCTTTTGCT TCTTTTTGCG TTTTAAATCC TCTTCTAGTA 3540

GGTTGTATTC TCTTTCCGAA ACGATCATAG CCCAGAGTTA AAGAAAAATA CCAAGTCCCC 3600

CTCTTTTCAT CCTTATACAC ATTTGCAATT CTTCTCATCA GTATTATTCC TCCCAAGAAA 3660

TGTGTAAGAC TTCTTCAATG ATACTGACGG GTATTCGTTT TATTTTTTTA TTTTTGTAAA 3720

AAGAGTACCC TTTAGTAATT AAAAGTGTCT TACATTCCTT GAAGATTTTA TCAGCGTAAT 3780

ACTGTGAAAA ACCCAACTTT ATTATATCTT GCCGAGTGGC TGTCTTCATT CTACCACCTT 3840

CTTTTTCAGC TGTCAGCTCT TATAAAAAAT GTATTCATCT ACCAACTGCT CTTGTCAGAA 3900

AGCACATAGG ATTTTCACCT CTCCCCAGTA ATGGCGAGCC AATTTAGTTA CGGAAGTATC 3960

ATTATTACAT TTGTAGATCC TGGCTCAAAT TTGACAATTC TTATTTTAGT CATTTATCGC 4020

TCTAGTAGCT CATCATGGCG CAACAAAAAA ACGCTCCCTA CAAATGGAAA GTATTGGTAG 4080

ATCATCTATT CAGTTTTCAA TGGTCGTTTT CTTTATACTA ATAAAGACAA TTTGAAACAT 4140

AAAAAAGCAA CTTCGTGTGT AAGAAAGTCA CTTCTTTCAG TAACTGATAT CcAAGCTTAC 4200

AGCATCTTGC CTTTCTTGTT TTGAAATATT TCCATtGCTC AwGGrCAACA GATTTGACCA 4260

ATGTTACTTG ATTCTTCGCA TTCTCTCTTG CGTAGCTATC ACAATAAAGA TAGTCTTtGA 4320

ATCcGCyACA AAGAaCTTCA GTGCTtAGGA AaCTTACGAT TtGGCtAATT AyCGaTGrAG 4380

TtACATCTAA GTGGTTCATG GGGAAGTAGT AAAATCGCAG AAAACTTCTT TTTATTCTGA 4440

CGTTTCTCTA CTTAAAAAAA CGGTCTTCTA GGC 4473

(2) INFORMATION FOR SEQ ID NO: 215:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3662 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

CAATTGTATG TGCCATCTGC ATACGTGCTG CGGCAATTGC CACTCTCAAC GTAATAGATA 60

ATATTTCGGC ATAATCAAAT TGATTTTGAT CGAGTGGATT AATTAAATTT CGCATTTTTT 120

TCAATGCACT ACGATAACAA CCTTGAAACA ATGTCCCTGA CCCAAATTTA ATTTTAGGGA 180

TTTTTTCAGC CCCTTTTGTC TTTGTCACAT AGTTCATAAT TTGGTAAATA TCGTATTCTG 240

TAAGTTCTTT TTGAGTGATA AATTCCATCA GTAAGCGTAC ATTGCTTTCT TCACCACTGA 300

TTGTAAACCC TTGCCGCGCT TGGCGAATTA AGGTTAAATC ATATGCTTGC AGCAACCGTT 360

CGACCTGTTC CAAATCACTG ACAATTGTGT TTCGACTCAC TAATAATTCA TCTGCTAGCT 420

GCTGTGCCGT TAAGTACCCT TTGACTGACA ATAACTGAAA AATCAATACA TTGACTCGTC 480

TTTTTTGATC AGGATAGGTT TCAAATCGAT CGACATCGAT AATTTCATTT TTTAACAGTA 540

AGCGTTTAGA GTCAGGTAAA TCTAGCCAGA TTCCCTTATT TCGTTTTGAA TAAAGTGTCT 600

CATCGTGCTG TTCAAACCAA GCACGAATAT CTTCTAAATC ATACTTAATT GTTCGAACGC 660

TGACAGTAAA GCTCTCAGCT AATTCTTGCG TGGTAATACT CCCTTCCAAA TCCAAAAGCA 720

TCAATAAAAT ATTTATCGCT CTTTTGGACA ACAACACGGT CATGCCTTGT CACCTCTTTT 780

CTTTTCCCTT TACACTTAAA ACTATAGCGT TATTTTATCT GTTCCACAAC ACAAAAAAGT 840

TGCACTTATA AAAGTGCAAC ACTTCTCTTT TCATACAAAA AAAATCCCTT AACAGTAGTA 900

ACTGTTAAGG GATACCACAC TATTTATTTG CATAAATAGA CAGAAACCTT CTCCACAAAA 960

GTATTCTGTC CGACACGTTA CAAATCCGGT AACTGATAGC CGTACTATCT GGTGTCGTTG 1020

TGTTTGACAA TCACTAGCCG TCAAAACGGA ATAGCCGAAG CTATCTGATT GAGTCAACTC 1080

GTCGCATAAA ACACACTTTA ACACAGAATA GTTTGGATAG CAAGTGATGG TAATTATTCG 1140

TTATCGAACC ATTCTTGATA AAAACTTTCA ACAATTTTAA CGGTTGCTAT CGTTAATTCT 1200

TTCGTCATTA CTGGGCTAGT TAACTTTTTG TCTTGTAAAC ATTGATTCAC ATGATTGATT 1260

TCATACGTAA ATTCACTAGT GAATTGTTCA GACCATTTGA CAGTGTTGCC TTGAGCATCT 1320

GTATAATAAG CACAATCTGT TTTCCAGAAA TTAGGAATAA CGATTTGCCC TTTCGTCCCA 1380

CAAATGGTCA TTTCaCTAGG TATCTTTAAA CCAACATTAA TAAAAATATT GCCCAGCGTC 1440

CCTTCAGCAA ATTTCAAAGC TAAATTACAT TGACTATCCG TCGCCCCTTG CTGATACGTT 1500

GCGGTTCCCG TAACCTCTTG AATCTCTTTC CCTAATACGT ATTGTAAGTA TTGTAATGGA 1560

TAACTTCCTG AGCCATGAAG CGCTCCACCA CCCGCTTCTC TCGAGTAAAA CCAGGGAATA 1620

TGATCCACAT TGGGATAGGC AGTAACTGAC TGAACCCATA AAATCTCTCC CAATCCCCCT 1680

TCTTGAATCG TTGCCTTTAC TTTTTGGGTA ATTGGTAAAA AGACAGACTT CTGAGCTTCC 1740

ATTAGAAAAA CCCCTTGCTC CTGTGCGATT GCAAAGAGTT CTTCTGCTTC AGCAGCATTT 1800

AATGTAAATG GTTTTTCCAG TAGGACGGGT TTTCCTTGTG ATAACGCAAG CTTAGCTGCT 1860

GAGTAATGTC CTTGATTGTA GGTGGGAATA TAAATAATAT CAATTGTTTC ATCCTTACAT 1920

AACTCTTCAT AACTGCCATA AGCCACCGGA ATCGCTAATT CTTTAGCCAT TTTTTGAGCA 1980

TTCTCTAAAC GCCTTGATGC AATCCCGCGC ACTTcAGcTT GCGCGCTTTC TCGAAGCCCT 2040

GCTACAAATC TTGGGACAAT TTGGGCAGTA CTCATAATTC CATAACGAAT TTTATCCATG 2100

TGATCGCCTC CTCTTTCACT ATACGAGAAA AACGAATATT TGGAAACCAT TTCAACAAAA 2160

AAAGAAACCG TTTCCCCAAA ACGATTTCTT TATATGTAAA TTAACGGATA CCAACGCTAA 2220

CTTTTAAAGC TTGATTGACC TGTTCCATCT CTTCACGAGA AAGTTGACAA ACCTTGTGTA 2280

GCATTCGTTC TTTATCCAAT GTTCGAATTT GTTCTAACAA AATAAGTGAT GGCGTCCTGC 2340

ACTCGTTATG GGGAATTTCG ATAAGGACTT GCGTTGGTTG CAGCCGCTTG CTTACGTTAC 2400

GGGTAATCGG TGCAACAATC AACGTTGGAC TAAATAAGTT ACCTTTGTTA TtTTGAATAA 2460

TCaGGACAGG TCGAATCcCA CCTTGTTCAG AACCAACAAC TGGTGAGAGA TTCGCATAAA 2520

AAACCTCACC TCTTTTTATC ATTCAGATTC CCTTCCCTCC TTTTCCTTTA AAAAATTATA 2580

GCATATTGTT TCTTGCAATA TTTTCAACTT TTTAGCTATC CCtTAAAATA TTTGATAAAA 2640

CCACCCTTTA AAAGTCAAAA AAAGCTTAAA CAGTCGACTC TGATATAGAG ACGGCTATTT 2700

AAGCTGACTA TTACATTTTT GTTAAAATAC TATCCACTTT TGTATTAATT AAATCAATAG 2760

CGACGTGGTT TTCGCCACCT TCTGGCACAA TGATATCTGC ATAACGTTTC GTAGGTTCAA 2820

TAAATTGGTG ATACATGGGC TTGACGACTG TCAAATATTG TTCAATCACA GAATCCAACG 2880

TACGGCCACG TTCTTCCATA TCACGCTTGA TTCGACGAAT AATACGGATG TCATCATCTG 2940

TATCCACATA CACTTTAATA TCCATTAAAT CCCTTAAGCG GCGATCTTCT AAAATTAAAA 3000

TCCCTTCAAG AATAATTACT TCTTTCGGTT CTTGTATCAC CGTTTCTGTA CTTCTTGTAT 3060

GCGCAACGTA GTCATAAACT GGCTTTTCAA TGGCTTGATA ATTTAACAGT TGTTCAACAT 3120

GTTGAATCAA TAAATCTGTA TCAAACGCAA AAGGATGATC ATAATTGGTA TTTAAACGTT 3180

CTTCAAAACT CAAATGGCTC TGATCTTTAT AATAAGAATC CTGTTCTAAC ATCATAATTG 3240

AATGATCGGG AAAATTATTA AAAATCGCAC GGCTAACACT GGTTTTTCCG CTTCCTGATC 3300

CGCCAGTAAC ACCAATGATA ATTGGCTGAC TATCTTTCAT GTATAAACCT CTTTCTTTTC 3360

AAAAACTTCA TTGTCTTTTA CTATTATAAT GAAAACTCTG AAGAATGCCA GACCTTTTCC 3420

AAGACTTTTC ATGTGATTTT TATCTTCTTA TGAAAAATCT GCCCACACTT CTTTCAATTA 3480

CAGTATACTT ATCAATAAAA GGAGAATCCT ACCTATGATT AAAAAAGTCA AAAAAGAAGC 3540

ACTTACAACG GCGCATTACG CTCkTCcTkT ATGAGGgCAG ATyCCTCAAA AAAAATGGGG 3600

GGCAGATTAC CATACCCGCG GGnATTTGGC TTTGGACTAT CCGACTGGAA GAACTTTCCA 3660

nG 3662

(2) INFORMATION FOR SEQ ID NO: 216:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7750 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

TGAnTATGAA CGACGTTATG ATGAGCACCA AGGACTAATT TTGGGTTCAG TGACAACAGC 60

CATTCCTTTA TCAAAAGAAC AACATCAAGC AATGGAAGAA AAAGCTGCCC AATTATTAGG 120

GTATGAACAA GCGCATTTAG TCAATTTAAT TGACCCGTCA ATTGTCGGCG GAGTGGTTAT 180

TGAAGCAAAC CATCAAGTAA TCGATGGCAG CATCCGCAAA CAATTAGAAC ATATGCAGCA 240

GAAGTTATTA AAATAATCTT TAGGACTACA AAGAGGTGAA TTGAATGGCT ATCAAAGCAG 300

AAGAAATCAG TGCCTTGATT AAAGAACAAA TCGAAAACTA TCAACAACAG CTTGCTGTTG 360

AAGAAGTTGG AACAGTCACT TACGTCGGTG ACGGAATTGC CCGCGCCCAC GGTTTAGAAA 420

ATGCCATGAG TGGAGAACTT GTTGAGTTTT CAAACGGTTC ATACGGAATG GCACAAAACT 480

TAGAAACAAA TGATGTAGGG ATTATTATCC TTGGCGATTT TGAAACCATT CGTGAAGGAG 540

ATAAAGTACA ACGCACAGGC AAAATCATGG AAGTACCCGT TGGGGAAGCT TTGATTGGCC 600

GTGTTGTTAA TCCATTAGGT CAACCATTAG ATGGTTTAGG CGAAATTAAA ACAGACAAAA 660

CACGTCCTGT GGAAGCAACA GCTCCCGGTG TTATGCAACG TCAATCTGTT GCTGAACCAA 720

TGCAAACTGG CTTAAAAGCC ATTGATGCCC TCGTACCAAT TGGTCGTGGA CAACGGGAAT 780

TAGTTATTGG AGACCGTAAA ACAGGGAAAA CATCGATTGC AATTGATACA ATTATTAACC 840

AAAAAGGACA AGATGTAATT TGTATTTATG TAGCGATTGG TCAAAAAGAA TCTACTGTTC 900

GTAACCAAGT AGAAACACTT CGTAAATTTG GTGCCTTAGA TTATACAATC GTTGTGACTG 960

CAGGGGCTTC ACAACCAGCA CCATTACTTT ACATTGCACC ATATGCAGGA ACTGCAATGG 1020

GTGAAGAATT CATGTATAAC GGCAAACATG TCTTAATTAT TTTTGATGAC TTATCAAAAC 1080

AAGCCGTGGC TTATCGTGAA CTTTCTCTAC TATTACGCCG TCCACCAGGT CGTGAAGCTT 1140

ATCCAGGGGA TGTCTTCTAT TTACACTCAC GTTTATTAGA ACGTGCAGCT AAATTAAGTG 1200

ATGAATTAGG TGGCGGTTCA ATGACAGCCT TACCATTCGT TGAAACACAA GCAGGCGATA 1260

TTTCTGCTTA TATTCCAACA AACGTTATTT CAATCACTGA TGGACAAATT TTCTTAGAAA 1320

GCGACTTGTT CTATGCAGGC ACTCGTCCAG CCGTTGATGC TGGGTTATCT GTTTCTCGGG 1380

TCGGAGGATC AGCGCAAATC AAAGCAATGA AAAAAGTTGC CGGAACGCTT CGTTTAGACT 1440

TAGCTAGTTA CCGTGAGTTA GAAGCCTTTA CACAATTTGG CTCTGATTTA GATGCGGCAA 1500

CACAAGCTAA ATTAAACCGT GGTCGTCGTA CGGTAGAAAT TTTAAAACAA AAACTACATG 1560

CGCCACTGCC TGTAGAAAAA CAAGTCTTAA TTTTATATGC GTTAACACAC GGCTTTTTAG 1620

ATAGCGTGAG CGTAGATAAA ATTTACACTT TGAACAAGA TTTATTCGAT TATTTTGATG 1680

GCAAACATGC AGACTTATTG GAAACCATTC GTACAACGAA AGACTTACCA GATACTGATG 1740

CATTAGATGC AGCGATTACA GAGTTTTCTG AAATGTTTGC AGCGGCAAAC AATAGCGGAG 1800

ATTCAGCCAA AGAAGCTTTA GAAAAAATCG ATAACGCATA AGAGAGGTGA GCGACATTGG 1860

GTGCTTCATT AAACGAAATC AAACAACGCA TTGCATCAAC AAAAAAGACA AGCCAAATTA 1920

CCAAAGCGAT GCAAATGGTT TCTGCAGCCA AATTGACCAA GTCAGAAGGC GCCTCTAAAA 1980

GCTTTCAAGA ATATTCTTCA AAAATTCGTA GTGTCGTTAC GCATTTAGTA GCTGCTCAGT 2040

TAAGTGAATT ACGTGAAACA GAACAATCTT CTCTTAGTGA AGGGAACTAT CATGTGATGT 2100

TAGCACAACG ACCTGTCAAA AAAACAGGCT ATATCGTTAT CACTTCTGAC AAAGGGTTAG 2160

TTGGTGGCTA CAATAGCTCA ATCTTAAAAC AAACCATGAG TATGATTCAA GAAGACCATG 2220

ATTCTAATAA AGAATATGCC TTAATCGCTA TTGGTGGTAC AGGGGCGGAT TTCTTTAAAG 2280

CACGCGGAAT CGATGTTTCC TATGAGCTAA GAGGCTTAAC GGATCAACCA ACGTTTGAAG 2340

AAGTCCGTAA GATTGTGACA ACGGCAACGA CTATGTACCA AAACGAAGTG tTnGATGAAT 2400

TGTATGTTTG TTATAATCAC CACGTTAACT CACTAACGAG TCAGTTTCGT GTGGAAAAAA 2460

TGTTGCCAAT TACTGATTTA GATCCATCAG AGGCGACCTC TTACGAACAA GAATATCTTT 2520

TAGAGCCTTC ACCTGAAGCA ATCTTGGATC AATTATTACC ACAATATGCA GAAAGCTTAA 2580

TTTATGGGGC GATCATTGAT GCGAAAACAG CAGAACATGC TGCTGGGATG ACCGCCATGA 2640

AAACAGCCAC TGACAATGCG CAAAATATTA TTAGCGATTT AACAATTTCT TATAACCGTG 2700

CTCGTCAAGG GGCGATTACC CAAGAAATTA CAGAAATTGT TGCTGGTGCA GCTGCACTAG 2760

AATAATCTAG TCTTACGTTT GAACGTTTAT AAAAATGTTC AAAGCGTGTG ACTTTATGAA 2820

GACTGGAGGA AAGAAATTAT GAGTTCAGGA AAGATTGTTC AAGTAATCGG TCCCGTTGTT 2880

GACGTGGAAT TTTCATTAGA TCAATCCTTA CCCGATATTA ACAACGCTTT AGTCGTTTAT 2940

AAAAATGGCG AACAAAACAA AAAGTAGTAC TTGAAGTCGC TTTAGAACTA GGTGATGGAG 3000

TGATTCGTTC TATCGCTATG GAATCGACAG ATGGTTTACA ACGTGGAATG GAAGTTATCG 3060

ATACAGGAAA ATCAATTTCA GTTCCTGTTG GTAAAGATAC ATTAGGTCGT GTGTTTAACG 3120

TTTTAGGAGA CACAATTGAC TTAGAAGCGC CATTCCCTGC AGATGCTGAA CGTAGtGrGa 3180

TTCATAAAAA AGCGCCAGCA TTTGATGAAT TAAGTACCAG TAATGAAATT TTAGAAACAG 3240

GGATTAAAGT TATTGACTTA TTAGCACCTT ATCTAAAAGG TGGTAAAGTC GGACTTTTCG 3300

GTGGTGCCGG TGTTGGTAAA ACCGTCTTAA TTCAAGAATT AATTCATAAT ATTGCCCAAG 3360

AACATGGAGG GATTTCCGTC TTTACTGGTG TTGGTGAACG GACACGTGAA GGGAACGACC 3420

TGTACTATGA AATGAAAGAT TCAGGCGTTA TTGAAAAAAC AGCCATGGTT TTTGGTCAAA 3480

TGAACGAACC ACCAGGTGCA CGGATGCGTG TGGCCTTAAC TGGGTTAACG ATTGCTGAAT 3540

ATTTCCGTGA TGTGGAAGGA CAAGACGTGC TATTATTTAT TGATAACATT TTCCGTTTCA 3600

CCCAAGCCGG TTCAGAAGTT TCTGCCCTTT TAGGTCGGAT GCCGTCAGCC GTTGGTTACC 3660

AACCAACCTT AGCGACTGAA ATGGGACAAT TACAAGAACG GATTACTTCA ACGAAAAAAG 3720

GATCAATTAC CTCTATTCAA GCAATCTATG TTCCAGCCGA TGACTATACC GATCCAGCGC 3780

CAGCAACAGC GTTTGCCCAT TTGGATGCAA CAACTAACTT GGAACGTAAA TTAACCGAAC 3840

AAGGGATTTA TCCAGCGGTA GATCCGTTAG CTTCATCTTC TAGTGCCTTG GCTCCTGAAA 3900

TTGTTGGAGA AGAACACTAC GAAkkGGCTA CCGAAGTGCA ACATATTTTA CAACGTTACC 3960

GTGAATTACA AGATATCATC GCTATTTTAG GGATGGACGA ATTATCAGAT GACGAAAAAG 4020

TACTTGTTGG TCGCGCACGC CGTGTTCAAT TCTTCTTATC TCAAAACTTT AACGTTGCTG 4080

AACAATTTAC AGGTCAACCT GGTTCTTATG TACCAGTTGC GGAAACAGTT CGTGGCTTTA 4140

AAGAAATTCT TGAAGGTAAA CATGACAATC TGCCAGAAGA AGCCTTCCGT AGTGTCGGTA 4200

AAATTGAAGA TGCCATCGAA AAAGCGAAAC AATTAAACTA CTAGAAAGGC GGCAAACACA 4260

ATGGATAGTC TTACTGTGAA TGTAGTAACG CCTAATGGTC TAGTTTATGA CCACCATGCA 4320

AAAATCGTTG TGGCGAAAAC AACAGATGGT GAAATTGGTA TTTTACCAAA ACATGCGCCA 4380

ATCATCGTTC CGCTAGCGAT TGATGAAGTC CGCATCAAAC GGACCGATTC AGATACACAT 4440

GTTGACTGGG TAGCTGTAAA TGGTGGTATC ATGGAAGTCC GTGATAATGT CGTTTCTATT 4500

ATTGCCGATA GTGCCGAACG TGAACGAGAT ATCGATGTAC CTCGTGCTGA ACGAGCAAAA 4560

CAACGAGCAG AACGTTTAAT TGAAGAAGCC AAAGCAAAAG ATGACCGCGA TCAACTACGT 4620

CGTGCGACTG TTGCTTTGCA TCGTGCAATC AACCGGATCA ACGTATCGAA ACATGGGTAA 4680

AAAAGGCTGA AGGATTCAGT CACTTTTGTA GAGTTTAAGG CGCTTGCTAT TGAGTCAGTT 4740

GCCTTAAACT CTATTTCTTC GTATTTTTTA ATAAATAGTG TCAAATTTTG GCATTTCAAG 4800

TGGATTAATT ATCGGTTTTT ATTATTTTGT GATAAAATGT TCCTACGTTA GAAAAAAGGA 4860

GGCTCCTTAT GCAATACTAT GGAGTAGATG CTATTGTTCG AATCGTTAGT CATCTTATGT 4920

TCATTTACAT AAGTTTTTGG GCGTTGCAAT CATTGCGTAT TGAGCAATTT TTCAAAACCC 4980

AATTTACACC GCAAATCAGA ATGTTGATGG TATTTTTCGC CATCGCTATC GGATACACAG 5040

TTAGCTCGTT TGCTTTAGAA TTGATTGCGC TTTGTCGAAA TCTTTTCATT GTGTATTTTC 5100

CTTAAAAATT GAAAAATAAA AAAAGGTACA ATTTTTAGTT GAATAATGGT ATAATAGTAA 5160

AGATTTTGAA TAGGAATACC TATTCTCATT TTTTTGGGGG GAACAAAATG GAACAAATTA 5220

TTGTTCATGG TGGTAATACA AAATTAGAAG GAACTGTAAA AATTGAAGGC GCCAAAAACG 5280

CTGTTTTACC AATTTTAGCA GCTACTCTTT TAGCTGAAGA AGGCGTTACT ACTTTAAAAA 5340

ATGTACCAAT TTTATCTGAT GTATTTACAA TGAATCAAGT AATTAAGCAT TTGAATGTAG 5400

CCATTGATTT TGATGAAGAC GCGAATGAAG TAACAATAGA CGCAACACAA CCTTTAGGGA 5460

TCGAAGCAAA CTATGAGTAC GTTAGTAAAA TGCGTGCTTC TATCGTTGTA ATGGGACCTT 5520

TATTGGCTCG TAATGGTCAT GCCAAAGTGG CAATGCCAGG TGGCTGTGCC ATTGGGAAAC 5580

GACCAtTGAT TTACACTTAA AAGGCTTCCA AGCATTAGGT GCAAAAATCA TCCAAAAAAA 5640

TGGTTATATT GAAGCAATTG CGGATGAACT AATCGGTAAC ACTATTTACT TAGACTTCCC 5700

AAGTGTCGGC GCTACGCAAA ACATTATGAT GGCTGCTGTT CGTGCGAAAG GAACAACTAT 5760

CATTGAAAAC GTTGCTCGTG AACCTGAAAT CGTTGACTTA GCAAACATTT TAAACAAAAT 5820

GGGTGCCAAT GTTATCGGTG CTGGGACAGA AACGATGCGT ATCGAAGGCG TCGATAAATT 5880

ACATGCCGTA GAGCATTCAA TCGTTCAAGA TCGAATTGAA GCAGGTACGT TCATGGTAGC 5940

TGCAGCAATG ACTGAGGGCA ATGTTTTAAT TGAAGAAGCA ATTTCTGAAC ATAATCGTCC 6000

GCTAATTTCT AAATTAACTG AAATGGGTGC AATCATTgAA GAAGAAGAAA ACGGCATCCG 6060

TGTAATTGGT CCAAAACATC TAAAACCAAC AGATGTTAAG ACAATGCCGC ATCCTGGTTT 6120

CCCAACAGAC ATGCAAGCAC AAATGACAGC GATCCAAATG TTTGCTGAAG GAACAAGCAT 6180

CGTGACAGAA ACAGTCTTTG AAAATCGTTA CCAACACTTA GAGGAAATGC GTCGAATGAA 6240

CGCTGACTTA AAAATTGACG GCAACATTGC CGTTATTAAT GGTGGCAACG AGTTACAAGG 6300

AGCAGCAGTG GAAGCCACTG ATTTACGGGC AGCTGCCGCA TTGATTTTAG TAGGGTTACG 6360

TGCAAATGGC ATTACACGTG TCTCTAATTT GAAATATTTA GATCGTGGCT ATTATGAGTT 6420

CCATAAAAAA CTTCAAAAAT TAGGCGCGAA CGTTGAACGT GTCAATGATG AAAAAATTGA 6480

AGAAAAACAA GCAACGACAG TAATCTAAAG GAGTCAAACG TATGAGTTCA TTTCGTTATA 6540

TTTTAGTAAC ATTATTGAAA ATTCTAGTGG TCmTTTCGTT AGTCATTATC TTATTTGTTG 6600

TGGGamCAAT GATTGGTTAT GGTTTGATTG GAAATGGGAA TCCAATGGaT GTCTTTGATG 6660

AAAAAATTTG GACTCATATT ATGAACTTCT TTAAATAAAA ACTTATTTTA AGGCGGTGAA 6720

CAAAAGTCAA TTGACGATTG TTCACCGCCT TTTCTTTTTC CTGAATATCT ATTTTATTCT 6780

ATGCTATAAT AATAGAAGTA AGATAATGAA GGAGTGGGAT GGAAGATGCA AGCAACCGTA 6840

ACAGAAATAG GCAAACATGC AATCGATGAT TCAGAAAAAA TGATTATTCT TTTTGGAGAA 6900

ACTGCTACCG ATACCTTGAA GCAACATGCA GTTATTCAAT CATTCCCAGA AAAAGACCAA 6960

GTTACCTTGG CAGAAGGAGA TCATCTGAAA ATTGGCGATA CAAACTATAC AATCACAAAA 7020

GTGGGTTCTT TTGCCAATAG TAACTTGCAA AGTATCGCTC ACAGCACTTT AATCTTTGCT 7080

GATGCACCAA CCGATGAGGA TGATGTTATC CGGAATGGTG TTTATTTAAC GCCCCATCAG 7140

TTACCAAAGA TAACTATTGG AACAACAATT GATTATTTAG TAAATGGAGC GTGAATCAAT 7200

GGATAAAAAA CAAACGACTG ATAAACGCCT GTCTTGGTTT TGGCGTTGGT TTTTAAATAA 7260

TCAAGTAGTA ACAGCCTTAC TGATTGTCTT ATTGGTGCTG TTAATTATTC TAACTTTTAC 7320

GAAAGTTTCT TATTTATTTA AGCCCGTCTG GCAGTTTTTC GGCGTTGTCG GCTTACCAGT 7380

AATTATGGCG GGGATTTTAT ATTATTTATT AAATCCAATT GTCGATTATT TAGAGAAAAA 7440

ACAAATCTCA CGTGTCTGGA GTATTATCGG CCTATTTATT TTAATTGTGG CACTACTCAT 7500

TTGGGGTGGT ATTGTAATCG TTCCTAAAAT TCGTGAGCAA TCTGTCAGCT TTGTGAATCA 7560

TTTTCCTCAA TATATTGATA CAATTGATCA AAAATCGCAG GAAATATTAA GTGATCCACT 7620

TTTCGCGCAA TTTAGAGAAC AACTTGAAGC GGCAGGCGAT AAAGTTGTCA GCTCGTTaGG 7680

TACCATCATT AAAAACGTCT CAACGTyCAC AGTACmAGGA ATTGGTAATT tCnTTGGGGC 7740

GGTGGCTACG 7750

(2) INFORMATION FOR SEQ ID NO: 217:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8395 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

GCCCCCAGAA TGTTCCAAAA GAAATTTTGT ATGCATTTTT AATTGTGCTG GCGATTGATA 60

GGGCTGGTGA CTTACTGAGT TGATATAATC AGCCAAACTT TCCGCTTTCG TTAACAGTTC 120

TCGATTTTGT TCATTCCAAT TATTGCCAAA GTCTTTAAAC GGATAAAGTA AACCTAGCTG 180

AACTTCTGGA TATTCCGTTT TTAATTCCCC AACTACTTCA GCTGTCCAAA GTTCAACACC 240

TAAATTGCCA GAAATCAAAA TCCATTCTAC TCCCGTTTCA ATATAACTGG CTAACTCTTT 300

TTTTAAGACA TTTTTAATAA CTGTGATTTt CGGGTCTTTt CCTtGAAATA TCCCCAATTC 360

AAAGCTGCGA TAACCGGTGA CATAGAGGGT TTTTATCATT TCCATAAGTC ATCCCTTTTC 420

TTTCAGGTCT TTTTTTtGGT ATAATACAGG CTAGGAGTGT GATAAACGTG ACGTTCCATT 480

ATCCTAACGG AATACCTTAT AATCATCATG AAGCACAATC ACCAAAAAAA CAAGTAAAAA 540

AGAAACGCCC CGTTGAATTT GGTAATCGTG GGATGGATTT CGAAGAAGCG ATCAATCAAA 600

GCAATCAATA TTATCTATTG AATAAGCAAG CAGTCGTTCA TAAGAAACCA ACTCCTGTTC 660

AAATTGTTAA AGTAGATTAT CCCAAACGAA GTGCGGgCTG TTATTAAGGA GGCCTACTTT 720

AGACAAGCTT CAACGACTGA TTATAACGGT GTATATCAAG GCCGTTACAT AGATTTTGAA 780

GCGAAAGAAA CGAAAAATAC CACGTCCTTT CCCTTTAAAA ATTTTCATCA ACATCAAATT 840

GATCATATGG AACAATGTAT AGCACAACAA GGGATTTGTT TTGTGCTATT ATGGTTTTCA 900

AGCTTAAAAC GCTGCTTCTT TTTTAGTGGT GAAAATCTAA TAACGCATTG GAAAAACCAA 960

GCAGTTACTG GAAAAAAATC TATGCCGCTT GCACTCATCG AGAAAGAAGG AATTGAATTA 1020

TCTTTTGGAA TCGCACCGCG AATCCCTTAT CTCGACGCAg TGGAACAATA TATTCAGACA 1080

CACTAGATTC ATAAAGTTTT GAAAGTATAA CAAAGGAGTT GCTATTTATG CCAACCGCAA 1140

ATTCAGGCTC GCGAGCTGCT AGACATGCAT CCAACACCTC ATCGAAGTCA AGTTACAAGA 1200

AAAAGACACA ACCGCCGAAA AAGAAAAAAC gGCTTGTTTT AAAAATCTTT CTTGGcTTAC 1260

TTATCGCTGG GaTGGTCGCT TTTTTAGCAG GCGTTGGTCT TTTTTGGtTC TATGCACGAC 1320

AAGCCCCAAA ATTAGAAGAT GACAAACTGA ATGCGACAGT TTCATCGAAG CTTTACGATA 1380

TTAATAATGA AATTTTTGAG GATTTAGGCG CAGAAAAGCG TGAATTAATT CAACCAAATG 1440

ATGTGCCTCA ATTATTAAAA GATGCCATTG TTTCTGTCGA AGACCGTCGT TTCTATAAAC 1500

ACATTGGTGT GGATCCGATT CGGATTATCG GTTCTGCCCT TTCAAATGTC AAAAATGGTG 1560

GATTACAAGG CGGAAGTACC TTAACCCAAC AATTAATTAA ATTGTCCTAC TTCTCTACAA 1620

AAGAATCTGA CCAAACCTTG AAACGGAAAG CACAAGAAGC ATGGATGGCT GTTCGTTTAG 1680

AACGTGAAAA ATCAAAAGAA GAAATTTTAA CTTACTACAT TAACAAAGTC TACATGGCGA 1740

ATGGTTTTTA TGGAATGGAA ACAGCGGCTG AAAATTATTA TGGTAAACAT TTATCTGAAC 1800

TTGATTTACC ACAAACGGCC CTACTGGCAG GAATGCCACA AGCGCCAAAC TCATATGATC 1860

CCTATACAAA GCCAGATACT GCTAAAGAAC GCCGGGATGT TGTACTTTAC ACAATGTACG 1920

ACAATAAAAA AATCTCGAAA GCCGAGTACG AAAAAGCCAA aGCTACCCCA ATCGATGAAG 1980

GCTTAGTGCC ATTAAAAGCC AGTGATGACA ATCGAAAAGT CGTAGATAAT TACGTCAAAG 2040

AAGTAATCAA TGAAGTGAAA GCAAAAACTG GTAAAAATGT TTACACAGAC GGACTAGATA 2100

TCTACACAAA CTTAGATATG AATGCCCAAA AACAACTTTA CGATATTGTC AACAGTGATC 2160

AATACGTTGC GTTTCCAGAT GATAAAATGC AAGTGGCTTC AACCGTTATT GATGTTGCCT 2220

CTGGACAAGT CCGAGCACAA ATTGGTGGTC GTCACATTCC TGATGATGTG CAATTAGGAA 2280

ACAATTTAGC TGTTAACACC CAACGTGATG TTGGTTCAAC CGTGAAACCA ATCATGGATT 2340

ATGGTCCCGC AATTGAGAAT TTAAATTATT CGACAGGTCG CCTCATGGTG GATAAACCAA 2400

CGAAATATCC TGGTACCGAC ATTGATGTCT TCAACTCTGA TTTGACATAC CAAGGCGTCA 2460

TTACTATGCG CCGTGCGATT ATGGGTTCTC GTAACACAAC CGCGGTCCAG ACGTTTGATG 2520

AAGTTGGTAA AGAAAACATC ATGCCTTTCA TTAAAGGACT AGGAATTGAC TATAAAAACT 2580

TAGAAGCGTC AAATGCCATT TCAAGTAACA CAAGCGATGT AGATGGCGAT AAATACGGGA 2640

TTTCCTCATT GAAGTTAGCA GCAGCTTATG CGGCTTTCGC CAATAACGGG ATTTATAACA 2700

AACCGTATTA CGTAAACAAA GTCGTTTTCA ATGATGGAAC TAGCGTTGAT TATCAACCTG 2760

ATGGTAAACG TGCGATGAAA GACTCTACCG CCTATATGAT GACGGATATG TTAAAAGATG 2820

TTTTAAACGG CGGAACAGGA TTTAACGGCG CCATTCCAGG CTTAATCCAA GCGGCCAAAA 2880

CAGGGACTTC TAACTATACT GATGAAGATT TAGCGCGAAT GGgCACAACA GAAAAAGGAA 2940

TTGCCCCGGA TAGTACCTTT GTTGGTTATA CAACACACTA TGCCGTTTCT GTTTGGACTG 3000

GATACAtGAT CGAAtACACC AATCTATCAA GAATACTATG GAATTGCCTC GGATGTTTAT 3060

CGTGAAATAA TGAGTTACTT GTCTCAAAAT GTTTCAAACG ATGACTGGGT GCAGCCAGAC 3120

AGTGTTGTGC GGGTAGGTAA TGAATTATAC GTGAAAGATG CCTATGAGGT ACCAAATGTA 3180

CAAGTTTTAC CAAGTACTAC CTCTTCAGCA CCACAGCCGG AATCTAGCAG TACGGTGGAA 3240

TCCTCTTCCA CAAAAGAAGC GGAAAGTTCA TCAAGTTCTA GTTCAGAAAG TGCGCCTTCT 3300

TCCTCTGAAG CGCCACCATC AACCGAACAA CCGGCAAGCT CTTCCTCAGC AGAGCAACCG 3360

GCAACATCAG AACAACCACC TGAACCTTCA AGTTCTAGTT CACAAGAACC GCCGCAACCA 3420

CCTGAAAGTA GTTCTAAACC AGATGAAAAT AAAGCAGCAT AAAAAGTACG GGGCTCATGA 3480

TGAGCCCCGT ACTTTTTTAG ATTAAAAATG GACGAAGTGA ACTGAATCAC TtCGTCCATT 3540

TTTGCTTATT TACTTTCTTT ATTTGCCAAA GCTGCACCAA CAAATCCTTT GATTAAGCGT 3600

TGTGGGCGAT TTGGACGAGA AATTAATTCT GGGTGGAATT GACAAGCCAC AAAAAATTGT 3660

TTTTCAGGAA TTTCAACAAT CTCAACTAAA CGATTATCTG GTGAAACTCC AGAGAAAACT 3720

AAGCCATTTT CTTCAAATAA TTGACGATAT TTGTTGTTAA ATTCATAACG ATGACGGTGA 3780

CGTTCTTGGA CAACATCTTC ATTTCCGTAT GCTGCCGCAG TTTTCGTTCC TTTTTTCAAT 3840

TTACATGGAT ATAAACCTAA ACGTAGTGTG CCACCTAAAT TCTCAATGTT TTCTTGATCT 3900

GCCATTAAAT CAATAATGTT ATTTGTCACG TCAGGATTAG TTTCTGCAGA ACCAGCATCT 3960
TCTAAGCCCA CAACGTTACG ACCAAATTCT ACACAGGCCA TTTGCATACC TAAGCAAATG 4020
CCAAGGAATG GTACGTCATT TTCGCGAGCA AAACGAATCG CTTCAATTTT TCCTTCAATT 4080
CCACGATCAC CAAAACCACC AGGAACCAAA ATACCATCTG CTGAACCAAT TCTTTCAGCG 4140
ACATTTTCAG CAGTTAATTC TTGTGAATCA ACCCAATCAA TTTCGATATC TGAATCAAAA 4200
TCAAACCCAG CATGTtTTAA TGCTTCTACA ACAGAAATAT AGGCATCTGG TAATTCAACA 4260
TATTTACCAA CCAGCGCAAT TTTTGTTTTC TTTTTCAAGt TArGAaCTTT TTCTTCTAAT 4320
GCACGCCACT CAGTCATGTC TGCAGCCGGT GCGTCCAATT TTAAATGATC ACAAACAATT 4380
TGATCCATGT TTTGTGCTTG AAGTGCTAAT GGAATGGAGT ACAATGTTTC AACATCACGA 4440
GATTCAATTA CCGCTTCTGG ATTAACATCA CAGAATTGTG CTAATTTGTT TTTAGTGTTT 4500
TGAGAAACAG GTAATTCTGT ACGGACCACT AAAATATTTG GTTGAATCCC TAAACTTCTT 4560
AATTCTTTGA CACTATGTTG TGTTGGTTTA GTTTTCATTT CGCCAGCCGC TTTTAAGTAA 4620
GGAATCAAGG TTGTATGGAT GTACATTACG TTGTCACTAC CCATATCAGC TTTCATTTGA 4680
CGCAATGCTT CTAGGAAAGG TAATGATTCG ATGTCGCCGA CAGTACCACC TACTTCTGTG 4740
ATAATCACAT CGGCATCCGT CATTTTCGCT GCACGCATAA TTTTTTCTTT GATTTCATTG 4800
GTAATGTGTG GAATAACTTG CACTGTTGCA CCTAAGTATT CCCCTTTACG TTCTTTACGT 4860
AAGACTTCTG AATAAATTTT CCCAGTGGTT ACGTTTGAGT ATTTATTTAA GTTAATATCA 4920
ATAAAACGTT CATAGTGACC TAAGTCCAAA TCTGTTTCGG CACCATCATC TGTGACAAAA 4980
ACCTCTCCAT GTTGGTAAGG ACTCATTGTC CCCGGATCCA CATTAATATA TGGATCGAAT 5040
TTTTGAATCG TTACTTTTAA TCCACGATTT TTTAATAATC GTCCTAAAGA TGCTGCGACA 5100
ATCCCTTTTC CGATAGAAGA AACCACGCCA CCTGTAACAA AGATGTATTT TGTCATAATA 5160
GTTAAAAACC CTTTCTTTGA TAATATTTTG TAGAGAGAAG TAAAACAACT TTGAGACTCT 5220
TACGTTTCTC CAAATAAAAA AGAAACAAAA ATAGAAAAGC TCCCTATCCA TTACGAATAG 5280
GGAGCTTGTA AGTTCGTTCA AACTTCTGAC CTTCTTAAAG GTGCCCAAGT AGTATAATAC 5340
AAGCAAACGG ACATCAGGTC AAGTACTTCT ATTCAAGAAA ATAATTTTTT CTATTTTCTC 5400
ATTTTTCTTG GACATTCAAC TAAAAAGCTT GACTCCTTCT TAGGAATCAA GCTTTTTGGT 5460
TTCTTACTCT TCAGAATATT CGTCATCAAA ATCGTCTTCA TCTTCAACAT CATCGTCATC 5520
AATGATCGTT AAATCTTCTT CGATTTGCGG AATGTCTTCT TCATCATCGT CTGAATCTGC 5580
ACCAATCTCT TGTAAGTCAG AGTTGTATGC TTTGATTTCT TCATCTTCAT CTTCTTCATC 5640
ATCGTACAAG ATGTCTTCTT CATCATCATT TGTTAATTCA GCATCTTCTG GATCGTCATC 5700
GTTATAGTCA ATTACATCTT CATCGTTTGG ATTAGTGATG AAGGCATTTA CTTTTTTACG 5760
TTTTCTACGA CGTGGTGCAT CTTCTTCGTC TTCTTCCAAG CCATGTGTAA CTTCTTCATC 5820
GATTGAATCG ATTGGATACC ATGAACGTAA GCCCCAACGA TTGTCTCCCA ATGAGATAAA 5880

GCTACCGTCG ATGTTTAAAT CTGTATAAAA TTGTGCTAAT TGGTCACGAA TTTCACTATC 5940
AGACTTGCCA AGATAATTTT GGATTTGGTT AACTAAATCA GAAAAATCCA TTACATCTGC 6000
ATGTTGTTCT AAAATAGCGT GTGCTACTTC AATCATGGAT AATTCTTTTT TGTTTAACCC 6060
TTCAAATACA TTAATTTCCA AGAGGTACAC GTCCTTTCAC AGTCTACTCC TTATCATACA 6120
AAAACATTGA AAAAAAATCA ATTCTATTCT GCAATATTTA TGCAGTAAAT TCTAATTCTA 6180
AATGATATTC TCCGATTCGC TCTGTTTGAG AGTATAAATC ATAATCAACT AAAATCTTAC 6240
CAGACATTGG TTGGTCTTTT AAACTAACAC GCAAATTATG TGTAAACGTA CTAATCTGTA 6300
GCAAGCCATA AGGGGTTCGA TAACTTGTAT CTAATTTTTC CTGATAGCCA AATCTTAAAC 6360
GCATCCGTAA TTCACCGGCA CGAATCAACT GAACATGCCC GTCTGGTTCA ATTTTAATGG 6420
TTACGGGGGT TGGCTCTGTA TCCTCTAATA ATTCTTCTTT GTAACGAATG TATAATGTAT 6480
CACCCATTTT TACTAATTGA CCTTCTAAGT CAAAGAAAAA ATCTTTTTGT TCGTTGCCTT 6540
GTTGGACAAT CGTTTTTAGT TGAATCGATA CAGGAACGCC TGTTGATAAA TCCATAATGG 6600
TCACTCCTTT TCCAAAATAT AACTTTCTTA TTGTTGACTT TTTCAATGTA TTTGTCAACT 6660
AAAGTCAATT TTTCCTTGTT TATGTACCTT TTCTTATTGT ATAAGCTCTT TTCACGGGTT 6720
GCAAGCATTT GTtTCATCTA TTTAATAAGA AAACCGTTTA CTATTTTAGA GAATTCCTTC 6780
AGAATtCGkT TATATAACAt GCTTTTTTGC GTTTTTtGTA TATAATATAA GAGAGAGTAA 6840
GAAAATAAAC GTATTAAATG ACGTAATTAT AAATAGACAG GACGTTTTTT ATGAGCACAA 6900
CTTTTCCCCT TCACACCGAA TCATTTATGC TCCTTGATTC AGGTTGTTTA ACAAACAGTG 6960
ACTATTTTTT ACCTTTTGCT TTGACCGATG TTTTAACTAC TTTTTCAGGA ATGCAGCACC 7020
AACCAATCAT TCATTTCTGG CAATTGGATC AAGCAATGAT TCTTGGCATG AAAGATACCC 7080
GTGTACCACA TCTAAaAGAA GGCATCGCTT CTCTGCAAgA AAACGATTAT TCTGTAGTCG 7140
TTCGCAATGC CGGCGGTTTA GGGGTGATTG CCGATTCGGG TATTTTAAAT GTTTCACTTA 7200
TTTTACCAAA TAATTCTGAA CACAAATTAA GTATCGATGC GGCTTATGAA TTAATGTGGG 7260
CTTGGCTCCG TCAGAGTTTT CCGGAGAAAG AGATTGACGC CTTTGAAATT ACTACATCCT 7320
ATTGTCCCGG TACTTATGAT TTAAGTATTG GCGGTCAAAA GTTTGCAGGG ATTGCCCAAC 7380
GACGCGTCAA AGACGGTATT TCTGTCATGA TTTACATCAG CGTTAACGGC AATCAACTTG 7440
CTCGTGGAGA AGTGGTTCGA GACTTTTACC TTGCTGGGTT ACAAGAACAA TTTGGTGAAA 7500
ATGGCTATCC ACCCGTTGAT CCAGCAGTGA TGGCAAATTT AGAAACCCTG ATTGAGACTC 7560
CATTAACAAT AGACGCCGTA AAAACCCGTT TAATCGAAGC ATTACCTCAA CAATTTGAAA 7620
AATCTATTGA TCCAAACTTA ACCGAACCAA TTATAACCTC AGAATGGTTC CAAACAAATT 7680
TAACTGTCCA GCTGGAAAAA ATGGCGCAAC GAAACGCCTT AATTAAAGGA GAGATCTTAT 7740
GACGATTCCG TATAAAGAGC AACGATTACC AATCGAAAAA GTTTTTCGTG ATCCCGTCCA 7800
TAATTATATT CATGTCCAAC ATCAAGTGAT TTTAGACCTT ATTAATTCTG CTGAAGTGCA 7860

ACGATTACGT CGCATTAAGC AACTAGGTAC TTCTTCTTTT ACGTTTCACG GTGCCGAACA 7920

TAGTCGTTTT TCCCATTCTT TGGGTGTTTA CGAAATTACA CGTCGGATTT GTGAGATTTT 7980

CCAACGTAAC TACTCTGTGG AACGATTAGG TGAAAACGGC TGGAATGATG ACGAGCGCCT 8040

GATTACACTT TGCGCAGCTT TATTACATGA TGTCGGTCAT GGTCCTTATT CACATACATT 8100

CGAGCATATT TTTGATACAA ACCATGAAGC AATCACGGTC CAAATCATCA CCTCACCAGA 8160

AACGGAAGTC TATCAAATTT TAAATCGTGT TTCGGCAGAT TTTCCTGAAA AAGTCGCCAG 8220

CGTTATTACT AAACAATACC CAAATCCTCA AGTGGTGCAA ATGATTTCTA GTCAAATTGA 8280

TGCCGATCGC ATGGACTATC TGTTACGTGA TGCTTACTTT ACAGGGACGG AATACGGAAC 8340

CTTCGATTTA ACTAGAATCT TACGTGTGAT TCGACCATAC AAAGGAGGCA TTGCC 8395

(2) INFORMATION FOR SEQ ID NO: 218:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5360 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

AAACTATATC ACTGGAaTCC mAAAGGCGGT ACCAGCACTT GGTCATTGGA TAAGAAAATG 60

CAAGGAAAAA CGAATTTACA TTTATATGAA TTAACAGATC AAGGGCGTAT TGACAAAGGC 120

GCAATTGCCA CTACAAATAA CCAAGTGACC ATCCAAGCCG AGGCTAATAC ACCGTATGTC 180

ATTGCTGAAC CTGACAGTAT TGAACCGATG ACATTTGGAA CAGGAACACC ATTTAAAGAT 240

CCTGGATTTA ATGAAGCCAA TACCTTAAAA mATAACTGGA AAGTTTTCCG AGGTGATGGA 300

GAGGTTAAAA AAGATGCCAA TGGTGATTAT GTCTTTAGTT CAGAAAAAGA AAGAACCGAA 360

ATCAAACAAG ATATCAATCT TCCTAAACCA GGAAAATATA GTTTGTATCT AAACACAGAA 420

ACACATGATC GTAAAGCCAC AGTAACTGTT AAAATTGGTG GTAAGaAATA TACGCGGACA 480

GTGAATAATT CGGTTGCCCA AAACTACATT CAGGCAGATA TTAACCATAC AAGCAGGAAA 540

AATCCGCAGT ATATGCAAAA TATGCGAATT GATTTTGAAA TCCCAGATAA TGCCAAAAAA 600

GGCTCGGTGA CATTAGCGGT TGATAAAGGC AATTCCGTTA CAAAATTTGA TGATTTACGA 660

ATTGTTGAGC GTCAAACGGA TATCATGAAC CCAGACAAAC AAACAGTTAT TAAGCAAGAT 720

TTTGAAGACA CACAAGCAGT TGGGTTATAT CCGTTTGTTA AAGGCTCAGC TGGTGGTGTA 780

GAAGATCCAC GGATTCATTT ATCAGAAAGA AATGAACCTT ACACACAATA TGGTTGGAAT 840

GGAAACCTTG TTTCAGATGT ATTAGAAGGC AACTGGTCCT TGAAAGCCCA TAAACAAGGA 900

GCAGGATTGA TGCTTCAAAC AATTCCGCAA AATATTAAAT TTGAACCGAA CAAGAAATAT 960

ACGGTCCAAT TTGaTTATCA AACTGATGGT GAAAATGTCT TTACTGCTGG GACCATTAAT 1020

GGGGAGTTGA AAAATAACAA TGACTTTAAG CCAGTCGGTG AGTTAACTTC GACAGCAGCA 1080

GATGGTCAAA CCAAGCATTA TGAAGCAGAA ATAATTGGGG ATGCTTCAGG AAACACTACG 1140

TTTGGTATTT TTACAACAGG TGCCGATAAA GATTTCATTA TGGATAACTT TACGGTCACA 1200

GTGGAATCAA AAAAATAAAT TTAGTCTATC TAAGCAACTt GGTCTACAAG TTAAACGTAC 1260

TTGTAGCCCA AGTTGCTTCT TTTATGAAAA GAAGCAAGGT CTAGTGCCTC ATTTTTATAC 1320

TGCaGAATAA ACCGACACAT TTTTTAAAGT GTCAGGCTTC TTTGTGTATT TCATCAAGAA 1380

CGAATGGACA AACTTGCCTT TTGCCACAAA TCGTTTaCmA TTACmATGAT AGGCATAATT 1440

TATTTAGAGG TGAAGTGATT GAAGAAATAC AATGTTGATT GGAAyTACTG GGTTGTCCGC 1500

TTTTTATtTG TAATGGCGCT GATTGTTGGC TATTTATTAA TCACGAATTA TCAGCACTTT 1560

GTCCATAGTG TTTCGGGTTT ATTAGGCATC TTATCCCCAT TCATTACTGG GTTTGTGATT 1620

GCTTATTTAT TGAGTGGAaG TCAGAAAAAA ATCGAGGGAT TACTAGAAAG AGTTCCTTTA 1680

CCAGTCGTTA AAAAAGCGAA ACACGGCTTG AGCGTACTCC TTTTGTACCT GATTATTTTG 1740

TTTATTTTTG TTTTGACATT AAACTATATC GTTCCGCTAC TCATTAGTAA TTTAGTAGAT 1800

TTAGCAAACT CTTTACCCAC TTTTTATGAT CACATGGTTC AGTTTGTAAT GAGTTTAGAA 1860

GATAAAGGGA TTTTAAAAAC AGCCGCGATT GAAAAATATT TAAATAGTGT CCTTAAAGAT 1920

TTGTCGCCAG AACGTTTTTT AAATcAATGG ACACAAGCGT TGTTTTCATT AGGAACGTTG 1980

ACTAAAAATG TTTCATCATT CTTTTTAAAT GCATTTTTGA CGTTGATTAT CTCAATATAT 2040

GCGCTGGTAT TTAAGCAATC TATTTTGACA TTTGTTGAAA AGGCTGCCCA CAAATTGCTG 2100

TCTGAAAAAG TGTACAAGCA AACACAAACT TGGTTAAATA CAACGAATAA AATTTTCTAT 2160

AAATTTATCA GCTGCCAATT TCTTGATGCT TGTATAATTG GCGTTTCATC AACGATTTTG 2220

TTAAGTATTT TGAATGTGAA ATTTGCTGTG ACTTTGGGGA TTTTATTAGG TATTTGTAAT 2280

ATGATTCCTT ATTTTGGTTC GATTTTTGCG TCAATTGTTG CTGGTGTGAT TACGCTTTTC 2340

ACAGGTGGTG TTACTCAAGC AATTACGGTG TTACTAGTTT TATTAATTTT ACAACAAATT 2400

GACGGCAATA TTATTGGTCC TCGAATTATG GGAGATGCCC TAAATGTCAA TCCGATTTTA 2460

ATTATTGTTT CAATCACGAT TGGGGGAGCG TACTTCGGCG TCTTAGGGAT GTTTTTAGCA 2520

GTTCCTGTTG CTGCAATCAT tAAAATTATT GTCTCGGAGT GGCTAAATGA ATCAAAAGAA 2580

AATGATAAGA TTGTAGATTC TATTGAATCC TAAAGTAAAA CTTTTCGTCG CGCGTTGACG 2640

AAAAGTTTTT TAGTGGAAAA GAAAGTTTAA TTATTTGGAA TTTTTGaTAT GCTATACTTA 2700

AAGTAAATAA AAAATCAAAA TTGAGGTGAA TAGATGCAGT TTTTAAAAAA ATATGGCGGC 2760

TATTTTTTAG TTTTAGGGGT CTTACTTGAT TTTTTTACAC CTTATTATGT CGGATTCAAA 2820

GATCAAGGTT ACAATCAATT GACAGAAGTA ATTAGCTTAT TAGGTGATGT GAATAGTCCA 2880

GTTAGGGAAA ACTTTAATCG GTTAACTATC ATTGCGGGAA TGTTGATGTT AGCAAGTTTA 2940

CCTAGGATAT ATGCTATTTT TTCCAGAAAA ACGAAAAAAG GTGCTTGGTT AGTAGTGGCA 3000

ATGATTGGCG CTTATGGTTT ATTTGATTGT ATTTTCAGTG GGCTCTTCAG CGTAGACACT 3060

```
TCAAGCGCTG GCACAGTTGC TGCCGCGCTT CATAACGGAG GATCAGCAGT CGGTTATACA    3120
GGCTTTTTGT TGTTATCAGG CGTCTTAACC ATTATCTACA GTAAATACGG TTCGCAAAAA    3180
AATAAAAACC TTTTTGGCTT TTTGTTTATT TTATGTATGC TTGCAGCCGG CTTATATGGC    3240
TTAGCAAGGA TTCCACAGTT GCAACAGGTA AAACCATTTA ATTATTTAGG CTTATGGCAA    3300
AGAGTGAGTT CaTTTTGTAA TTATCTACCA ATGCTGGCAC TGTGtTTACA AACTAAAACC    3360
AATGACAAAT TTGATTGAAC ATTTGTCATT GGTTTTTTAC TAGTTAAAAT TGTCTGGTAA    3420
TAACTTAATT ATTGGCTACG GTGACTTCAC CAGTAACAGT GGCACCATCT TTTTCTAAGT    3480
CAGCAGAATC TTTGTATtCT TGTTTATCAA AAGTAATATT GCCATTAACT TTGGTACCAT    3540
TTAAAGTAAA GCCATTTGCT TTCACCTCAA TATCACCTTT GACAGTCCCG TGAACGATGT    3600
TGAAATTTTC AGAAGAAACG ATTAGCTTAG GAACCGTGAT TTCATATTCA GCAGTTACTT    3660
TTTTATTATC ATCTTGGGAA TAAAGTGCTA ATTTACGATA GACATCGTTG GAATCTTTCC    3720
CCTTATCATG GAAAGTACCA GCAACTGTTA CCTCTTTATC AAAAGTGACA TTGTCAGTAG    3780
CTGCGACAAT CCAATTTCCT TTATCACTTA ACGCCATTtC GAGCTCTTCA GGCTTGGCAC    3840
TAATTGAAGC ACCAGAGACT ACTTCTGTAC TTGAATTAGC TGTTTCTTTG CTACTAGAAC    3900
TACTATCTGT TTTTTTATCA GATGAACAAG CACTAAGCAA CACAAGAGCA GCAGTTGCCG    3960
TCAAAACAGT GGTGAGTACT TTAGACATTT TCATGAAAAT CCCTCCAAAA AATTTATTTA    4020
TTTCTTCCTT TATCATAACT GATTCAAAGG GAAATGAGTA ATCATACACT CAAAGATTGA    4080
GAAGTTTGGC TAATGGGAAA TTAGGCTTAT TCTCTCAAGA AGGTTCAGGA AATATTCAAA    4140
GTAATCATTT CTTTTTCAAA TGTTCCTTTT CTGTAGTAAG CTAGTAATAG ACAGAATTTA    4200
AGGAGGAGAT TTTGATGCCT TGGAATATGA ATGATTATCC TACATCAATG AAAAATTTAG    4260
CACCTTTGAT TCGGAAAAAA GCGATAGATA TTGGCAATGC GTTGCTTGCG GATGGCTACC    4320
CTGATGATCG CGCCATTCCA ATTGCGATTA GTCAAGCGGA GAAATGGTAT CAAGAGGCAA    4380
GTGCAGCAGA CAAAAAGGCA TTTGAACAAG AAGCAAATCC TACGAAACAA GATTCGCATA    4440
AACAGGATAA ACATGCGGGA AAATTATTAA CGGCGGCTGT CAACGTTwmA TmTAAAGACG    4500
GCCAATGGTT GGTGATTTCA GACaGCGCTG AAAAAGCAaG TAATACATTT CCCCACAAAC    4560
mAGAGGCGAT TGAACGAGCC CaAGAAATTG CTCGGAATAA GCAAACAAAG CTGAAAATCT    4620
ATAAGCAAGA TGGTACGTTA CAAGAAACCA AAGAGTATAC TGAATAATTT TTCGAAACCT    4680
TGATATTCCG TTGCAATTCA AGAAACGGTT ATCAAGGTTT TTTTGCGAGG GAGAAATGGA    4740
CAAGTAAGAA AAAATAACAA AAGTTAAGAA TGAGAATAAG ACTTGCTTTT CTAGTTTTTG    4800
GTGGGTACAA TTACACTAAT TAAGATAAAA AATCAAAATA ATTGAGGTGC TACGACAATG    4860
GAAAAAGCCA ATCTTGATTG GAATAATTTA GGATTTTCTT ACATTAAAAC GCCCTTTCGT    4920
TATATTAGTT ATTGGCGAGA TGGCAAGTGG GAAGAAGGCA CGCTAACAGA TAACAATCAG    4980
TTAACGATTA GTGAAGGTTC GCCTGCTCTA CATTATGGCC AACAATGTTT TGAAGGATTA    5040
```

AAAGCGTATC AATGTGCAGA TGGTTCTGTC AATTTGTTTC GTCCAGATGA AAATGCCAAA 5100

CGGCTACAAA AAAGTTGTGC GCGTCTACTG ATGCCACAAG TGCCAGTAGA AACCTTTGTT 5160

TCTGCATGCC AAGAAGTGGT GAAAGCCAAC CTTGCTTATT TACCGCCATA TGGAACGGGC 5220

GGTACGTTGT ATCTTCGCCC TTATATGATT GGTGTCGGCG ATAATATTGG CGTAnACCTG 5280

CCAAAGAATA CATTTTTTCC AATGATTGCG TGCCGGnTGG TTCGATTTTA AAAATGGGTT 5340

AGCACCAACC AACTTTATCG 5360

(2) INFORMATION FOR SEQ ID NO: 219:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1644 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

TGGCCATGCT TATCTATTAA TTCCCGCTGT GTTTATCGGC ATTTGCCTTT ACTTTATTTT 60

TAACTCACGT ACGATGCAAC GAAAGAAGGC AAAAGTGTCA CGGCTAAATT GTCGGCGGGC 120

TTAGGGCTGA ACTTATTAAT TGCTTTACCA GCCTTTCTAT ATTTATTATC aGTTGGCACA 180

GCGCAAATAC CTTATGTGCT TTTTCTGTTC TTACTTTTTC TATTGTTAAT GGATCTTTTG 240

CTGACCTTTC TATTTGCTGC TTACGTCTTG TATTCATGGA TGTACCAAAT GATTCCTTTA 300

AAAAAAGCGG TTGATTATAT CATTGTTTTA GGTTCGGGAA TTCGTAGCGA GGAAGTACCT 360

CCACTTTTGA AGAGTCGGTT AGATAAAGGG ATTGAATATT ATGAAAAAAA TCCCACTGCT 420

AAATTTGTAG TCAGCGGTGG TCAAGGTCCT GACGAACCTG TGGCCGAAGC TTTCGCAATG 480

AAAAAATATT tGCTTTCACA AAATATTCCT GCAGAAGCAA TTTTGATGGA GGACCAATCC 540

ACAACCACTT ATGAAAATAT GgTtGTTTTC TAAAGCAATA ATTCAGGCAG ATTGGCAAAA 600

GATGCCgtCt GAtTCTAAAc AACCCTCTGT tATTTTTtCA ACAAATAACT ATCATGTCTT 660

ACGAGGAGCT ATGTATGCCC aTCGTGTCGG CTTAAAAGCT GAAGGTGTCG GGGCGCCAAC 720

TGCCTTATAT TTTTTACCAA CTGCTCTAAT CCGAGAATAC ATTGCCTTAC TGGTTCATGA 780

TrAGCGAATT GTGCTTTTTG TTTTTCTACT TGTCACTCTT CTTTTAGGAA TCAGTATCTT 840

ACCCATCTAA AAAAACArcG ACtCACATTT TTGTGAGTCG CTGTTTTTAT TCTTTTAATA 900

TAAAAAGGTT TAATGAACGT AATTGGCTAT CAATGCAATC AGCCATAAAT GGAGCGGATA 960

GAATCCGTAA AATAAATATT TGAAAAATGG TGCTTTGCTC CCTCTCTCAC CATTGTAAAG 1020

CGATATAAAG GGTAACACAG TGATAAACAT AAAATCAGAA TTATAAGCTA ACATTTCAAT 1080

TGTCGTTGGC CAATCTCCCA ACCATTGAAA ACTTGTAACT AAGAAAAAGA GCGCTAATGC 1140

ACCATATAAA CAGTTCCGTA AGACAAGTCT CTTTCTGGcT AAATACGTAA tCAACATAAA 1200

AGGTAACATC ACAATGCCAC CTTCAGCAAA CATTGCTCCT ATTCCTAAAA TTGCTAGTAT 1260

ACTAATCAAT AAAACAATTT TTAAACTGAT TTTTGGcATC TCTAGCATCG CTTTTGTCAC 1320

AATTAmCATA CTGACACCAA GGGCTAAGGT GAAGAAAATA TTGTTGTGAA CAGCGACAGy 1380

GGaTTATTCA CTAAATGATT TAGAAGCGTA TTTCCAACAA ACATGATTGC TGCCCAAATA 1440

TACAAACGAC CATTGTAACG aTAAACATTC CGTGTATAGT TGAAaCCTTC TACAGCCATG 1500

TAACCmAAAA ACACACCGAC ACmACGAGTA ATTACATGAA AAATCAGTGC CCATTCTGGG 1560

CGGAACAAAA TAACTAATAT GGGCCAGGAC CATTAACCCC ATCATCAATA ACTTTAAGCG 1620

ATTTGCGTCC AGGAATCGCC CCTT 1644

(2) INFORMATION FOR SEQ ID NO: 220:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5926 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

nCCnCGTcAC TArGAAtCAG CTACgTCACG AAGGGAAAGT aCCTGCCATT GTTTACGGTT 60

ACCAAATTGA AAGTACACCA ATCTATTTTG AAGAAAAAGA CTTATCAAAA ATTTTACGCG 120

AACATGGTGC TAACACCGTT ATTAAAATGA CAGTTGATGG CAAAAACATC AACACATTAA 180

TGTCGAAAGC ACAGTTAGAT ACCTTTACTG GTCAAATGCT TCATGTTGAA TTCTTATCTG 240

TAAATATGAA AGAAACAACA GAAGTCGAAG CAGAAGTTCA ATTAATTGGT GAATCTGCAG 300

GTGTGAAAGC TGGGGGTACC TTAGCTCAAA ACTTGTACAC TGTTTTAGTG GCAGCCACAC 360

CAGACAAATT GCCAGAATCA ATTGAAGTAG ATATTACAAA CTTAGAAATT GGTGATGCTC 420

TTACAATCGC TGATTTACCT GAACACAAAG ACTATGAAAT TTTAACTGAT CCAGAAGAAC 480

AACTAGTTGC TATTGTTGAA GCACAAACTG CTCCTGAAGA AGAAGAAGGA ACAGCTGCTG 540

AAACGACTGA ACCAGAATTA GCCGAATAAA GAGTACACTT TGTTCAGCT GATTCTCTTA 600

AGTTTTAGAA ATGGGATGTA CGTTGAAACA ACGTACGTCC CATTTTCTTT TTACTGCATG 660

TCGTGATTTC TTCACATAAA TGATGACACC TGCAAAGCTA GGAAGCTACA GAGGAATATT 720

AAATTTTTGC TAATCTACGC AATATTTTCA AACGAAACAA TTTCTTTTTT TTAAATTTGC 780

AGTATACTAA AAGTAAGTAA ACAAAAAGGA GGTCATCCAT ATGTTTGAAC GTTTTGATAG 840

TGACCGTAGC CGCTATGCTT CATTAGGTGT TGTTTCTAGC TTACCTAGTG GCTTAATTGA 900

TAGTATTTGG TTGATTATTG ACTTAAATCT AAAAGGGGTT ATCCCTTTAA ACGACTTACT 960

TCACTTTGAT TTGTTGAACA ATAATGGCAA AGTGACTGTC CATTTTTCTC AAGAAAATAG 1020

TTCAGTAGAA ATGGCCATCG ACTTACCATT TAGTTATTCT ACGGCATATC CGTCGCGAAT 1080

TTTTGCTTTT GATGATGGCC ATCGTGAAAC GATTTTATTA CCTGCAGAAA TGTTAGAATC 1140

TTAAATAATT TCAACAAAGT AGCCTCAAAT TCCGATTAAA CTGGAATTTG AGGCTACTTT 1200

AATTTCTTTT AGTTATTCTA CAATAATTTG TTTACTACTT CCTGCAGCAT TTTCAAAATT 1260

AATAGTAAAG GTTTCTGTTG CATCATCATA AGTGATTCCG TCACTTTGAC TAATTACTGT 1320

GCTGTAATCT TTTGGTATTT TTAAAGTCAG TGTCTTAGTT GTTTGTGTGG GATCAGArAA 1380

AGTATACGTT TTTTGGTGTT GTTGTGTCTT TCTCATTAAA GAAATAGCTT TATTTGAGGT 1440

AATCCCTGCT ATTGTACCGC CAGTTTCTGA CCAAATGTTA ACTCCAAGAT AGCCTTCTTC 1500

TTCTTGCTTA ACTGCTTGAA TTTCTTCTGT ATTTGAGAGA ACTTCTACTG TTTTTTTGGC 1560

AGCATATGTT TTTAAATCAG TTTCATCGAT TCCTGGTAAA ACTACATAAG CATAATGCTC 1620

GTTCGTTGGG TGTTGACCAT GTTGAATCGT AAATGTTCGA TAATCACCTA CATATTCTTT 1680

ATCACTTGGA AATGCTTCAT TAATTGCTTT ATAGGTTCCT TTTCTAGTTT CACTTTTAAC 1740

ATTTACCGTT GTTTCTTCTG GAAAATAATA ACCCATATTA GCATTGGAAT GATCCGACTT 1800

TAGTAATAAC CATTGTTTTT TACTCGTTTC ATTTTTCTCA TGAATCTCCC CAATATTTGA 1860

TAGCACTTGA TACTGATACG CATCATTTAA CAAACGATTA TCAACAACTG TCTCAATCGA 1920

AGCTTCTGTA TCCCCTTTAA TTCCTGCGCC TAATGCGATA ATTTGCCCAT TTAATACAAA 1980

CCAAGATTTT TTAGCTTGTA GGTTCATAGG TAAAAGTTTC CCATTATTTT TGGTCCCGTC 2040

TTTATTAAGA GCCATCCCTA CTACTGCCTG ATTGTCTGAC GTTACTCCAC CAACCCATTG 2100

CTCTTTCGAT GTGATTGTTG TAAATGCTGA AACTTCATCT GCCAAAGAAA TGGTATCTAC 2160

CGTTGTTCCT GGTAGTCGAT AGGGATCTAC AGTTGGCCAA TAGCTGGAAT TAAATTGAAC 2220

CTCGTCATCA TTATATACGT ACATCATTCC GTCACCAGTG TGCCAGCCAC GTTTATTTTC 2280

TTTATTCCCT GCTTCAAAAG ATGCAGTGCG CTTTGAATAC AAGCCTAGCC CAAACATATA 2340

ACTAGGTGTG CGTTGGACAA ATCGATCCAT CGAAGCATAA AGTTTCGTGC CTGTAAAAGG 2400

AAGTTGCCCT CCCGTAATTT CTGGATTAGT TAAAAGTTGC ATAGTCATCT GTAAATCATT 2460

AAAATCACGA GCATTTGTCA AATAATAATC TGGGTTCTCC TTCATCCAAT ATTTAACGGC 2520

TTCTTGAAAT TTCTTTTGAT AATTATTAGG GGCAAATTTA GCCACAATTA ATAAATTATA 2580

CATGGTAGTT GAGCCGTAGC CTGTTTTTCC AACTGCTGGG GCTCTTGAGA TGGAACGGCC 2640

ATTCACGGTT GGTAGCATTT CACCTTTATA AATCAAAGGA AGAAAGGCAC GATCAACATT 2700

TTCAACAAAT TCTGTCACAA GAGTTGCATC CATTTGAAAA GATGAATCTG CTGTAATGGC 2760

CAAAATTTGT CCAACCCCTT TCACCAAAAC ATTGCCATAC GAACCTGTAT AAGGAATGTT 2820

ATTGTGCTGG ATAAAAGAAC CATCTTGGTA AAAACCATCA CCTTTTGTGA CTAATTTAAA 2880

TACATCGACG ATACTACTGG ATGCTTGGTT AATTTTCCCG CTATCCTTTT GTAAAATTCC 2940

TAACCCCAAA ACAGTTAAAG CTAAGTCCGT TCGGTTCGCT CCTGATGTCA CAAAATTTGG 3000

GATAAAAGCT AAATCCACAA ACGTTCCTTG TGGTTTTGTA TACAACTGtT TAAACGGATC 3060

TGGGACATAA CTGCTCAGCG CATTTGTATA GATTTGTTGC TTTTCTGGAG AAACTTTATC 3120

CTCTAAAATC ATTAGAATAT TAAGAAACTT CTGTGGCACA CCAATTTGCC AGTCCCACCA 3180

ATTTCCATGG TACTTTTTCC CATCATATCC TTTTTGCGTT ACCATAAAAT CTAAGCCATC 3240

AATAATGGTT GCTGCCAATT TGTCATCTTG GTAAAGCGTA CTCCCCTTCG TTCCATAGGn 3300

TAACGCTAAT TTTTGCAATT TGGTAAATTG AGTCGTTAAA TCAGCCGATG GCGTATTGCC 3360

AGGCTCTAGT GGCCATAAAT AGCTGCGGTT CGGCTCTTTA TTCATCGTTT GATCGAGTTC 3420

TGTTGCTTCA TTGGAGATTT TTTGGACGTA GTCTACAAGC GCTTGGTTAG TTGAATCATA 3480

TTCTTTTGAA ACAAGCTGAT TTCGCCATTT tTCTTTTAAT TCTCGGTATT CTGCAGATGA 3540

ATCGATAACA GGAATGCTAC CAACGGGGCC ACTTTTCATA GAAATGATAT TCAGCTTTGA 3600

TAGATTACCA AATAGTCTAA TTCCTACAGC ACCCATCCAG TCATCATTAA TGGACGTACC 3660

ATCAGGATAA TGAACCTCTT CATTTTCATA AAATAATTGA TCATTCAGGA ATGTTTGAAT 3720

TTTTTTCCCA TCATAGCGAA CTAGCAGTTT ATATTGTTGA CCACTTAGTA ATGTTGGTCC 3780

CGGAATATTA GTTAGCCATT TACCGCCTGG TTGTCCTAAC TGCCAGCGAC CGTCACGATT 3840

ATAAGCAAAA GATTGCCATT GACTCGTCTT TTGTTTATCT CCGCGAAAAA CCAATCCGAA 3900

GTTACTTTGC CCTTCATATA AAAAGGTTAG TTCAAGATCT CCTGAAGATT GTTCGCGTGC 3960

ATCAAGATTG ATAGAGACAG ATTCGAGGTT TGTCCCTTGC TTCGTGTTTG CGATCCATAA 4020

ACCAGCAGAA TCATTCTTTT TTTCTGCTGT TCCTACCAAA TCTTGCCAAT GCCCAACTTG 4080

ATTGGGGAAA TCTGAAATGT AAGAGGGTTG CTCAATCGTT TGGCTCGCTT CGTTTTCTTC 4140

TGCAAAAGTA AGGGTGGTTC CCATAAATCC ATTCGTCAAA GCAGTAAACA GTCCGACCAG 4200

CATTACCCAT ATTTTCCTGT TTTTCATTCT TTCTCCTCCT AAAAAATTAT CGTAAAACGC 4260

TTACATCCGA ATTAGAGAAA AAGAGTGAAG AAAAATACTT TCTACACTCT CCGACTCTCT 4320

GAAAATCACT TAATATACTT TTTACGCCAA AATACCAAAG GCACTCAAAA CAATCGATAA 4380

AATTAAAATA ATCCAAATCA CTCGTGTGGA ATTTAGTTTC TTTTGACCTA ACATCCAATA 4440

AGTCAAAAGA ACAAGTAAGA TTGGAACTAA GGAAGGCAAA ATCGCATCTA ACGTATCTTG 4500

AATAACTAAT TCGACCCCGT TCTTTTTATA AACAAAAGGA ACTGTCGCCT TGATAACTGA 4560

AGGAATTAAG GCACCAATCA CAGTcACACC AAGAACAGTT GCTGCATTCG TAAAGGCAGA 4620

TAACTTGTCT TTTAAGGTCG TCACTAGTTT GACCCCTTGC TTATAACCAA GTGGTAAGAG 4680

CGCGGCCCGA CCAAACAATA GTGCAATATT TGCAATAATC CAAATAACAC AGCCAACCAC 4740

CGAACCATTT TGAGCCAGTG TCCCTGCAAC AGAGCCGAAA ATCGTCCCCC AAATAACGTG 4800

GAATAAAGAG TCCCCAACGC CTGcCAAAGA ACCCATCAGA GCTGTTTTTA GCGCAATAAT 4860

TGTTTCTTTC GCTTTATAGC CATCTTGCTC TTCAATCGCA ACATCAATCC CCATAATTAA 4920

ATTCCCAAAA ATAGCATTGG TGTTAAAAAA TTGATTGTGT GTGCGCATCA TATCTTGAAA 4980

TTGTGCTTCA TCATGACCAT AGAATTTTTT TAAAACAGGT AAGATGCCGT ATAAATATCC 5040

TGAACTCATC ATACGTTCAT AGTTCCAGTT TAATTGGCTT CCTAAAATAT AACGCCAGCT 5100

AATTGCATTT AATTCTTTTT TTGTCAGTTT ATGCTTACTC ATCGCCGTCA TATCCTCCTT 5160

CAAAATCACC ATTTCCTGTT GTTGCTTGTG TCGTGCTTTT TGCCCCATCA CGTTTAAATA 5220

CAAGCATCGC TGCTACCATC CCAACTATCG AAACACCTAA CATTGGAaCT TTTAAATAAG 5280

CCGCTAAGAA AAAGCCTAAT AGTAAATAGG GGATATATTG CTTCGTTGGC AAGTAACGTA 5340

ATAAAATCCC TACACCGACA ACAGGTAATA CTCCTCCAGC TGTTTTTAAG CCCGTCATTA 5400

GCCATTCTGG CATGTTATCG GTAATCGTTG TTACTACAGC ATCCCCCGCT AATAACATTA 5460

ATAAAATCGG AATGGCACGA GACAGTCCCC ATAGAACAGA ACCAGATAAA ACTAAACGTG 5520

GAATCCGTTT TACTTGCATC TTATCAATCG CTTTATCGAC ACTATGAAGT AAAAAGACAT 5580

TACAAAAACG GGCGACTACA TCTAGTTGCA ACATCAACAG AGmAACAGGA ACAGCTAAAC 5640

CAATGCCATA CTCCGCTCCT TTTCCAGAGA TTACAGCGAA AGCGGTTCCT AAAACAGCCC 5700

CTGTGAAATA ATCAGGTACC GATGCTCCAC CAAAAGCAGC AATTCCTAAA CGCATTAATT 5760

GAAGAGTGCC GCCAACCATT AAACCAGTTT TTAAATCTCC CATAATACAT TCCAGAAATC 5820

ATTCCAGCAA TAGCTGGTTG CGTCAACGTA TTTGAAATAA GTGGTCATTA ATAGCnATAA 5880

ACGCTTAATA GTAATTAAGA TAATTTGATA CGGCTGCTAA ATGCTT 5926

(2) INFORMATION FOR SEQ ID NO: 221:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9121 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

GGATATaGAA aCCCAACGGT TnCAAACAGC AAAAAAAATG GCaAAAATAT TACAAGTTCC 60

GTGGACAATT TTTTTTGATG AAGAAGTACT CGAAACGTAT GATTTTTGAT aGGrGGATAA 120

AAtGACaACC AAAGAAAAAA TTGAATTTAT CAAACAAGTA ACACyTCATT CAGATTCAGA 180

AGTAGAGAAA ATTATTAAAG GAATGAGCGA TACCAGTATT AACCGCTGGT ATGAAATAGA 240

GAAATATCGC ATTGATCAAG AATTAGAAGA AGCGGTGTTA ACTATCTATT GTTAAAAGAA 300

TACACCTATA TCAGTCGGAT AGAAATAACA AATTATACGA AATGAGGTAA AAATGACGAA 360

ATCTATAAAA CTAAACAAAC AATTAAGAAA AGCAGCTTTT AAAAAAGGAT TAAGCCAAAA 420

ACAAGTGGCC CAATTAGTAT ATTTTGCACA TACTACAACA AATGGTCACT TTAATGGCTA 480

CCCAGTACCG CCAGAAAGTG CTATAGCCTA TAACGAATTA TTTAACGATA GTGAGTTGGC 540

ATTTGCATTA GGCCAAGAGT TATTGGGCCT AATAGGGCTA GCAACTGGAG TGAAAGTAAA 600

AAAAGAGCCA TTAGCGCTTT CGGTACTAAA GGAAAAAGAG GAACGAGAAA GAGAAAAAAT 660

CGAAGTGGAA AACGAAATCG ATTATTTAAT GGCGATTCCA GTTGAAGAAT TGACCGAAAA 720

ACAAAAACAG GCAATATTAC AGTATTGTAG CGAATATTCA GATGAATTGT TGTTTGAAGT 780

TTCTCTTATT TGTAAGCAAT TAGAACTAAT AGGGATGTCC TTTATGGATT TAATGATATT 840

AAGAACACCT TATTGGAAAA ATAAAGAATG GATTGaATAA AGGAGGGAAG TAATTTGGAG 900

TCAAACATTA GTATGCAAGA TGTGTCTAAG AGACTTTCTA GGAATCTATA TATGATGATG 960

GGAACTCCTA GAGAGGATAC CAAAAAAATT TGGAAGGTAT CTGAGCTTTC AAAAGCAAGC 1020

GGAGTTACAC CATGTGTTaT TAGTCGGATT AAAAATGACA CTGAAGGAAA AGAAAAACCG 1080

ACTATTGAAA CAGTTGTTAA GCTTGCTAAA gCATTAAATG TTGATCCTGC AGAATTATTG 1140

AAATAAGAGG TGAAAATAAA TGGAAAAAGA AAATATCCAA GAAAATTATC ACGAATCATT 1200

AAAAGAGCGG AAAATCAAAG ACACTTTACA AGAAATTAAA GAGAGATTAA ATGGCGTAAA 1260

ACTAACCTAA AATGAAAGGA AGTTArAAAT GAATAAGAGG AAACCACGTa GTTTATATGA 1320

AGCAAGAATT TTAGGGACTT TGTTAGTATT CTTCATACTT GGTCTAGTAA TAAAGAACAG 1380

TATTCCGTTT AACTACTTAC TTTTAATACT TGGACCTATT ATTGCAATTT GGTTTATGAA 1440

ATATGACGAT GCTAAaTATC rAGAAAaTTT AAATAAAAAT GAAAAGGAAG TGTATTAATT 1500

GAATAGAAAA ATCAGAAGTC TAATAAAAGA ATTAACAGAG GAATGTGACA AAGAAAAAGT 1560

TTCTCTTATT TGTACAGCTA ATAACCAAGG TGAGACAGTT AGCGCAATTT GTGGCGGATT 1620

AGTAGATCTA TCATTTTGTT TAGGAGTCCA AGAAAAAAAG CTAAGTGAAA AATTGCCGAT 1680

ACATCCAGAA ATTTTGCGCA AATCAGCAGT TGAAGCATTG GAAGAAGTTA AATCCGATAA 1740

TCATAAACAT ACTTTTGTAA TCGAAAATGC AGAAGATTTA CAAGATATAT TAAACAGAAT 1800

TGCTTCAGGT GAGTTTGATG AATAAAAAAA TGCCCAATCG TTTGCAGACG ATTTAGGCAC 1860

TAACAAAAAT ATACTAGGAA AATTATAACA TAAATTTGGA GGGAGTTAAA TGCAAAAAAA 1920

CAACAATAAG CTAACCATGT CAGCCCGATT CGGTGAATGG TACAAAGAGG CGACTGAAAA 1980

TTGTGCAGAA CGGGAATTAT ACGCAACGTC GTTAATTGCA CGTATGGATT GGGTAGTTGA 2040

TCCAATAAGT AGTAAAAGTT ATAGATATGA TTTAACAACA GATCTGAATA AAAATCATAG 2100

CAGCGCTTGG TATGATGTTG CCACTGAAAT TTGTAACAAA CGCAAAGAAA CAGTTATACG 2160

AGCAATCCTA GAAGACAATA TCAAAATTGA AAATTTCTAA TAAATGGAAG AGGAGGATTT 2220

CAATGTATTA TCCAACTACA AAATTTGATC AGTATGGTGC TCGCCAATTA CCTAAAGAAG 2280

ATGTATCAGA AGAGATAGAC GAAACAGCAG ATTATGACAA GGATAGTATG GGaGAAACTA 2340

TCGCCCAAGA AGATGGTGTA TTTGAAATAT TGTATACAGT AACTACTGTC ATGAGTAATA 2400

GATGTCTCAA AATTAAAACA GAGAAAGCGA TTGTTACTCA ACAAGATATT GTAGATTTTA 2460

TTTTAGAAAT AGGTGAAGAA AAAATAACAT CTATAAAATT TGTAGGATTA GGAGAAAAGT 2520

ATTTAGAAAT TGGAGGAAAT TATAATGGGG AATGAATTAA TAGTAAGCGT TCAAAATAGA 2580

ATTCAGGAAA TGCAACATGG TGAAGGATTG AGATTACCCA CAGGGTATTC GGTAGGAAAT 2640

GCTTTAAACT CAGCATATCT AATTCTTAGC GATAATTCTA AAGGAAAATC ATTATTGGAA 2700

AAATGTCACC CTACATCAGT GTCAAAAGCT TTACTAAATA TGGCAATACA AGGACTTAGC 2760

CCAGCAAAGA nTCAATGCTA CTTTGTTCCC TACGGGGATC AATGTACATT GATGAGATCA 2820

TACTTCGGAT CAGTAAGCAT ATTGGAACGA CTTTCTAATG TAAAAAAAGT ACATGCTGAA 2880

GTAATCTTCG AAGGAGATGA ATTTGAGATA GGATCGGAGG ATGGTAGAAC CGTCGTTACA 2940

```
AACTTCAAAC CTAGTTTTCT GAACAGAGAT AACCCAATTA TTGGGGCATT TGCTTGGGTA     3000
GAGCAAACAG ACGGAATCAA AGTTTACACA ATCATGACAA AAAAAGAAAT ATATAAATCA     3060
TGGAGTAAAG CTAAAACAAA AAATGTGCAA AATGATTATC CGCAAGAGAT GGCCAAACGC     3120
ACAGTTCTAT CAAGAGCAGC AAAAATGTTT ATCAATTCAA GTAGTGACAA TGATTTACTC     3180
GTAAAGGCTA TTAATGAAAC AACAGAAGAT GAATACGATA ATAATCAGCC ACGCAAAGAT     3240
ATTACACCTA ATCCACCAAA TATTGAAAAG CTTGAGAAAT CAATTTTTAA TCAAGATGAA     3300
AATAAAAAAA TTGCTCAAGA TATGATTGAT TCCATTGATT TAAATCAAGC TGACAAaGAT     3360
CTACAAGAAG AGCTAAATAT TGAATTTCCT GATCCTAGCA AAAACTATTT AGCTACTGGG     3420
GAGGTAAATG GAGATGTTGA AAACGAAGAC GGCCCTTACC CTTTCTGATA AAAATTATTA     3480
TTCAACTGAA GCTGACTGGC ATTATATGTC GGTCAGCCAA TATAAAAATT TTAATGAATG     3540
CGAAgCAGCC ACAATCGCTA GATTAAAAAA AATTTGGTTA CCCTGTTCAG ATCCAAAAGC     3600
ATTGCTTGTT GGTAATTATG TTCATTCTTA CTTTGAATCA TCAAAAATTC ATGATCTATT     3660
CAAAAAGGAA AACAAAGACA AAATGTTTTC TAGTAGAAAA CCATATGGGC TTCTAAAAGA     3720
TTTCCAAATA GCTGAACAAA TGATTCAACG ACTAwAAGAA GAACCTGCTT TTATGAATTT     3780
ATATAAAGGG GAAAAAGAAG TAATTGTTAC AGGGCAAATT GAAAATGTAA ATTGGAAAGG     3840
GAAAATCGAT TGTCTAAATT TAGAGGATGG TTATTTTGTT GATATAAAAA CAACAAAAGA     3900
TATTCATGAA CGTAAATGGA ATGATAGCTA CGCGGAGCkT TGtACATTCA TTGAAAACTA     3960
TGGATATGTA CTTCAAATGG CAGTTTATAA AGAATTGCTT AGTCAAAAGT ATAATAGGGA     4020
ATTTGTACCA ATTATAGCCG CTGTTTCAAA GCAGACTCCT TCTGAAGTAA AGCTAATTAC     4080
CCTAGATGAA GACAAGATGC ATTTTGAAAT GATCCAATTA AAAGAAAATA TTGAACATAT     4140
TATTAAAGTA AAAAATGGCG AAGAAAAACC AACACTGTGT GGTAGGTGTG AATATTGCAG     4200
AGGGCAACAA CGAATTACTC ACTTTACTAA TATGAATGAT TTATAGGAGG GAGGCAGATA     4260
TGGCAAGACC TTTAAAACAA GGAATTGACT ACTTTCCGCT TTCTGTAGAT TTTTTAAGAG     4320
ATATAAAGGT TCGTAAAATC AAACGTGCGT GTGGACCTTA CACTGTCGAA ATACTACTTT     4380
GCCTGCTGGG TAATATTTAT CGGGAAACTG GTTATTACAT CGGGTGGGAT GAAGATACGA     4440
TGTTCTTAGT TGCTGACGAA gTTGGGGCGA AAGAGGGTCT AGTTGAAGAG ACGGTAAATA     4500
AGGCCGTTCA AGTTGGATTT TTTAATCAAG AGAAGTTCAA TGAGTATAGA ATTTTGACTT     4560
CTAACGGAAT ACAAAAAAGA TATCTCGAAG CCACTAAAAA AAGGAAAGAA GTAGTTATTT     4620
CCGACATTTA CTTAGTTAAT GACACCATAA AGGATGAAGA AACCCTAGTT AATGGTGTCG     4680
ATAATGAACA AAGTAAAGTA AATAAAAGTA AAGTAGATAA AAGTAAAGTA AATAAAAACA     4740
ATAAAGAAAA ACCTGCCGTT CAGTTGTCGT CTGAAAAGGA TTTTTTGGAG AATCCTTTAG     4800
GAGACAAAAA AACAGCGGAG CTAATCGCTT ATTATTCAAA AAATGTTAGT CTTGCTACTC     4860
CTGTAAATAT GACAAATCTT GCaTATGATT TGAAAGACTT TAATGGCGAT CTGGATTTAC     4920
```

TAAAAGAAGC GATAAACATT TGTTCAAGCA ATGTAAAAAC ATATGCTTAT TTTGCTGGAA 4980

TATTAAAACA GTGGCGAGCT AACGGAATTA ATACTCATGC AGATTTTTTG AAGAAAGTAA 5040

ATAAAAAATC AAAGCAAAAA AATAAACGAG ATAGTGAACC ACCGCTAAAT GATTATTCAG 5100

GATTGTTTTA AGGAGGAATA GCTAATGATG AATACTAAAG TTGCATTAAA GCAATTAATG 5160

GCTACACATT TTATTACAGT TGATGTCCCT TGTCCTGAAT GCGGAGGAGC AATGACAGCA 5220

TGGAAAGAAC CAACTCCAGA TACTCCACCA AGATGTCCTC CCGTATGTAT GGAATGTGGC 5280

TATCGTTCAG TGAAGaAAAA AGAaGCAaCC ACTGCTAAGA aCCTTTATGr ATCTAGTCTG 5340

ATAAAAAAGG CAAAAGACTA TTTTTTAAAC GGATCTGTAT TAACGGATAG AAAATCACTG 5400

CATAAAAAAA TTAGTAATTA TTATCCAAGC AACCAAGAAT CAACCAAAGC TAAAAATATT 5460

AGTATGGAAT TTTGCGAAAA AACACTGCTG GGAGAAATAC ACCATCTGAT TTTAACAGGG 5520

GGAGTTGGGG TTGGGAAATC ACATCTAGCA ATTGGATGTC TCAACGAAAT ACTTGTTAAA 5580

AGTAATTATT CTAAGAAGGT ACTATTTGTC AGCTATCGAG AGCTTTTAGA ACAGTTAAAA 5640

TTTGCGATGA ATGACGAAGA GGCACGAAAA GCAATTACTG GTGTAGCAAT GTCCGAAATC 5700

AAAGCAGCAG ACGTCGTAGT TCTAGATGAT ATAGGGGCTG AACTAGGCGA CTTTAATAAA 5760

AGAGATGAAA AAGGCAATCT GATTCTTAGT AGAGCATCAA ACTTTGATAT TGATACGCTG 5820

ACAGGAATAT TGGAAAGTCG CATAGATAAG CCAACAATTA TCACCTCTAA CCTAAAATCA 5880

AAAGAAATTA AATATTGTTA CGGCGAAAGA ATAGTCTCAA GAATGGCTAC CCATTCAAAT 5940

GGATTTATGC ATAGGTTTAC AAATACAAAA GATTATCGAT TGAAAACAGC GTAGGAGGTC 6000

ATATGAAAAC TACTATACCA ATTACACCAA AACCACAAAG CCGTCCGAGA TTTGCTAGAC 6060

GTGGAAAGTA TGTACAAGCC TATGAAGATA GCACCATGAA GACGTATAAG CAACmAGTAA 6120

AAAACTATCT TAGAAGAGCA AAACCTAAAT TGATAGAAAA AGGGCCAaTC ATGGCACACG 6180

TGACGTTTTA TGTACAAGCG CCTAAGTCAG CCCTAAGTAA TAAACAGAAG CGGTTAGACG 6240

TAAAATTAGA GCGCAAATAC TGCGATAAAA AGCCAGATTT AGACAACTAT TTCAAAGCAG 6300

TCACGGATGC TGCAGAAGGT ATTCTTTACm AAAATGATGG GCAAATTGCT GTGATAGTTT 6360

GTCAAmAACT ATACAGCTTA AGACCTCGAA CAGAACTCGA AATAATGAAA TTGGAGGTTG 6420

ATTCATGAAA GCGCAAATGG AATCTTGGCA AGATATACAA GAGTATGTAA AGAACAACAA 6480

GAAAAAACAT AAACGTAGAA TGGCTAACGA TGCAATGAAA AACGTCAACG AGTTAGGATT 6540

TCCTTTCACC TATTTCGAAG ATACTTACCA ATTTGCAATT AAAACATCTA TCGGGATGGT 6600

CGACTATTTT GGAACAACTG GCACATGGGT TGTGCGTAAG AATCAAGATC GTGGTAGAGG 6660

GCTAAAAAAA TTAAAGAAGT ATTTGGAACA TCCAGTTCCA ACAAAAGAAA TAAAAAAAAC 6720

AAAGGGTAAT CAGTCATGGA TGAAAAATAA TAGATAGGAA TGGGAAAAAT GAATAGGAAT 6780

AAATGGCAGA GAGTGCCTGC ACCAAACACA CGTGGTGCAA AAATGCATGG ATGGGTTAGA 6840

GAAATGGATG AAGCCTACAG AGAGGTGAAT GGGGaATACG CTGTTCTTGT CCGAACTTTG 6900

AATACTAAAT TAGGCACAGT TAAGCATGCA GCTATCAGAA ATGCTACAGA AACTGACATT 6960

ACTTGGaGTG AAAAACcAGC GTATTAAAAA TGrAATTTTT gGaGAGGaAA AACAAGCGAT 7020

TGAATTTTTT CCAAAACAAT CaGAATTAAT AGATGAGGCT AATATGTATC ATTTATGGAT 7080

TTTAGAGGAT TGCGAATTGG AATTTGGAAT ATACGAATAA TTTGGAGGTA ATACTTTTGG 7140

GTAAAAAAGG TTCAAAGATT AAAAAGAAAA AACGCCGACT TAAAGAAAAA GCTATTGCAA 7200

ATGGCACATA CAGCAAGCGA GGGAAAAGCG ATGGTATGTA TAAAATGCAA GGGACAGATG 7260

ATTGTTTGGG AAAAAGATAA GTTGAAACAT CAAAAAGGTT TTTATTGTCC GTTATGCAAC 7320

AAAAATGGGC AAAGTGTTGC GAAAAAGTTA GCTGAAATAA AGAAAAAATA AGCGAAAGGA 7380

GTGGAGTTTT GCGGCCGCAT AAAAAGCTTT TTGCTCCTTT GAAATGATGA AAAAAATACT 7440

TGATGCATGC TGTGGTAGCA GAATGTTTTG GTTTGATAAG CAAAACGACC AAGTTTTGTT 7500

TATGGATAAC aGAAAGCATT ACgAAAAATT AGAGAGTGGA CATATTGTTG CTGTTAATCC 7560

TAACaTTGTT GCAGATTTTA GAAAGATGCC GTTTGATGAT AATTCATTCT ATCATGTTGT 7620

ATTTGATCCT CCGCATTTAT TGAGAGCTGG TAGTAACAGC TGGTTGGCCA AAAAATATGG 7680

GAAGTTGAAT GAGCAAACTT GGAAAGAAGA TATCCAAAAA GGTTTTTCTG AGTGTATGAG 7740

GGTCTTAAAA CCTAACGGAA CATTAGTTTT CAAATGGAAT GAGGATCAAA TCAAGCTATC 7800

TGAAATATTA AGCATAATTG ATTACGAACC ACTTTACGGC AATAAAAGAG CGAAAACGCA 7860

CTGGTTAGTG TTTATGAAAG AGAGTGAAGA AAATGAATCG AATGGTTGAA GTTGGATAGG 7920

TTTATAAATT AGTCGATGCA AATCCACCTG CATTTGGAAA AGTAAAAAAA TCTTTGGAGG 7980

CTAATTATGA AATTAATCTA TGTTTTATCT GGAAAAGAAG AAAATAAAAA CTATGTAAAA 8040

AAATTTGTTG GAAATTATTG CAGTTTTGGA CCTAAAGAGG ATGCAAAAGC ATTTACTAGT 8100

GAAGAAGCTG AACAGATGAG AAGACTGTTA GAGAATAGTG TAGGCAATGC GTTTGTTATT 8160

GATGATGACA GAGAGGTAAA AAATGGGTTT CAAGTTTAGC GATTATTATA ATGCTGAAAA 8220

TATTTGTCTG TCAGAAGAAA ACATTAGAGA GTTACTTATT GCTTATTTAG AAAAAGGAGA 8280

GGAAGAAAAT GATTTATATT AAAAATTTTG TACATGATTT TGATTCTTCA ACAATTACCT 8340

TTGAAGTAGA GCGTGATGGG ATCACGAATT ATGTAGGAAC TAGAGATACT GGTTATGGAA 8400

CGACTTCTAC AGATATTAAT GATTTTACTG AAGaTTGGTC AGATTCTGAA TACGATCAAT 8460

TAGAAGAGTT TTTAAACGGT TGTCAAGAAA TAGTACATTC ATTTTACCAT TAATTTTTTA 8520

GGTTGGAGGA ATAACAAATG GCtTATGaAA AATTACaTTT AGTaACAGCG ATAGTAAGTG 8580

GAGATATTTA TGTTGGTAGA GTCAAgACGG CTTGATGGAC ACAAGATATC GTCGAATAAT 8640

TACAGGTGAG GCTATTCAGT CGGTGGCAGA TTGGTTTTAT GTAAATAAAA AAAAGACAGT 8700

CCAATTTAAA GGTATTGATG GTAAAGAGCA TAGCTTGTTT TATACATCTG ATAAAGAAAA 8760

AGCAAAAAAA ATTCTAGCTA TTTTAGAGGA GGAAAAACAA ATGAAATTTT ATAAAATGAA 8820

AGAACCTTAT TTTGCATTAA TCGCTGCTAA AGATGAAAAA CAATGTTTAA AACTTTACAA 8880

GGATATTGTT TGCGAAGTAG AAGACGAAAA AGAATTTTTC GATGATATGA AAACAATTGA 8940

TAAATACGAA GCGTTCAAAA TGCTTGCTAA AAGTCGTATT GAAGATGGTG GAGAATTGGG 9000

GGCAGAAGAA GCTTTCAATC AATTAGAAAA TCTTGAAGAA AACGGCGAAG TATTGTTAAT 9060

CGACGGTAgC TTGCTTTAGG AGGrGTAATk GTgCtATCTA TTCTAAGAGG TTATAAAGAA 9120

G 9121

(2) INFORMATION FOR SEQ ID NO: 222:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11597 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

AGCCCAAAAT TAGGGAAGGT CTTCTTTGAC ACCTACtTTA ATTTTCCCyG cTTTTTTTAA 60

TTGTTCTACT GTGACCGTTT TGTTTGATGA AGCACTTTTG GAAGTTGAGT CATTGTTGCC 120

GCTACCACAG CCaACAATGA TGCCTGCTAA TACAAGTAAG GAGAGTCCAG CTAGCCATTT 180

ATTTTGTTTT TTCATTGTTC ATTCCTCCAA TTTTTTGTTT TTTTATTCAT TACCCGTTAT 240

GAATAATTTT ACTTAGAAAT TCTTTTGCAC GAGGGTTTTC AGTGCTTGTA AAGAAATGTT 300

CAGGCGTTCC TTGTTCAATA ATCTGTCCTT GGTCCATGAA GATTACACGG TCAGCCACTT 360

GACGAGCAAA GCCCATCTCA TGTGTGACAC AAACCATGGT AATATTTTGT TTGGAAAGAT 420

CGACCATTAC ATCTAAAACT TCTTGAATCA TTTCTGGATC AAGTGCCGAA GTCGGTTCAT 480

CAAAAAGGAT AATCTCAGGA TGCATATTTA AGGCACGAGC AATAGCGACC CGTTGTTGCT 540

GCCCGCCAGA AAGCTGCGAA GGATAAGCAG AAGCTTTATC AGCCAAACCG ACTCGTTCTA 600

AATACTCCAT GCCAGTACGA ACCGCTTCGT CTTTGGAAAC GCCCATCACT TTAATCGGTG 660

CTAACGTTAA ATTATCAATA ATCGTTTTGT GAGCATATAA GTTAAAATTT TGAAAGACCA 720

TGGCAACTTT CTGGCGAACC TTTTGAATCG GTGCTTTCGG TTTGGTGATG TCAATGTTAT 780

CAATTACGAC GGAACCATCA TTGACTTTTT CCAATTGGTT CATACACCGA ATTAATGTAC 840

TTTTGCCAGA GCCAGAGGGT CCGATAATGA CAATTTTTTC TCCTTGCTCT ACATTCAAAG 900

TAATATCTTT TAAAACGTGA TGCTCCCCAA AATACTTATT TACATGTTCT AAAGAAATCA 960

TAGGCAGTCT CCTTTGCAAT CATTTGGTAG AAAAGATAAC AGAGGATTCG GTGAGAAACG 1020

TGTTATTTAA ATTAAATTAA CCAAATGTAT TTCTAATTGT ATTCTCAATG TCAGGAATTA 1080

TAACATACAA AAAAATGAAT TACTAGTTTT TTAATCGAAA AAGATTTCCT TCAAAAAAAG 1140

AAATAAATGA GAGCAAAACA GGAAAAATTG ACAATAGTCC TTAAAAAGTA GAAAATAAAG 1200

GCAGTGAATA AAAGGCAAAG GTAGGATGAG AATGAGCAAT CAAGAAGCCA TTGGATTAAT 1260

TGATTCTGGC GTTGGTGGAT TAACTGTTTT AAAGGAAGCG CTAAAACAAT TACCAAATGA 1320

ACGATTAATT TATTTAGGAG ATACAGCCCG TTGCCCATAT GGTCCACGAC CAGCCGAACA 1380

AGTCGTTCAG TTTACTTGGG AAATGGCCGA TTTTTTATTG AAAAAACGAA TAAAAATGCT 1440

AGTAATCGCA TGTAATACCG CGACGGCTGT CGCATTAGAA GAAATTAAAG CTGCCTTGCC 1500

AATTCCAGTT GTCGGTGTTA TTTTACCTGG CGCACGAGCA GCCGTTAAAG TCACaAAAAA 1560

TAACAAGATT GGTGTCATAG GTACGTTAGG GACAATCAAA AGTGCTTCCT ATGAAATCGC 1620

CATTAAAAGT AAGGCACCAA CAATTGAGGT GACTAGTTTA GATTGCCCTA AATTTGTCCC 1680

CATTGTTGAA AGTAATCAAT ATCGTTCTTC CGTAGCAAAA AAAATTGTGG CAGAAACACT 1740

TCAAGCACTA CAATTAAAAG GACTTGATAC GTTGATTTTA GGTTGTACCC ATTACCCGTT 1800

GTTACGTCCG GTGATTCAAA ATGTGATGGG GAGTCATGTG ACATTAATTG ACTCAGGAGC 1860

CGAAACAGTT GGCGAAGTCA GCATGCTTCT CGATTATTTT GACATTGCCC ACACGCCTGA 1920

AGCGCCTACA CAGCCCCATG AATTTTATAC AACTGGTTCT GCAAAAATGT TtGAAGAGAT 1980

TGCAAGCAGT TGGCTkGGwA TAGAGAACTT AAAAGCACAA CAGATTCACT TAGGAGGAAA 2040

CGAAAATGAT TAGACATGAT AGTCGTGCAC CAAAAGAATT GCGAAAGATC ACTATTGAAA 2100

CCAATGTTTT TAAACATCCA GAAGGTTCGG TGGGTTATTT CGTTTGGCGA TACAAAGTAA 2160

TTTGCTCAGC CCACTGTCGA AGAAAAGnTT CCCCCATTTT TACGAGGCGG GGGAACAGGT 2220

TGGGTAGCAG CTGAATATAG CATGCTCCCT CGCGCGACGA ACACAAGAAA TATTCGGGAA 2280

AGTGCGAAAG GAAAATTAAC AGGGCGAACA ATGGAGATTC AACGTTTGAT TGGGCGGTCA 2340

TTGCGAGCTG TTGTTGACTT AGAAAAGTTA GGGGAACGGA GCATTGTTGT CGATTGTGAT 2400

GTGATTCAAG CAGATGGGGG AACCCGTACA GCTAGTATTA CAGGCGCTTT TGTTGCTCTA 2460

CGCTTAGCAG TCAATCATTT ATTGATGACT GGTGCACTAG CGGAAGACCC AATTAAAgAG 2520

CATTTAGCTG CTGTCAGCGT GGGTATTTTA CCCGATGGTT ATTGTGTGCT GGACTTAGAT 2580

TACGCAGAGG ACTCGAATGC TGCAGTGGAT ATGAACATCG TGATGACCGA GTCTGGTGAA 2640

TTTGTGGAAA TTCAAGGAAC GGGTGAAGAA GCTACCTTCT CAGGAGATGA GTTAAATGCC 2700

ATGCTCTTTT ATGGAAAAAC AGGGATTGAG GAATTAATTG CGTATCAGAA AGAAGCGCTT 2760

TATGCTTTAG CCTCTGAAGA AGTCCCTTCG CAAGATTCAG AAGAGAAAGT GATTGTCATT 2820

GCGACTAGAA ATCCTGGTAA AGCGAAAGAG TTTAGTTCAA TTTTTGGTGA AAAAGGATAT 2880

ACAGTTAAGA CGTTATTAGA TTATCCGAAT TTGCCAGATG TAGAAGAAAC TGGTCGAACA 2940

TTTGaAGAAA ATGCTCGTCT AAAAGCTGAA ACAATTGCAG AAATTTTACA GAAACCAGTA 3000

CTAGCAGATG ATTCAGGATT AATTGTTGAT GCCTTAGGAG GAATGCCAGG GATTTATTCT 3060

GCTCGTTTTG CTGGTGAGCC AACTAACGAT GCTTCAAATA ATGCGAAACT ACTACATGAG 3120

TTAACGGGCG TGCCAAAAGA AAAAAGACAA GCGCGTTTCC ATTGTACATT GGTTTTTGCT 3180

GAACCAAAGA AAGAAAGTCT GGTGGTCGAA GCTGAATGGC CAGGAGAAGT CGGTACGATT 3240

CCTCGTGGTG AAGGAGGCTT TGGGTACGAT TCACTTTTCT ATGTTCCAGA GCTTGGTAAG 3300

```
ACGGCAGCTG AATTGTCTGG AGAAGAGAAA AATAAAGTTA GTCATCGTGG ACAAGCAGTG      3360
GCAAAATTAA AAGAACAATG GGAAGAATGG CTAAAAAAAT AGGAGTTGAA ACGTATGAAA      3420
ATCTTGGTAG TAAGCGATAA TCATGGCGAC CGTGACTGTT TAGTTGATTT AGTGAATCAT      3480
TATGAAGGAC AAGTCGATGC ACTCTTTCAC TGTGGTGATT CTGAATTAGA ACCGACAGAT      3540
GAACTTTGGC AAAAATTCAT TGTAGTCCAA GGAAATTGTG ATTTTTATGA TGAATTTCCC      3600
AAAACTGTGA CTAAAAAAGT GGGGGATCAA GTTATTTATA TGACACATGG TCATTTGGCC      3660
AATGTTAGAA TGGGGCTCAC GACACTTGCT TTACAAGCGG AAGAAGCAGG AGCGACAATT      3720
GCCTTATTTG GTCACACTCA TGTACTCGGC GCAGAGCGAC ACAATAACAT TCTCTTTGTT      3780
AATCCGGGAA GTATTCGTTT ACCTCGTGGA CCAGTTCAAG AGAAAACATA TGCCATAATT      3840
GAAAGTACAC CTGAACAATA TCTCGTTCAG TATTATAACA AAGAGCACCA ACCGTTGGAA      3900
CGATTAAAGG CAACGTTTAC GAAGTAAGTG AATGTTAAGA AATAAGTGCT ACTTTTTTAT      3960
TTATCAGATT TTGAGCGTAG GACAAAAATC TCTTTTTGAT TTTTGTCCTA CGCTCTTTTT      4020
TTAGTCGACT TGGATAACCT TTAAGCCGGT CTTACGAAGT TCTGTAATTA AAGTTGATGA      4080
GCAATGTTTG TCTGTAATCA AACAATCAAA AGCGGTAATA TCAGCAATTT TAAATTTAGA      4140
CGTCAAGCCG ATTTTTCGAT GGTCCGCGAC CAGAATTTTT ACTTTTGTTG TTTGTTTTAT      4200
CATAAGCTCA TTAATTTTTG CTTCATTGAA AATTTTTGTA GTGACGCCAT GTTCTAAACT      4260
GACCCCGCTA CAGCCAATTA CGCAAACATC GGCATGCGTT GCGGTGATTG TGTTAATGGC      4320
AATATCTCCT ACCAGAACTT CTTTGGGAAA ACGTAATTcG CcTCCGGTCA AAaTATAATT      4380
GTATtGAGaA ACATGAGGCT TGGTGGCAAT TTtTAAATTA TTCGTTACGA TGGTCAGATG      4440
GTCgGCCTTT AAATAATGGA GAATATTCaA AGCAGTGGAA CTTGAATTAA TGAAAACGGT      4500
ATCATAATCA CTTAGATAGC TACAAGCAAC ACGAGCAATG GCTTGTTTTA ATAACTCCGG      4560
TCCTGATTGA TCAAAGTCTT TCTGATTTTG GTCGTTGTAA ATACAATACC CATATTTTCG      4620
AATGATGTCA TTCTTAGTTT CTAAGACATT TAAATCACGA CGAATTGTGC TTAAGGAAAC      4680
ATTTAAAGCA TCTGCTAATT TATAGGTAGT CATTTGTTTG TTTTTTTGCA GCAGCGCGAG      4740
AATGTTTTGT TGCCTTTTTT CAATGGTTTC TAATGAACTT TTCAAGTAAT ACAACCTCCT      4800
TTAGTTATTT TTTTGGACAC GGTCCCCCAT GAAATAGTAT AACTCATAGG AAAACAAAAC      4860
ACGTAAAAGA ATCATTAAAA GTGCAAACAA TCATATAAAA GGGTTTTCAT TGTGATTTTT      4920
TGCGGCTATG ATTTTCTTAT AAAGCATAAC ACAATCATTT GAGAAGAGGA GTGGTAGCAT      4980
GAGCAAATCG ATTCATGAAA TTACAAAAGA AAGTTGGTTG AAGGCGACAT TTCCAGAATG      5040
GGGCACTTAT TTAAACGAAG AAATTAATCA GACAACTGTG CTTCAAGGAA CGGTTGCACT      5100
CTGGTGGCTT GGTTGTACAG GCATTTGGCT GAAAACGCAT GAGAACACGA ATATTTTATG      5160
TGATTTATGG TGTGGAACAG GAAAACAAAC GCACGGGAAC GGCTTAATGA AAAATGGCCA      5220
TCAAATGATG CGCATGAGTG GTTGCCAAAA AATGCAACCT AATTTAAGAA CACAACCTTT      5280
```

TGTGATTGAT CCTTTTGAAA TAAAAGAGGT AGATGCTTTA GTTGTCACAC ATATTCATTC 5340

AGACCATTTG GATATTCACA CAGCTGCTGC TGTTCATCAA AATTGTCCTA AGGCTCTATT 5400

TATTGGTCCA AAAGAAGTAG TGAAAACATG GCAGCGTTGG GGTGTTCCGG CAGAAAAAAC 5460

ACGGGTTATA GAGCCTGGTC AAGAAATAAA AGTGAATGAT GTGAATGTAG TGGCCTTGGA 5520

AGCTTTTGAT CGAACCGCGC TAATTACTTC TGATGATCCT GAAAATACAT TTAAAGGTAA 5580

AATGCCTCAG GACATGAATG AACTAGCTGT CAATTATCTA TTTGAAACCT CAGGTGGTAC 5640

GGTATATCAC GCCGGAGACT CTCATTATTC TAACTTGTTT GCGAAACATG GTAATCAATA 5700

TGCAATTGAT GTTTGCTTAG GAGCATATGG CGAAAATCCT CGCGGTATCA CAGATAAAAT 5760

AACTTCTGTT GATATGTTGA GAATGGCCGA ATCTTTAAAT GCGAAAGTAG TCATTCCTGT 5820

TCATTATGAT ATTTGGTCTA ATTTTATGGC GGATCcAAAA GAAATCATGG AAATTTGGAA 5880

ATTTAAAAAA GATCGTTTAC ATTATCAATT TAAACCATAT ATTTGGCAAG TGGGTGGCAA 5940

ATTTGTTTAT CCAGATGACA AAGACAACTT AGAATTTAAC TTTGATCGTG GCTTTGATGA 6000

TGTTTTTGTT AAAGAAAATG ATACACCATT TCCTTCATTC TTGTAAGAAA GGCTGGATGT 6060

ACAATGTTAA CTTATTTTTG GGAGCAAGAG CTTATTCATT ACCCTGATAA AGAACCGGTT 6120

AGTTGGCAAG AGGCGATTCA GGAAAGTTGT TTGATTTTAC TCCAAAAACA CATCATCGAT 6180

CAGAGTTACG TCGATGAGAT TATCCAGTGT GTCGAAACGT TTGGACCTTA TATCATCATT 6240

GCTCCTGAAG TAGCCATGCC ACATTCATCA GAGGAAAGCG CAGGCGTTTT TGGAACGGCC 6300

ATTTCTTTTA CAAAATTTAA GCAAGCAGTC ACTTTTGCTG GCGATCAAGA AGCAAAAACA 6360

GCAACATTAT TTTTTACATT GGCAGCCCAA AATCCCGCCG AACATCTGGA AAATATCCAG 6420

CAATTGATGG ATTTATTGAT GACGGATGGC GTGATTGCTG ACTTGTTAGA AACAAACACG 6480

CCAATGGATT TTAAAGAAGT TATGGAGAAA TATCAATTAT AGGGGGAAAA ATCAATGGGG 6540

GAGTTCTTGT TAAATAGTTG GACTTATTTT GCCACGAACG TTTTAACACA GCCAGCATAT 6600

TTGATTGGTT TCATCGTGTT ATTAGGTTAT CTCTTGTTAC GGAGGCCCTT TTATGAATGT 6660

TTAGCAGGTT TTCTGAAAGC GACAGTTGGT TATTTTATTT TGAGTGTTGG TTCTGGTGGT 6720

TTGGTTAATA ATTTTCGCCC TATTCTAGTA GGATTGAAAG AGCGATTTAA TTTACAGGCG 6780

ATGGTTACCG ATCCATATTT CGGGCAAAAT GCAGTCGATG CAGGATTGAT GAAAACCTTT 6840

GGACGAACCT TTGGTGATGT GATGATTTTG TTATTGATTG CATTTATTAT GAATATTTTG 6900

TTGGTTCGTT TTCAAAAATA TACAAAATTG AGAGCCGTTT TTACAACAGG GAATGTTCAG 6960

ATTCAACAAG CAGCTACCGC TTTCTGGATT TTACTCTTTT GTTTTCCTAA CTTAGGTAGG 7020

ATCGAAGTAC TAATTTTTAT GGGCTTAATT CTTGGGTGTT ATTGGGCAGT TGCTTCAAAT 7080

TTAACGGTAG GGATTACGCA GGAATTAACA GAAGGTGCTG GTTTTGCGGT AgCCCATCAA 7140

TAAATGTTTG GCATATTTAT TTTTGCAAAA TTAGCAGAAT GGTTGAAAAA ACGTGATCAA 7200

AAGAAACATC GTTCAGAGAA GGTCGATAAA AAATTGGAAG ATATCGAATT ACCAGGATTT 7260

CTATCCATTT TTAATGAAAA TATGGTTGCT ACGTCCATTT TGATGCTCTT TTTCTTTGGC 7320

ATTATCTTGC TTGTTTTGGG ACAAGATTAT TTAATTCAAG CTAAATTTAT GCAAGAAGGG 7380

CAGAGTTTCT TCTTTTATAT CATGACCACT TCATTAAATT TTGCTGTCTA TTTAGCTATC 7440

TTACAATTAG GCGTTAGAAC CTTTGTCGAC GAGTTGACGC AGtCATTCCA AGGAATTTCA 7500

AATACCATTT TACCAGGCGC TGTTCCTGGG ATTGATGTAG CAGCGACTTT CGGGTTTGGC 7560

TCGCCCAACG CAGTAACAAT TGGGTTTCTC TTTGGCGCAT TAGGCCAGTT TTTAATGATT 7620

GGCTTACTGA TTCTATTAAA GTCTCCGACA ATCGTTATTG CGGGTTTTAT CCCATTGTTT 7680

TTTGATAACG CTGTCATAGC TGTTTATGCa AATAATCGTG GCGGAGTAAA AGCTGCTTGT 7740

TTATTCCCCT TCCTTTCAGG TTTAATCCAA GTAGGTGGCT CAGCTTTATT TGCTACATGG 7800

ATTGGATTAT CTCAATATGG CGGTTATTTG GGTATGTTCG ATTGGGCGAC AGCTTGGCCG 7860

TTATTTACAA TCATTATGAA AGTTCTAGGC TATGTTGGCG TAgCAGTGGT GGTGATTGGG 7920

TTACTAATAA TTCCGCAACT TCAGTATCGA AAAAATCCAG AAGGTTACTT TATGATTGTC 7980

AATGATTATG AACAATATGC TAAAAAATTT CAAGAGAATT AAAGGAGGAA TTTGAAATGA 8040

GAATACTTGT TTCCTGTGCC AATGGTTCTG GCACGAGTCT AATGATGATG CGGAGTGTTG 8100

AAAAAGCAAT GAAAGAATTA GGTGTACCGA TTACAAAAAT TCATCATTGC GCCATTTCAG 8160

AAGGAAAAAG CTCTGCTAGC CAATACGACG TTGTTTTTAC TCCAGTCAAT TTTTTACAAA 8220

TGTTTCAGCA AGCAGAAAAA CGCGGAGTCA CAGTGATTGG CATTAAAAAT GTAATGTCTG 8280

CAAAAGAAAT TCAAGAAAAA TATATGGCTT CAGATCGTTA TCAAAAATCT GAATAAAAGG 8340

GTGGGAAAGA TGAAAAGACC AAACTTACAA ATCGCTTTAG ATCATAATAG CTTAGAAGAC 8400

GCTTTAGCAG ATTGTATGAA AGTCGGAGAA ATTGTGGATA TTATCGAAGT TGGCACTATT 8460

TTGTGTTTAC AAGAAGGCCA AAAAGCAATT CGTTGTTTAA AACGCATGTT TCCTAATAAA 8520

ACGATTGTTG CTGATACAAA ATGTGCTGAT GCGGGCGGTA CAGTTGCTCG TAATGTCGCA 8580

CAAGCAGGCG CTGATTTCAT GACGGTCATT TGTTGCGCCA CCTTACCAAC GATGGCTGCA 8640

GCTCAAAAAG AAGTCCGAGA ATTACAAGTA GAACTATATG GAAATTGGAC AATGCAACAA 8700

GCTCGACAAT GGCGAGAATT AGGGATTAAT CAGGTGATTT ATCATCAAAG TAGAGATGCT 8760

CTTTTAGCTG GCGGCAGTTG GGGAGAGAAA GATTTGAATA AAGTTCAAGA ACTAATTGAC 8820

TTAGGCTTTG AAGTGTCCGT CACAGGTGGT TTAACTGTCG AAACGCTTGA ATTGTTTCAA 8880

ACAATGGCAG TTGCAACATT TATTGCGGGC CGTGGGATTA CTGAATCAAA AAATCCTGAA 8940

CAAGCAGCCA AAGATTTTCA AAAAAAAATC GATCAGATTT GGAAGTGATA AAATGGCACG 9000

AATTGGACTG TATGAAAAAG CACTCCCGCA AAATTTAACT TGGGCAGAAC GATTAACATG 9060

GGCCAAAAAG TTAGGGTTTG ATTTTTTAGA AATGTCAATT GATGAATCCG ATGAACGTTT 9120

AGCACGATTG GCTTGGACAC CAAGCCAGTT ACAAGAATTT TCTCAATTAA TGGTGAAAGA 9180

AGATTTTTTT ATCCATTCTT TGTGCTTAAG TGGTCATCGG CGGTTTCCTT TAGGCTCTTT 9240

AAATAAAGAA ACACGTGAAA AAGGGCGGAA AATTTTATCT CAAGCGATTC GTTTGGCTCA 9300

TCAATTAAAC ATACGGGTGA TTCAAATAGC TGGCTATGAC GTCTTTTACG AAGAAAAAAC 9360

TGCCGAAACA AGGGAATTTT TTCTTCAAGG TCTGAAAAAG GGCGTTGAGG AAGCTGCACA 9420

ATACGGTGTC ATTCTGGCTG TTGAGATTAT GGATGACCCT TTTATGAATA GCATACAAAA 9480

ATTTCTAGAA ATTAAGGAAC AAATTCCGTC CCCTTTTTTG CATGTTTATC CTGATTTAGG 9540

TAATTTGTCT GCTTGGCCGG AAAATAATCC GGCGGTTGAG CTTGAAAAGG GAATTGCTGA 9600

AATCGCTGCT ATTCACTTAA AAGATACCTT CGCTGTAACA GACACATTTG AAGGGAAGTT 9660

TCGGGAAGTT ACTTTTGGAG AAGGCTGTGT TGATTTTACA GGCTTATTGA AAACACTCAA 9720

GCGGTTAAAC TATAGTGGTC CTTTTTTAAT TGAAATGTGG AATGAAACGG ATTTGAATTT 9780

TCAAGAAAAA ATTCAAGCCG CGCAgcAGTA TCTCTATCCC AAATTAGCAG AGGTGGGTTA 9840

TTATGAACAA TAAAGCGATC ATCCAAGCAA TGAAACAACG TGTTTATGAA GCGAATCTGG 9900

CTTTGCCAAA ACTGGGCTTA GTCAAATTAA CGTGGGGCAA TGTTAGTGAA ATCAATCGGT 9960

CATTAGGGAT TATCGTAATT AAACCAAGCG GCGTCAAGTA TCAAGAAATG ACCAAGGAAC 10020

AGATGGTAGT GACTGATTTA AAGGGACAAC TATTGGAAAC CAATGCATTA AAGCCGTCTT 10080

CTGATTTACC CACGCATCTT TATCTTTATC AAAAAATGCC AGAAATTGGA GCGATTGCGC 10140

ACACGCACTC CCTGAATAGT GTTACGTGGG CTCAAGCAGG GCGAGCCTTA CCTCCTTACG 10200

GTACCACCCA TGCGGATGCT TTTTATGGCG CTGTACCTTG CACAAGAGCG CTCTCTGAAA 10260

GTGAAATTAA AGAAAACTAT GAAGAAGAGA CTGGCAAAGT CATTGTTGAA ACATTTCATG 10320

AACAAGAATT AGATCCTTTA GCAGTTCCAG GAGTGTTAGT TTATGGACAT GGCCCATTTA 10380

CTTGGGGAAT GACTCCGGAA AAAGCAGTTG AAAACAGTTT GATTTTAGAT GAAATCTGTT 10440

CAATGGCACG TTTGACGGAA ATAATTAATC CAGATGTTGA ACCAATTGAT CATTTTTTAT 10500

TGGATAAACA CTATCTTAGA AAACATGGGA TTTCCGCTTA TTATGGTCAG TGATGGCTGT 10560

GATGAAACGT GTGAAAAAGA AGGATAATTG ACTGATTAAT TGGTCTTTTA TAAAAACTTA 10620

GTTGTCTGTT ATGGGACAAC TAAGTTTTTA TTTTTTTAAG AAAGCGTTCT CTTTTTTGGG 10680

GGAGAATAGA ACAAAACGTT AATTTTCGTT TGTTAAGTTA TGAAGAATTT GCGAATTCCA 10740

AGACTTTTTC CGCAGAATTA GGTACAATAA GTTACGTGTT AAAGAATTAG AAATGGAGGC 10800

AACCCAATCA TGATAGGGAC TGCTGTAAAA GAATTATTAT TAGAAAAACA AGAAACGTTT 10860

TTAGTACCCG CCGAAAATGT CGCCAACGTA ATGTGTTTAA ATCCATTAAG TCACGCGGCA 10920

TTAGTCTTAT CGCAAGTTCG CTATTCCAAG ATTCCTGTGT TAGATAAGGG AGATCGCTTT 10980

GTTGGCTTAA TCGGCTTGAC AGACGTTGTA GATAAAATGT TTGATTTGAC TTCAGTGGAT 11040

TTTGAAAAAT TGAATGAATT TACTGTGGCC GATGTAATGG AAGTCAATGT TCCTGTAATT 11100

GGTGAATCGT GGGATTTAGA AGAAGTGTTA CATTTATTAG TAGATGCTTC TTTTTTACCA 11160

GTGGTTGATG ATAATCAACG CTTTAAAGGT ATCATTACAC GCAAAGAAAT TTTAAAAGCG 11220

GTCAATCATA TGGCGCATGA ATTAGAAAAA CATCATCTTG TTCTGCCTAA ATCAGAAGAA 11280

GAAATGAAAG TAATTTAAAT AAAACGCAAG AAACTCTTTA CTTTTTTCTC ATAAATGCGT 11340

TATTCTGTAT GATAACACGA AGTAAAGGAG TGCTGTGATT GGACGCATAT GAGATTATTC 11400

AGTATATTGG CGATGCCAAA AAGCAAACGC TTGTTAAAGT AACACTGAAA GGGCAATTAA 11460

AAGAAGTCAC GTTTCCAGAA ACAATTAAGG TCTTTAATAA TTGTAAAACT GGGACATTGT 11520

TTGGCGATTG GGCTGACGTA AAACCTTTTT TAGAaGCAAm CAAaGAaAAA TTGAGaTTAT 11580

GTTGTAGAAA ACGATGC 11597

(2) INFORMATION FOR SEQ ID NO: 223:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24601 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

TTATTTTTTA ATTCGTAATA GTTTTTTTAA ATCGTTTCAC AGAATTACTT TAATTGTTAT 60

ACAGTTTTTT TAGTTGTnAA AAnGGGGAAA TnGGCCTTTA GGAGGTAATC ATGATTGAGT 120

TAGTGAATGT GTCGAAAGTC TATCAAGAGA CACACGCGTT AAAAAATATT CACTTTTCTG 180

TCCAACCAGG AGAAATCGTT GGGATTGTTG GGAAAAGTGG CTCTGGCAAG TCCACGTTGT 240

TACGTTTGTT GAATTTAATG GAACAACCGA GTGAAGGGGA GATTCGTATT GATGGCAGAA 300

ATGTTCAAAC CTTCTCTAAA AAAGAAGTTC GCCAGCAGCA ACAACAAATG GGGATGGTGT 360

TTCAGCATTA CAATTTGTTA GAAAATTTGA AAATTTATGA TAATGTTGCG TTGCCTTTGA 420

AACTTTTAAA AGAGAAACAG CCTGAAAAAA TTGAGCGTTT ATTAACGTTT GTTGATATGG 480

CCCATAAGGC AGAAGCGTAT CCTGCTCAGC TTTCTGGTGG GGnAAAACAG CGTGTCTCTA 540

TTGCTCGAGC GTTGAGCCGT AATCCAAAAT GGTTATTATG TGACGAGGCA ACGAGTTCAT 600

TAGATGarGr AAATwCAGrA AGTGTCgTtC GACTATTACA TAAGACGCAT CAGGAATTCC 660

GACCGACTAT TTTCTTTGTT AGTCACGAAT TAGAAACCGT GAAGCGCTTA TGTAATCGTA 720

TCTTAGTGAT GGAAAAAGGT CAGTTAATTG GCGAATTTCT GAATAATCCT CAGCAATATG 780

AGGAAGAACC GCTTTCCTAT TTAGAAAAAG TGGAAAGGAG TTTGCGTCCA TGACTTTTTC 840

AGAAACCTTT GCCGAATATT GGCCATCTAT GCTGATTGGT TTAAAAGAAA CAGGGATTAT 900

GATGTCGATT TCGATGGCTA TCTGTTTGCT TGTCGGCTTA CCTTTAGGCT TAGGACTATT 960

TTTGGCTAAT CCGAAaGTCA AAGGgCATCA ACCTTGGCTG TACTGGGTTT TAAATTTCAT 1020

TGTAACTGTT ATTCGCTCGT TTCCTTATTT ATTATTTGTG ATTGCGTTGA TTCCAGTTAC 1080

TCGTTCTGTT TTAGGAAAGG CGTTTGGCCC AATTCCTTCA TCTGTGCCAT TGAGTATTGT 1140

TGCGATCGCG ATTTTTGCTC GCTTGGTTGA ACAAGTGCTG TTGGATGTTC CAGAAGAAAT 1200

TTATTTTTTA GCTAATTCGT TAGGTACCTC GAAAGTCCAA TACATTTGGC ACTTTTTATT 1260

GGTGGAAGCA CGTAGCGGAT TAGTCCTTGC TTACACAACC ACGACAGTGA GTATGGTTTC 1320

ATATTCCACG GTGATGGGTG TTGTTGGTGG CGGCGGGATT GGCGATTTTG CAGTTCGTGT 1380

CGGGTATCAA CGGTATGAAT ACGGTATTAT GTATGTTGCC ATTGTGATTA TGATTGCACT 1440

GGTCTTCATT TTACAAATGC TTGGGAACTT CATTGCCCAA AAACTAGACA AAAGAAAGTA 1500

GGAGGACAAA AGAATGAAGA AGTTTTATTT AgCmACaTTC GCTGTTATTG CAACAGTTAT 1560

TTTAGCTGCC TGTGGGGGAA ATAAACAAGC AGACCAGAAA GAAGACAAGG AGATTACCGT 1620

TGCCGTGCAA TTGGAATCTT CAAAAGATAT CTTGGAGATT GCCAAGAAAG AAGCTGAGAA 1680

AAAAGGGTAC AAAATTAACA TTATGGAAGT GAGCGACAAT GTTGCCTACA ACGATGCCGT 1740

GCAACATGAC GAAGCGGATG CTAATTTTGC GCAACATCAA CCCTTCATGG AAATGTTTAA 1800

CAAAGAGAAA AAAGCTGATT TAGTGGCTGT GCAACCGATT TATTATTTTG CTGGTGGTTT 1860

CTATTCAAAA GAATACCAAG ATGCGAAAGA TTTACCTGAA AATGCCAAAG TGGGGATTCC 1920

TAGCGATCCA ACCAATGAAG GTCGTGCTTT AGCAATTTTA AATGCAAACG GCGTGATTAA 1980

ATTAAAAGAA GGTGTCGGCT TTAACGGCAC GGTGGCAGAT GTCGTGGAAA ATCCTAAAAA 2040

CATCACTTTT GAAAGCATTG ATTTACTGAA TTTAGCTAAA GCCTATGATG AAAAAGACAT 2100

CGCTATGGTG TTCTGCTACC CAGCCTACTT AGAACCTGCT GGTTTAACAA CGAAAGATGC 2160

GATCTTGTTA GAAGATAAAG AAGCAAGTAA ACATTACGCA TTGCAAGTTG TGACACGCAA 2220

AGGCGAAAAA GATAGCGAAA AAATCAAGGT TTTAAAAGAA GCGATGACAA CAAAAGAAGT 2280

TGCTGAATAC ATCAAGAAAA ATTCTAAAGG CGCCAATATT CCTGCGTTTT AATGAATAGA 2340

AAAAATGAAG CCTCTTTTTT GAGAAGCTTC ATTTTTTTCT TTTCTTTGAT TTATAGTATA 2400

GTTAATAGGA AGGAGGTGCC GGTGTGGTCG AATTATTTAT TTTAATGATG GAACGGGTGG 2460

GCTTAATTAT TTTACTGGCC TTTCTTTTAG TGAATGTGCC TTATTTTAAG CGTGTTTTAT 2520

TAAGTCGTGA AAAAATGTCT TCAAAAGTTC AGTTGATTTT GATTTTTGGT TTATTTGCTA 2580

TTATTTCTAA TTTTACAGGT ATCGAAATCG CTAAGAATCA GATTGTTCCC AATAATTTAC 2640

TAACCTATTT AAGCAGCAAT GCTTCCATTG CGAATACGCG TACGTTAGTA ATTGGCGTTT 2700

CTGGATTAGT GGGTGGTCCG ATTGTCGGCA GTGCGGTTGG TTTAATTGCT GGCTTTCATC 2760

GGGTTATTCA AGGCGGGGGG CATAGTTTCT TTTATGTGCC TGCTTCTTTA ATTGTTGGCT 2820

TAATTGCCGG TTTTTTAGGG AGCCGCATGG CAAAGCAAAC TGTTTTTCCG TCTGCCGGCT 2880

TTTCTGCCAT TGTCGGAGCG TGCATGGAAA TGATTCAGAT GATTTTTATT TTCTTTTTCA 2940

GCGGTGATCT ATCGGATGGG GCCACGTTGG TTCGTTTTAT TGCGCTGCCG ATGATTTTGT 3000

TAAATAGCGT AGGAACGTTC ATTTTTATGT CGATTCTAAC GACGACATTG AAACAAGAAG 3060

AGCAGGCCAA AGCGGTTCAG ACGCACGATG TTTTGGAATT AGCAGCGGAA ACACTGCCAT 3120

ATTTTCGAGA AGGCCTCAAT AAAAATTCCA GCAAAAAAGT TGCGGAGATT ATTAAGCATT 3180

ATACAAAAGT CAGTGCGATT AGCATGACAA ACAGTCATCA AATTTTAGCT CATGTTGGCG 3240
CAGGAAGCGA TCATCATATT CCAGAGCTAG AAGTCATTAC GGAGCTTTCT CGGGAAGTGT 3300
TACGGACGGG CCGAATGACC ATCGCGCATG CCAAAGAAGA AGTTGGTTGT TCGGACCCTA 3360
ATTGTCCGTT GCAAGCGGCG ATTGTGATTC CATTATTTTC CCATCAGCAA ATTGTTGGCA 3420
CCTTAAAAAT GTACTTTACA GATCCAGCGC AGTTAACGCA TGTCGAGGAG CAGTTAGCAG 3480
AAGGTCTTGG GACAATCTTT TCTTCACAAA TTGAATTAGG TGAAGCGGAA GTACAATCAA 3540
AATTACTTAA AGAAGCTGAA ATTAAATCCT TGCAAGCGCA AGTGAACCCG CATTTCTTCT 3600
TTAATGCCAT CAATACTATT TCCGCATTAA TGCGGAAAGA CAGTGAAAAA GCGCGGAAAC 3660
TATTGCTCCA ATTAAGCAAA TACTTTCGAG GCAACTTGCA GGGAGCCGTG CAAACTACGA 3720
TCCCAGTTTC ACAAGAACTA GAACAAGTAA AAGCCTATTT GTCATTAGAA CAAGCACGGT 3780
TTCCGAATCG TTATCAAGTA ACATTTGACG TGGAAGAAGC GGTTGCTTCA GAAAAAGTAC 3840
CGCCGTATGC GATTCAGGTG TTAGTGGAAA ATACCATCAA GCATGCATTT GGCAGCCGGA 3900
AGGAGAACAA TCAAGTCCGG GTAGTGGTCA AAGCAAAACA ACACAAATTA CATGTGGATG 3960
TGTATGACAA TGGGCAAGGG ATTCCTGAAG AACGCCGCTT GCTATTAGGA AAGACGACAG 4020
TGACTTCGGA AAAAGGCACC GGCACGGCAC TGGAAAACTT GTCCCGCAGA ATGGCGAACC 4080
TTTATGGATC AGAAGGTTGT TTTTTAATTG AAAATTTAGA GACAGGCGGG AGCCATGTTC 4140
ATTTAGAAAT CCCGATGGAA CAGGAGGAAT TAGATGCACG TATTAATCGT TGATGATGAG 4200
CCCTTAGCAC GAGAAGAATT GAGCTATCTT GTTTCCCAAC ATCCGCAAGT GACGAGCGTT 4260
GCAGAAGCGG ATTCTGTGGC GGAGGCAATG GAAGAGATGA TGGATCAAAA GCCAGATTTG 4320
TTATTTTTAG ATATTCATTT GACCGATGAA AGCGGGTTTG ATTTAGCGGA AAAATTAACC 4380
CATCTTAAAA AAGCGCCCTA CTTGGTCTTT GCAACCGCCT ATGATCAATA TGCGTTAAAG 4440
GCCTTTCAAG TGAATGCAAA AGATTATATT TTGAAACCTT TTGAAGAAGA AAAAATAACA 4500
CAAGTCATTG AAAAAGCATC GAAAGAAATG GGTCAAGCTG TCCCAGAAAA CACTGCCGAA 4560
AAAGGACCAA AATCTGAAGC AATACCTATT CAAGGGGAAG ACCGCATTTA TCTTGTGGCG 4620
CCAGAAGATA TTTACCTCGT TTCTGTCGAA GAACGGCAAC TGAGCATTTT TGTAGATCAA 4680
CAGGTCTACA AGATGACAGG GACCCTCAAC AGTATCGAAC AAAAATTACC TGCCACGCTC 4740
TTTATTAAAA CGCATCGAAG TTTCATCTTG AATCGAACAA AAATTCAAGA AATTCAGCCG 4800
TGGTTTAATA ACACGTTGCA AGTAATTTTA ACCAACGGCT CAAAAGTACC TGTTAGCCGT 4860
TCATACGTCA AAGAATTTAA AGAAAAACTT GGcTTAAGTT AAAAAATCAC CACTTGAAAT 4920
GTCGAATGCA TTTCAAGTGG TGATTTTTGC ATTTGGTGAA GAAATTCTGG CATTTCATGA 4980
AAAAAAGCCC TTAAAAAAAC GTTTTCTTTA TATACTTAAC CCATAGAAAA CAGACAAAAG 5040
GGAGCGAAGA AGAATGGAAA AGAAAGTATA TTCATTTTTA CAACAAGCAG CAGTTTTTGC 5100
CATTATTATG TTAATCGCCA ACGGCTTGGC GTTCATTTCC CCAATACCAA TGCCTGCTTC 5160

AGTAATTGGC ATGGTGTTAT TGTTCATTGG TTTATGTACA AAAGTTGTTA AATTGGAGCA 5220

AGTGGAAGGG CTAGGTAATA GCTTATCAAC GGTTATTACA TTTTTATTTG TTCCTTCAGG 5280

GATTTCCTTA GTTAATTCTT TAGGGATTAT GCAACAATTT GGCGTTCAAA TTATTTTTGT 5340

TATTTTAGTA GCAACATTAG CGTTACTAGC AACAACTGGT TGGACGGGCG CCTTTTTATT 5400

GCATATGAAA GATGCTCGAA AAGAAAAAGA ACCCGCAGCA ACGAAGAAAG CGAAAGGAGA 5460

GGTTCGCGCA TGAATCCGTA TGTTGGCATT ATTATTTCAT TAGCTGTATT TGGCATAGGT 5520

ACTTGGTTGT TTAAAAAAAG TAAAGGTTTC TTCTTATTTA CGCCATTATT CGTTGCGATG 5580

ATTTTAGGCG TTGTCGTTTT AAAAGTAACA GGAATTAGTT ACGAGCAATA CAATGAAGGT 5640

GGCAAATATA TCAGCTTCTT TCTAGAGCCT GCGACAGTAG CGTTTGCGAT TCCTTTATAT 5700

AAAAAGAGAG ACGTGTTGAA AAAATATTGG TTAGAAATTT TAACGGCGTT AACAATTGGT 5760

TCTTTTGGTT CGCTAGTGGC TGTTTATTTT GCAGGAAAAG TGATTGGCAT GGATAATCAT 5820

TTAGTGGCCT CAATCTTACC ACAAGCAGCT ACAACAGCGA TTGCCGTACC TATTTCTCAA 5880

ACTGTTGGCG GCATTGCTTC CATTACGGCG TTTACTGTAA TCTTCAATGG TGTACTAACC 5940

TATGCCTTGG GTCGAATTGC ATTGAAATGG TTTAAGATTA ACAATCCAAT CGCCAAAGGT 6000

TTAGCTTTAG GTGCGGCCGG ACATGCCCTA GGTGTGGCAG TTGGCATAGA AATGGGCGAA 6060

ACAGAGGCTG CTATGGCTAG TATTTCTGTC GTAGTGGTTG GCGTAGTTAC TGTTTTAGTA 6120

GTTCCTTTCT TTGCAACGGT CATTGGTCTA TAAAAAAGTG GGCTTCCTCA CTTTTTATGA 6180

GTAGCTAATT TCTATCAGCT AACAGTCTCT GCTTATTCTG TGAATGCTAG TTATCACTAA 6240

TCCGTAGGCT CGGAGTTAAA ATGAGTGTAA AAAAGCCAAC CTTATCCATA GTAAGGTTGG 6300

CTTTTGCGTT GTTCACAGGA TTTTTTGAAT ATAGATAGGT CGGATTGACC AGTCTTTTTA 6360

TAGAATGGTT GCTCCTAAAG TTTGTTTAAA GTGGCGTAAC GCCCATTCGT GTCCTTGTTC 6420

ATCAAAACTA GCAACGGCGC GTTCAGGAAT CATTAATTGA TAGCCTAAAT TATAGGCATC 6480

GACAGCCGTA TGCAACACGC AAATATCTGT ACAAACACCA GTTAAACAAA GCTCTGTAAT 6540

TCCTCGTTCT CGTAATCGAA TGTCCAAATC CGTGCCACTA AAAGCAGAAT AGTGCCGTTT 6600

GTCCATCCAA TAAATGGTTG CCTCATTTTT GTGAGCTTGA TAAAAATCAG CTAGCGAGCC 6660

AAACAGATTA CGCCCGTCAG TTCCTAAGAC ATTGTGGGGC GGGAAAAGTT TATTTTCAGG 6720

ATGATAGCGA TCTAGGGGGT CATGCCCATC AATGGCAAAA ACCACAAAAT CGCCATTGTC 6780

AAAATAGTGC TTTGTTTGTT GAACAAGGGC CGTTTCAATG GCTTGCCCAG CGGCACCAGT 6840

GGTTAGTTTT CCGTCTGTGG CGACAAAATC ATTCGTATAA TCAATGGAAA TCAGAGCTTT 6900

CATTTGGCAA CCTCCTAAAG GTTAAAGTCA TTTTGTTCAT TTTACAGCTT TTAAGAAAGC 6960

ATTTCAAATG ATAAAAGTAC ATCAAATTCT AAAAAAGCTG TTTCAAAATA AACGCTTTGG 7020

CTAAAAAACA TGCTATAATT TTGCTACAAA TTAAaGCAGG TGAATAGAGT GAAAAGAAAA 7080

ATCGTACGTT TTATCCTTTG GTTGTTAGGA ATCATTGTGG TTATTTGCTT AGGCGTGTTT 7140

```
CTCGCCTtTC AATTATCCCC TAAGCCAGGC GCTTGGGTCA TCAACCAGTT ATTCGCAGGC    7200
GAAGTTGAAA TAAAAGATTC AGCTTCCTAT GATAAAGCAT TGCCCAATGT CCAATTGCAA    7260
GAAAaTCmAA CATACCCTyC TyCArGAAAA aaGAATACCT TtGATTTGTA TTATCCCCAA    7320
ACGTCAAAAC AAGCGGTACC AGTCGTTTTA TGGGTCCATG GTGGTGGCTA TGTGGGAGGA    7380
GATAAGTCTG GGATGAAAGA GTTTGCGACA CGTTTAGTGG CTGATTCCTC AGTTGCCtTT    7440
ATTTCAATGA ATTATGAATT AGCACCAGAC GCACCTTATC CAAGCCAGCT CCAGCAGGTC    7500
AATGAACTTG TTCAATTTTT ATTAGAAAAG AaACAAGCGT ATCCCATGTT AGATTTATCT    7560
AAGCTTTTTA TTGGCGGGGA CAGTGCGGGT GCACAGATTG CTCTACAATA TGCAACAGTT    7620
CAAACCAATG CCAACTATGC CAAAGAACTA GGGATGACTG CAGCATTGCC GGCTAGCCAT    7680
TTAAAAGGCA CGCTCTCGTA TTGTGGTCCG GTGGATTTGA AACAAATGGC CAACCAACAA    7740
AGTGATAATC GTTTTATGAA ATTTTTCGTT AAAACAGTGG CTTGGTCCTT GCTCGGCACT    7800
AAAGACTGGC AGACGAGTGC AGAACTGCAA GAAGTGAGTT TAGTGGATAA AGTCACCAAA    7860
GAATTTCCCC CGACGTATCT AACAGATGGC AATAGCTTTT CATTCCAAGA GCAAGGACTA    7920
GCCTTGGTTC AGCGGTTAAA AACTTTAAAT GTTCCAGTTT CAGCTCTCTT TTTCAAAGAT    7980
AAGAAAGCAr CGATCACCCA CGAATATCAA TTCGACTATC AAACAAAAGA AGCTAAACAA    8040
TGTTACCAAG AAACAGTCCA GTTTGTGAAT ACGTATAAAT AAACAGAGGC CTTAGAAATA    8100
TTTCTAAGGC CTCTGTTTAT TTATTTTTGT CATTTTTTCA GCTGAGAAAT ATGTATAAAT    8160
TAAACAAAAA TGTTTATTTA ATGATGTTTA CATGAGAAAA AATCCGTGCT ATACTACGAA    8220
CTCGTTTGTT AATTTTAATG TTAAAAACAA CGTTTTTGTT ATTTTTGTGA ACAAATTAAC    8280
ATTTTTATCT TACAAAAACA GCAAAAAATT CGTTTAGTTA ACAAAAAGAA AATTAAAAAA    8340
TAAATTTTTT TTTTGAAATT AAGAAAATAT CTTTAAAAAA ATCAGAAGAa AAcTTCGTAT    8400
ACTTGAGCCA TAGGTTGGTT GTGGAAAGTT TACTTCGAGG GAGGAGTGAT GAAGTTAATG    8460
AAAAAGAAAA TCGCTGTTTT ATTGGCAACT TGTTTACTTG TACAACCATT TTTATCTGTT    8520
TCAGCGTGGG CGATTGAAGG GGAGCAACAA ACAGAACGTA CGCAGACACA GCCAAAAACA    8580
GCAGAAACCA GTACAAGTGA AgCAACAAAA GAAAGCAGCG CACATTCTAC AGAACAGTCA    8640
TCAATGACTG AATCTAGTGC TATCACAGAA AAAGTAACCA CCTCTTCCAC AGAAACAAAA    8700
CCAAGCGAAG AGCAGAAGCA AATCACGTTG ACCTTTGAAA CAACGGATCA AGCTTTGTTT    8760
CTAAATGATG CCAAAAGTTA TCAAGTGGTA AAAGAAAAAA ATCAGCCACT TCGGACAGAA    8820
GAATTACCAA AATGGGCGAA TAGTGAGGAA AATGCGACTT TCGTTGGCTG GTCTTATCAA    8880
GGCAAAAACAT ATACAAATGA AGAACTTCTT CAATTAGAGT TTTCAGAAGA TACACAACTT   8940
ACGGCTGTCT TTAAACAGAA ACTGACACGC ATGGCCAGAG CTGTTTCAAT TGATCAGTCC    9000
ATTGCGGACA CAATTGCTCC TGCAGATAAA ACAAAAGTGA TTGTTGTGCC AAGTCCTGCA    9060
GGAGATGGAG CAGCGTACAA AGAATATGCA TCGACAGAGG CTGGCTTGAA AGAAGCCTTG    9120
```

TTTGATTTGT ATCAACAGGG GAACAATGGC GATTTCACTC TTTATATAGG AAACAATATT 9180

ACGACGAGTG CGGCGACCAC CGCTAAAGTG ATTCCTGACA CCGTTACGGC AAGTAATATG 9240

ACCTTTTACG CCTTGCAAGG TAAGGTCAAT CATTTAGTGA TTACTGGAAA CAGCGCGGAT 9300

CCGATTAGTA TGGATCAAAC GGCGCCAACG GGTTCTAAGA CGTTAGGATT TAATCAAAAT 9360

ATCTATTTTG GCTCAAACAT CACTTTGAGA AATCTAAACT ATACAGGTAC GAATATGTAT 9420

TTAAATGGTT ACAGTTTGAA CCTGAATGGC GGCTCTAGTG GAAACGGTCT GACCGTATAC 9480

GGTGGAACGG ATACAGGAGA TGTTTCAGGA AACCCAACTT TAACAGTAAA TAGTACAGGT 9540

ACAGGGACAT GGAATTTCTA TGGCGGAAAC CAAAATGGTG GTAATTTAGC TGGAAATCCT 9600

ACGATTGTTA TTAACAATAC ACGGAGCGGA CTAAATACAT TAAGTGGGGG CGCCAATATT 9660

GGGACAGTGA CTGGGAATAC TTCATTGGTT GTCAATGATT CTGGTGGCAG AATTGCTTCA 9720

ATTTATGGGG GCGGTTATGG GACCAACGCA ACCAACACAG CGAATGTTAC TGGTAACGTC 9780

TCAACAAAGG TTGCGATTAC GAATGCTGCT ACAGGTTTCC AATTAAGTAC GTATTATGGC 9840

GGTGTTCAAT ACGGAAATAT TGGTGGAAAA GTAACAAACG ATATTTCTGG ATATGGCCGT 9900

TGGTATACAG CGGGTCAACG TTTTATTGGT GGCTCTTCCC GCGGCGATAT AGGAACGAAT 9960

CGTGCGACTG ATGGAATTAC TACAAATTTA AATACACAGT TGTATAGTGC GGGTCGTGCA 10020

GATTTTGAAG GAGGCAACCA ATATTCTGGA ACGATCATCG GAGATATTAC GAATGTGGTT 10080

ACAGCAGGAA CGAATTCCGC TGGTGGTATT AATGATTTTA ACGGCGGTGC TGGTAACAAT 10140

GTTTCGAAAT TTAACAAGAG TCAAATTGGG GCCTCAAATG AAGCAACGTA TGATGCTTAT 10200

ACACCACAAC AAAGAGCAGA ATTAGCCAAA TCAGCAGCAG CTTTTAAAGT TTTTGGAAAT 10260

ATTTCTTCTA AATTAGTCAG CGGTTCGTTT AATAATGGTG CAATTTACTC AACTGCAGCT 10320

GGCCGCGGCG GTTATATCGA AGGAAATACA ACAATTGAAG TGGGAACAGC GAATGGTGAT 10380

GGGTCCTTAG GTGGCGATGG TATGGCTTAT TCAGGCGCAA AACCGACAAG CTTAGATTAT 10440

TCAACGACCG ATAAAAGTCG TGGGTATAGT TCAGGTTGGG ATATTGTTGG TGGCGGCGGC 10500

TATCCTGCTA GTAATGATAC TTGGGATATC TATATAAAAG GAGATACTAA GACTGTTTTA 10560

AATAATACCA TTGCTCGCTG GACATATGGT GGCTCTTTTT CAGGTGTCGT TGAAGGCAAT 10620

ACTTCTAACA CATTGAATGG CGGGATTGCT GACACATTAG AAGGAACAGG CTATCAAGCA 10680

GCTCGAGTTT ATGGAAACGG TCAAACAATC GTTAATAATG GTCAAGTCGA TTGGTTTCTT 10740

TCAGGCGGTG GCTGGAATGA TGCCAAAAAT GTTGGGAATG TCGGTGTAAC GGTTTATGAA 10800

GGAGTAATTA ATGCTTCGAT GGGTGCTTCT TATGGAGCAA GCGGTGGCCA TACAATTACT 10860

GGTAATTCAG ACAATCGCAT CTACGGAGGA AATTTCTCAG GAACTCCTCG GACAGGCGCG 10920

AACGGTTTTT CAGGTGGGAT TACCAATGCG GGCTCACTTC TTGGAAATGC CAAACTCTTA 10980

ATTGATTTAA GAAATTACAA CGGAGAGTTT AAGTTGCCTG GTAATACCTA CATTACTGGT 11040

GGTCGGCCCT ATGGTCAAAA TACGACACTA GGAACGGATG AATCAAACAC AATTACTTTA 11100

AATATTTTCA CTAAAACAGG TGTGGAATCT CTAAATGGGG CCAGCATTTA CGGTGATGGT 11160

GGAACCAATG CTGCGTATAC CAAGAACGGC AAGATTACAA TGAATATTCA AGCAGCGGGT 11220

TCTAGTATTG GAAATTTATA TGCGACACAG TATTCCAATA TCAGTGGCGG AAAAATCTTA 11280

CGTGATGTTA CCGCAAATGT TCAAGGTGCA GTTAGTATTA ACGGTTTATC CGGTGGTTCA 11340

TCAACAGATA ATTTCACTAA TGCAATTGTG AATGCTAGTT CAAATAAAGT GGCCTTTAAC 11400

TTTGGGACTA ACGTTGACGG CACCAATAAC TATCAAACAG AACCACTAAA TGCGACAGGC 11460

TTAGGCGTTG TCAACTTTAC GGAATTAAAT GTCACCAATG GCATTAAATT AATGGCTAAT 11520

GGTGGGAACA TCAAAAATGG CGTTTCTGCT ACTGCCGCGA ACCACAGTAC GACTTACAAT 11580

GAATTCGGAT CAATTCATCT ATCTAAAAAC GCTGGAATTG GGATCACCAA CACAGCCAAT 11640

TTAATCTCAG CTAGCAAACT AACGGTTCAG GATCATGCAA CATTAGAAAC AATTCCTGGT 11700

ACAGGTAAAG TAAACATTGC GGATTTTGAA GCTGTCGATC CAACAAAAGA TGAACTTTTA 11760

TGGATTAAAC CAACGACAGA TACAACTGCT TTGGTGGATA GCACAGGGAC TTGGTTTGGA 11820

AATGTCAAAG CCTATCAGGT CTTGACGATT AATCCAACGG TCAATAATGC TACAAAACTT 11880

ACGCCAACAA CGTTTCATGG GATTGAAAAA GCGACTGGAA AAACATTTAT TGGCGATAAC 11940

GATGTGACCA AGGGCGCCAA TGGCTATGGG ATAGCAATTC CTGGCTCCAT CATCGATTAC 12000

GAAGTGGAAA CGCCTGGAAT TGTGGCAGGT GCAGGTACTA TTTCTCATGA TGTAAAAGAG 12060

GTCAAAGCCG GCAATGCGCC ATTAACCTTA CAAGCGTGGG GCACCGAAGT AGCTGGACAA 12120

AAAGTTCAAA AAGGACGTTT GATGATTCCG ACCTCAAGTG CTTTAACACC AACCCTAAGC 12180

TTCCTTCCTG ACGAAtAACA GGTTCTTGGT TGCACCAAGG AACGGTTAAG TCTAGTGAAG 12240

TTGGGAGCGC GATAGAGCAA ATTCCAGAGC AAAAAGATAC CACCCCGCTT TCATGGAAAG 12300

CGACGAATAT GAACTACAGT TATCAGGTGA AAGTTCAATT TTCAAATAAA GTGGAACTTT 12360

CGGGGCAATC GGTGATTGTC ACAGAAGATG AGGCAGCACA GCTAACAACT GTTGATAAGG 12420

TAGTGGAAAT GCTCGGCGCA AAAGGTCGGC CGTTTTTCAA GACCTCATTA AAAGAGACGG 12480

ATTTACCACA AATTCAAGCG CCGTTAGCAG AAAAAACTGT ATCGCGGACA CATGATATTC 12540

AATTAGAAGC AGGGACGTCA GGAGATAATT TGCAAAACAA AACGGTTCAT TTAGTTGTTG 12600

TGAAAAATGA AAGTATCCTT GCCAAAGATC GTTCGTTTGC CTTGTATGCG ACAAGTGCGC 12660

ATTTGAAATT AAAAGAAGCA AATGGCTTAA CGAATTCAGA CGAGTTGGCA CAATGGACAC 12720

AAGCAACGGT GATTTTCGCA GATGGTCGGG CCAATCAAAC ACCAAGTCTT GATGCCGCAA 12780

CCTTCCAAGC CATTCAGCAA GCGAAACCAG AAGAATTAGC CAAAACGGTG CCAGCCAATT 12840

ATCAATTTAC AGAAGCTGGT GAAACGTTAA ATAAAAAAGT CAATGTGTCT ATTTCAGGTG 12900

AATTAACGTT AGAAAAAGTA CCTAAAACAA TTGATTTTGG CAAACAAAAA ATTTCTGCCA 12960

AGCCAGAAGT TTACTGGCCA ACTCTTTCAG ATGATTTAGT CGTTCAAGAC ACACGCGGGT 13020

CAGAAAGTAC GCCTTGGAAA CTGAATGTTC AGGTAACTAA TCCCTTAACG AACGGCACAG 13080

ATCAATTGGA AGATTTGTCA TTGGTTACAG ACAAAGGAGA ATTTTTACTC AATAAAGGAG 13140

ATACGGTTGT TACAGAAAAT GAAGGCTCAG GTTCAGGTAG TTATACAATT AACCAAGGTT 13200

GGGGTGCGGC AAACCAACGG GGACTTAAAC TATCTGTGCC AGTCGAGAAA CAAAAAGTTG 13260

GCGAATTCAA AGGGACATTG TCATGGTCAT TAGTCATGGC ACCATAACAA GGAGGAGAGA 13320

AGTGAGATGA AAAAATACCT TTTGCTTAGT TGTTTTTTAG GTCTTTTCAG CTTCTGTCAT 13380

TCAGACACTG CGTTTGGAGA AGCAGCTTAT GAAAATAGTG GTGTTGTCTC CTTTTATGGA 13440

ACGTATGAAT ATCCCACAGA AGAGTCGACA ACAGCGACTA GTAATTCTTC CACAACGACC 13500

GAACCCACCA AGCCAGCTGA CGGAGGCGCT TCATCCGTCC TTTCTTCTGG CGTATATGGA 13560

TCGCGACAAG GAAGATTACC AGCGACAGGT ACCACCAATC AAGCACCATT TATTTATTTG 13620

GGAATCAGCC TTATCACTAT AGGCATATTA TTTATTAAAA GGAGAAGAGA AGATGAAAAA 13680

AACAGTATTA GCAGTAGTAG GGATTGTAGG ATTTAGCGGT GTCCTAGCGA CACAACAAGC 13740

ATTTGCGGAA GACATTAACG TGAACTATAC AAGTAATGGC GCCATTACAT TTGAACCAGA 13800

TACAGATCCA ACGAAGCCAG TTAACCCAAC CAATCCAGAT GAAAAAGTTG AACCAGAAGA 13860

TCCAACTGAC CCAACTGGAC CAAAACCAGG GACTGCAGGC CCCTTATCAA TTGACTATGC 13920

TTCAAGCTTT CAATTTGGCG CCCAAAAAAT CACTTCCGAT ACGAAAGATT ATTACGCACA 13980

AATTCAAACA TTTAAAGATG GAACAACAGG TCCCAACTAT GTTCAGGTAA CGGATAAACG 14040

TGGGACACAA GAAGGCTGGA CGTTAAGCGC TGTACAAAAT GGGCAATTCA AAACAGCCCA 14100

AAATGAAGAA TTAGTGGGGG CCGCTTTATC GATTGCCAAT GCAGGTGTGA CATCGATTGT 14160

TGATGCCGCA TACGCGCCAA CACCAACAGC AGCACATACC TTTGTTCCTG GAACAGAAGT 14220

AGAATTAGTC AAAGCAGAAG ATGGCAAAGG AATGGGGACT TGGGTCTATC GTTTTGGGAA 14280

AGATGCGACA GAAGGTGCTA CTGCCGTGAA ATTAAATGTC CCAGGTAAAG CGATTAAATT 14340

AGCAAAAGAA TATCGTACGA CCTTAACATG GACATTGAAA TCTGTACCAA CGAATGTTGG 14400

AGGCTAAGTA AATGAAAAAA ACAGTCGTCT ACTCCTTGTT ATTCGGAACA ATGTTGCTTG 14460

GCGCCACTGT TCCTGCTGAA GCGGCGACGG TCGTTTTTGA TAGCGAACAG TCGATTGTTT 14520

TTACCCCAAG CACAGATGGG ACGGATCCAG TAAATCCAGA AAATCCCGAT CCAGAAAAAC 14580

CAGTTCGACC AGTCGATCCA ACGAATCCTG ATGGACCTAA TCCAGGTACC CCTGGTCCAC 14640

TTTCCATCGA TTATGCCTCA AGTTTGGATT TTGGGAGTAA TGAGATATCG AATAAGGATC 14700

AAACGTATTT TGCCAGAGCG CAAACCTATA GAAATCCAGA TGGTTCAGCA AGTGAATTGG 14760

CAACTGCTAA TTATGTACAA GTAAGTGATT TACGGGGAAC CAATGCTGGC TGGGTTTTAA 14820

AAGTGAAACA AAATGGTCAA TTTCGTAATG CAGAAACATT ACACAAAGAA TTAACAGGCG 14880

CCACCGTCGC CTTTACTGAG CCCAGTGTTC GCTCAAATGC GACGGACGTA TTGCCGCCAA 14940

CTGCTACCGC AAACATTCAA TTAGATGCTG CGGGCGCAGA AACTGTTGTC ATGCAAGCCC 15000

CAGAAAAGAC CGGCGCCGGA ACGTGGATCA CGCTGTGGGG GCAAGCAGAA AAAGTGACCG 15060

AAAAAAATCA ACAAGGACAG CAAGTAAATG CCACAATCAC ACGGGCAATC TCACTAACTG 15120

TTCCTGGGAA AACCCCTAAG GATGCAGTAC AATATAAAAC AACATTGACT TGGCTACTTT 15180

CAGATGTACC AGTAAATAAT GGAGGGAAAT AAATGAAAAA GACAAAAGTA ATGACATTGA 15240

TGGCAACCAC AACTTTAGGC GCACTGGCAC TTGTACCAAT GAGTGCATTA GCAGTCGACG 15300

GTGGTGAATA CCAAACAAAC GGAGCGATTC AATTTGCACC AAATACGAAC CCAACGAATC 15360

CAGTTGATCC GACGAATCCA GACCCAGATA AACCAATTAC ACCAGTTGAT CCAACTGATC 15420

CGACAGGGCC TAAGCCAGGG ACAGCAGGTC CGTTATCCAT TGACTATGCA TCTAGCTTAT 15480

CTTTTGGGGA ACAAACGATT ACCTCAAAAA ATATGACCTA CTATGCAGAA ACACAAAAAT 15540

ACAAAGATAA CGCTGGTGCC GACCAAGAAG GCCCAAACTT TGTTCAAGTC TCAGATAATC 15600

GTGGGACTGA GACAGGTTGG ACGCTAAAAG TAAAACAAAA TGGTCAATTC AAAACTGAAG 15660

CCAACCAAGA ACTAACAGCG GCCAAAGTAA CTTTAAGCAA CGGACGCGTG GTTTCAGCTT 15720

CACAATCTGC AAAGCCAACG ACAGCGCCAG CTACGATTGA ATTAAACCCA ACTGGGGCTG 15780

AATCAGTGGT CATGGCTGCT GGCGATAAAG AAGGTGCGGG TACGTACTTA ATGAGCTGGG 15840

GCGATAGTGT AGATACCGCT AAAACAAGTA TTTCATTAGA AGTACCTGGT TCAACCACAA 15900

AATATGCGAA AAAATACACG ACAACTTTTA CTTGGACTTT GACAGATACA CCTGCTAACA 15960

CAGGAAACTA AGAACGACAG CATACACGCA AGATCGGGAT ATAGGTCAAA AATTTTTTGG 16020

CTTATCTTTC GGTCTTTTGG TGCTTATAAT ACAACAAAGA ATGACAGACA TAGGAGAATG 16080

AATATGAACA GATGGAAAGT ATATGCAACG GTAATCGCTT GTATGTTATT TGGCTGGATT 16140

GGCGTGGAGG CGCACGCTTC TGAATTTAAT TTTGCGGTCA CACCAACAAT TCCCGAAAAT 16200

CAAGTGGATA AATCAAAAAC CTACTTTGAC TTAAAAATGG CGCCTGGTGC CAAACAAACC 16260

GTAGAAATTC AGTTACGCAA TGATACAGAT GAAGACATTA CCATTGAAAA TACGGTGAAC 16320

TCAGCGACAA CAAATTTAAA TGGCGTAGTA GAATATGGCC AAAACGGGAT CAAACCTGAC 16380

AAAACCTTAC GTTTTAACTT AAAAGATTAT GTGGAAGCAC CGAAAGAAAT CATCTTGCCG 16440

AAGCATTCCC AAAAGACCTT ACCTTTAACC ATTACGATGC CTAAAGATTC TTTTGATGGC 16500

GTGATGGCTG GCGGTATAAC ACTCAAAGAG AAAAAGAAAG AAACAACGAC TTCTGCGGAT 16560

CAATCAAAAG GGTTAGCTAT TAATAATGAA TACTCCTATG TTGTGGCTAT TATTCTTCAG 16620

CAAAATGAGA CAAAGGTTCA ACCAGATTTA AAATTACTGG GGGTTAAACC AGGCCAAGTC 16680

AACGCGCGAA ACGTCATCAA TGTTTCTTTA CAAAACCCAC AAGCGGCCTA TTTAAACCAA 16740

TTACATTTAA TCAACACTGT TTCAAAAGGA GGCGAAACGC TTTACCAATC CGATACTGAG 16800

GATATGCAAG TGGCGCCAAA CTCTAACTTT AGTTACCCAA TTTCTTTAAA AGGGGAACGA 16860

TTAACGCCAG GAAAATATGT CTTGAAATCA ACGGCCTATG GTGTAAAAGA TGAAAAGGGC 16920

ACCTATCAAG TCAAAGGCGC CAATGGTGAA GAACGGTACC TGTACAAATG GGAATTTACA 16980

AAAGAATTTA CTATTTCTGG GGACGTCGCT AAAGAATTAA ATGAAAAAGA CGTAACCATT 17040

AAAGGAACCA ATTGGTGGTT GTATCTACTG ATTGCATTAA TCATTCTAGC GCTGCTCTTA 17100

TTGATTTTCT TCTTGTATCG TAAAAAGAAA AAAGAGGAAG AACAACAATC TGAGCAATAA 17160

AAAGGGCTAT CTATTGACAA AAAGGGTTTT CCTTGTAAAA GTAAAGGTAC TTTTACAAGG 17220

AATTGCTTCC CATAAAGTAG AGGTAGAGAG AGTGAGAAGC CATGAGAAAA AAGTATCTTT 17280

TTTTATTTAG TTCTTTGCTA ATTTTTACGG CTGTCGAAAT AGAACCAATT CAAACATTGG 17340

CGACACCTAT TTATAATCCC TTAGAACCAA CGGAAGAAAT TCAACCAAAA CCAGAATCAA 17400

TCACCCCTCA GGAAAAGTCA GAACCGGAAC AACCTGCGCC AACCCCTCCG GAAGATCAAA 17460

AAACTACGGT TCCTTCTGAA GAAAAAAAGC CTTCACCCAA GAAAAAAGAC AAACCTGTGG 17520

AGAAACCGGC AATGTCAAAA GAAAAAAAGA AACAAACAAA AGAAACAACG GCTCATCCCT 17580

TACAAGTGAA TAAAGATTTT GTCTTTCCTA ACGGACTTGG TGCAGGCGGC GATGAAGAAT 17640

CTGAAAGCTA TTTAGCAGCG GTTGGGTACT TTCTTTCAGG AAATGGATTG TATGGGAAAA 17700

AAATGGATGT GAGAGAAGTC TATGATATTG ATGATGTTTT TAAATAAAGT GGATAAAAGA 17760

AAAATCGCCT TATTTCAGTT ACTGGAAAAT GCACCGATGT TAACAGAGTC AAAAGAAACC 17820

GTAATGCAGG AGTTAGAATT ATCTGAATTT ATTGTGAACA AAACGGTCGC TGAATTAAAT 17880

GCCGATTTTA TTGAATTTGG TTTATCAGAG GATTTTCAAA TTGTGAGCGA TGGTATTTTC 17940

TTAACTCTTA ATGAATCAGG AACGGAAACT TCCGCTACCT TGATTGATTG CTATATTAAA 18000

GAGTCGTTAC GTTTTGCTAT GTTTCAAGAT TTCTTTTTTC AGCGTTTTGA TTCTGTGAAT 18060

GAGTTTGCGT TAGAACACTT TGTTAGTCAC ACATTGGCTT ATAAGGAGTT CAAGGAATTA 18120

AAGAAAATGC TGGAAAATTA TCAAATTACA ATTAATAAAG AATTTCACTT AGTGGGAAAT 18180

GAAACCAATT TACGCGAAGT AACAACACTG GTTTTTTTAC AATTATACCA ACGTGATTTT 18240

TCATTGTACG GTAAAAAAGT TTTAGAGCAG ATTCGGAACT TTGAGCTATT TTTTTCTGAG 18300

AAGGTTCATC TTAATCAGCG TGCGGGCTAT GCACAAGCTA AATGTTTTCA TTTTTTGGCG 18360

GTTTCTTTAG TACGAACGCA GCAACAGTTT TATGTAGAGA AAAAAGTAGA TTCGCAAGTG 18420

TTAAAACAAT TAGCCGGTGA CAATCAGCAG AATGTTTTCA ATTGGTTAGA GAGTCTTCAA 18480

ATCCCAGAAC CACAACGAAC AAACGAAGGC AATTATTTAC TCTTATTTTT AATTGCTGAA 18540

GAGTGGcTTC AAGTAGCTTC AGCAGCACTT TGCCAAAGAT TTGCAGAAAT TCAACAATTG 18600

AATCGTGACT TGTTAAGCTA TTTAAAAATA CAATTTCAGC TTTCCGAGGA AGTCCTGCAA 18660

GGTTTTGAAC AAAAAGTAAC AATCGTACAT TTTAAATTGT TCCATTTTCC AATAGAAGCA 18720

AACTATAGTT ATCTCAAAAT GGATATAACC TACTTTGCAG AAACGTATAC AGAGTATTTT 18780

TCTGCTAGTC GACAATTTAT TCAAGAACAA AAAACGAATC CCGCTGTTTG GGATGCCAGA 18840

CGGTTTTTAT TTTATCGGTA TCTGTTGTTA ATGGTGGATG CGTTACCATT GAAAAAAGTT 18900

TTGGCTCCGA TTACATTGTG TGTCGACTTT TCCTTTGGTG AAGCCTATAA TCGTATGATA 18960

CAAAAAAATA TTGAAAAAAT TACGGAATGG AATATCATTT TCAAAGCACA TGTAGATGAA 19020

AGTGTACAAT TAATTTTATC GGATATTCAT TTTGTCGATT GGCCTGAGAT TACTCAAATT 19080

ATtTGcTAGC GCCACCACGT CCTGTTGATT GGTCGCATTT TATTAAAACG CTGAGTGCAC 19140

TAAGACCGCA CTCGGGAGAA GAGAAAAAGA AATCATCGTA GACTAGACGG TGATTTCTTT 19200

TTCtTTTAGA ATATAAAAAT TAAAAATAAT AAAAAAGTTG AAACTTTTTA TGCTTGTGAT 19260

TCGTCTATAT AAGTGTACAG GAAAGGAGGT TCCAGAAATC GATGGTTTTC GATGGTTGGA 19320

AAAGAAAAAA ACAAATTCGC GCGTTAGAAA CGCTTGTGGA TACCAAATAT GAAAAAATGT 19380

ATCGCATTGC CTTTAACTAT ACACATTCGA AAGAAGATGC CTTGGATGTG ATTCAAGAAA 19440

GCTTTGAAAA GGCACTGAAG GCCATCCAAA AAGGACCCGA TATTAAGGAT TTTGACGCGT 19500

GGTTCTTTCG AATTCTGCTT TCTGTAGCGA CAGATTATTG GCGTAAAAAG AAACGAGAAG 19560

CTGTTCATAT TGAAAAAGAT GAACATGTTT TAGCCCTTTT AAATAATCCC CAACGCAGTT 19620

TTTTAGAAAT TGAAGAAATT ATTAATCAAT TAGATAGTCC CAATCGTGAA ATTATTCTAT 19680

TAAAATTTTT TGAAGGATTT ACGCTAAAAG AAATTGCCTT GATCCTGGAT ATTAACGAAA 19740

ATACGATTAA AACGAAGATG TACCGCAGTT TGAATGAACT CCGAAGTCTA TTGAATTAGT 19800

AAAAGAAAGG AAGCGAAAAA ATAATGGAAG ATTTTGTAAA AAGTGTTGTT AAAAACTTTT 19860

CAAAAGAATA CCGCCAACAA CCGTTGCCAA CGGATCTACA AGCGGAAGTC AAAGCGCGTT 19920

TTCATAAAGA GAAAAAGCGG TACACCTGGC TTCGTTTTAG AAACCGCTCA CTGCAAAGTA 19980

GTCTAGTAGT GGCTTGTGGC TTTGTCTTAT CAGTCAATTT GTTTCCAGGC TTTAGTGAAG 20040

CAGCCCGAAA CATTCCTGTG TTAGATAAAA TTGTTCAATT AGTCACCATC AAAACGTTGA 20100

CAGCGAAAAA ACAGGAAAGT GAAGTAAACA TCGATGTCCC TAAAATTCAA ACATCCCAAG 20160

AATCTTCAGT AGCCGACACG CTCAACAAAA AATACTTGAA AGAAGCCCAA ACAGAATTTC 20220

AACAAGTCAA AGGGCAATTG TTTGATGGTT CTCGTGTTTC AGTCACTGGA GATTATGAAA 20280

AAGTCGTGGA CGACCGTCGC TTCCTAGTGG TCAAACGAAC CTTCACAGAA ATTAAAGGCA 20340

GCTCTGCGAC GACAACGAAA TATGACACCA TCGACAAACG AGCCAATGTT GTCGTTTCAC 20400

TCCCCTTAAT TTTTAAAAAT GATTTTTACA TTGATGTCAT TAGTAGTGAA ATTAAACAAC 20460

AAATAGCAGA CCAAATGAAG AACGATAGCA ATAAAATATA CTGGTCCGAA AAAGAGGATG 20520

CCCCCAGTGA TGTTCAACCC TTCAAAAAAA TTAAAAAGGA CCAACCTTTC TATATCAATG 20580

AAAAACACCA ATTAGTGATT GTGTTTCCCC AAGGAGAGAT TGCGCCTTAC TATATGGGCA 20640

CTCCTGaATT TGTGATTCCT GATCAAGTCA TTGAAAATGA ATTGGCAGCG CCTAATTATT 20700

TAAAATAAAC tCAGTAAAAA ACACGCGACA TTTCTTAAAA CTGTCGCGTG TTTTTTGCTG 20760

ACCTAACCAA GTCAGTTGCG GTTTTACTGA GAGAAGCGTA CACTAGAATT AAGCATAACA 20820

AGAAAGGGCG AATGGCAATG GaAAAATCAG AAAAAATTAG TATTTTACAA GATGTGGTTA 20880

AAATTAAATC TGTAAATGGa AACGAAGAAG AAGTGGcAAT CTATTTACAA AATCTTTTGA 20940

AGAAATATGA GATCCCTTCT GAACTTGTTT CTTATGCACC AAATCGTAGT AGTTTGGTTG 21000

CTTATTTAGG TGAAAACCGA GAAAAAGTCT TAGGGTTTAG TGGCCACATG GACGTGGTTT 21060

CTGAAGGGGA CGAAAGTCAA TGGACTTTCC CACCATTTGC GGCCCATATT GAAGGAAATA 21120

AATTGTATGG CCGTGGCGCA ACGGACATGA AAAGCGGCTT AGTCGCAATG GTTTTAGCAA 21180

TGATTGAATT AAAAGAAAAG AAAGTCCCAT TAAATGGAGC GGTTAAATTT TTAGGTACAG 21240

TTGGTGAAGA AGTCGGTGAG CTAGGTGCTG GCCAATTAAC TGAAAAAGGC TATGCGGATG 21300

ATTTAAGTGC CTTAGTCATT GGAGAACCAA CCAATTATAA CTTGATGTAT GCGCACATGG 21360

GTTCAATTAA CTACTCCGTT GTGTCACACG GAAAAGAAGC CCACAGTTCG ATGCCAGAAG 21420

AAGGCATTAA CGCGATTAAT AATTTGAACG AATTTATTAC AGAAGCAAAC CAACAAATGG 21480

CGGAAGTCAC AGCAAACTAT GAAAATCCTG AATTAGGCCG AACGATTCAC AATGTAACCG 21540

TCATTAAAGG TGGCACACAA GTGAACAGTA TTCCAGGTCA AGCAGCCTTG CAAGGGAATA 21600

TCCGTTCAAT TCCAGAATTC AGCAACGACC AAGTCATTGC ATTATTACAA AAAATTGTCG 21660

ATGAATTGAA CAAAAAAGAA AAACATCAGT TAGAATTAAC GATTGATTAC AATAAAATTC 21720

CTGTAAAAGC TGAGAAAGAT TCTGCGTTAA TTCAAGCCAT TCAAGCCCAA TTTGATCAAC 21780

CATTACCATT AGTGACAAGC GCTGGGACCA CAGACTTAGC TGAATTTACT AAATCAGACA 21840

ATACATTTGA TTGCGTTGTC TTTGGACCAG GTGTGACCAC TACTGCACAC CAAGTCGACG 21900

AATATGTCGA AATTGATAAT TATTTAGATA TGATTGATAA GTATCAAGCG ATAGCAAAAA 21960

GTTATTTAAA TTAAAAAAAT GTCGCGCCTG AATCATAAAA ATCAGGCGTG ACATTTTTTA 22020

TAAGGTTTAT TCCACCACAG TCCATTCATT TAAGGTTTTG CCGTTCTCAT CAACTAGCTC 22080

TAACCAGCTG TTTGTTCCAG CAAAATAGAG AaACAAGCCG ACTTCATTAA CGCTTTTTAA 22140

GACGATATCT TCCGTTGTGA CAAAGTCTTT AAATTGGCCC TTGCGTTCAT CACGAAGTTT 22200

TTCACCAAAT CTAGCGGAGA AACCGACAGA CCCGTCTTTA TTTAAAGGGG CTTCGGGCAT 22260

CATTGTCGCG CCAGCTTTAA TAAGCATTTC ATTGCGCTTT TGCCAAAAAA CAGTTGCTTT 22320

GCTTTGTCGA GTATCAATAT AAAAAGGAAT TTGGCTAACT TCTTTTGTCC ATTTGTGGCG 22380

TGCTTTGGCT GGTCTGGCTT TCTTCGGCGC CGGTTTTTTA AACAGAGAGT AGCCAAAACG 22440

TTCCAAGATA AAGGTGACTG TTTCTTTATA GTTAGTCACT GTCGCTACCG CACTTTTCGG 22500

AATTTTGACT GTCGTTGGGG TTGCTTCCAA TGTTAAACCA CTTTCTTTCG CATCATCTAG 22560

GAAAGCTTGC GCTAAATAGG CAAGTTCTGT GTCCCAAGTC GGTTTCAATA AAATCGAGTG 22620

ATCAAAAGAA ACTTCTTTAT ATGTAGCTTT TTCTTCAGCA AGTCCTAAAG CAACTAGTTG 22680

GTTATCTATT AGCAAGAAAA GAATCGCTGC CTCTTTTTTC GGTAATTTAG TTGTTTGTTC 22740

AATGATTAAG TGGGCCTGAT CGGCTAAGTC AATGAGTGTC TTTGTTTTTT TTAGATCAAC 22800

GGTTAAATGT AAGTCCATGA AGTCTGCTCC TTTATCGTTA GTTTCTTCAT TGTATCATGA 22860

AACAGCGTCT TCGTCTTTTA ACGAGTCAGC GACCTTACAA TATTGTTTGA AACAAAAAGA 22920

AGCTTTCGTG GTATACTTAC TCACATACAT AATTTTTAAT GAGGAAAGTA GAGGAACAAG 22980

AATGAAAACA CTTGGCAATA TTTTATGGTT CATCTTTGGC GGCTTTTTTG GTGGCTTATC 23040

TTGGCTATTC GCAGGCGTGA TTTGGTGTAT TACAATTATC GGTATTCCGA TTGGTTTACA 23100

ATGTTTTAAA TTAGCGGGGT TGAGCTTTTG GCCGTTTAAA AAGCGAGTCG TGTATAGTGA 23160

TAGCGGCGTA TCTTTGGTTG TAAACATTAT TTGGCTGATT ATCAGTGGAC TACCATTAGC 23220

AATTGGGCAT TGTATTAGTG GCTTGTTTCT TTGTCTGACA ATCATCGGTA TTCCGTTTGG 23280

GCAACAGTCC TTTAAATTGG CGAAATTAGC GTTAATGCCG TTTGGGGCCC GTGTTGTTTC 23340

AACAAATGAT TTTTATTTTG AATAGTCAGA GGTTTCTCAT TGACAAATAA AACGAAACAG 23400

CGTATTCTGT AAATGTAACA AATAACGTTA CATTTAGTTT GGAGTGATTA GACATGGCCA 23460

TGTCAACGAA ATTAAGTGTT GCGATTCATA TTTTAAGTTT GATTGAAACA GGTCCTCAAG 23520

AACAAGTGAC TTCTGAATAC ATTGCTTCGA GCGTGAATAC GAATCCAGTC GTTGTTCGAC 23580

GACTAATGAG CCAGTTAAAA AAAGCCGGCT TGATTCATTC CATGCGTGGA GCCAATAAAA 23640

ATACCTTATT AAAGAAACCA GAAGAGATCT CACTGTATGA AATTTATACA GCGGTTGAAT 23700

TGGAACGAGA GATTTTCAAT ATTCATCAAA ATCCTAATCC AAACTGTTCC GTTGGGGCAA 23760

ATATCCAATC GGTACTAGAA ACTGAATTTA CTAAAGTGCA ACAAAAAATG GAAGAAGAAT 23820

TGAAAAGTAT CACCTTAGCA GATGTGATTC ATGAAATAGA AGTTAAGAGA GATAAGTAAT 23880

CTCTCTACTT TTAAAGTATT AAATGTAACT ATTTTAGTTC TATTTAAAAA TGAAAGGATG 23940

ATTTACATGA AAGTAGCAGT ATTAGGAGCA ACAGGAAAAG AAGGACAATT ACTTTTAGCA 24000

GAAGCTAAAC GTCGCGGGAT GGACGTGACA GCCATCGTTC GTAATGCCTC AAAAATAACA 24060

GATGGCACAC CGACCATTGA AAAAGATGTT TATCAATTAA CAACAGAAGA TATTCAAGCG 24120

TTTGATGTAT TAATTAGTGC CTTAGGATTC CCTGATGTGC AAGATTTCCC AAAATCTACC 24180

CAGCATTTGA TTGAGATTTT AACCAATCAA AAAACGCGCT TGTTTGTCGT TGGGGGCGCT 24240

GGAACCTTGT ATGTGGACCC AGAACACACA ACTCAATTAA AAGATACACC ATCATTCCCA 24300

GCAGAAATTC AGCCATTAGC TTCGGCGATG GGCGAAgCTT TAAACTTATT GAAAGAAGCT 24360

CAAAACATCC ACTGGACCTA TATTAGTCCT GCAGCCATGT TTGATGCTGA TGGTCCTGCG 24420

ACAGGCCATT ATGAAGTCGC TGGCGAAGAA TTAACAACCA ATAGTCAAGG TGACAGCTAC 24480

ATTAGTTACG CAGACTACGC CGTTGCCATG TTAGATGAAG TGGAAAGTAA TGCACACCCA 24540

AATCAACGAA TTAGTGTGTA TCAATAAaAG AAmCAAGAnA AATGGGGCTC CGTTAAACAA 24600

C 24601

(2) INFORMATION FOR SEQ ID NO: 224:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1405 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

CTTCCATTTT GGCACACAAA TTTTTmCAAa GrATTgAGCT TCAGGAAaTA ATTTTTTTAT 60

CAGGAAATTa TTAATTATTA CGAAAAAAAT GTTCATTTTC TCACAATwCa TTGgTATAAT 120

AACGTTATAT CGAAAGGAGA GAAATGAATG AGTGTAGTAT TAGTACTTTT GGGTCTATTA 180

GGGTTACTTG TCAGCGGGGG ACTGACTATC AAAGCATTAA TTAAGAAACA ACCGAAAAAA 240

CAATATGGCT ATGCCGCGTT AGCAAGTGCC GCTCTATTTA TTGGTGGGGT TGCCACGACA 300

GATTCTTCCC CGACACTTTC ACTTGATGCA GAGAGTGTTG AAACCGATGC CAATGGCTCA 360

GCCACCATCA AAGGCGAATA CGACGGCAAC AGCAAACTAA CGGCCAATGG CAAGGAAGTA 420

AAAACGAAAG ACAAAACGTT TGCTTACAAA GTAAAACTGT CCGATGAAAA AACGCAAAAA 480

ATCACTTTTG TTGCTGAGAA GAAAGATACC AAAGTTGAAA AATCAGTCGA AGTCACACCG 540

TCAAAAGAAT TTATTGCGAG CATCACAGAT TCTGAGCAAG AATTGGCCAA AACAGAAAAA 600

GCCTTAGCTT ACGCTGAAAA ACAACCTTCT CAGAAAAACT ATGATGAAGC CGCAACCTTA 660

GTATCTTCAT TAAGTCAAGA ATACGAGGAA TACAATGATC GCCTTGAAAA AATCAAAGAA 720

GCAGTACCTG TTGATGAAGC AGTTACAACT GCTGAAAAAT CTAAAAGTAA AAGCGATTAT 780

CAAGCAGCTG AAAAATTAGT AGCTGCTGCC CCAGTTGGCA AAGAAGGATT TCAACAACGA 840

TTAACGACTG TTCAAACCGC TATCGCTGAA AAAGAAAAAA ATGAACAATT AGTCGCCAGT 900

GCAAcTGCCG CTGTTGAAAA AGCTGAACAA GAACCAACCA ATGAAGCTTA CTATAATGAA 960

GCGGTCAAAC AAATTGACGC TTTAAACTCC CCAAATCAAG CTTTAACAAA ACGGGTAGCC 1020

GTTGTTAAAA CTCAATTAGA TGCCCACAAA GAAAAACAAC GAAAAGAAGC AGAAGCTCAG 1080

AAACTAGCCG CTGAAAAAGC GCAAAAAGAA CAAGCCGAAG CGGCAGCAAA AGCGCAAgCT 1140

GAAGCAGAGG CGCAACAAGC CGCACAGGCG CCGGCGGAAG TCGAGACGGC TGCTGCGGAA 1200

GCACCAAGCG GTAACGCCTT AATTAAAGGG TCACGTAATG GAATTTATCA TGTACCTGGA 1260

AGTAGaTATT ATAATAGAAC TACGAATcCA GTAGCTTGGT TCTCTACTGT TGAAGAAGCA 1320

GAAGCAGCAG GATACCGAGC ACCCAAACAG TwAATACtTA GCAATAATTA ATCAAmCAgn 1380

AAATGGGcTC CGTTGTTTAg GAGGC 1405

(2) INFORMATION FOR SEQ ID NO: 225:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6092 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

TATTGGATTA CAACAGTACT TTTTGTTGCA ATTGGTGGAT TTTTACGTTT TTTTACAATA 60

TTTGATGCTT CAAACGAATT ATATCTAAAC GTCCAGAATT TTTTAGAAGG CTCAGAAAAC 120

TCAATTAGTT TTGACCACAA TTTATTGATT ATCTTGTATA TTATAGTTCT AACAGTCATC 180

GCGCTGACTA TTGCAAGGGT AAATAATAAA CGTTGGCTAA GACTAGGAGT ATAAACGACA 240

CCCACTTCGC AGATAGGATT CATCTACTTA TTCGAGACAA TAGATTAAAT CAAATAAAAT 300

ACAACTAGTT GAAACTAAAA AAGCATCAAT TTTGATACGC CCCAAACTGA TATTTTtCTA 360

AACTTTGGGG CGTGTAATCA AAATACGAGC TTCTTAGTGT GGCACTGGAT ATTTCAACGC 420

ATTTTTCTTC AACTTTTCTT CGATAATTTC GTCCAAATCA AACCCTAAAT TATCAGCCAT 480

CATGTAGGAG TAAATGAGAA CATCAGCCAA TTCTTCTTTG ATTCGTTCTT CTTGTTTGAT 540

GCCTTCTTCT GCTGTTTTCC ATTGAAAAAG TTCTAACAAT TCTGCTGCTT CTAAACAAAT 600

TGAGAGCGCC AAATCTTTTT CATTATGGTG GGGGCGCCAA TTTCTTTCAT CACGAAATTG 660

ATTAATTTTT GCCATTGTAT CCATTTGTTC CCCTGCTTTC TTATTCATTT GATGGCAACT 720

GAAACCCAyC GAACGTTTCA CCATCATAGT ATTTTAATAA CTGTTCTGTA GACAATAATG 780

AGGAGGTTAT TTCTTCCTCT TCAACAACTG CTTGATTCTC TTCTCTATAC AACAAATAAG 840

CCTGTAAAAG CTGTTGGTAA AGTTTTTTTC GTGAAGAGGC TTCCGCTGAA AGAAACGtTA 900

ATTCTTTTAA GTGTCCGCGA TAGACAGGCA AATCATCAAC GAATTCGATT GGTAATTCTA 960

CTACAGTTAA CTGAAATTCT TCAAAGACTT sTTCGTACAT CTGCCATTCC CTCTTTCTGT 1020

CGATATTGTT TTGGGGTCAT GCCCATTTCT TCTTTAAAAC GTTTATAAAA AAAGGCTTTG 1080

TCTTGATAGC CGATTCGTTC aACAATCGTT TGAATCGATA AATCtGTTTG CTTCAAGTAG 1140

TTTTTCGCCA ATTGGATTCG GTAGGCTAAG ACAATTTCCG TAAACGTTTT GTTTGTTTTT 1200

TCTTTTAAAA AATTCGATAA GTAATTCGGA TGAAAATGAA ATTTAGCGGC GACTTCTGCT 1260

AAAGTGACAT GCGCAATCTG TTGCTCAATA TACGCTAAAA TCACCTCGAC ACGTTGATGA 1320

TTCGACAAAT AATTATCTTG CATCTTTAAA GCTTCTCTAG GCAGTTGACC TAGTTCCACC 1380

ACTAGCAAAA CAAAGGCCGC TTTGGTCACT TTATTATTAA AATAGGCCAT TTTCACAATT 1440

TGTTTTAGTA ATAATAGCGA ATAGACGTGA ACATCCGAAT TTGGCTGGCT GGCAAAAACG 1500

AAATATCGAG GAATGTTTTC AAATTCACTG CTAGCAATTT CTGTAAAAAA ACGATACATG 1560

ACTGGTTCTT CCGCCAGCTG ATTTAAGAGC GTATCAAGAA AATAGTGCTT TTTAAAATAA 1620

AATATAATCA TAGACATTCC TTGCCCCTGA TAGCTAACCG AGGTGGCCTC ATTGACAATC 1680

ATGACTTGGC CAGCCGAAAA CGAACACCTT TGTTGATCCA AATAAGCATA GCCGTCTTGC 1740

TCTAAACAAT AGAGCATCCG AATAAATCCT TTCGGAACGT CAATGGTTTC CGAAGATGCC 1800

AGCATAAAGC CCGCAATTTC TCGATGTGGA TTTTCCTTAA ATTGCGCTAC TTCATTCTCA 1860

AGAGCATTGG TAGTGGTTAA ATAACAATCT AAGGTCTCTA AAAACTGATA AAATGCTGCA 1920

TATTGTTCTT CTTTGCTTGC CAATTAATCA CCTCATAATC CAACACTTTT TACTTTACTC 1980

CCTCTATTTT TTTCTTTGGC GTTTCCTTCT ATACTAGATA CATGGTAACA CGAAAAAAGA 2040

GATTCAACAG TTTTTAGGAG GAATTTTTTA TGAAAATTAG TTTTATTGGT GCTGGAAATA 2100

TGGCATCCGC TATTGCCAAA GGTGCGCTAA AAAAGCAATT TATTGCCGCA GAAAACCTTT 2160

ATTTTTATGA TATTCAAGTC GAAAAAACCG CGGCTTTTGC ACAAGAAATT GGTGCACACG 2220

CGGTTGCTTC GCCTCAAGAA GCTATTACAG TAGCTGACTT AGTTATTTTA GCAGTGAAAC 2280

CGCAATTCGT TCAAGCGGCT TTACTTGAAG CAAAAGAAGC AATTTTAGAC AAGCAACCGT 2340

TAATTGTTTC CATTGCGGCA GGTACGACTA TTGCAGAATT GTATCAACTG TTTGAGACTA 2400

CCCAACCGCT TCGTCTCGTT CGTGTGATGC CCAATATGAA TGCGTTGATT GGCGCGGGTG 2460

CAGCAGCTGT CTGTGGTAAC GCTTTTGCAT CCCCTGAAGA CATCCAGACT GTTTTATCGT 2520

TGTTTCGAGC AGTTGGCCAA GCCTGGGAAT TAGAGGAACA GTATTTTTCG ATTTTTACTG 2580

CTATTGCAGG TAGCACTCCT GCGTATGCTT TCTTATTTAT TGATTCGATT GCCCGAGCAG 2640

CCGTTAAAAA TGGTATGAAC AAAGAGTTGG CCTTAGAAAT TGCCACACAA GCTGTACTAG 2700

GCAGCGCTCA AACTTTGGCA AATAGCACAG AGAATCCTTG GACTTTAATC GACCAAGTCA 2760

GTTCTCCTGG TGGAACCACG GTTGCTGGGA TTGTCGAATT AGAAAACAAT GCGTTTATTT 2820

CAACAGTCAT CAAAGGGATT GAGGCCACCA TCCTTCGGGA TCAAGAATTA GCCAAACAAG 2880

ACTGATAAGA ATAATTACAT GAGCGAATAA ATAAGAAAAA GCGCTGAAAA TTCAGCGCTT 2940

TTTCTTATTT TAGTTGAATG AGTAAGCCAA AATGATCACT GACTATCGGT GTTTTCCTAC 3000

CATCAAAAAT TACTTCATAT TTTCGTACGT GCCACTGTTT TGGAACAAAT GCGTAGTCAA 3060

TCCGCAAGGC TGCTTCATTT TCTTCCCAGC CATCTATTTT TTTCTCTACT GTCGCTTCAC 3120

CGCTTGTCTC TTCCGCTACA ACAAATGCAT CTTGAATCGG CAAGTAACTA TTTTCCACTA 3180

ATTGATAGCC AGTTGTACCA GCTGGGTTGT TTAAATCGCC TAAAATGACT AACGGTTGTT 3240

GCCCCGTCGC TAAGTATTTC TCTAATTGTA GCCATTCATA GGCAAAACCT GTACAGGGTG 3300

TTTCCCACCA AGAAAAGTGA CCAGAGACAA CAGTAAGCGT ACCTTTTGAA CTCTCTGTCT 3360

CCGCTACTAA GATTTTTCTT GTTCGATAAT TCGTTGGTTC TTGGCTTTCT GAAATAAGGA 3420

GAGCTTCACT AGTCAACGGA CATTTCGAAA GTAGTGCGTT CCCTTCTTCA TAAATAGCAT 3480

AACCAATGTG GCTCATTTCC CAACTCCAAT AATAGTGTTG GCCTTTTTCA GCTAGATATT 3540

GAACGATTAG ATATGCAAAA TTATCTTCAT GGATCGGCGT TTGGGTCGCT ATGGGACAAA 3600

AAGTAGTTAA CTGTTCTAAC GGCACTTTCT TGGAAGCGAC TTTCTGATTC ACTTCTTGTA 3660

AGGCAATTAT TTCACTTTCA GAAGATAAAA TCACCTTGGC AATCTCCTCT AATTTTTTTA 3720

ATGGTTCTTC TTCTAACCAG CTATGAGTAT TGATTGTTAA AAGATTCATC TTTTCCCCTT 3780

CTCTTACATG TTATTAGGTC TGTGAGTGGG ACAAAAATCA CTTTGGATTT TTGCTTCACG 3840

CGCAAAACCT GATAAACGGC GGGAACAGAA CCAACTcCTT CGGAAATAAG CCGAAATTAT 3900

CCAAAAATTA AAAAGCAATT TCCGGAAATT CCTTCTTATT TCTCGGAGTT AAACGGTTCT 3960

GTCCCGACCT CATACCTAAT CATTTAAATT TCAATCTGTC CGATCACAGT TCCCGCAGTG 4020

GTGGTTCCTA GTTGATTAAT AACTACTTGT TCCACTTTGT CGCTGTTTGT AAAGGCGACA 4080

ATAATCGTTG TTTCTTTGCC AGCTTGTTCA ATCACATCCA AATTCATTTT AGCTAGTTTC 4140

GTCCCAGCTT TCAATGATTG TCCTTCTTTA ACAAAAGATT CAAAAGCAGG TGTTGCCATT 4200

TCAACTGTAT CAATTCCCAT ATGGACAAGT ACTTCAATAC CTTCTTCTGT TTGTAGCCCA 4260

ATCGCATGTT TGCTTGGGAA AACACTTAGA ACTTTCCCCG AAATTrGTGC CACCACCTCT 4320

CCTTCTAGAG GCTTAACTGC AAATCCATCC CCCATCATTT TTTGTGAAAA GACAGGATCA 4380

TTGACTTGCG CCATTGGAAT AACTTCCCCT GTCGCTACAG CCACGAAATC TTTTTTCTTA 4440

CCTAAGAAAT TCGTTGTTGG CACTCCTGCT GTACTTTCTG CTATAACTGG TTCTGGAATG 4500

TCAACGCCGG AAGCTAGTAA ATCTTCAATA TCGGATTTTA AAACATCGGC TTTTGGTCCA 4560

TAAACGGCTT GAACGCCATT ATCTTTCACA ATCAAGCCCA TTGCGCCTGC CCGTTTCCAC 4620

GCCTCTTCCG AACCGACTTT TTCTCGATCC TTTACACTGA CACGTAAGCG TGTCATACAC 4680

GCATCAACTT CTTTGATATT TTGTTTACCA CCAAGTAAGT AGACAATTTG CGCGATTTGC 4740

TGATCGACGA CGCCTGAACC AGCCGCGCCA GATGCGATTT CTTCGCTGTT ATCATTGTCA 4800

TAGTTTCCAT TTCTTCCAGG TGTTGCGTAA TTAAATTTTT TAATTAAGAA ATTCGCTAAG 4860

AAATAAGTGA CAACACCAAA AATGATGACC ATTAAAACAA AATTGATTAA ATCGCCACCT 4920

AAACCAGCTT TTATTGCTAG CGGTGTTCTT GTTAACAATT CAATATTTCC AAACGAATGC 4980

ACACGCAATG GTAAAATATC AGCCATTGCA AAAGCAGCAC CTTGGATCAC CGCATAAATT 5040

ACATAAAGCG GCACTGCAGC AAACATAAAC ATGAATTCTA GCGGCTCTGT CACACCAGTT 5100

AAGAAAACAG tAAAGCAGCA GAAAAATACA TTGATTTATA TTTAGCTTTT TTATCTGCAT 5160

CGACATTACG ATACATCGCC AAAGCCATCC CCATTAAAAT CCCTGAAGAa CCAATCATTT 5220

GGCCTACTTT GAATCGAGCT GGTGTCCAAT TTTCTAAGAC AAACTGATAC TTGCTCATAT 5280

CACCAGCGCC TTTAAGATTC ACTAAGTCTG TTGCCCAAGC CAACCATAAA GGATCTTGAC 5340

CAAAGACTTG GGTTCCAGCT TGTGCGCCAG ATAAAATTTC ATACGTTCCA CCTAATTGTG 5400

TATAGTTGAT TGGAATCGTC AACATATGAT GTAACCCAAA AGGTAATAAC AAACGTTCCA 5460

ACGTTCCATA TAAGAATGGC GCTAAAATAG GTGCACTATC TTGAGATTGC GCAATCCATA 5520

ACCCAAAATT ATTAATTCCC GCTTGAATAT TTGGCCAAAT AAGTGCTAAT GCAATTGAGA 5580

CAATTGTTGA CCACAAAATA ACGACAAATG GGACAAACCG TTTACCATTA AAAAaTGATA 5640

GCGCATCTGG TAATTTCCGA TAATTGTAAT ACTTGTTGTA AGCCATCGCA CCGACAAAAC 5700

CAGCAATAAT CCCTACAAAA ACACCCATAT TTAATGCCGG AGCTTCTAAG ACACTTGTAA 5760

AGAATCCCTT CACCATAATC TTGGTGCCAA ATAAGGTATG CGTGAAAGCT TGTTCATCAG 5820

CCAACATTTC ATTTGTTACA CCAAAAATGG CTCCAGTAAT ACGGTTGATT AAAACAAAAG 5880

AAATCCCCGC TGCAAATGCG CCACCTGCAC GATCTTTGGC CCAACTGCCA CCTATgTAAC 5940

GCAAATAGTA AATGTAAGTT GCCAATAATT GCCCAACCGA TACTCTCCAG AACTCCGCCT 6000

GTTGTTACTA ACAAGCCTAA ATTAGGATCG ATTAAAGGAA GCGATTTCCC AATACTAATC 6060

ATTAAACCCG CTGCGGGCAT TACTGCCACA AC 6092

(2) INFORMATION FOR SEQ ID NO: 226:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1089 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

ATAAACTATC AGCATATTTA TTTAAAGTAC GAAGGCTTAG AAGGAGCCTA GGCGAAAGCT 60

GAAAATAATG CACCCTCTTA TGAGATATTT CAACAAACGT GCTTTTTGCA CGTTTGTTTT 120

TTTTTATCAA AATGAAAACG TAGAGGAGGG ATGGAATGGG TGAAGCTAAA GTAACTAAAA 180

CTTTCCGATT AACAGAATTA GTGGCAAGTG ATTTAAGTTG TAAAGATTTA TTTGATGAAC 240

TTATTGCCGT TTTAGAAATT TTAGATGTAG AAGTTGTTAA TAAAAAAATA GTTTATTTCT 300

ATAAAAGAAA CGTGAAAAGT AAACAAGAAA AATTAGAAGT TTGTTTTACT AGAGAGAGTA 360

TTCTGCACTT AATTGGGATA AGTTATTATG ATATTAATGA AAGAGAAACC TCTAAGTCAA 420

GAATGAAAAG TAAGTATGCT ATTGAATTTT ATAGAGACTT TAAAAGrAAT AAATTGAACT 480

TTTCAAAATG TTGGGTAGAA AGTACAAGAA AAGTAAAAGA TAAGCtcAGG TATTGAAATA 540

CATTAAAAGC ATAAAGACAG ATGTGGTCAG AATAGGCGCT GACGGACAAT TAAGAACAAT 600

TCCTATGACG AACACAATCA GTACGCCTAA GATTGGTTTA GGGATTGGCT TATATCATGA 660

TCATCCAGAA TTCTCAATTC CTCGTTCTTG TTTAAATCTT GCTCAGGATA AAGAAGCGAA 720

ACAACATACC TCGTTTCGTA ACGCCTGTCG GTGTACAAAA ATTTGGATAT ATGAAAGAAC 780

AGAACAGGGT ACTTGGAAAT TAGAAGATAG GCAAGAATTT TTTAAAAAAA TACAGGCAGA 840

AAAAAGTAAG AAAAAAAGAA GAAAAAAGTA AGATTACGTC ATTAACTTAG TTTGTAATTT 900

GCTATATATC GTCAGCTAAT GATGTCGTAG AATGAAAAAA GGGGTTCTCT TATACTAATT 960

TTTGTAAAGA ATTAAATTTA TTTTCTATTA CAAGTCCAGT ATAGAACTTA TTACTTACmT 1020

AACAGTACTT TTAGAAAAAt CAGAAcCAAA TTAAAGAGAG TAGCCCCTTT GnATTTAAAA 1080

TTTAATAAC 1089

(2) INFORMATION FOR SEQ ID NO: 227:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3567 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

AGGGTCTTTT ATAaTCAAGG AACAATGAGT GAGAACGATT CTCAATAAaT GTAAAATAAC 60

GACAGAAAAA TTAAAGAGGA ATGGTTATGA ATTATTaGAG GCAGAGGACT TAACTACAGG 120

ATATGGTCGT CAAGTGATTA GTGAAGCGCT GAACTGTTCA TTTGAAGCGG GAACTATTAC 180

```
AAGTATTATT GGACCAAATG GTTGTGGCAA ATCGACCTTG TTAAAAACAC TTGCACGTAT     240
TTTGCCAGCT GAGAAAGGCA AGGTGTTCAT TAAGAACCAA GAACTTCAAC ATTTTTCTTC     300
AAAAGAAGTA GCGCAAGAAC TAGCGATCTT ATCACAAACG CCAGAAAATA CAATCGATTT     360
ATCTGTTTTT GACCTCGTCT CTTTCGGGCG ATATCCTTAT CAATCGGGCT GGAAAAGTGT     420
CACCACAGAA GATCAAGAGT ACATACAATG GGCATTAGAA TTAACAGGCT TAACTGAGTT     480
AGCTAGAGAG TCTGTGGCAA TTTTATCAGG TGGACAACGG CAACGGGTTT GGATTGCTAT     540
GGCGCTAGTG CAGGATACAG ATATTTTAAT TTTAGATGAA CCGACCACCT ATTTGGATCC     600
AGCGCATCAG CTAGAAATTT TAAATGTGTT AAAACGAATT AATCAAGAAC ATCAGAAAAC     660
CATCGTGATG ACGATTCATG ATATTAATTT AGCTTCTCGT TTTTCTGATC GCGTGATTGC     720
GATGAAAGAG GGAAAGATTA TCAATGAGGG AACACCGCAA GAAGTGATTA CTGTTACTGG     780
CTTAGCAGAT GTTTTCGCTA TTCAAGCAGA AATTAGTTAT CATAGTCAGA CACAGCGGCC     840
ACTAGTGCTT TCTTATGAGT TGATGGAAAT GGAGCTGGAT GGATGAAAGG CATAAGCGGA     900
TTTTATAGAC ACCAATGGCA AGTGTTGCTT GTCTGTCTTT TTTTGCTACT GGGGACAATG     960
CTTCTGTCAT TGATGTTTGG TGCGACGTCG GTTTCGATTC AAACGATTAT TGATTCATTT    1020
ACGCACTTCG ATGCGACCAA TCAAGCGCAT CAAGTCATCC AAACGGTTCG GTTGCCGCGG    1080
ATTTTAGGGG CCGCTTTTGT GGGGGCAAGT TTAGCGGTGA GTGGGGCGTT ACGCAaGGGA    1140
TTACACAAAA TCCATTGGCA GATTCTGGCT TGCTGGGTAT CAATGCTGGG GCTGGCTTGG    1200
GCTTGGCGCT TGTTTTTGCT TTCCTCCCTC AGGCAAGTTA TTGGTGGCTT TTAGTTGTTT    1260
CTTTTCTTGG TGCCGGCGTG AGTGTGGCAC TGATTTAwTA TCTTTCTAAT CATTCAATTC    1320
GAGGAGCAAG CCCCATGCGT CTGACCTTAG TGGGAGCTGG TATTAGTGCC CTTTTTTTAT    1380
CATTTAGTCA ATTCTTAGCC ATTCAGTTCA ATTTGAGTCA AGAGTTAACG TTTTGGTTTC    1440
TTGGCGGTGT CAGCGTTATT TCTTGGGCAC AATTAAAAAT AGTCGTCCCT ATCTTTCTGG    1500
GCGCATTTGC ATTAGCAATT CTGATTAGTC CGTCGGTGAC TATTTTACGG TTTGGCGATG    1560
ATGCAACAAT TGGTTTGGGG CGAAATCCGC AGCGTATTCG GTTGTTTGCC AGTATAACGA    1620
TTTTGCTTTT ATCAGGATTG TCTGTGGCGC TAGTTGGCTC GGTCAGTTTT GTTGGTTTGG    1680
TGGTGCCTCA TGTGATGCGA ACAGTTTCTG GAGAAAATTA CCGGCGCTTG ATTCCGTTCT    1740
CGGCGTTAGG TGGTGCGCTC CTCGTTTTGG TGGCAGATTT AATCGCTCGA ATGGTGAATC    1800
CGCCTTTTGA AACACCATTT GGGATTATCA CGGCCTTAAT TGGCATCCCA TTCTTTCTTT    1860
ATTTATTTCG TAAAGGAGGG AAACTAGGAT GACACAGAAA AAAGTGTCTC GTTGGGTGGC    1920
TATCTTAATA GTTGGGATTA TCCTAACTGC TTGTATCAGT TTAACGCAAG GGGAATTATC    1980
TGTTTCCGTC ATGGATTTAA AAGCCCTTAT TTTGGGACAG GGAAGTCAGG CCAAACAACT    2040
GTTAGTTTGG GATTTTCGGT TACCACGGAT TGTGCTTGCG GTCCTAGCGG GCGCAGCTTT    2100
AGGCTTAAGC GGCGCTATTC TTCAAGGGAT TACTCGGAAT CCGTTAGCAG ATACTGGGAT    2160
```

TTTAGGTATT AATGCGGGTG CGGGTTTGGC AGTGATTTTG TTTATTAGTT TTGTTGAGAC 2220

GACAAGTATC ACAACGTTTG CACTTCCACT CGTGGCCTTA GTTGGCGGCT TGCTAGCGGC 2280

TGTTGGGATT TTGGTAATCG CGTATCATCA TCGTTTAGGA TTGACGCCAG TTCGAATGGT 2340

CTTAGCGGGT GTCATTTTGA GTACGGGTAT TTCAGCACTC ACCTTATTAA TTACATCAAA 2400

AATTGACACT GAAAAGTACC GTTTTGTTAC AATGTGGCAA GCAGGCAGTA TCTGGGGCAG 2460

TAATTGGTTC TTTGTGGGCG CTTTGTTGCC TTGGCTAGTC ATCGGGTTCT TTTTTGCGAT 2520

GAAGCGTCAT CGTCTGTTGG ATATTTTACA ATTGGGGGAC GAAACAGCTA TTTCTGTAGG 2580

TATTGCTGTG AAAAAAGAAC GAATCGTTTC GTTGCTGTTA GCGGTGATGT TGGCGGCATC 2640

AGCGATTTCC GTTACAGGGG GAATTAATTT TTTAGGTTTA CTTGCGCCAC ATATTGCTCG 2700

CAAATGTGGT TTTACGCAAC AAAAACAGGT ATTGCCATTG GCGGCGTTGT TTGGGAGTTT 2760

ACTTTTATTG GTTAGTGATA CGATTGGTCG GCTATTGCCT GGCAATGGAG AAATGCCTGC 2820

AGGGATTATT ATTGCGATCA TCGGTGCGCC TTATTTTCTT TATTTGTTAA TCAAAAACAAG 2880

CGATTCGTTA TAAAAAAGAA TCAAAAGCTG GGTCAATCGG TCGTTGTGAA AGCAACTGTC 2940

ATTTGCTCGG CTTTTTTATT GGCGGATTTT TGAAAAAATA GTATGATAAA TGCGTAAGAG 3000

GATTGAATGG AGGATGAAAG AATGGAAAAC TTTAATTTTT ATGTACCAAC AGACATCCGT 3060

TTTGGGAAAG ATCGTTTAGA CGAATTACCT GCTGTGTTAG ATCAATTTGG CAAAAATGTC 3120

TTACTTGTTT ATGGTGGCGG TAGTATCAAA CGTAATGGCT TGTACCAACA AATTACTGAT 3180

TTGGCCCAAA AAAACGGGCA TACACTGATT GAGTTGAGCG GTGTCGAACC AAACCCGCGC 3240

ATTGAAACCG TTAGAAAAGG AATTGAGTTG TGTAAAGAAA ACCAAGTCGA TGTTATTTTA 3300

GCAGTTGGCG GTGGTTCAAC CATTGATTGT TCCAAAGCAA TAGCTGCGGG CTTTTATTCG 3360

GAAGAAGATA TCTGGACAGT TATTGCAGCC AGAAAAGGGT ATACTGGTCC TGCTTTACCA 3420

ATCGTGACCA TCTTGACTTT AGCAGCGACA GGTAGTGAAA TGAACGCAGG TGCTGTCATT 3480

TCTAACTTAG AArCCAATCA ArAATTAGGC TTTGGTGGTC CGAATATGAT TCCAAAAGTT 3540

TCGTyCTGGG ATCCAACGAA TACTTTT 3567

(2) INFORMATION FOR SEQ ID NO: 228:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7893 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

TTTAGTGGTT TGTTGATGTT TTTAAGTAAA ATTGGCACAT TGTTTGGCTG GCTTTCACTT 60

CTGGCAGCTG TTATTGGCTT TGTTATTTTT TATAAACGAG CAACGACCGC TGAACAGCCA 120

CTAGTTCGTT TGACAATTTT AAAAAATCCC GCCTTCGTTT TATTTTTATG TGGCTTTTTA 180

GTTTGTCAAT TCTTATTGTT AGGTATTTCC TTTGTCTTAC CAAATTTTGT TCAAATCGTC 240

CTTGGAAAAA ATGCGTTTGT AGCTGGGCTA GTCATGCTGC CAGGTGCTAC AGTTGGGGCC 300

ATTTTGGCAC CACTTTCTGG ACGAGTGTTG GaTCAATACG GCGCTAAAAA ACCTATATTA 360

TTTGGCTTAA GTTTGGCGAC AATCGGTTGG CTAGCTTTAA CAATTTTACT AGAGATGCCT 420

GTTTTATTAG GATTCGTCGC TGGACATGTT ACCTATATGA TTGGTCTAGG TTTTGCATAC 480

AGTAATATGA TGACGACTGG TAtGAGCTTG TTAGAGGAGA AAGATTTTGG CGACGGCAAT 540

ACGTTATTTA ATACACTTCA ACAATTTTCT GGCGCTATTG CGACGGCCAT TGTAGCGACG 600

ATTATTAATA TCGCTCAAGA CCACGCAGAT AATTTTGCTC AGGGGACAAC GATGGGCTCC 660

CTCATTTCTC TGATTTTCTT GTTAACCTTA TTAGTAATTG TTTTAATTGC TTGCTGGAAT 720

TATTTCCGCA AAAAAGCTTA AAAAAAAGCA GCTGGTCAGT TCCATAGTGA CCAGCTGCTT 780

TTTTTAGTTT TCTAAAGATA GTTGTTTTTT TGTTCTTCTT GCTTCCGAAT ATAAAGTACG 840

AACAAGACAG CAATTAGCAC GATAAAACCA CCAAAAATCG GTGTATATGT TACTGATAAA 900

TGATCTGTGA CCACCCCACC TGTCATTGAG CCGACGGTAA TCCCAATATT AAAGGCTGCA 960

ATATTAAAAG CTGAGGCCAA AGTGATATCT TCTGGAACAT ACTTTTCCGC TAATTCTACA 1020

ACATATAGTT GTAAACCAGG TACATTCATA AAGGCAAATA GGCCCAACAT CAAACTTGCT 1080

AATAAGCCTA GAAGACTGTT ATCCATTAAA ACAGTAATGA ATAAAAAGAC TAAGGACAGA 1140

ATTAATAAGC TAAACATTTT TACTAACGAA TCTAAAGGTT TTTTATTTGC CCAGTGACCA 1200

CCTACTGTAT TGCCAaTTGC GACCATTACG CCaTAAaCAA CTAAAATAAT CACTACCGCG 1260

TTGGCTGaAA aGCCTaAtTT TCCmwTAACG GTGACAgTAC GTATATGCAG CAAACGTTCC 1320

ACCATAACCA AAAGCAGTAA TTAGAAAAGA GAAAACGAGG GGTTTGTTCG TAAAAATCCG 1380

CGTGATTCCT TTTATGTCAA CCTTCCCTGG TATCGGTAAA TTTTTGGGAA CTAAAAAATA 1440

ACTAGCAATC AGTCCCACCA AGCCAATACT GGCTATAAAA AGAAAGGACA TATGCCAAGC 1500

TGTTTGTTGC CCAATAAACG TCCCTAAAGG TACGCCAGTC ACTGTTGCAa CCGTTAAGCC 1560

CGTAAaCATA ATAGCAATGG CGCTGGCACG ACGTGATGGC GGCACAACAT CTGCGGCGAT 1620

GACAGTCGAA ACAGACATGA AAATGCCATG CGCCAATGCA GAAAGTATCC GCCCTACCAA 1680

TAAAATAATA AACGTTGGCG CAAaCGCAGC AATTAAATTC CCACTAATGA AAGAAGCAT 1740

AATCCCCAGC ATTAATGAGC GACGATTCCA TTTTCCTGTC AAAATTGTTA ATAACGGCGC 1800

CCCAATCGTC ACACCTAAAG CGTACAAAGA AACAGTCAAA CCTGCTTGGG AAAGACTAAT 1860

ATGAAAACTT TTAACAATCA TTGGTAATAA ACCCACACTA ATAAATTCTG TTGAACCGAT 1920

TGCGAAAGCA TTGATAGCTA ATGCAAAGAG CGTTAAATTT GGTGAAATTT TCTTTTCCAT 1980

TCTCTTCCTC CTAAAAAACA TTTGAATTTC AACAAAAATT ACTATATACT AAGATTGAAA 2040

TTTTGATAAG TACCTACTTT TTTGTATCAT AGGTACCTTT TAGTAACTTT TGGTTATTAT 2100

CAACCTATTA AGGAGGAATG TTTTTTGGAA ACTCGTACAT ATGCGATTGG TGTCGAAGCA 2160

ACGATGGATG TTATTGGCGG TAAATGGAAA CCAATTATTC TGTGTAATTT ACGGCACGGC 2220

GCCATGCGCC CCAGCGATTT AAAACGCGGT ATTACTGGCA TTAGTCAAAA AATGCTGACA 2280

CAACAATTAC GGGAACTCGA AGAAGACAAC ATTATCGAGC GGAAAGTTTA CAACCAAGTT 2340

CCACCAAAAG TCGAATATTA TTTAAGCGAT TATGGTCAAA GCCTCTCTGC TGTCTTAGAT 2400

AACTTATGTA ATTGGGGAGA ATCCCACGTG GCTCATTTGC AAGAACAAGG CCAAGAAATT 2460

TATTTAAAAT GTGAAGATTA GTGAACAAAA CGTTTATAAA CTATATTAAG AGGGAGCTTT 2520

AAGCTATTTA AAGCCTGGGG AAAAATCATT TTAGATTTTT CCCCAGGCTC TTCTTATGCT 2580

TTAATAACTT GAATATCTTG CGTTAATTGG GCCCAATGAT TAGTAGGTCC GTGACCATGT 2640

CCAACAGAAA TTGCTTGGCT AATGGcCCCT TGTATAAACT GTTTGCCAAT CACAATCGCC 2700

GCTTCTAACG GCGTACCTTT TGCTAATTCG GCTGCGATAC ACGCTGAAAA AGTATCTCCT 2760

GTTCCATGAG TATTTTTCGT TACAATTCGT GGTGCACTTA GCCAAAAGCT CGAACCGTCT 2820

TCCATTAACA CATAATCAGC CGCTTGTTGA TTTGTGCTAT GTCCGCCTTT CATAATAATA 2880

TTTTTGTGC CCAATTGCTG TAATTTTTTT GCTGCGGTCT GCATTTCTTG TTCTGTCTTG 2940

ATGGTCGTCT CTAGTAAAAC TTCCGCTTCT GGTAAATTAG GTGTCACGAC CGTTGCTAAT 3000

GGCAACAGTT GCTCTTTAAT CGTTTGAACT GCCTCGGCTT CTAACAAATG ATGTCCACCT 3060

TTAGCCACCA TTACAGGATC GACAATCAAC GGACCAAAAT CAACTTTCTG CAAATTGCGA 3120

ACAACAGCTT CGACACGCTC ACTATCTGCT AACATCCCCG TTTTCGCTGC CCCAATTTTA 3180

AAATCAGCCG CCAGTGATTG GAATTGTGCA TCAATAAATG AACTAGGAAT CGGTAAGCTA 3240

TCTTGCACGC CATACGTATT TTGTGCCGTC AATGCAATGA CAATATTTAA GCCAAATGCC 3300

TGCCGTGCTT GAAATGTTTT TAAATCTGCT TGCATGCCAG CTCCCCCACC AGAATCAGAA 3360

CCAGCAATCG TGACTACTTG TGGGGTCAAA TTATATTCGG TCATTTTCTC GCCTCCAACA 3420

TTCGCTCCGT TACACGCATC AACCCTTGAA CTTTCTCGGT TATCTGTTCA GCCAGCATAA 3480

TTTCACTGAC CAGTGAAACA CCAGCCACTC CCGTTCCCAT AAGTTGTTCT ATATTTTCTT 3540

CTTTGATTCC ACCAATCGCT ACAACAGGAA TCGTGACTGC TGCGGCAATT TCAGAAAGTG 3600

TCTGCAAAGA AGTTAACGGA CTGTCCTTAG TCGTTGTGGG GAAAATTGCG CCCACGCCTA 3660

AATAATCGGC GCCCTCGTTT TCCGCTTCAA CTCCTCGTGC CACCGTTTTG GCTGAAACAC 3720

CTACGATTTT TGTGGACCCA ACCAGTTTTC GCACAAGAGC AACCGGTAAT TCATCATCCC 3780

CTATATGCAC ACCTGCTGCA TCCACTGCCA AGCAAATATC GACACGATCA TTAATAATCA 3840

ACGGAATCTG ATACGCATCT GTTACCACTT TAACTTTCAC TGCTAATTCG TAAAAGCGCC 3900

GTGTTGACAC TTCCTTTTCG CGCAGTTGGA CCAATGTCAC TCCACTGCGA CAAGCAGTTT 3960

CAATTCGTTT TAAAAATTCA GTATCACTAA AATCATACCG GCCTGTCACC AAATAAACTT 4020

TCAATTTCTC TCTCATAGCA CTCGTCCTTT CACCGCCTCA AACCAATCCT TCTCCTTCAT 4080

TAAAAGAGAT AACTGATTTA ACGTATTTTG GCGAAAATCG GcCAGCCCTT GGCTCTTAGT 4140

CTTGGCTTTT TCTCCGCAAA GATTAAAATA GCTCACTGCG GCTACTGCCG CAGTcATTGG 4200

CGCGTTGCCT TCACCTAATA AAGCGGCAAC TAATGCGCCG ACTAAATCGC CCGTGCCAGT 4260

GAAACAATCC AACTCAGGAA CGCCATTTTG TAAGACAATC ACTTGCTCTT GGCTAACCAA 4320

AACATCTTGA ATCCCTGTTG CTAAAAATAC CGTCTGGGGA AACTTTTGCG TTTGTTGACG 4380

TAAAGCCTGA ATCAGTTCCT CAATTGCTTC TTCACTTTGA TCCAAAGGGC TGCCATCCAC 4440

TCCTCTGCCA TGACTCACCA ACTGGCAAAA AGTCCGCATT TCTGAAAGAT TTCCTTTCAC 4500

AACGGTTGGC TGATTGTGCA CCAGCTTTTC CCCGACTTCA TTTCGAATAT CACTGGCGCC 4560

ATAACCGACT AAATCGACTA CGGTCAGTTT ATTCACTTGC CGAGCATAAT CACTAGCTGC 4620

TAACAAGCTT TGCTCTCGTT CTTGTGATAA ATGGCCTAAA TTCAAAACCA AAGCACTGGT 4680

TTGCTGAAAC ATCTGCGGAA ATTCCCGGGG ATCATCTGCC ATAATGGGTT TAGCATCAAT 4740

ATAAAGTAAT GCATTCGCCA TAGACTCACA CGTAATTTCA TTCGTAATGC ACTGAATTAA 4800

TGGTGCCGTT GTCAGTGGAA AAATCGTTTC AAACTTAACG CTTGTTTTCA TAAAAGTACC 4860

CTCCTAAAAA TAATTCCTGT GTTTGTTTAA ATACCCGAGT TTCTTTCAGT TTTACTAAGA 4920

CAAGCCAAGC AATCACTGAA CCAATCAGTG TGGCACCAAT AAATCGTGGT GTATAAATGA 4980

ACCAATAAAA ATTCTGCTTA CTGCCAGTAA ACCAAACCAT TACAGGATAG GACAATAACG 5040

AGCCAATGAT GCCTGTCCCA AAAATTTCCC CTAGCATCGA CCAGCCTAAT TTTTGTCCCC 5100

ATTGATAAAA TAAACCAGCA AAAAACGCGC CAAAAACGGC ACCTGTTAAC GCTAATGGCG 5160

GAATACCTAA GAATGTCATA CGGATTAAGC CGCAGACAAG TGCCATGAGC GTTCCGTAAA 5220

GAGGACCTAA TAAAACTCCG GCAATAATGT TCATGACACT GGACATAGGG GCCATTCCCT 5280

CAATTCGCAT TAACGGCGAT AATACAACGT CCATCGCAAT CATTAAAGCC AGTAACGTAA 5340

CTCTGTGGAT TCGATTTTTT TCCACGTAAA AAAACTCCTT TCTGCGAGAA GCAAAAGGAG 5400

GCGCGACTGA AAAAGACAGA AAAACCCAAC CACAAATAAA CGTAGTTGGG TGCAATAATC 5460

GGTCAAAATT TCAAGACATT TCCCTGCGCC AGTCCTAACT GTTTCAGGTT CCATGGGTAT 5520

AATCTCAGCT AATAGCACCC CAAATGTTTC TCTTCAATTG GCTTTAGTAT AGCATAGTTT 5580

TTTGGAAGCG ACTACTAAAA TTGCCAAAGA GTTTTCCCAC AAAAACAGAA AAAACATCAT 5640

GCCCAAAAAG TTCGATTTTT TGGGCATGAC GTTTTTCATA CTTTCGTTTG TCTTATTTTT 5700

TAGAATCCTG GCATTCCTTT GGTGTATTTG CCTAAAGTTG CTTCGGTTTC TTTTTCAATT 5760

TTTACTAACG CGTCGTTGAC AGCCATAATG ACTAAATCTT GTAGCATATC AACGTCTTCA 5820

GGATCAACCA CGTCTTCTTT AATGATCAAA TCTTTCATTG TGCGATCACC AGTAAATGTC 5880

GCTGTCACAA GTTGATTGGT GGCTTCACCG ATAAATTCTT TTTCATTTAG TGCTTCTTGC 5940

GCTTTGGCCA TTTCTTTTTG CATTTTTTGT ACTTGTTTCA TCATGCCTTG CATGTTGCCC 6000

ATTCCTCGCA TCATAATTCC TCGACTCCTT CATATTAATC ATCTAGTACT TCGACAAGCT 6060

CTTCACCAAA CATAGCGATT GCTTCATCTA CTACTGGATT TTCTTGTGAA TGTTCaTTTG 6120

TTAcAGGCGG TTCGTCCGCT AGACGAGCCa TTTCATTTTC aGGTTCAGAA TGATTCAAAC 6180

TGCCCTGATT TTGGTTAATA AAACTTTGTC TTAGTTTTGG CCAGCTTTCT CTAGTAATGC 6240

AGACCATCTC TGGTGTATAG TCCATCAAAC GGCTTAAATT ATTATTGAAA GCTAACTGCA 6300

TCTCTTCATC GTCTGTTGCA CGAGCACAAA CAATCTCATA GTCAAATGCC ACCACGATGC 6360

CCTTAGGACT CGCTGCGACA GGTTCACTAG CCTTGAGCAT CGCTCGTTGA GTCACAGATA 6420

ATGTTTGCAG TAAGTCTTCC CAGACATTTT TCACATTCAT TAAATGTGTT CTAGTCGCTT 6480

CGTTTAGGAC TTGATATACG CGTTCTGTGG GTACACGGAA CGAACTTTTT GGCGCTTGTG 6540

GTCGTGCTTG TTGACGTGGC GCATCTGCTT CTTTGGCAGC TACACCGTGT TTCTTTAATT 6600

CAGCCAGTTC TTTCTTCAGT TGCCCGATTT GATTTTGTAA ATCTGCTAAT TCTGGATTTC 6660

CCTCTGCACT ACCATCTTGA TTAGCTGTTT CTGGCGTGTT GTGTTTATTG GGTTGCACCG 6720

TTTTAGCTAA TTTGACAGTG GCTACTTCCA GATAGATATT CGCATTGTTG GTAAAGCGGA 6780

TCTCATTTTG TGTGTCACTT AAAATTTGGA TTAGTTGGTA GATTTTTTCA GCAGGCGTTT 6840

GTGTAGCTAG TTCCTTAAAT GCTTCTGTCA AAGTTCCTGC TTTTTCTGCT AATAACTTCG 6900

GTGCTTGCTG ATACATCAGC AAATCTCGAC AATATAATAA CAAATCTTCT AAGAAACGGC 6960

GCGCTTCTTT GCCTTCGCCT AAAATACTTT CCAAGCCTTC TAATGCTCGT TCAACATCAC 7020

CGGCGACACA GCACTGAATA TAATGATCCA TCATTTCATA GGTTAGGCTG CCTGTCACTT 7080

GCATTGCATC TTCCAGTGTT ACTTTTTCAT CACTAAAGGA AATCGTTTGA TCCAAGATAC 7140

TCAAGGCATC TCGCATCCCA CCTTCCGCAG CACGCCCAAT CACATAAAGG GCTTGTTCCT 7200

CATAATCGAG AGCCATTTCC TGCATGATAT GGGCCATATG ATCGACGATA TCTTGCGTAC 7260

TAATTCGCTT AAAATCAAAG CGTTGCGTCC GTGAGATAAT CGTTAACGGA ATCTTGTGTG 7320

GTTCAGTCGT TGCTAAGATA AAAATAACAT TTTGTGGCGG TTCTTCCAAG GTTTTTAAAA 7380

GTGCATTGAA CGCGCCTGTT GATAGCATAT GAACTTCATC AATAATATAA ACTTTGTACT 7440

CGGCTTGTGT CGGTGCATAC TTTGCTTTAT CACGAATATC ACGGATTTCT TCCACGCCAT 7500

TATTACTCGC CGCATCAATT TCAATGACAT CATTTAAACG ACCTTCCGTA ATGGCTACAC 7560

AGGTTTCACA AACATTACAA GGTTCACCAT CTTGACTATG TTTACAGTTA ATCGCTTTAG 7620

CGAAAATTTT TGCGGCACTG GTTTTACCTG TACCACGAGG GCCAGTAAAT AAATAGGCAT 7680

GAGATGTTTT CTTTTGCACT ATCGCATTTT TCAGTGTCTG AGTAATTGCT TTTTGTCCTA 7740

tACGTCATCG AAACGTTGTG AACGCCAGAC CCGATAAAGT GCTTGATAAG CCATTTTTTA 7800

CCCTCCCTTA TTTTTCATTC TTATCTATTA TACGTGATTT TnAGTCATTC GGCATCTTTT 7860

TTCTTTTAAT GAGTCACTGG CGCAACCATT GCT 7893

(2) INFORMATION FOR SEQ ID NO: 229:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 17082 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: double
    (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

GCGCAAATTG GCTACCAATC TAAATnACTT TCAGAAGCGT ATTATACGAG CATGATTGTG 60

GTAATTATCT TAACGACCTT AGTTGCACCG TTCTTATTGA AATACTTTGT TAAAAAACAA 120

GAATCTGCGT TGCAATAAAA AAGATGAGCA GAAGTTGTTG ACAAAAAAAT AAAATAGCTA 180

TATGATGAGT TCAAGATAAA CGACGGCATC AGGACCCGTG GCCCCATGCG TTTTAATGAC 240

TGGCGCTTTT AAGCCAAATA AAACCGCTCC CCCATGTTTT GAGTAGTCCA TTTCGTCTTT 300

CATGCCATGC AACGCATTCT TTAATAATAA GGCGCCCATT TTTCCTTTGA CACCTTCAGA 360

AAGAATCGCT GTTTTCAGCA AGCTCATCAT GTTCATAGCT GTGCCTTCGr tGGATTTTAA 420

AACGGCATTT CCTGTGAAGC CATCTGTCAC CACCACGTCA GCTACGCCAT TTAATAATTC 480

TCGTGCTTCA ACGTTTCCGA CAAAATTAAT CGTTTCATCT GCTGCTAACA ATTCGAAGGC 540

TTTTTTCGTT AATTCACTGC CTTTTGTTTC CTCAGTTCCG TTATTTAATA GACCGACACG 600

AGGATTCTGA ACGTTCCGAA CTTTTTCAGC ATAGAATGAA CCTAAAACGG CATACTGAAC 660

CAAATGTTCT GGTTTATTGT CTGCATTGGC TCCTAAATCT AACATGTCAA ACCCTTTGTC 720

TGGTTCTCCC ATGACAGGTA AAGTTGACAT CAAACCGGGA CGTTCCACAT TTTTGATGCG 780

GCCGACAATG AACAGACCCG CTGCTAATAG CGCTCCGGTG TTGCCTGCAG AAAAAATCGC 840

ATCGGCTTCC CCATTTTTAA CTGCTTGCGC TGCTAAAACC ATGGAAGCAG TTTTCTTACG 900

ACGGATTGCT TTAACTGGCT CATCGTCACT AGCAATTTTT TCATCGGTAT GAATAATCGT 960

AATGTTTTTT TCATCGGTGA TGTATTTTTT TATTTCGGCT TCTTTTCCAT ATAATTGAAA 1020

TTCAATGTCT GGGAAGTCTT GTTTGGCTAG CATAACCCCT TCAACAATTG CTTGTGGGGC 1080

ATTATCTCCA CCCATTGCAT CTACAGCAAT TTTCATGAGT GATTCCCTCC TACTTTTTTG 1140

TTTCCTAACA AGTATAACAT ATTTTTTTCA AAGAAATACG CATTGGCTTT TACTTTCTGG 1200

CTTCAAAAAT TACCTCTTCT AATCGAAAAA TTGGGCTGCC TCATCTTGGG GTTTCACTCG 1260

ATTGGCTAAG CCTTGATAAG CAGGATATTG CCACCACTCT TTCACTTTCC ATAGGGCACT 1320

TGCTTCTTGG CGCGCAACTT CTAAAATGTT AAAGTCCGTC ACGATATCTC CTACCGCAAA 1380

TTGAGGCACC CCTGATTGCC GTGCACCGAA CACTTCCCCA GGTCCGCGAA GTTCCAAATC 1440

ACGCTCACTC AAGACAAAAC CATTGGTGGT CTCTGTCATG ATTTTCATAC GCTCAACGCC 1500

CATTTCATTT TTCGGATTGG CCACTAAAAT ACAATAAGAT GCACTGGAGC CACGCCCAAC 1560

ACGGCCCCGT AACTGATGTA ACTGTGCTAA CCCAAACCGG TCCGCATCCA TAATTAACAT 1620

CACCGTGGcA TTTGGTACAT TCACGCCAAC TTCAATAACG GTGGTAGAAA CCAATAATTG 1680

CAGCTGATTG TCTTTGAACT CCTGCATAAT GTCATCTTTT TCTTGGtTTT TCATTTTACC 1740

ATGTAGTAAT CCCACTTGAT ATCTTGGGGA ATAAAAGCTT TGCATATGTT CAAAAATTTC 1800

TGTTGCGTTT TTCACATCTA AGGCTTCTGA TTCTTCAATC AATGGACAAA TAATATAGGC 1860

TTGATGTCCT CGAGCCAGCT CTTTTTCCAT CCATTCTAAA ACAGTATCTA ACTGCGGCGG 1920

CCGAATCCAA CGTGTTTCGA TTGGAATCCG TCCAGCTGGC ATTTCATCAA TGATTGATAC 1980

ATCCATTTCG CCATAGGCTG TAATAGCCAG CGTTCGAGGA ATCGGAGTGG CCGTCATAAA 2040

TAGAACATCT GGTTTTAAGC CTTTTTCTCT TAATATTTTT CTCTGATTCA CACCAAAACG 2100

ATGTTGCTCA TCGGTGATTA CCAAACCTAA TTGATGGAAA CTGACATCTT GTTGTATCAA 2160

AGCATGGGTC CCTACCACAA TGTCAATCTC GCCATTGGCC AGTTCTTCGA GAATTAGACG 2220

ACGTTCTTTC GTTTTGGTGG AACCTGTTAA TAAGGCCGTT CGAACTTCTA ACGGATCAAA 2280

CAACTGCTGT AAGCTTTCCA TGTGTTGcTG CGCTAAAATT TCAGTCGGAA CCATCAACGC 2340

TCCTTGAAAG CCAGCGGTCA TGGTGGCGTA TAGCGCAATC GCAGCCACTA CCGTTTTCCC 2400

ACTCCCCACG TCACCTTGAA GCAATCGTTG CATATGCTTC GGACTTCGCA AGTCTCGACA 2460

AATTTCATTG GTCACTTTTT TTTGGGCGCC TGTCAATTCA AAAGGCAGTC CTTGCGTGAA 2520

CGTTTTTAAG CGATCCACGT CATATTGAAT AGCTAAACCA TTTTTTTCCG CTTTTTCTTG 2580

TTTCTTCAAT CCTTGCATTT TAAGTTGGAA AAGGAAAAAC TCTTCAAAGA CTACTCGTCG 2640

TTTAGCTTGA TGGCTTTCTT CTGGATTACT GGGAAAATGC ATCGCCCACA TTGCTTCTTT 2700

TCTAGGCATC AAGCGGTACT TTTCTAATAA ATCATTCGGT AAAATCTCTT CAACCAGTGA 2760

ACCATACTCT TCAAAAGCAG TGCGGATTAA CTGAACCAAC GTGCTTTGGC GGACTTTTTT 2820

ATTCACATGA TAAATGGGGG CGAAATCTTC ATTGTCTCCT TTAGAAGCTA AGATTTTCAT 2880

GCCATTTAAT GATTTTCTCT TGGCATCCCA TTTGCCATAA ACAGCAATTT CTTCGGAAAG 2940

AACTACTTTG TCTTTTAAAA AAGGTTGGTT GAAGAAGGAA ACATTAATGA CCGCATGTTC 3000

TTGCATCATT CGAAACGTTA ACCGGCTCTT CTTATAACCG TACCGACTAA CGACTGCTTC 3060

CGAAACCACT AGCCCTTTTA GCGTAACTTT TTCCTGATCC TGAATCTCAC TTAAATCCTT 3120

TTCTTGGATA TCATCGTAGC GGAAGGGATA ATACGTCAAT AAGTCTTCAA TTGTAGCAAT 3180

CCCCAGTTCT TGTAAATTTT CAGCACGTTT TGGGCCAACA CCGGGCAATA CGCTGACTAA 3240

ATCTGTGATG TCCATAAAAT CCCCTCCTCA TTAATATACG CAAAAAAATG GATCATTCAT 3300

TAAATAGGCT GGTTCATTAG CCCTAACTAA TGAACCAGCC TATTCGTTTT TATTTTTTCA 3360

TCACGAAAAC CTATTCTGCT GAGAATAAGT ATGGATACAC AGGTTGTCCA CCTTCATGAA 3420

GTTCTGTTTC TAGTTCATCG TTTTCAGAAA CTAAAATAGC GATTAGCTTT TCAGCTTCTT 3480

CAACCGAACC ATCTTCGCCG ACAATAATTG TCACGATTTC AGTGTCTTCA TCGATCATGC 3540

GGTTCAACGT GTCTAATGAT GCATTAAAAC GATCTGCTTC AGAAACGACG ATTTTACCAT 3600

CAATCATTCC TAAGAAGTCA TCTTTTTTAA TTTCAACGTC ATCAATGGTT GTGTCACGAA 3660

CAGCTGTGGT TACTTGACCA CTCACAACGC TGTCTAACAT TTCCGTCATA CTTGCTTTGT 3720

TTTCTTCTAA TGATTGTTGA TCATTGAAAG CTAACATTGC TGTCATTCCT TGAGAAATGG 3780

TTTTAGATGG AACAACCGCT ACTGGTACAT CGGCAACTTC TGCTGCTTGA TCAGCTGCCA 3840

```
TAAAGATGTT TTTGTTGTTT GGTAAAATAA TGACTTGCTC TGCATTCACT TCTTTGACTG     3900
CTTTTAAAAT ATCTTCCGTA CTTGGGTTCA TTGTTTGGCC ACCACTAATA ATGTAGCTTG     3960
CACCTAAACT TCGGAATAAT TCTTGAACGC CTTCACCGGC TGCAATAGCA ATCACAGCAA     4020
ACGGTACGCG AACTTTTTCA GTCAACGCTT CTTGCTGAGC ATCGCGTTCA ACTAACGTTT     4080
CATGTTGTAA ACGCATATTG TCTACTTTTA TTTTTACTAA CGAACCAAAT TTTTGGCCGT     4140
AGTTCATCAC TTCACCTGGA TGCTCAGTGT GCACGTGAAC TTTGATAATT TCGTCATCGT     4200
TAACGACTAA TAGAGAATCG CCTAGTTCAT TTAAGTAGTT TCTGAATGTT TCATAATCAA     4260
AAGCGCTGTC CACGGTTGGG CCTTCGCCGA TACGAACCAT AATTTCTGTA CAATAGCCAA     4320
ATTTAATGTC TTCTGTCGCA ACGTGACCAC TAACACCACG ATGATGTTCA GCATTAACCA     4380
TTTCGTCCAT TTCGGCAGGA CTTGGTTCAT ACGTTTCAGT TGGTAAAAAT TCACCACTAA     4440
GCGCTTCTAA AAATCCTTCA TAGATAAATA GTAGTCCTTG ACCCCCACTA TCTACAACGC     4500
CGACTTCTTT TAGGACAGGT AATAAATCAG GTGTTTTCGC TAAGGCACGT TTAGCACCTG     4560
CTACAACAGC TGTCATAACT TCGATACAGT CATCTGTTTC TTTCGCTTTa CGTTCaCCAG     4620
AacGCGCCGC TTCACGAGAA ACAGTTAAGA TTGTTCCtTC aACGGGTTTC aTCaCTGCTT     4680
TATACGCTGT TTCTACACCA TGTGTAAaGG CAGCGGCCAA GTCTTTGGCA TTTAATGTTG     4740
TAACCTCAGG AATTTGTTTT GAGAAGCCtC GGAATAATTG AGATAAAATA ACCCCAGAGT     4800
TTCCACGAGC ACCCATTAAT AACCCTTTGG ATAACGCACC TGCTAATTCT CCTACTTTTT     4860
CAGAGCGAGA ATCGGCCACA GCCTTTGCAC CACTGGTCAT TGATAAGTTC ATATTTGTTC     4920
CTGTATCACC ATCGGGTACA GGGAAAACGT TTAATGAGTT GACATATTCT GCATTTGTGT     4980
GCAGACGAGT CGCACCTGCC TGGACCATTT CCTGGAACTG ACCTGCGCTG ATTTCTGTTA     5040
CATTCACCTA AAAAATCCTC CTTCAGATTT ACCGCTAACG GCAACACACG ACTTCCTTGT     5100
GCAATTGCAT CGTTGACAGT ACACTACCTC TAATCCGGTA ATACACGAAC TCCTTGAACA     5160
AAAACGTTAA CTGAATTCGC AGTGACACCT AATAATGTTT CAAGATTATA TTTTACTTTT     5220
TCTTGGACGT TACGAGAAAC TTCAGAAATT TTTGTACCAT AACTAACAAT CGTGTAGACA     5280
TCCACTGCGA TTCCGTTGTC TTCTTGACGA ACAACGACAC CACGTGAGTA ATTCTCACGA     5340
CGTAAAATAT CGTTTAAGTT ATCTTTAATT TGATTTTTGC TTGCCATTCC GACAATTCCG     5400
TAAATATCTG TCGCAGCTCC GCCAACTACT GTAGCGATCA CATCATTGGT AATCTCAATG     5460
GTACCTGCGG GTGTTTGAAT TTTTACAGCC ATTGATGAAA GCCTCCTTAT AAGAGCACAT     5520
TCGCCCTATT TnCTTCCATA ACATTTTATC ATATCCAGCC TTTAATTAAA AGGAAGTtCG     5580
TCAGATATkT TCATTCTTCC ACTTTTAAGT GAACATTCTA GCAAGATTTC AATTTACTCT     5640
TTTTGATTGC TAAAGTCAAA TCATAGAAAA AAATCTTCTT GCAAATCAGT GCCTAATATG     5700
ATAAATTAAG TTAGTATGAG CTGAATATTA TCGCTCCAAG ATCAAGGCAA AGGAGGAAAC     5760
TTAAAATGGC AAAAGTATGT TACTTCACTG GTCGTAAGAC TTCAAGTGGT AATAACCGTT     5820
```

CACATGCGAT GAACTCAACA AAACGCACTG TGAAACCTAA CTTACAAAAA GTTCGCGTAT 5880

TAATCGATGG TAAACCTAAA AAAGTTTGGG TGTCAACTCG TGCTTTGAAA TCTGGTAAAA 5940

TCGAACGTGT TTAATAAACT TAAATAAGCA AAGCTGGACG TTCGTTGTTC GGCTTTTTTT 6000

ATTGCATAAA AAAAGAGGCT GGACCTTCAT CGCATTGATG AAGGTCCAAC CTCTTTTTAC 6060

TTGTCTTTAC TTTGAATAAC CGCAATCATC CCTGCATCGA AGGAAAAAGC AGCGGCTTCT 6120

TCGATAAATT CATTACTTGC ATACGATGTT GGGTAAGGAA CATCTTGATT GGTTAATAAA 6180

TATTTACTGC GTCGTAGCGT CAAGTTTTCT ACAGGCGTCA AACAGCAGTA GGCCAGATAT 6240

TCTTTATCTG GCTCTTTTGG TACGATATAT GAACCAGGGG CGTAATACTG GATGCTATTT 6300

TGACGATCGC ATAGACGGAT TTGACGAAGT ACGCCTTGAA ATCGTGGTTC AAAGGGTAAC 6360

CAAAGATTGG CTAACAGGTG GTCAATTCTG CCCCCTGTCG CACCGATAAT TGTCATCTCT 6420

GCTTGTGGAA ATCGTTGTAA AGCTTCTtGA AGCGCTAATT GTGTATCTGt ATCATCTTTT 6480

TCAGCAGGTG CTTGAATGAG TGTTTCGGTA GTTTCTTGGA CAAAATGGTA TTCTTCTCGT 6540

GATAATGAAT CAAAATCGCC CACCGCTAAT TGTAACGGTA AATCGGCCTC TAATAAATGG 6600

AGACAGCCGC GATCAATACC TACGAAATAA TCATACGTAG CTGGTTCAAT GGTTGGCCAA 6660

TCACTTGGAT TGCCGCCAGC TACCAGTAAG ACACGACTCA TTTGTCTAGC ACTGCTCTCA 6720

GTGCATCGAT TCGTTCTTGA GGATTTTCAG CATTATAAAT GTAGGAGCCT GCAACGAACA 6780

CTTCCGCACC GGCTTCTTTG CAGCGTGCGG CAGTTTCAGG AACAATCCCG CCATCGACTT 6840

CAATTTCATA GGTGTATCCT TTTGTTTCTT TCCATTCTTT TAATTGAGCA ATTTTTTCTA 6900

GTGAGTTTTC AATAAATGAT TGTCCGCCAA AGCCTGGATT GACAGTCATT ACCAAGACTT 6960

GGTCTGCTAA ATCTAATACA TTTTCGATCA TCGAAAGTGG TGTCCCAGGA TTGATGGTTA 7020

CGCCAGCTTT CACACCCGCA TTTTTAATCA TTTGTAAGGC ACGATGAATA TGGGGTGTTG 7080

CTTCTTGATG GACGGTAATA ATATCCGCAC CGGCTTCCGC AAAAGCATTG ATATAATTTT 7140

CAGGTTGTAC AATCATCAAA TGAACATCCA ATGGTAGTTT TGTCACTGGA CGAATCGCTG 7200

AAACGATATT CGGTCCTAAG GTAATGTTGG GCACAAATTG ACCGTCCATG ACATCAACAT 7260

GAATATAATC TGCTCCTAAT TCTTCTACTA AACGGATATC TCTTTCTAAA TTGGCAAAAT 7320

CGGCACTTAA AATGGATGGT GCTAGTTTCA TTTGGCTCGC TCCCTTTATA TTTGCTATGA 7380

TTTTTTCTTA TAAACAGGTC GTCGATTCTC GATTTCCATT AAAAACTGTA AATAATTTTC 7440

GTAACGACTC GTCGCAATGG TTCCCGCTTC CACTTGACGT TTCACTTCAC AACCAGGTTC 7500

TTTGTGGTGC ATACATTCAC GAAACTTACA GTGAGAGGCC GCGGCAACAA ATTCAGGAAA 7560

TTGCTTTGGT AATTCCACTG CTTCCATCTC CAAGAAATCG ATGGCACTAA ATCCTGGTGT 7620

ATCAGCGACC AAACCGTCGT ATAACGGAAT CAATTCCACG TGACGGGTCG TATGTTTGCC 7680

CCGTCCTAAA GATTGGGAAA TTTCCGCAGT TGCTAGTTGT AATTCTGGTG AAATTTGATT 7740

CAACAATGTT GATTTTCCAG CGCCTGATTG TCCCATAAAA ACAGTCAACC GTTCTGGGAA 7800

AAAGCGTTCT AATTCTTTGG TGGCTTCTAC GTCTTCCGAA GCGATTACCG CATAACCTAA 7860

CGCTTCATAA ATCTGTTTAA TTTCAGTGAC TTGTTGACGT TGCGGCTCAT CTAATAGATC 7920

AACTTTCGTT AAATAAATAA CTGGTTCTAT ATCCTTATAC TCTAAGCTGA CTAAGAAACG 7980

ATCCAACAAA TTAAACGAAA AATTTGGCGA AACCATACTC ATGACAACCA CACCTAGATC 8040

CACATTTGCA ACAGGCGGGC GCACTAATTC ATTTCGTCTT GGTAAAATTT CCAACACATA 8100

ACCATCTGTT AAATTATCGC TTTCAAAAAG TACTTCATCT CCCACGAGTG GCGTAATTTT 8160

TCGATTACGA AAATTCCCTC GCGCTCTTGT TTGATATGTT TCTCCATCTG CGTATACATA 8220

ATAAAAACCG CTTAACGCTT TTCTGATTTG ACCTTTCAGA TAAACCACTC CTAACTACTT 8280

TCGAATATAC AGGATAGTGG CAGAGAAATC AAGGGAGCCT CTCGATTCTA TTAATTAAAA 8340

AGAGCCTAGG ACATAAGTCG AAATGACTTA CCTTCCTAGA CCTAAATCTT AAGATAAATG 8400

GCGAGACAGG AGCAGCTTCT TCTTATTGCT GAGAAACACT ATTTGAAAAG CAAATAGATG 8460

TTAAAcAGTA CCTACTTGGT TCTTGAAGCT AAACACTCCT GTCCCAACCT CTTGTCTTAA 8520

TTACTCGTAC TACTTTCACT AGTGGTGCTT TTTGAATCAT TTGCTTTTGA ATAATAAACT 8580

GTAACAGAGT CGCCTTTTTT GACTTTGCTG CCAGCGGATG GGCTAGTTCG TTCAACGACT 8640

TGGCCATCGC CAGAGCCAGA TAATCCTTGT TCATTGATTT TAAGGCCAGC GTTATTGATG 8700

CTGTCTTCTG CAGCTTTTGG CGAATAACCT GAAATATCAG GAACAGTTAC TGAGTCACTG 8760

CCTTTGCTGA CATGTAAAGT AATCGTGTCA TTTTTCGGAT CAACAGCGGT ACCAGAAGCT 8820

GGTTCTTGAC TAATCACGGT ATCTTTTTCA ACTTTGTCGC TTTCTTCGTC CACTCGTTTA 8880

ATTTGAGATT CTGGGATACC TAAAGCAATT AAGCGACTTA CCGCATTATC GTAAGAAATT 8940

CCGCTATAAT CAGAAAGTGT CACTTTGTCG CTACCTTTAC TTACATATAA CGTCACTTGG 9000

CCATTTTTAG GTGTCACGGT GCCACCTGGT GCGGGGTCTT GCGTAATGAC TAAACCTGGT 9060

TCTACCGTAT CGCTTGCTTG GTCGACACGC GTAATTTGAG AGTCAGAAAT GCCTAATTGT 9120

GCCAGTGCAT TTACTGCATT TTCGTAAGAA TAACCGGCGT AACTAGGCAA AGTAACCGTT 9180

TCTGGTCCTT GACTGACCGT TAAAGTGACG TTGTCTTTTT TCGGATCAAC TTTTTTACCT 9240

GCAGCTGGTT TTTGTTTAAT AATGCTATCC GTAGACACAG AATCACTGTA TTCTTTTTTC 9300

GTTGTAATTT GGTCTTCTGA AAACCCTAGT TTTTTCAAGG CTTCGACAGC AGATTCATAC 9360

GATTCCTTTG TATAATCGGC CATCTCAATT TTTTCTGTTC CAGAGCTGAT GTATAAAGTA 9420

ACAGATCGGC CTTTTTTCAC AGATGATTTT GCTTCGGGGT CTGTTTTGAC CACCTTGCCT 9480

TCTTCAATCT TATCGTCGGG AATTTTTTTG GTTTCACTAT CGACTTTCAG CCCGGCACTT 9540

TTTAGCGCTT GTGAAGCGTC CGCTTTCGTT TCATTTGTAA CATCAGGAAC TTCAACGTCT 9600

TTACCGCCCG ACATTGCAAA GGCTAAGCCA CCTATTGCTA AAGAAATTAA TGCCGCAATT 9660

GCTAAGCCAA TCCACAATTT TTTACGGTTT TTCTTCGGCG GTTGTTCAGC TAAAATGTCA 9720

TCTGCGATAT CTTCTGGGAC TTCGACAGGC GTTGTCTCTT CCGGTTCAGC GATGTCTTCG 9780

GGAATCGGAG TTAATACTTT CGTTTCTCCT AATAAAGCCG TTGGTTCCCA CGCAGGTTCG 9840

TTTAAACGAC TGGCTGACAA GGACGTGTAT AAGTCTTCTG CCATCTCATT CGCTGTTTTG 9900

TAGCGATCGC TTGGGTCTTT TGCGGTTGCA TGACGAACCA CATTTTCCAA TGATTGAGGA 9960

ATCCCTGGAT CAAACATTTT GACAGAGGGA ATTTCTTCTT GAAAATGTTT TAAGGCAATC 10020

GTTACGGCTG ATTCACCATC AAAAGGTACA TTCCCTGTTA GCATTTCATA GAGAATAATT 10080

CCCACAGCGT AAATATCTGA TTGGTTGGTC GCCATGCTTC CGCGCGCTTG TTCTGGCGAT 10140

AAGTAATGCA CCGAACCTAA CATTGTGTTC GTTTGCGTAA TTGACGTTTC TGACAAGGCA 10200

ATCGCAATCC CAAAGTCAGT AATTTTGACT GTGCCGTGTT CGTCAATCAG AATGTTTTGC 10260

GGTTTTAAAT CTCGGTGAAT AATTCTATGT TCATGTGCCA TTGCGACAGC AGATAAAATT 10320

TGTTGCGTAA TGTCTACAAC TGTGGAATAA GGAATTGGGA AATGCGTTTG GATGTAACGT 10380

TTTAAGTCCA TTCCTTTCAC ATATTCCATG ACTAAATATT GTAGTCCATC TTCTTCGCCT 10440

ACATCGTACA CACTGACGAT ATTCGGGTGA ACCAGCTCAG TTGCGGCTAG TGCTTCACGC 10500

TGAAAACGTC GGATGGCGGC TTGATCGTTT TGGAAGTCAA AGCGCAAGAC TTTTACTGCA 10560

ACGTCTCGGT CTAAAATTAA ATCGTGTGCT AAATAGACGT TGGCCATGCC GCCGCTTCCG 10620

ATGCTGCCAA TAATGTGATA TCGACCATTC AGCTTCTTGC CGATTTCTAT CATTGGTTTT 10680

CCTCCTTCTG TTCGTCAAAG TGTATCACTA ACACGGTAAT GTTATCTAAA CCACCCGCTT 10740

CATTTGCTTG AGCAACGAGT TGGCTTAATT TGGATTCTAA AGGATCTGAC GTCTCTAAAA 10800

TTTCTAAAAT TTTTGTCTCA GGAACCATAT TGGTTAAGCC GTCCGAACAT AACAGTAAAT 10860

AGTCATTAGG CAGCCATTCA TGATTGGTTA CATCGACTTC AACAGTGCCA GGCATTCCCA 10920

AAGAACGTGT TAGGACATTT TTTCGTGGAT GATTTGCAGC CATCTCACGA GTAATCTCCC 10980

CTGATTTTAC TAGTTCATTC ACTAAAGAAT GATCTTCTGT TAATTGTAAC ATGTGTTGGT 11040

CACGAACTAA ATAAGCACGA CTGTCCCCCA CATTTGCTAA GACAAATGAA TCCGGTAAAA 11100

GAATGGCACC AACCACTGTT GTCCCCATTC CTAAATACTC TGGTTTCGAT TGTCCGCGTT 11160

CATAAATTTT TTCATTTTCC TTTTGAATTG CTTGAATTAA CCACTGTGCA GATTTTTCAG 11220

CTGAGTCGAC AGACGCATTG GACCAACTCT CACCTAAATT ATTGACTGCC ATTTGGCTAG 11280

CGACATCACC TGCTTGGTGA CCACCCATGC CATCTGCTAA AACAGCAAAG GTTATGTGTT 11340

GTTGGTTCTC AAAAACGTTG GCATAATCTT GGTTCGTATT TCTTTTGCGA CCGACATCTG 11400

ATTGAACGTT GATTTCCAAA TTATTTCACC TCATTTGAAC TAACTTTCCG CATACAGCAA 11460

ATAAAGAAGC CATCTGTCAT GTATTGATGG GGATACAGAA CCAGTTCTTG CTCTTTTACT 11520

ACTGCTTGGA CATTTTCATT TGCAACAATT TCAATTTTTT CAAATTCAGG GTGTTTGGCT 11580

AAAAAGGCTG CCACCACTTC CTGATTTTCT TCTGGCGTAA TAGTACAAGT GCTATAGACC 11640

ATTATACCCC ATTGTTTTAG TGTTGGGGCA ACACTTTCTA AGATTTCTAA CTGAATTTTA 11700

GGTAAATTTT GAAAATCATT GGCCGTTTTA TGGTATTTAA TGtCTGGTTT ACGGCGCATT 11760

AGCCCTAAGC CTGAACATGG TGCATCAACT AAAATCCGAT CaAAGGTTTc TGCGGGGAAT 11820

TCTTCGGCCA CTTGTCGCGC ATCAAGCTTT TCTGCTTGGA CGACATCCGC CACATGGAGG 11880

CGTTGCGCAT TTTCATTAAT TAGCTTAATT TTGTGCGCAT GAATATCTAA AGAAGTCACA 11940

CGACCACCTG CTTCAGCCTC TAAAAAAGTT GCGATATGAG TCGTTTTACC ACCAGGAGCC 12000

GCGCAGGCAT CCAAAACACG ATGATGCGGT TCAATTTGCA TCGATGGAGC CACGAGCATA 12060

GAGCTTTCAT CTTGAATGGT TAAACAACCG TTAATAAACA GTTCACTGCC AGCTAAGTAG 12120

CCTTTTTCAG CGACAACACC ATAAGGGGAA AGTTGACTTT CCTTCGCGGC AATCCCTGTT 12180

TCTTGTAATT CTTCCAGCGC TTCTTCTCGT GAAATAAACC GTGTGTTTAC GCGGCCACTC 12240

GCATGGCTTT TTTCAAATAA TGACAAACCC AACTGCCGTG TTTCTTCTTC GCCTAATTGG 12300

GCAAGTAATT TTTCTGTTAA CCAACGTGGC ATACTAATTT CCGTTGCTAA ACGATCAACT 12360

GGATCGCTAA TTGCGGCCAA ACTAGGTGCT CCATTTCGTT GGAACGCCCG CAGCACGCCA 12420

TTCACGAATT TACCAATCCC AGGATTGcCA CGACGTTTGC CAATTTCAAC CGCTTCATTA 12480

ATGACTGCGT GATCAGGCAC TTTATCTAAA TAAAGTAATT GATATAACGA TAAAATCAAT 12540

AAATTTTTCA CCCAATTATC GACTTTTTGC GGTTTCTTAA CAAATGGTGT TAAGTAGTAT 12600

TCCAACAATA ATTTGCGACT AATTGTGCCG TAGACAAGTT CTGTAAACAA GCGGCCATCT 12660

TTCTCGCTTA GTTCTGATTT TGTCATCATT TCATTAAGTA AAAGATTGGA GTACGCCCCA 12720

CCTTTATCCA CACGTTCGAT GGTTTCTAAA GCAACGTATC GAACAGAACG TTTTAATTTC 12780

TTGGGGATTT TTTCAGCCAT TTTTGTTAGT CCACCTTATC TTTTTCTGCG ACATTTTGAC 12840

CAACGCCATT TAAAAATTCT TGAATCGTTA ATTGACCTTT ACCTGCTGGT TGCAGGCGAT 12900

CAATTTGAAG GACTGTCCCT TCACCACAAG CAATCCACAG CGCTTTTTTA CTACGTTGGA 12960

TAATCGTTCC TGGGGCGCTG GTTGTTGTTT CAGTCAGTGG TGTCACTGCC CAAATTTTCC 13020

AGTTGGTCCC TTGATAGGTT GTGAACGCTG TTGGCCAAGG ACGCATACCA CGCACTTGGT 13080

TATCAATTGC CTCAGCAGTT TTTTGCCAGT CAATTCGTTC TTGTTCTCGA GTAATATTTG 13140

GTGAAAAAGT AGCCTCTTCT TCCACTTGTG GAATAGGCGT GATTTCGCCA GCAATTAGCT 13200

TAGGTAACGT CTCTAATAAA AGTTCTTTTC CTAGAATGCT TAATTTTTCA AACATGGTGC 13260

CTACATCATC TTGTTTCGTA ATCGGAATTG CGCGTTGCGA TAAAATCGCG CCAGCATCCA 13320

TTTTCTTCAC CATTTCCATA ATCGTGACGC CTGTTTCTTT TTCTCCTTCA ATAATTGAGT 13380

AATGTACAGG TGCGCCGCCA CGATATTTTG GTAAAAGTGA GGCATGGACA TTAATTGCAC 13440

CTAGTTTGGG CGCTTTTAAG ATTTTCTCTG GTAAAAATTG TCCAAAGGCT GCTGTCACAA 13500

TTAAATCTGG TGCTAAGTCA ATCACTTTTT CCATTTCAGG AGAACCTGAA ATTTTCTCTG 13560

GTTGTAAGAC TAATAAGTTA TGTTTTAAAG CTGCTTCTTT GACAGGCGTT GGGGTAATCA 13620

CTTTTTTTCG GCCAACTGGT CGATCTGGTT GCGTGACAAC CGCTTGAACG TCATAGCCTG 13680

CCTCGATTAA ACTTTCTAAA ATCGGAACAG AAAAGGCTGG TGTGCCCATA AATACAATTT 13740

TAGTCATCAA AATGTTCCTC CATATACGCT TCTAAATCTT CTGGTGCAAT TTTCTCAATC 13800

ATTTTATCGA TAAATAACTC GCCATTCAAA TGATCAATTT CGTGTTGAAA AGCTCGAGCT 13860

AAATAGCCGT AAGCCGTAAC TTCCATTTCT TCACCTTCAC GGTCAAAATA ACGGACAGTC 13920

ACTTCATCTG CTCGTTCAAC GGTTCCATAT GTTTCTGGAA TACTTAAACA GCCTTCCACA 13980

TCAATACTGG TTCCCTTTTT TTCAATAATC ACTGGATTAA TTAGTTCAAA ACGACCAGAT 14040

TCTTCATCGA TTTCAATCAC TGCCATTTGC AAATTTTTAC CAATTTGCGG GGCTGCTAAG 14100

CCAATCCCAT CATGGGCCAG CATTGTTTCA TACATATCCT CTAGTAACTG CACAATTTCA 14160

TCGGTAATAA TCGTCACAGG TTGAGCCTTC ATTTTCAAGT GTTCGTTCGG GTGAATCACA 14220

ACTGGATAGC GCATAAGTGC TCCTTTCAAA GCTGTATTTT TTAAATAAAA TTCATGGGTT 14280

CTGCATCAAT GGAAAGCTTT AAGCCACGAG CCGTTGCTCG TTGTGTATCG GTTAAAATTT 14340

CTTTCAATAA TGGTTGAAGC ATTGGTTCTT GTTTATACTT GATAATCACT TGATAAAAGT 14400

ATCGATTATT GACACGCATA ATTGCGTTGG GTGTTGGTCC CAACAAAATA GCTTGCGGCG 14460

ATAAGCCTTG TTTTAATTTC GTCGCGATTT GAAACATTTG TTTCGCTGCT TCATTTTCTT 14520

CAGGATGGCT AGCCGTAATT TGGACCGTAA AATAGTACGG CGGATAATCC CCACGATGAC 14580

GAATATACAT TTCTTTTGTA TAAAAATCTT CATAATCTTG GGCTTTTGCC AATTGTATCG 14640

CATAATGCTC GGGGTTAAAA GATTGAATGA TGACTTCCCC CGGTTTTTCA GCGCGACCAG 14700

CTCGACCGCT GACCTGTGTC AATAATTGGA AGGTCCGTTC ACTGGAACGA AAATCGGGTA 14760

AATTCAAGGC GGTATCAGCA TTTAAGACAC CGACTAACGT CACATTTGGA AAGTCCAGTC 14820

CTTTGGCAAT CATTTGTGTG CCTAATAAGA TATCCGCTTG TCCTTCACCA AATGTTCGCA 14880

AAATCTTTTC ATGTGCACCT TTGCGACGGG TCGTATCAAC ATCCATTCGG AGAATACGAC 14940

TGTCTGGTAA CAACGTTTGT AATTCTTCTT CTACTTTTTG TGTGCCTGTT CCGTAGTAGC 15000

GAATTTTATC TTGGCCACAG TTAGGACAGC GATAAGGAAT CCGTTCTTCA TGTCCACAAT 15060

AATGACATTT CATTGTTTTA CTATCCATAT GCAAAGTTAA GGAAATATCA CAATTCGGAC 15120

ACGGCAACAC ATAGCCACAA TCGCGACACA TCACAAATGA CGAATAGCCA CGGCGATTAA 15180

GTAGGAGAAC ACTTTGTTCG TTTTTCTCTA AGCGTTCTTG GAGTTTTTCT TGCAACGACA 15240

TGGAAAAAGA AGAGACATTG CCATTTTCTA CTTCTTGCCG CATATCCACT ACATCAATCG 15300

TAGGCAAGGT TGCCGCTTGA TTGGCACGCT GTGTTAATCG TAACCGCTGA TAAACATTTT 15360

TTTGTGCGCG AGCGCGTGAT TCCAAGGATG GCGTGGCACT GCCTAAAACA ACAGGGCAAT 15420

GATGGTACTC CGAACGCCAA ATCGCTAAAT CTCTCGCGTG GTAGCGCGGG GTTTCTTCTT 15480

GTTTATAACT TGCCTCGTGT TCTTCATCAA CAATGATCAC GCCGATGTTT TCGATTGGCG 15540

CAAAAATGGC TGAACGCGCG CCAACCACCA CTTCCGCTTC TCCTCGTTCA ATCTTCCGCC 15600

ACTCATCGTA TTTTTCACCT TGCGATAAGC CACTGTGCAT TACAGCAACA TGTTCGCCAA 15660

AGCGACTTTT AAAGCGTtGC ACCATTTGCG GAGTTAAAGA AATTTCAGGG ACAAGCATAA 15720

TGGCTGTTTT GCCTTGGTTA AGAACTTCTG CAATGACCTG TAGATAAACT TCTGTTTTCC 15780

CGCTTCCTGT AATACCTTCC AATAAGTAGG TTTGGCTTTG CTGTTCTTGG ACCGATTGTA 15840

AGATTGTTTC CACAGCCACC TGTTGTTCTG CATTCAAAGA TAACGCAGTC GTTTTTTCAA 15900

ACGTCTGATT AGCAAACGGA TCACGATACG CTTCTTTTTC AATGAACGTT AACCAGCCGT 15960

TTTTGGCCGC TTCATTTAAA AGCGCGGTAC TGAAACCTAA TTCTTTCATT TCCTTCACAG 16020

CGGTGACTTT TTCAGTGCCC AAGCGTTGTA AATAGTAAAG AAGTTGCTCT TTCTTCTTAG 16080

CTCCCTTGCG TAACCCCAAA CGAATTTCCT CTAGTTGTTC AAATTCCTTA GCGGCTTGAA 16140

TGAAGCGGAC CATTTTGACT TTATTTCTCG TCGTAAcTTC GTAGCGAATA TCGACTTTTT 16200

GCTGTTTACG CAACGCCATT AGTTGAGGCA AAAGCCCACG TTCTTGCGCT TGGTCCCACG 16260

AAATTTCCTC TAAACCATAA AATAACTGAT CTTGTAATTC TTCAGAAAGC TCATCCGTTA 16320

AATAAATATA TTTTTGGTAA TCGGCACGCA TTACGCTAGG CAACATGGTT TGCAAACAAG 16380

TGATTTTAAA AGCAAACGTT TTTTCCTTCA TATAATCAGC TAAAGCCAAC ATTTCTGTAT 16440

TTAAAACAGG TTTTAAATCC AACACCGCCA CTAGTTCTTT CAACTGAACA TTGGTTTCGT 16500

CAAGAACCGT TGCAGCCACT GGTTCAATGG CTAGAACAAA CCCTTGAACA TGCCGATTAC 16560

CGTTACCAAA AGGCACTTCT ACACGCATGC CGACAGCCAA TTGCTCATTT aAATTTTCAG 16620

GAACTAAAtA AGTGAATGGT TGATCGGTTT GcATCGTTGG TACATCCACA ATAACTTTGG 16680

CTACTTTTTG CATGGTGCAT CCTCCTTTTT ATGTAAATTC TCTTTGCCTA TTGTACTAAA 16740

GATTGACTGA TTTGTCTTGT GTGTCTCGTT GGCACCTTTC AGAATAAACG AAGAGAAAGC 16800

CTCTGAAATA GACAGACGGC TTTCTCCATT CGTTGACAAT CATTCAATTA ATTTTTTTCG 16860

TCACGTAAAC GGTTTTCTAA TTCTTTTTGT TCTTGTTCAC GTTGGCGTTT GCGTTCTTCA 16920

CGTTCAATGC GTAGACGTTC GCGTTTTTCT TCAGGATTTG GATCACTGAT CACTGTTCCT 16980

GCTTCAATTT CTTCTAACGC ACGACCTACA CTTTTTACAG AATCAAAGGA TTCAACGGTT 17040

GGTTGAACGC CTTCATCTAA TTCATGCGCA CGTTTGCTTG CT 17082

(2) INFORMATION FOR SEQ ID NO: 230:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4594 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

CTTTTAACAG AGAACCTGAC TATTAGTGAA AAAAGAAAGC AACAGCTAGC CTTTGAAGTG 60

TCACCAGCTG GAGTAGATAT CACAGAGGAT CAGATTGAGT GGTCGAGTAG TGATCCTACC 120

ATTGTGACGG TTGACCAAAC GGGTAATTTA ACAGCAGTTA AGAGTGGGTG AAGCGAAGGT 180

AACrGTCAAA ATTAAaGGAA CAGAAaTTAG TGATACAATT CCTGTGACTG TTGTAGCAGA 240

AAATAAACAA TATGCGGAAA TGCGAGCAAA ATGGAAAATG CGATTGTTAG GTACAACGCA 300

```
gTATGATAAC GATGCAGATG TGCAACAGTa TCGTGCCCAG ATTGCTACGG AGAGTTTAGC     360
ATTATGGCAA ACGCTTAATC AAGCAGCAGA TCGTGAGTAT CTGTGGGAAC GAAAACCATC     420
GGATACAGTG TCTGCTGATT ACACGACTCA ATTTACCAAT ATTAAAAAAT TAGCGTTAGG     480
CTACTATGAA CCATCAAGTG AGCTTTTTGA AAAACCAGAA GTTTATGATG CGATTGTTAA     540
AGGCATTGAA TTTATGATTG ATACAAAAAA ATACAATGGA ACGTATTACA CAGGTAACTG     600
GTGGGATTGG CAAATTGGTT CCGCGCAgCg CTAACAGATA CATTGATTTT ATTACATGAT     660
GACCTATTGA ATACAGATGC AGAAAAATTA AATAAATTTA CTGCTCCGCT GATGCTGTAT     720
GCaAAAGATC CAAACATACA ATGGCCAATT TATCGTGCAA CAGGAGCTAA CTTAACAGAT     780
ATTTCAATCA CCGTTTTAGG TACTGGACTT TTGTTAGAAG ATAATCAACG CCTAGTACAA     840
GTACAAGAAG CTGTTCCGTC CGTTTTAAAA AGTGTTTCCT CTGGTGATGG CTTATATCCT     900
GATGGTTCCT TGATTCAACA TGGTTATTTT CCGTACAACG GCAGTTACGG GAATGAGTTG     960
CTAAAAGGGT TTGGACGAAT TCAGACTATT TTACAAGGTT CCGACTGGGA GATGAATGAC    1020
CCTAACATTA GTAATTTATT TAATGTTGTG GATAAAGGTT ACTTACAATT GATGGTAAAT    1080
GGAAAAATGC CATCGATGGT TTCTGGTAGA AGTATTTCCA GAGCGCCAGA AACGAATCCT    1140
TTTACTACAG AGTTTGAATC GGGTAAAGAA ACAATAGCTA ATTTAACCTT AATTGCAAAA    1200
TTTGCACCAG AAAATTTAAG AAATGACATT TATACATCTA TCCAAACGTG GCTTCAACAA    1260
AGTGGGTCAT ACTATCATTT CTTTAAAAAA CCAAGAGATT TTGAAGCGTT AATTGACTTG    1320
AAAAATGTAG TGAATAGTGC GTCACCTGCC CAAGCGACAC CAATGCAATC TTTAAATGTA    1380
TATGGTTCGA TGGATCGAGT CCTACAGAAA AATAACGAAT ATGCGGTGGG GATCAGTATG    1440
TATTCACAAC GTGTCGGAAA CTATGAATTT GGGAATACGG AAAATAAAAA AGGCTGGCAT    1500
ACAGCAGACG GCATGCTTTA TTTATACAAT CAAGACTTTG CTCAGTTTGA TGAAGGATAC    1560
TGGGCAACGA TCGATCCATA TCGATTACCA GGAACGACAG TTGACACAAG AGAATTGGCA    1620
AATGGTGCTT ATACAGGGAA ACGCAGTCCC CAGTCATGGG TAGGTGGCTC AAATAATGGA    1680
CAGGTTGCCT CTATAGGAAT GTTTTTAGAT AAAAGTAATG AAGGAATGAA CTTAGTTGCT    1740
AAAAAATCTT GGTTCTTATT AGATGGTCAA ATCATTAATT TGGGAAGTGG CATTACTGGT    1800
ACGACAGATG CTTCGATTGA AACAATCCTC GATAATCGGA TGATTCATCC ACAGGAAGTG    1860
AAGCTTAACC AAGGTTCAGA CAAAGATAAT TCTTGGATTA GTTTAAGCGC AgcGarTCCA    1920
TTGAATAACA TTGGCTATGT TTTTCCTAAT TCmATGAATA CGCTTGATGT TCAAATAGAA    1980
GAACGCTCTG GTCGCTACGG AGATATTAAC GAATACTTTG TTAATGATAA AACCTATACA    2040
AATACATTTG CTAAAATTAG TAAAAATTAT GGCAAGACTG TTGAAAATGG TACTTACGAA    2100
TATTTAACAG TGGTTGGGAA AACGAATGAA GAAATCGCAG CTCTTTCTAA AAACAAAGGC    2160
TATACTGTTC TAGAAAATAC AGCAAACTTA CAAGCCATTG AAGCAGGTAA TTATGTCATG    2220
ATGAATACAT GGAATAATGA CCAAGAAATT GCAGGACTGT ATGCGTATGA TCCAATGTCG    2280
```

```
GTTATTTCAG AAAAAATTGA TAACGGTGTT TATCGCTTAA CTCTTGCGAA TCCTTTACAA 2340
AATAATGCAT CCGTTTCTAT TGAATTTGAT AAGGGCATTC TTGAAGTAGT CGCAGCGGAC 2400
CCAGAAATTT CTGTTGACCA AAATATTATC ACTTTAAATA GTGCGGGGTT AAATGGCAGC 2460
TCGCGTTCAA TCATTGTTAA AACAACTCCT GAAGTAACGA AAGAAGCGTT AGAAAAATTA 2520
ATTCAGGAAC AAAAAGAACA CCAAGAAAAA GACTACACCG CAAGCAGCTG GAAAGTCTAC 2580
AGCGAAgCAT TGAAACAAGC ACAAACTGTG GCAGATCAAA CAACAGCAAC GCAAGCAGAA 2640
GTAGACCAAG CAGAAACAGA GTTACGTTCG GCAGTGAAGC AATTGGTAAA AGTGCCAACT 2700
AAAGAAGTAG ATAAAACCAA CTTGTTGAAA ATCATCAAAG AAAACGAGAA ACACCAAGAA 2760
AAAGACTACA CCGCAAGCAG TTGGAAAGTC TACAGTGAAG CATTGAAGCA AGCGCAAACT 2820
GTGGCAGATC AAACAACAGC AACGCAAGCA GAAGTAGACC AAGCAGAAGC AAAACTACGT 2880
TCGGCAGTGA AGCGATTAAC ATTGAAAAAT AGTGGGGAAA ATAAAAAGGA GCAAAAAAAT 2940
GGGGGGAATA ATGGACACTT AAATACTAGT ACAGGAGTTG ATCAAACTGG TACGAAACAA 3000
GTTAAGCCAT CAAGCCAAGG TGGTTTCAGA AAAGCTAGCC AATTTTTACC GAGCACAGGA 3060
GAAAAGAAAT CGATCGCGCT TGTGATTATT GGTCTTCTAG TTATCGCCAG TGGGTGTCTT 3120
TTAGTTTTTC GTAAAAGTAA ATCGAAGAAG TAAAAAATTA CGGATTCACT AAAAACATAA 3180
TAAAAAAGCA GACTTCCATA CAAGAAGTCT GCTTTTTCAT TAATTACTCC GCTTCTGCAT 3240
TTTGCAAGTC AGCGACATCA GTCAGATAAG TGGCTTCATC CAACAAATCT TCCGAATGTG 3300
CGACATAAAT CGCAGTGGCG CTATCACCAA CCACGTTTAG GCTAGTAACG AGCATTTCGA 3360
TTGGACGATT AATGGCTAAA ATCAATGAAT ACGTCGCAAT CGCAATCTCA TTGTTATAGT 3420
TTAAACCAGT TAAAATAGTG AAAAGGATAA TCCCGCCAGC GCCAGGTGCC GcCGGTGTAC 3480
CAATGGATGA CATTAAAGCT AAAATGGCAA TCAAGGCAAT ATTAGCTGGT GTCACTTCAT 3540
AAcCAGTACG CCCGCCACAA AAATAGCCGC AATTACTTGC ATAATCGCTG TGCCGTTCAT 3600
ATTGATAGTC ATTCCTAACG GCAAGACAAA GGCGGCAACA TCGCTGTCGA CACCTAACTC 3660
TTCAACCGTT GTTTTTGTAT TTAACGGTAA AGTGGCTGCC GAAGAAGAAG TAGAAAATCC 3720
AAAGACCACA ACTTTCATAA TCTTTTTAAT AAAAGGAACA GGATTTAATT TTGTCGCAAA 3780
TAAGATAAAT AATGGGTAAC CGATCATCAA AAAGGCTAAG AGCGTCGCTG TCGTGGTTAC 3840
GACGTAGACT AACGCTGGTT TTAAATAATC AmTCCCGTAA GAAGCAAACG TCCGAACTAA 3900
CAACATAAAG ACGGCAATAG GAGCAAATTT AGTAATTACA AAGGTTAAAA ATAACGTGAC 3960
CATCGCATTG ACTTCTTCAA TTAATTTTTT GAAAACACGA ATTTTATCCC CTAATGCATT 4020
AATACATAAA CCAGTAATCA CTGCCAACAC AACAACGGCT AAAATTGCAC CATTGTCGCT 4080
AAATGCAGAA CCGATATTTT TCGTGACAGC ATTGACAATC ACCATTAATG GGTTGCCGCT 4140
CGTACTTACT TCGACAGGAG CTAGCGTTTC AGATAACGAC ACATTAAAAG CGCCAGTTAA 4200
ATATACGCTA TAGCCCACAA TTCCAGCCGT TAATAGAGCA ATCAGTGACA TCGATAAAAA 4260
```

GCCAAAAATT GTTTTATAAG AAATTCGTCC TAATTTTTGC GTATCGGTAA TATGACAAAT 4320

GGCTAAAGTA ATTGAAGTAA AGACCATCGG CACAATCACT AATTGTAGGC CGTTGATAAA 4380

TAGTTGACCA ATAATATAGA AAAGCCCAAT CGCAGATTGT CCTTTTTCAG GATCTGAAAT 4440

ATCTTGGAAA AGAATCTGAT TGATTGTCTG CCACACAGCG CCTTGATTGC TGCGAATAAG 4500

GTGCTCTCTT AAAAATAAAA AGCCGAGCCC AACGATTAAT CCTCCCACCA ATGAAATCAC 4560

CATTTTTTTG GTAAATGAGA TACCTTTCAC TTCA 4594

(2) INFORMATION FOR SEQ ID NO: 231:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2991 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

TTTCGGTCAC TATCACCATC TCTTGAAAGA AGTTCAATGG CTACATCTCC GTTTAGCTCT 60

GCTTTTGCTT TATCTAAATA CTCTTTGGCT TTTTTGGCGT CTGTTTTGAC AAGAGCGCCT 120

GCTTCTTGAC GAAAATCTTC ATTCGTTTCT GGGTTATATA CAAATTTTTC TGGAATGTAC 180

CCATAAATTT CTTTTGAGCC ATCTGCTAAA ATATTATCGA CTAAACTTTT TTTATCCAAA 240

GCATAAGCTA AAGCTTTGCG GACATTCTCA TTTGCAAAAA TGGATGGTTT TTCGTTTCTT 300

TTTTGGTTTA AACGTAAGGA ATAGACTTTT GACCGTTCGA TTGCTTCATA GTCTGGGTTG 360

TCTCGATTTT GTTTAGCAAA ATCTCCTGTT AAGACAGCCA CATCTAATTC ACCTGATTCG 420

TACAAATTAT AAACGGTATT GGTTTCTTTA AGAACTTCAA AATGAATCGT TTCTGATTTT 480

ACTTTTTCTT TATCGTAATA GTAGGGATTA CGCACAAAAT CCCACGACAT CGCTGACTGA 540

TCCCAATTTT TCATTAAAAA CGGACCATTA GATACAATCA TTTCACTAGA AGTGCCATAG 600

GCTTGTCCTT TTTCTTTCAC GAATGCTTCG TTTTGTGGGA AAAATGGAGA AAATGCCAAT 660

AACGACGTAA AATATGGTAC AGGCTTTTCT AAAGTAACCT CCAAAGTATA ATCATCAAGC 720

GCTTTGACAC CCAATTCCTC TGGTGCTTTT TCCTCTTTTG TAATAGCTGT TCCGTTCAGA 780

ATCGTTCCTT CTAACAAGAA AAAGTAATTG GCTTGGTTTT TGGGATTCGC TAATTTACGC 840

CAAGCAAAAA CGAAATCATT GGCTGTTACA GCATCACCAT TGGaCCACTT GcCATCTTCT 900

CTTAGTTTAA TGGTATATCG TTTGCCATCC TCAGAAATTT CTGGCATCTC TTTAGCAGCA 960

GCAGGTATTA GTTGATCTTT TTCATCCAAT TGATATAACC CTTCAAAGAC ATTGGACTGT 1020

GCGTTAATAC TGGCTTCATC TTGTGTAAAA ATAGAATCCA TTGAACCGAT TTCCGAGTTC 1080

TCCATTAAAT TTAAAACTTT TTTCGAGACA CTGGTTTTCT CAGTTGCTTT CTCGTCTATT 1140

TTTGTwTCTG CTGTTGTGCC ACAACCTACC AAGACCGTAC TAACAAATAA CATAATCCCC 1200

ATCATTTtTA ACTTTTTCAT ACTATTCCCC CTATGTGTCA ATCAAACGGC AACTCTTCCA 1260

TGGAAAGCCG TTCTTGCGAA ACTATCAAGC ATGCGTAGGG CAAGCCCTCA TTTTCGTACT 1320

CAAATTCTCC CCGCTGGTTG TCAATTGGGT CTGCACGCAA TTGCATTTTT GCCGGCTGAA 1380

AGTCTGCCTG GATCTCACTA TTTGCGACCA TTCTAAAAAC ATCTAAATTG CCGAAATAAT 1440

GATTCCACCG CTGTCTATCC CCGTCTCGGT AAGCACCTCC GCCAAAAGAC AAATCCGCAA 1500

AGAGCCAGCC GTACGGTTTT ATATAAAATT GTGCCCAATC ATGACAGCCT GTATAATGTG 1560

TAGAAACATA TAATCCCGAT TGCCATTTGG CTGGTATCCC TGCAATCCGA CAAAGCGTGA 1620

TAAACAAAAT CGCTTGCACG CCGCAGTCAC CTTTTAAATT AGTTGCAGCA TACTCTGAAA 1680

TATTGGAAAT TGTAAAATAT TCACGCATGA AAGAATAATG AATCTTTGTC GTCACAAAGT 1740

CATAAATCAA GCGTGCCTTT TTTAAGGGaT TACTTTCATC GCCCACAATC TCATCTGCTA 1800

AATTTCTTAA ATAAGGCGTG AATCGAATAT GCGGTAACTG TTCCTCTAAA TCAAAGCTTG 1860

GTTGCATTTT GTCTACGTTA TTTTCTTCTA ACTGTTGGTA AACCATCCGA TTAACATAAC 1920

AATATTCAAC TGTAAACGTT TGATTTTCTT TTAAGATTGT TTCAAAACAC AGTGTTCGTT 1980

GAGGCGCATC TTCTGCCGCA ATATAGGTTG CTTCAGGTGA CGTAGCTAGG ATTTCAATTT 2040

CTGATTGTTG TTGACACGCT TTGGGCAGCG GCAAATGGAC ACGAACACGC TCACCTGGAC 2100

GCTCAAATTC TTTTTTCACC TGAATGGTGG CTTTAATTGT TGTTCTCACT TTTCTACCAC 2160

CGTTTTTTTG CATTTGTTCG ATGTTCTCTT TTAGTTCTTT TTTGCGGAGA ATTTCTAATT 2220

GATTGCTTTC TGCTTCCCGC AAACGTGCAA CATATGCCGC ATTTGTTTTA ATTAAATTGG 2280

CCAAAAATCG AAGTTGGAAA TACGGTTCAC CATCAATATA AACCCAGTCC ACTAATCCCG 2340

TTTCTTTTAA ACGAATTAAT TCCTCTGTTT GATAATCCTG AAAAGTATCT CTCATCATTT 2400

GATCAGCTTT TGcAAAGcTA TGAGGATACT CATtCATACG CATCATTTtA ATGgACAtCT 2460

TTTTCaATTA TCAAACGTTC TTTTaATGCC TGCGGAAGTT CTTTTGTCAA CCAAAAmTCA 2520

ATGGCAATAT CCGCACTTGC AAAATCGCCA AATTCTTTCA CCCTTTGAAT TTCCTCAGGC 2580

AAAGGAACAC TAAGATACTT TAAATCATAC TTCATTTACT TTACTCCTTT TTTAATGAGA 2640

AATGCCTAAT CCAGCTGCTT CTCCTAATTC CAATAATGGT TTTGCTTCTA AGGATACGCC 2700

CCCAGTTACG GGCGCAGTTT CTAAACCGAA TGTCGCATCT AAATCTGCTC GTGTGATATT 2760

CTTTTGAGCA GCAGCTAAAT GGGCAGCAnC CGTGATGCCA ATGGTTGTTT CTTCCGCCAT 2820

ACAACCAATC ATGCATTCGA TGCCTGCTGT TTCACAAATC TGATTAATTT TCAGCnCTTC 2880

ATGAATGCCn CCACACTTCA TTAATTTTAT ATTGATAACA TCGACTGTGC CTTTTTTAcT 2940

AATTCCAACG CATCTTGAGC ATCAAAACAG CTTTCATCAn CCATAATTGT C 2991

(2) INFORMATION FOR SEQ ID NO: 232:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4346 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

TTGGACCCCC CAAATTCTTC CTTCGTCGAA ATTAACCACT CCCCCTTGGn TGATTGGATC 60

AAGGTCTGGn TTCCTTTACG CCCCAGGCCA TTGGGGTGTT TCCACCAAAn CAAACCTTGG 120

ATCCCGCTGG TTAAATTTCA CATATCACCc ATTGATTGGA AATCaAAAGC TATCTTcTTT 180

AAATAAGATT ACTTATAAcG AGAATTTATT GACAAATTAA GACAAGAAAA CAAATATACA 240

GAATCAACTA TTCAAACGTG GCATAAAATT GTAATGAGGA TGATTAACGC TGCGGTACAC 300

AATCAAATCA TCCCTAGCAA CACGCTAACA GGCTTTAAAT TTGATTTAAG TAATAATGTT 360

CGTTCGTTCT CTAAAAAGGA ATTACAGCGA TTTGTGGCGG TTTTAGAAAA CGAAGATATT 420

CAAACGCAAG TTATATTTTT AACTCTGCTA AAATCTGGAA TGAGGAAAGG CGAACTGATG 480

GGGCTGCGTT GGAGTGATAT TGATTTAACC GAAAAATATT TCGATATCAA TTCTACACGT 540

GGTGATTACG GTGAAAATAA ACCTAAAACA AAAACCAGCA TACGTAAAGT TTATTTTGAC 600

AACTCGTTAC TCACTTTAAT AAAAAAATAC AAAAATCATG AGAAAGAACG GCTTTTCAGA 660

GAAGGGATAA TTTTAAGCGA TAAGGACTAT TTTATTTTAA GTTCTCGAAA TTTACCTATC 720

AAACAATCAA GAATTACGTA TATGTTTCGC CTGTTATGTG AAAAAGCAGA AGTTCAAAAC 780

ATAACCGTAC ACGGCCTAAG ACATACCCAT GCAACGTTTT TAATTGAAGC AGGAGCAAAC 840

ATTAAGTACG TTTCAACTCG GTTAGGACAC AAGAATATTA ATATAACTTT GGATGTTTAT 900

AGCGATGTGC TAAAAGAAGA AGAAAAAGAA ACAGCTGATA TGATGGATAA ACTTATTGAG 960

AACTTGTGAT TTGTGGTAAA CCTGTGGtAA AAACATTAAA AACCATTGAT ATATAAGCAC 1020

TTTACCACTC CCCATCTGAA ATTGTAAAAT AAGTTGtTGT TTCTTTTCAG TCCAAGTAAA 1080

AATATATTTC ACTAAACGGC CATTGTTCCC TGACGAATGA AGAAAACTGA GTTTACTTGG 1140

GCCACTCACT GTCATTTTTT CAGGTACAGC TAGTAAAGTC CTCCGAGCTG ATTGGTTCAA 1200

TTTATCAGGA AATGAAAAGC TTTGTTGTTC TTCAAAAAAC CAAACTTCTT GTGTTTCCTC 1260

TAAAACAATT GCTTGCTGTT GCGCAAAACG AAGTTGTTTT TCTACCGTAG CAAGAAACTG 1320

TTCAATTTTT AACGTTTCTT GCCAGCGGTG AATAGCGAGT GTCGGGAGCA TAGTCATAAA 1380

GCAAACAATC GTTAAACTAA CAAGTGTTTC CATTAATGTG AATCCTGCTG ACTTGTTCAT 1440

TGCTTTTCTG CTGTATATTT ATCTAACTGC TCTTTAGTAA TGTAGCCTTC GTTGACTAAT 1500

TCATTTAAGG AAGGCGTCTT ATTTTTTTCT AGTGTGTAGA GCTCGATTTG TGATTCTACA 1560

ATTTTTACGA TTGCTTCATT GCCTTTTTTA TCAACTGTTT CTTTATGTTT CGCTAAGTTA 1620

GGGACAAAAA GTAAAATCAA TACGGAAATA ATCAATAAGA CAATCAACAT TTCTAATAAT 1680

GTAAACCCTG CGTATTTTTG tTTCTTTTTC ATTATAAAAC TCCTTCCATA TTGcCATAAw 1740

TAGGCAGCAA CaTTGCCCCA TAAACACCGA TAATTAAACA AGCAATTATG ATAAAAATAA 1800

GTGGTTGAAC GAGAGTAATG ACTCGTTCAA TTTTTTTGAA AAAaCCTTGG AAACAGACAT 1860

TACTATAAAC CATCAaCTCC TTGCCTAATT TTCCTTTTAC TTCCCCTTGT TGAATCATTA 1920

AACCCAATGC TTTCGTAAAA AAAGGATAAG CAGCTAATTG TCTATGTAAG GTTTCACCAA 1980

CTAGCAGTCC TTCACTAATT GCTCCAGCTA ATTCTTTCAT TAAGGGGTAA TTGGTTGTAT 2040

GGCGCATCAA ATCGATGACC TCTTTCATTT CTAATCCTTG TGCAAGCAAC TTCCCCCATT 2100

CTAGTGCAAA AAAAGCACTG ACGTATTCCC GATAAAGGCC ACGGATAATT GGCAGTTTCA 2160

GATATACGAA GGCCCGCTGA ATAGCAGACT GATGCTTCAA ATAACGGTTT AAAATGAAGA 2220

AACTAAATGA GAGGACAAAA ACTAAGGTAA CCAGAATGTA TGGACTGTAA ACAATCAGAA 2280

GAATCCCAAT ATTTTTAGAA TAACTATTGG CATCATTTTC CATTAATTGG GGCAATAAAA 2340

ATTGCCGAAT ACCAACAATG ACGAGAATCA AAAAAACAAT CAGGAGCAGC GGATAACTAA 2400

TTATTTTATA GAACTGTTGC CGATATTTTT CCGTTTGAAA AAAATGTTCG GAAATCGCTT 2460

GTAGTGTCCC CACCAAATCC CCGTGTGCAG TAGCAAACGA CAATTGTGCC GAGACCATTG 2520

TTGAAAATCC TAATTCTTCG ACTAAAGCTT GTAAGGCAAT GCCTGAGGAA AAGCCTTCAA 2580

CTAAAAAAGT TATACTGTGC TGAGGTAAGG CGTTTGATTT TTTCATGAAG GAAAAACTTT 2640

CTTGTATATT AAAACCATTT GCAAGCAAAT CCGCTAATAG TTTCAGAAAA AGCTGCTGTT 2700

GTTTTTTCGT TAATTTATTC GATGGCTTGA AAATACGTTT TTTCTGTAAT AAACCCATAC 2760

GCCCACGCCT TTCTTAGATT CTTTTGCCAA CTACTCTTTA TAACGGCTTC CTTAAAGAAG 2820

AAATAATCAT AAAGAATTCC CGCTACTTGT TCCTTATTAA TCGTGGTTGT CAAAATCTGT 2880

TGATAAATGA TACCTTGCAA GCATTCAGCC AATTCATGCT CTGGGCCACC TAATTCTATC 2940

AAACGCCGTT GCACCCCCAA TAAGTTACGG GCGTGAATTG TCGCAAAAAT AGTATGACCC 3000

GTTAAGGCGG CTCGAATAAC TGCTTGAGCC GTCTCGCCAT CACGAATTTC ACCAATAATC 3060

AAAATATCAG GATGATGGCG TAAGCAAACT TTTATCAAGG CTTCGTACGT TAACTGAATT 3120

TTTTCATTTG TTTGCAACTG TAAAAACTGC TTTTCCTCAA TTTCAACAGG ATCCTCAATC 3180

GCTATCACTT GTTGTTCGCC ACTTCTTTGT TTAGCCACAC GATACATGGT TGTTGTTTTC 3240

CCAGAACCAA CTGGCCCACT GAATAAATAA AGTCCTCGTT TCTGACAAGA TGCTTCAATT 3300

TCATTCAGCT GGTGGGGAAA GAAAAAATGT TCAGCTTGTt CACCaAATCG GTGCAAAAAA 3360

CGAATTACCA AGCTCTCTCG ATTCTGAAAA TCTCCCACAG ATGATAAACG CAAGCGTTGC 3420

TtCGTCTCGC CaATTAGATA AGTTGTCGCG CCTAATTGTG CTTTTCGTCT TTCTCCGACA 3480

TCCcATGCCG CcTATAAATT TAAAGTGGAA AATTAACTTT TCGCCTATTT CTGCGGAAAC 3540

TTGtGTGTAA AGTGTCCGTT GCATCCCTGT CCGAAAAGAG ATGGCTATTT CGGTATCATT 3600

GGGTAAAAAA TAAACGTCTT GCGCCCCATG AAGGACGCCC CATTCAATTA ATTGATGAGC 3660

CAATTTTTTT ACATTkTCAG TCATTCATTT CCTCCTCTCA CTAATAAACT ACGCATTTAT 3720

TTTATGATTT CAGCAAAAAA AAGAACCTTC ACCATTaAGA TGAAGGTTCT TTTTTTGAAG 3780

TTTGCGCTGT CCATCTATCT TCCCCAAAAT AGTGGAACTC CTAGACACTT AGTATTATAC 3840

TAAAAAAATC TTGCGTTTAG TTTTCAAAAA TTTCTCTGAA TCACAAAATT TTTGAGAAAA 3900

ACTATGATAA TCCGATTGGA TCGTTTACGT TTTTGGCTTC GTTTTTTGTT TATTAACTAA 3960

ATAAACTGGA ATTCCTAGAA GCGTAATCCC TATACCAATC AATGCTAGAA TAGGCTGTGT 4020

AATACTCGTT GTAACCAAGA TGAAGATACC ACCTAGACTT GCGATAATTG GAATAATTGG 4080

ATAGCCAGGA ACTTTATACG GTCGTTTTAA TTCAGGTTCT CGTTTTCTTA AAATAAAAAC 4140

AGCTAAGAAA ATTAATAGAT TAAACAGCCA CATAACAAAA ACTAACATAT CCGTCAATAA 4200

ATCGAAGGTT CCGAAAAACA TCATGATAAT GGCTACAGCT AACTGGAAAA TTCCAGCAAC 4260

GTAAGGGcAC AACAAAGCGT TTAGgACAAC TTTTGAAACT GAACGACTAA AAGGCATCAG 4320

ATTATCTAAG GCTAAGGGCA TAAGGG 4346

(2) INFORMATION FOR SEQ ID NO: 233:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10555 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

CTTCATTCGA TTAATAAATT CATCCCAATT TACTTCATTG CCtTTAAACT CAGGTCCATC 60

AACACAGGCA AATTTCATGG TGCCATCAAC AGTGACCCGA CAACCGCCAC ACATGCCTGT 120

TCCGTCAATC ATAATAGGAT TcAAGGAAAC ATAAATTGGT AATTGATGTT TTTCAGCTGT 180

CAAGGTACAA AATTTCATCA TGATGCTGGG CCCAATCGCC CAACTGCACG CTACAGATTC 240

ACGTGCAATA ACTTCTTCCA TCGCTTCTGT TACTAAGCCC TTTTTACCTA ATGAACCATC 300

ATCCGTTGTA ATGATTAGCT CATCTGaATA mCGACGGCAT TCTTCTTGTA ACAGAACTAA 360

CTCAGCTGTT CGTGCTCCCA AAATCGTAAT CACTCGGTTG CCTGCTTCTT TTAAAGCCtT 420

GATAATTGGA TAAATAGCCG CGATACCAAC CCCGCCACCG ACAACCATCA CCGTTCCATA 480

TTGCTTCACT TCTGTGGGCA TACCTAAAGG TCCAGTGATA TCCAACAAAG AGTCTGAAGT 540

TTTAAGTTGG GAAAGTGCAG CTGTTGTCTT ACCGATTACT TGGAAAATAA TTCGAATACT 600

GCCTGCTTGT GGATTTGTCT CCGCAATCGT TAAGGGGATC CGTTCTCCCT TTTCGTCAAT 660

TCTTAAAATA ACAAACTGTC CAGCCTGAGC ATGTGTCGCT ATTCTTGGCG CTTCCAGCCA 720

AAAATCAAAT GTATTACTAG ATAATTGAGT CGATTgAAGA ATATTAAACA TTGAAATACC 780

TACCTTCATC GTTCCTTTTC ATTCTGACTC CTACTTTCTT AAACCCTTAG GCAACGCTAT 840

CAGATAAAGT TGTGGCGACA CTTTTAACAA ATGTTTTTAG CATTTCGATG TCTTCTTCTT 900

CTGCATTGTT TTCAATTTCA ACTGTTTCAG CTACTTTTTT AGCACCAGCT TTTTCAAAAG 960

CTCCTACAAA ATCATGCGCT GATTTACAGA AAAATTCACC ATATTCACGA TCACCGGAGC 1020

CAACAACACC AAATGGCTTA CCTGCTAAAT CCTTCTCAGT TAATTCATCA AAGAAATCTT 1080

CAAATTCAAA AGGCAGTTCC CCATCACCAT AAGTATACGT TGCGATAATA CAAGCATCTG 1140

CATCATCAAA GAAGTCTACA TCAACTTCTG AACATTCTTC ACGTTCTACT TCAACTCCAG 1200

CTTCTTGAwA CTGTTCTTCC AAAATCTCGG AAATTTCTTC TGTGTTTCCA GTCATACTTG 1260

CATAAACAAT TTTTACTAGT GCCATCTTCA TTCATTTCCT TTCTTTTTCC TAAATATTTC 1320

AATGCATTCA AATGCAATTG TAACCCCTTA CATTAATCCT TAAAAAAATA ACTTTTACAA 1380

ATGATTTTAT TCACAAcTAA TCTAACTAAG TATGTACTCT TCCCCTTTCG AGATAGAGTA 1440

TTTTTTACTA CGTCTTTACT ATACCACATT CCAAACCTTT TTAACAATAT TTTTAAGGTA 1500

AAAAAAATAT TATTTAGGTT CTTTTTTATA AAACGTTTTT CCTTCTTCCA ATCATAAAAA 1560

TTGCTTCTAA AACCGCTCCC CCTAATGCAC GAGCTTTGTC GGAGATCGTC AAGCAATCGA 1620

CTTTCTCAAC TGGTCGAGGA TCTACATCTG CACACTTCAA CCCTTGATAA CATGTTACCT 1680

TTTCTGAAAT CAATCCTCTC AAAGTCCCTG TCAGAGGCGA ATAAACTGGA ACCTGATCGA 1740

TTAAAAATAA TGCTTCCCCT TTCAACACGA GATCCCCTAT GTTTTTAAGG TGGGTAACTT 1800

GTCCTGAAGC AGGGGCATGG ACAACTCGTT CCGCACTTTT TCCGCCAATT TCACCTGGAA 1860

TACCTGTATT AGGTAAAGCT GTTCCTTCAA AGTATAAACG TCCCAACCTA TGGCCACGCA 1920

TTGTCTCAAT CACAACATCG ACATCTTGCG GTGCTGAAAA GCCTGGTCCT AAAGCAATCG 1980

TAATCGGTGC CATTCCTCGA TGCGTACCTA AATTTTTCTT AGCCAAAATC GCATCGATAA 2040

CAATCAGTGG TTTCAATTGT TGAATAGCAC TTGCTGTTTG ATCGACAAAA ACAGGCAGCT 2100

TATTTTGCTG CCAACATTTC GCGCAATCCT TTAATGAAGT AATTCGAACA GCAACCAAAT 2160

CTTCCACTCT TTGTTCTTTC TGAAAAATAG CATCACAAAC TGAAACAGTC CGCCGAATGG 2220

CCAGAGGACA TTCTGTTTCT AATACTAGTA TTTTAAAGCC AACGTGCCAT AATTTTTGGA 2280

TAACACCTGT CGCTAGATCG CCCCCTCCGC GAACAATCAC TAATTTTTCT TGGATGTTTT 2340

GACTGATTTC TCCCATAAAA TAAGTCCTTT CTCTAATTCC ATATTGCAAG CAATAATCTT 2400

GTCTAATTTT TCAAAACAGC TTTTGGGTAA TTGAAGACAG ACATTCATCG CTTGAACGAA 2460

CTCTCTTCGC GTATGGGTTT GATTTAAACA TAAAATTCGT TGGCCTTGGC TCTCTTTAAA 2520

CAATCCTTTT GGATGATCAA TCAGTTGAGC TAACAATGTT TCTGTGATTG GTTCTTGTAG 2580

TTGCGCATCG GTTAATTCCA AAAAAAGCGG CAACCGATGA ATGGTCGTTT CATCTATCTT 2640

TTGACCAATC ACAGAAATCG GAAGCACACC AATTGTGGTG GTTGTTTCTT TTAATATCAC 2700

GGGTTCATAA GAGGACCAGC CTTTCAAAGG TCGTTGTTTG GACCCATCCG CTTCTAAAAA 2760

AACTTTCGTG AAGTGCTGAA ACAATGTTTG AAACGAGGTC ATTGCAGGCA TCGTtAGTTT 2820

TGTTTGGTTG TTGATAGCGG TTCCTGCTAA AGTGATTCCG CAGATATTCG GCGTTATTTT 2880

TTCAAAATGG TTTGTATAAA AATAATCATA CAATTGACGT TGGGGAAACC CAATTTTTGT 2940

AGTTGTACTA ACCAGAACAG CTTCCTTTCG ATATGTTTTT GCTAGCGCCC ACATTAAGCT 3000

TGTTTTCCCA CCGCTACCAA CAATCGACAC AATCTCCTTT CCTTGTAAAT CAAAACACTT 3060

TTTTAAGGAC TCCACAGTTG GTACCTCCTT TCCCCAAGGT CATTTTTCAC AGAAAAAATA 3120

GAgGCTACGC GTACTTTTGC TTCTTCTAAA AAACATACAC TGTGAGTTAT TTCCGCTTCA 3180

TTAATGAGTA AACCGATCCC ACAGCCATCC GTCAATGATG GGGGCAAAGT AAAGACGCGA 3240

ACAGCCACCT TTCTTGCCAA CAGTACTTGT TCCGCCTGAA CAGCATAATG TGTATTCGGA 3300

AAGGTTAGCA AATACTCCAT CTTTTCACCT TCTTTTTTTG TTAAAGCGTA ATTACTTTAG 3360

CCGCTTGATT CATTGTTTCA ACAATTTCGT ACATATTAGT AATAGAGCCG ATAGCTAAGG 3420

TGTCTGTTAA ATGATAAAAA TCTAAACAGG CCCCACACGT TTGAATCGTG GTTCCCTTCT 3480

CCGCTAATTG TTGCAAATCA GCTAAGGTAT TCGCTGCTTG ATTCGTTAAA AACGCCCCTC 3540

GATTAAAGAA TAATAGCTGT GTAGGAACAG TTTCTGCTTC TGATAATGAT TGAAGATAAC 3600

TTTTCATTAA TAGTCGTCCT AGCTGCTCGT CTCCAGTACC TAATTGATCG GAGCCAATCG 3660

TAATTACTAA ATCTGTTGCT GACTGACTTG TTTGTTGAGT CGCTTCTTCA GGTACCGTAA 3720

TTTTGACAGA GAAAACCGTG GCTTCTAtnT GctTCGCAGT CAAAGTGGCC TGTTTTTTAG 3780

TAGCTAATTT TTTTAAATTT TCAATGGCTA CTTCGTTATC GACTAATACC TCGATAGTGC 3840

CTCCAGCTAA ACCTAACTCC GCCAAAGCTT TTTTTGTTTC AATCACAGGT ATTGGGCAAG 3900

GTTTCCCCAA TGCATCTACT AATTTCATAT TTGGCCCTTC TCTCTATAAA AGTACTACTG 3960

CTTGTTGTTC TTTTGGCAGA ACCTCGCCAA TAATTTTGGC AACAGGGTCC TCTTTTTGAA 4020

TAGCTGCTAA ACAGGCTGCC GCTTCGTCAG CAGCGACACT AATTAACAGT CCGCCAGAAG 4080

TTTGTGGATC AAACAAAAGT TCTTGCATTT CCATAGAAAT TTTTGAATAA TCGATTGTCG 4140

TTCCGATTGT TTGGCGGTTT CTTTGCCCAG CTGCAGTGAC TAAACATTCT GCGGCGTAGG 4200

CCAACGCTCC TGTAATGGTT GGTAATGCTT CTGTATCAAT CAGTAATGTT ACCTGATTAT 4260

CAGCCATTTC TTTTGCATGA ATTAATAGAC CAAACCCAGT CACGTCAGTA CAAGCATGTA 4320

CTGCGTAGGA TCGAGCATTC GCAGCGGCAT ATTTATTTAA ACGTTCCATC GAATCTTGTG 4380

CGATTTTTTG TTCCGTCTGc TTCGCCAATT GCCCCCGAAC AGCCGCTTGA ATAATTCCTA 4440

CACCTAATGC TTTAGTTAAA ATCAAGACAT CCCCCACCTG TGGCGTATTA TTATGCCACA 4500

AATTTTCTGG ATGAACTTCT CCGGTTACCG CTAGACCATA CTTGGGTTCA TGGTCATAAA 4560

TCGAATGGCC TCCCGCCAGT AAAGCACCCG CTTCCTTTAA TTTTAAGCTG CCACCCAAGA 4620

GCATTTGTTC CAACCATTCT TTTGACATCT TTTGTGGAAA ACAAACAAGA TTTAGCGCAT 4680

AAAGGACTGT TCCGCCCATG GCATAGACAT CACTTAAAGC ATTAGCGACC GCAATTTTTC 4740

CAAATGATAA CGGATCATTG ACCATTGGTG AAAAAAAATC TACCGTTGAA ATCAAAGCTT 4800

TTTCAGACGT TATTTGGTAA ACAGCGGCGT CGTCGGCATG TTCAAATCCG ACGAGTAGCC 4860

GCTCATTTGC CGGTGTTTCG GGTAAGTGTT GCAACAATTG GTTCAGTACA CGGGATTCAA 4920

TTTTTGCCCC ACAGCCACCT GATGTACATT GTGATAAAAA ATCCATGCTG TCCGCTCCTT 4980

CTCAAATTAG TAGTTCTTGT AAAGCGTGAA TTGTCCAAGT AATTTCTTCA GGCGTTGTCT 5040

GCCAACCAAA GCTAAAACGT AAGGTGCCAG TTGCTAAAGT CCCAGCTGTT TGATGGGCTA 5100

ATGGTGCACA ATGCAAACCT GCGCGCGTCA TAATGCCATA TTGTTCTGCT AGTTGCTGTG 5160

CCACTATTGT TTCTTCCTGA TTCCACAGTG TGATAGAAAC AACAGGGACA GTTTGTGCCA 5220

CGTCCTTGGT TCCTAAAATT GTGACAGGTA ACCCTGATAA ACCATTTAAA AAATTCTGCA 5280

TTAAGGTTCG TTCATGCTTT TGAATCGCTG CTAAACCAAT TTTATTCAAC TCTTTAATAG 5340

AACTATTTAA GCTTAGAATA CCTAAACTGT TTAACGTACC AGCTTCAAAT TTATCTGGTA 5400

AAAAAGAGGG TTGATCAAAG GAATTAGAAT GGCTACCAGT GCCGCCCGCC ATCAGTGGTT 5460

TGACCGCTTC AGCCCCTCGC TCACTAAAAG CTAACCCACC AATACCTGCT AATCCATATA 5520

AGCTTTTATG TCCAGTAAAG GCTAACACAT CAATGGCCAT CTGGGTCATT TTTATTGGTA 5580

AAAATCCCGC CGTTTGAGCC GCATCCAAAA TAGTTAGTAA TCCTTTCTGT TGAGCCCATT 5640

GAAAACATTC TTCAATGGGT AAAATGGTGC CTAAGACATT GGATGCGTGT GTCATCACCA 5700

AGGCTTTTGT ATTCGTTCGC CAAGCTCGCT GAATATCTTC TACTTCTAAC AGGCCAGTTT 5760

TTTGACAAGC AACATATGTC ACTGAAATGC CTCGTTCTTT CTCAAGTAAG TGAAGAGGTC 5820

TAGCTACTGC ATGATGTTCT AACATAGTCG TAATCACATG GTCGCCAGGT TGTAAAATCC 5880

CCGCTAATGC AAGATTTAAG GAAGTTGTTG CGTTATTCGT AAACGTAATC TGAGCCGCAC 5940

TGGGTGCCTG AAAAAAGTCT GCTAACAATT GACGTGTCTC TAAAAGAGGA AGCCCCTCTC 6000

TAAGCAAAGG GGCATTTCTT CCGTAATTCA TGAACTGATT TTCTGCTAAA TAGGCAGTCA 6060

AAGCCTCAAT GGTTGCTGGA AATTTTTTAT AGGAAGTCGC TGCATAATTT AAATAAACCA 6120

AGTCCTTCAT ACGTTTTTCC TCTCTAACTA TTGGTTTCTT GAATTAATTT TTGGTACTCT 6180

TCTGGCGTGT CGATATCTTT CAAACGGCCA GGAGTCGCTA TTCTCAGCAT CTGAACTGCT 6240

TCTGGAAAGC GTTCATAAAT CATGCGTCCG CCTTTTTCAC CAGTCAATGT TAATAGTTCT 6300

TGTCGAAATT GATTACCGAA TAATACTGGA CTACTAGGAC AGCCATCTTT TTGTAATGGC 6360

ACTACAATTT TATTTCTCTG ACATTTATCT AACACGAGTT GTAGCATTTT AGGTGTTAAC 6420

AATGGTTGAT CGCTAGGTAA GTATAAATAG CCAGCTCCCC GAGCTTGTTC GGTTCCTAAA 6480

CGAATACTGC TACTTTGTCC TTCTTGCCAT TGATGATTGA GAATCAACTG AATGTTCTTA 6540

GGTAAAGGAA ACGCTGGCTG ATTCTCAGGC GAAACAACTA AAATCCGTTC TAATACGTCT 6600

AATTGATTCA CTAAGGTTAA CGTGTGCTCT AAAAATGTTT TTCCTTGATA TTTTAAGAAC 6660

AATTTATTTT GGCCCATCCG TTGTGAATTG CCCGATGCTA GAATAATGAC ACTAACGAGT 6720

GACATGAGTT GTCTCCTTTT ATCTTTTTTT CTTTTGATAA GGGGTATTTT CTAAAGGCAG 6780

ACTTGGTCGG AGCTGACCAT CCAAGGCAAA ATATGCGCCT GCTAAAGCTG GTGCGGTAGG 6840

GATCGTTGCT AACTCACCTA CGCCTTTAAT TCCATAAGCT AAACCTTCGT GAAGATTGTC 6900

AGCATGAACC AAAATTGTTT CGATTGGTGG TACTTGAGTG GCGTTCACTA ATCCTAGCGT 6960

TGCATATTTT GCTTTAACAT AGCCTTCTTC TAAGGCAAAT TTTTCAGTTA AAGCATACCC 7020

CATGCCCATT GCAATCCCAC CTTCAATTTG ACCTTCAGCA GCTCGTGGAT TCACCACTTG 7080

CCCCATATCA TACGCAGCCA CAAATTTAGC TACTTTTCCT TTGTCATCTA GAATAGCTAC 7140

TTCAGCAGCA TAACCATAAC CAGCATGACT AACTGGACTT TTTTTATCAT TAATTAATGG 7200

ATCCGTTTTC GCTGAATATT CCCCATAAAA TTCTTCGCCT TCCAAATCGC TTAAAGCACG 7260

TCCCATATCA AGCTCATACC GTAATTGCAT TGCCGCTCTT CTAGTTGCTT CTCCTGTGAA 7320

GAGTGACTGC CTCGAAGCCG TTGTCGTTCC TGAATTCGGT GTTCTTCTTG TATCAGGGGc 7380

TTCCGCAACA ATCATTTCTG GTGGTAAATT TAATGTTTCA CAGGCAATTT GCGTCGTAAC 7440

GGTTGCCATC CCTTGACCGA TACAAGCAGC ACTTGTGCGG ACATGTATTT TTCCTTCTTC 7500

AACAGAAATA ATACAACGCC CTGTATCAGG TAGCCCCACA CCCACGCCAC TATTTTTAAA 7560

GAAACTAGAA ATTCCAGCAA CTTCCGCTTG TTCATAGACG TCTTTAACCG CTAACAAGGC 7620

TTCTTTTAAC GCTGCATTTT TCGAGACCAA TTGTCCATTC GGTAAAGTAT CTCCTGGGGC 7680

GACGGCATTT TTAAAGCGTA TTTCCCATGG TGATAGTCCG ACTTTTTCTG CTAACAAGTT 7740

TAAATTACTT TCAATTGCAA AAGCCGTTTG GCAAACACCA AATCCGCGAA ATGCACCTGC 7800

TGGCGGATTA TTAGTATATA CAGCAAACCC TTCCACATCA ATCGTTtGAT ACTTGTAGGG 7860

ACCTGCCGCA TGTGTACAAG CCCGTTGCAA TACCGGTCCT CCTAAAGAAG CATAGGCCCC 7920

AGTATCCGCA TAAATCACTG CTTTCATTGC AGTTAAATTG CCTTCTTCAT CACAGCCGGT 7980

CGTAAAGTCC ATTTCCATGC CATGGCGTTT GGGATGAACC ATCAAACTTT CTTCACGGCT 8040

TAGCAGTACT TTTACTGGTT TCTTCAATAA CCAAGCGGCT AAGGATGCAT GATGCTGAAC 8100

ACTCATGTCT TCTTTTCCAC CAAAACCACC ACCAACAAGT TTTGATTGAA CATGAATTTT 8160

TTCCTTGTCT AGCCCTAGCA TCCGAGCCAC TTCCCGTTGT TCATCATAGA TACTTTGACC 8220

CGCACTAAAG AGTAAAATGC CCTCACCTTC AGGTTGTGCA ATCGCACATT CGGGCTCCAT 8280

AAAGGCATGT TCATTAAygG TACGGAATAA TGTyCTGTCA CTACAAACGC AGAwTTTTCT 8340

AACACTTCCT CTGCATgCCA CGTACCAAGT GTTCaTGaGA CAAGATATTG CCCTTTTCAT 8400

GaATAGCTGG TGCACCTTCT GCTAAAGCAG CTTCACAAGA AGTAAGCGGT ATCATTTCTT 8460

CGTAATCAAC TTCCACTAAA TTTTTTATCT CTGGTAAGGT TTCTTTTCTT TTTGAAACAA 8520

CGAGTGCTAC TGCATCCCCA ACATACCGTG TTATTTCTCC TTCTGGAATC ATTACATCCC 8580

AATCGGAAAT AAATTCTAAA TGACCAATTT TATTATTTCC TGGAACGTCT TTAGCTGTGA 8640

ACACAGCCAC ACAGTCAGGA TGCGCCAATG CTTGGGTACT ATCGATTGAT AGAACTTTTG 8700

CGCGAGGATA GGCACTCCGC AAAGCAGATG CATGaAGCAT TCCTTCTATG TCAATGTCAT 8760

CTACATAAAT CCCTGTTCCT AAGGTTTTCT CCACTGCATC GACCCGTTGG AAGTCTTCCC 8820

CTAATTTACC ATTTGAATGk TCTTCTGGAA TGCTCGCTTC TTCACAAAAC ATTTTAGCAG 8880

CTAACTGAAT GGCTTCTATT ATTTTCACAT ATCCTGTACA ACGACAAATG TTTCCTCGTA 8940

TAGCTTTCTG AATGTCTTCT TTTGTGGGTT CGGGTTTCTT GTTCAATAAT CCTTGTGCCG 9000

AAATAACCAT CCCAGGGATG CAATACCCGC ACTGAACTGC ACCCGTTTTT GCAAACGCAT 9060

AGGCATAGAC ATCTTTTTGC CGTTGGCTTA AACCTTCGAT TGTCGTGACT TCTTTACCAG 9120

CTAATTTTTC TAGAGTAAAC AAACATGCTT TGGAAGCCTT TCCATTTACT AACACCGTAC 9180

ATGCACCGCA AGAGCCTTGG TTACAGCCAT CTTTTGTTCC AGTTAGGCCC AGATCCTCCC 9240

GCAAAAAGTC CATTAATTTT TTGTTTGTTT CACAGGTTTC ACTTTGACCA TTTACAATAA 9300

TTGAATACAT TTCTTGCACC TCGATTTCTG ATATrGTTTA GTGTTCAATC ACCkrGATTA 9360

TACCACAAAT CTCCGTTCTA AAAATTATTT TTTAGCAAAA AAAGTTTTTT ATTTAATTTA 9420

AAGGTTCGCA TTATTGATAG CGCTGTTGAA TTAATTCTCC CGCGATGCTA ATCGCTAACT 9480

CTTCTGGTGT TTCGGCTGCA ATGTCTAAGC CAATTGGCAT GAAAATTCTT TCAATTTCAG 9540

TCATTGTATA ACCGACCTCT TTTAACCGTC GAACTTGTGA AGCAATTTTG TGTTTGCTCC 9600

CCATCACACC AATATATTTG GCTGGTGTTT CCAATAACTG TTTGGCAACG ATGAAATCAA 9660

AATTGTGTCC TCTGGTAATT ACTATACCAT AATCGTCCGC CGTTATTTCA ATAAACGCAG 9720

CTAAATTTTC TAAAGGAATC ACCCGTAGTT CTTCAGCTTG TGGAAAAGCT TGTGCCGTCA 9780

AAAATTCTGG ACGATCGTCC ACTACAACGC AATAAAACGA TAAATAGTTT AAAATAGGTA 9840

CCAATGCGCG TGCAACATGA CCACTTCCAA AAATATACAC TTTCCCTACT TGGAGGATCG 9900

CTTCTGTTAA AAAAATTATG TCCTCTTTTT GCTCACAATG GATACCTTTA GGTAGCAAAC 9960

TTAATTCCCC CTCCCAAGAA GTGGAACCAA TCAATTCGCC TGTTTCACTA TAAAAGGATA 10020

CTTTTCCTAA ATTGTTTTGA TCTAGTTCAG TCACTAACCA ACCTTGTTGG TTATTTTGAA 10080

AATGCGTTTC TATGGATTCA CACAACTCCT TTAAAAATGG CGTTTCCCAT GAAAGATAAT 10140

GGAAAAAAAT CGTGACATTC CCGCCACAGA CCATACCTAA ATCAGCAATA TTATTCGGGG 10200

CTAAAATAAA TTGACGTTGT GCATTTTCTT TCTCAAGAAC TAGCTTTTTT GCCACTTGTT 10260

CAGCCTGAAA TTCAACCGCC CCACCACCAA CAGTTCCGGC TATTCTACCT TCGGCGGAAA 10320

CCAGCATTCT AGCCCCTTGA CTTcGTGGTG TGGAtCCtGa ACTTGTACCa ATCGTTAATA 10380

ACaTCGTGTC TTGCTTTTCC AAtATTGCCT GtTTCAAACG CTGAAATACT TCTCTCATTT 10440

CAGTTCCTCC GATACTACAT CCTGTTCTTT AAATATTTCT TCAAAGCATG CCTCTACCTG 10500

TTGATACACT TTTTTCTCAA ATAAAAGATA TTGATCAATC AACTCTTTTC CTTGA 10555

(2) INFORMATION FOR SEQ ID NO: 234:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 752 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

CCCACTCAAC GTACATTTAA AACATACCCT ATTACTATAC AGGATTCCCT TTACTTTTTC 60

AAGATGTTTT TATGATTTCA AGACAATTTT TTTATCCGTT ATTTTCTTCA TGAAATGCGC 120

ATCATGCTCA ATCACTAGCA TTGCAGGCTT CACAGATAAG ATTAGCGCTT CTAATTGTTG 180

ATGATTAAAC ACATCCAAGT AATTAAGGGG TTCATCCCAA ATATAAAGCT CTGCTGATTG 240

AGACAATGAT TTGGCTACTT CGACTTTTTT CCGTTGCCCC ATACTCATTT GTTCAATTCG 300

ATTAGTGAAA ACGGnCACGC TCCATCCCAA GTTTTCGTAA GTTATTTAAA AATTGGGTGT 360

AATCTAACTG ATTTTTCTCT GCAAATTCGG aTAAAGTTCC TTGATTGTCT TCATAATCTT 420

GGCGCACATA AGAAATGGTT AATTGGTGaG CCAAAGTGGc TTCGCCTTCT GaATCCCCAG 480

AAAAATTaTC CAATAAAtAC TGaATCaAGC TCGrTTTTCC TGaGCCATTT TTTCCTGTTA 540

TTCCGaCAAT TTCTCCCGCG tTTATTGaAA AAGaAATTGG cGCmAACAGC CAATTTTTCT 600

CGTAGCCTAG ACGAAGCTCT TCCACCGTCA ATAATGTTTT GTGATGCGTT GGCTGATAAT 660

CCATTGACAA AGAATCAATA TACTCAAGAT CTTTTAATAG TnTTTCTTTT TCTGCTAAnn 720

GTGTTTCGGC GCGTTGTTGG AATATGTTTC GA 752

(2) INFORMATION FOR SEQ ID NO: 235:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2779 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

TTGATATCCA GTTTGTAATA ACCATGGCTA AAACAATCGA AACAATCCCT GGATCAAAGA 60

CTGTTAGCAT CAAGATCATA ACAACTAAGT TAGGAATACC AGAAAGAACT TCTAAAATAC 120

GTTGCATTAC TGTATCGACA CGACCGCCGA CTAACCCGGA AATCAGTCCA TACGTTACCC 180

CAATCGTAAT ATCTAGAATT GCCGCGATAA ACGCAATTAA TAATGAAATA CGTGTTCCCA 240

TAAATAAACG ACTTAGGACA TCACGACCTA AACCATCTGT TCCTAGATAA AAGTTTACGT 300

TATCTGGAAC GTTTGCTTGT GCATATTTAT CAACTAATTG ACCAGCAACT GTTGTTTTAC 360

CATTTAAACC ATTCACGCTT TCCAAGCCAG GAATTCGTGG TGGTAAGTTA ATGTAAGAAA 420

CATTTTGTTT CGTTGGGTTA TGTGGTGAAA CCCAAATGGT AATAATAGAG ATAAAGATAA 480

TAACTGCTAA GAAACTCATT GAGATAACTG CAGCTTTATT CTTTTTCAAA CGACGCCATG 540

AGTCTTGTAA GAAACTTAAA GACGGTGTGG CAATACGCTC TCTTTCTTTT AGAGTTGAAG 600

TGTTTAGAGG CTGAAACTCA TCTGCCGGAA TATCTTTAAT CGCTTGTGGT AGATTATTTT 660

TATTGACTGT TTCCATTAGC CTTTTGCACC TCCTGATACG CGAATTCTTG GATCAATTAA 720

ACCATAAAGT AAGTCAACAA TCAGAATTAC AACAACTAAC ATTGTTGAAT ACAAAATAGT 780

TACTGCCATA ATTGTTGGAT AGTCGTTTGT CATAATTGAT TTTACAAACT GTTCCCCAAT 840

CCCAGGGATG GCAAAAATAT TTTCTACTAC TAAAGAACCT GTgcATTAAA CCAACGGCTA 900

AAGGTCCTAA TAAGGTAATC AATGGAATTA AACTGTTTCT TAATCCGTGG CGGAATGCAA 960

CTTGCCAGCG ACTCAAGCCT TTTGCTTTGG CTAATTCGAC ATAATCACTA TGCAAGACCT 1020

```
CTACCATCTC GGTTCGAATA AACCTAGCCG AATCGGCCAT GGGACTCATT GCCAAGGCAA   1080

TAGTTGGGAA GATTGTGTAG GCAAAACCAT TCCACATAGC AATTGGTAAG ATTCTTAATT   1140

TCATTGCAAA AATGTATTGT AATAATACCG CGAAAACAAA GTTTGGAATA GAACGACCTA   1200

AAATGGCTAA CAATGTTGCC AACGTATCTA CCCAAGTATT TTGACGCATC GCTGCGATAA   1260

TCCCCAATAA AATACCGACT AACGTACCAA AGATAATCGC TTGTCCTCCC AGTTGAAGGG   1320

ATGGCCCAAT TCGACCAGCT AATAACTTAG CAACAGGTTG GTTTTTAAAT TGGAATGAAA   1380

TACCAAAGTC TCCTGTTACC AGGTTCTTTA AGTAAATTCC GTATTGAACA ATCACTGGTT   1440

TATCTAAGCC AGATTGTTTA TTCAACATAG CGATTGTTTC TGGACTAAGT TTTTCTTGGT   1500

TGGTATAAGG CGTACCTGGT AATAATTGCA TTAAGAAAAA CGTAATCGTT GCGATTAACC   1560

ACAAAGTGAT AATCATGAAA AAGATCCGTT TAAGAAAATA TTTTCCAAAA CTATTCAATA   1620

ATGACACCTC CGAGATTTTT TTGACTTCCT CAGTTATTTT CGTTAAAGTA ATGACTAGAT   1680

AGACCACAAT AATAGATAGG GTGAACACAT GAraACTTTT AAATGGCCAA tTaTTACAGC   1740

CGTTATTTCC AGTATCGGCA TTTTTCTTTA CTTATTAATA AGTAAAGAAC CTATCACTAC   1800

CTCCTCATTA TCAGATACCT yCTTTATCGT TTCGCTTTtC TTTTTAATTA TCGGCATTGC   1860

TTTATGGATT ATATCCTCTG GCTTTTTCGA TAACTTTCAG CGCTCAATGA AAAACGCGTT   1920

TCGTTTTAAA AAGAAAAATG AACCGAAAGA GTTTATCCCG CTTTCAGTCA TTGGCGATGC   1980

ACATCGCTCT TTCTGGCTAA AAACTGGGGG CATCTTACTG ATTCTGTCTT TAGGCTTTCT   2040

ACTAnTTTAC TTGGTCTAAG ACAAGATAGC AAGAGGACCG TTGCCTTTTA TTCCATGGCA   2100

AAGGTCCTTT TTTGTTTGCT ATCGATCTTT TGTCAGGTTA TTCTGAAATG TATGCCCATT   2160

TGTAGTCATA CGTTGCGCCA GCAGGATGAA CTGCGACATC TTTGACTTTT TCAGCACGTA   2220

AATGCGCTTC TGATTTTTGG AATAATGGTA CCACACCCAT GTCTCCCATA ATCGTTTTCT   2280

CTGCATTTAA CATATCATCC CAACGTTTTT CAGGATCTGT CGCATCTGCA CTACTTGCTG   2340

CTTTCACGAA CTTGTCAAAC TCTGCATTGT TATAACGTCC ACGGTTATAT GAGTTATCTG   2400

AAGCAAATAA GTCTAAGAAA CTACTTGGAT CAGCATAGTC AGCACTCCAA CCACCGATTA   2460

CTGCATCAAA GTCGCCTTTG TTTGAACGAT CTAAACGAAC AGAGAATGGA ACCGGACTAA   2520

CCGTTACCTT CACACCATCT AAAGCATCTT GAATTGACCC TTGAACAAAT TCAACTGTTT   2580

TCTTAGAAGA ATCTGCATCT GAAGAAAGAA TATCCATTGT TAAGGTTGAA ATCCCCAATT   2640

CTTTTTTCGC TTTTTCCCAG TATTCTTTAG CTTTTTTCGt ATCATATTCA ATTTGaCTAC   2700

CTGCTTCTTT AGCAAAGTCT TTGCCGCCAC TTGGGtCTTT TGCCATATCT GCTGGTACTA   2760

AACCGTTTGG cTCGATAGA                                                2779
```

(2) INFORMATION FOR SEQ ID NO: 236:

(i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 10955 base pairs
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

GTGGCAGCCC AGCAAGTCAA AGTGCGCCCG CACCTGATGA AGTGTTGGCG ACAAGTATTG 60

GGACGTATGA AAATCGAACA TTGTTGGTTC CAGATGGGCG TTTAGCCTTA ACTGCCTTGC 120

ATAAAGGAAA AGATTATCAA GGCAAACCAA TGTTTTATGT CCTGTTTGAA TTAACCAATA 180

CAACCGAAAA AACACAAAAT ATTCAGCTAA TGATTCAAAG TTTTATGGAA GTTAGTCAAA 240

CTGTCCACGG AAAAGCCCAA AATTTACAAT ATGCAGTCTT AACAGATTCA CCTTTCCAAG 300

ACAAATTAGA TCGTTTAGCA GATGAAATCA ATCCTGGTGA AACAATCCAA GGCGCTTACC 360

CGTATGAATT TATTAATGAG AACAAGCCGG TCCATTTTAA ATTTAGAGAT CGCCTTCTCA 420

GTCTGGATGA ACCCATTGCA TCAGAAGAAA TTACCATTAC AGAAAAAAGT TTGGTCCGCT 480

AGCTTAGCTA GCATTGGACC AAACTTTTTA TTTATTATTC AGACGGAAAA GATTCTCCTT 540

CTATAAAACG TGTGTAAATG TGTTCGCTCT CAGCAGCTTC TCCCCGCATT AAATGAATAA 600

TTTTCTCTGC AGCCACCATG CCCCATCGAT GTTTAGAATA GGCAATTGTT GCTAATCGCG 660

GTTGCACAAA CGCACCTAGT TCAGAATTAT CAAAACCAAT AATCCGAATG TCTTTCCCCA 720

TTTGGTAATT TGTTTCTGCA ACATATTTAT AGACACCAAT CGCCATTTCA TCATTAAAAG 780

CAAAAACATC GACTGGTTCT GTCTGTGGTT GACTCAAAAT TTTTTTAGCA GCCGCATAGC 840

CAGAAGGTTC GGTAAAATCG CCTTGGATAA TTTCATAAGG TATCCCAAAA CGAGTTAATT 900

CTCGTGTACT AACAGCCAAT CGTTCTTGGC TATCATAACC TTTTTCTGGT CCAGAAAGCA 960

GAAGAACTTT TTTTGAACCC ACATTGACAA ATTGCTCAAT TGCCTGCGTG GCACCGCCAc 1020

GGTTATCCAA GAGAACTTGC CGAATATTCC TATGCTCAGT GGTTCGATCA AGGACAACGA 1080

TCGAATGGCC TCGTTCAGCA AACTTTTCAA TTTCTTTTGT TGGAAATGTC CAATCTAAAA 1140

TAATTGCCCC ATCCACCATT TTTTCAGGAA TGAAAAGATG TGACTTTTTA CCGCTGCAGA 1200

CAATCATCTC ATAATCAAAC AGTGCTAAGC CTTTCTTAAT TCCCTCCAAC AATTCACCAT 1260

AAAAACTACC GCCATAATCA GCCAAATAGA CACCAATAAT ATTGGTTTGA CGACGTTTTA 1320

ATGTGCGAGC GGCCATGTTA GGAACATAGT TTAGCTCTTC AGCAATTGCT TGGATGCGCG 1380

TTCGTGTTTC TTCAGTTACC TTTGAACTAC CATTCAATGC GTAAGAAACG GTCGAGATTG 1440

ATACGCCTGC TTTTTTTGCA ATATCTTTAA TTCCTaCCAT GGACTACTCT CCTTTCTAAT 1500

GGTAAATTTC AATAACACTT GCTTCATTTA ATAATGAAAG CACCTCGTCT TTTTGAACTT 1560

CCATTCCCCC TTCTCGATAA GGGTTTTCTT CAAACTTAAA CGGAACCTCT TGACCATCGA 1620

TAACTATTTT CGCTGACTGA TTTGGATAAA ATTTATATGT GACAGGTTTA TCAAACAATT 1680

GGTAAGTTAG TGTTAAACCA GACAATTTTT CTGGTAACAT GGGATCAAAA ACTACACTTT 1740

GGGCCTTTTG CCGAATCCCT AAAACAGCTG TAATTAATTG ATTTAAATAA ATTCCCGGCC 1800

CGCTTGAATA AATTCGCCAA CCACCTTTTA CCCCAATACT TCCATCTTTT AGTTTACCAA 1860

AATTTTCTTG TGATTCATAC CTTGTTTTAA AATCGCCATC GGAACTACTA AAATAGGCAT 1920

TGGCTTGACG TAACTCTGCA TTGTGAACCA CTTCTTTTAA TTGAACTGGA TTGATTTGCA 1980

TTAATGCTTG CCAGGTTTTG TCTCGCTGAC CAATTTTCGC CATCGCTTCT GTATAGCGAA 2040

TGTGCGCATG AACATACTGT AAGCCAATTT CGCGACCAAA ATTAGCCGCC TGTTCTGCTC 2100

GTTTAAAATT AGTGCTCACA CCGCCTCGAT AGTTCGCTGG TTTGTTCATT AAGCGCACGC 2160

CATCAGGAAA CGTTAATTCT TTTTCAATAA TTTCTAAATG ATGACTTACT TGTTCCGGTG 2220

TTAATAATTC AGCTAACATA CTACGAGTCA TTGGCAATAA TCGATATTGA ATCCCTGTTT 2280

TTTGGTCATT TGGGTGAATC ATTTGTTCAA AAGTTTCTGG CGTTTCCATA TAAATAAACC 2340

CTGGAATCGT TTCATCTGCC AACATATATT TTTGAAAATC GTTCTTTATT CCTGCGACTA 2400

AAATTGCTAA GTGCTGACTA TAAGTAGCAT CGATTTCCGT AATTAAAGCC GCAAATTTTT 2460

TCAACACTTG ATACGTTAAG GCAACGGTCC AACTACTAGC CATTTGCTCT TTAAGACGAT 2520

TATCATAAGG TTGTAACGTG TCATCCCAAT CGCCGTCGCC ATAACAAGAT AGAAAAGTTT 2580

CTGGTAAAAA ATGTTGTTCA ATATAATTTA TTTCTTTTTT CAAATGTTCA AATAAAGATG 2640

CTGTTTCTTT TGTTTTGTGA TTATCTTTGC GACTTGTATA AGGGATTTCC GTAGCTAAGA 2700

TTGAATAATC TGCGGTTTGT TCCAAATAAT CAGTCACTAC TTTCAATGGC CAGACGATAA 2760

TATCGCCATG GCTTTCACTG GCTTTTTGTT CTTCGTATCG GTCAAACATA AACCATTGCG 2820

GCCAGTTGCC ATCGTCCTCA AATTGATTTT CAAATAGATG CTGAATAATG GAGGCCACGA 2880

TCTTAGGCTG TTGCGTCGCA AAGAAAAATT CTGTCGGTCC TTGCGACACA TCTCGTGTGC 2940

CCCAAGCTGC TCCACCATAC TGTTCTAAGC CATGTGGCGA CAAATAATGA ACGAGCATAT 3000

TATGAGTATA CCAACGAGTT AATTGATTCA TTTGCTGCAC ACTTGAGTCT GTGTGTGTTA 3060

ACTCAAAATG ATGTAACAAG CCATTAATGA ATGCCAAATA TTGACTTTCT GCTTTTTCAA 3120

AATCTAATGG TGTATTGACT AATGATTGGT CCTCTATACT GCCCGTAATA AGTAGTGAAA 3180

ATTTTTGGGT CTTTTCGATT GCCAAGACTA ATAATTCTGC TTCAGCTCCA CCAAATAATG 3240

TGCTATCGGT TACTTGAAAT GATTTATCAA CTGTCAATGC ATAGGTTAGT TCTGGGTAAT 3300

AATTTTCTGT GGCAGAATGA TTTCCCGTCA CTCGAAGCGT TTTGCCAGTT TGTTCTAATT 3360

GATAGGTTGG TACTTCTTCC GGTCCTAATA ACAAATGGTT AGAAACAGCC CACATATATG 3420

CATCCTTTTG ACTGGTCACT TCCAATTGAA TCACTCTTGA ATCAGGCGAA GTGAAGGTGC 3480

GTACACGAAT TGTATCCTTT GGCAACTTAT AATACCACGT AGCAGAGTTC AACCCCATTT 3540

CAAAAGCAGA AGGCATTGTT AGTAAGCGCC ATAAGCCATT CTCTAAAATG TAGATTCGTT 3600

GACCAGAGCG TTTCATAACG TTTAAAGCGT TTCGGCTATT ACTCATTAAT TTGTTCATGG 3660

TCGTATTTCC TAAAACTACT TGTGAATTGA AAATGCCATA CATATACACT GTCGTACTAA 3720

GCAAAGGTTC GTCAACCGTT AAGCCTTTAC CACTTAATAA AATGTGGCCA TGCGCTCTTT 3780

CCATGGCAAT TTCTTTTTCT TGAAGAACAA CATGATGGTA ATTTTCCGTG AAAAAGGAAT 3840

ACGTCTGACC TGCTACTTTT TCAATTTGTG TCTGAAGGGG ATAAAGTGCG GCCAATTCTT 3900

CTTCTGTTAA TGGATTTCCT GTGATAGGTT CTCCTAAAAG TTTCGGAGGA TAGTCCTGTG 3960

TTCCTTCTAA TGAATCGAAT GTCACAGTTT CATAAATAGC TTGTAAGGCT TCTTTCGATA 4020

AAAAAGGTTC AGCTAATGCT TGCGGATGTG TTTCTGTTAC CGCACCATAA AAAACAAATT 4080

CCTTACTTTC ATTCGTTAAT TGGACGGCTT CCGTTTGTAA AGCGGTGTAA GCAAATTCAT 4140

ATTGATAGAC TTCATTTGCT AAATATCTTT TCCCCAATGC TTCTGGTGTG TTTATTTGTT 4200

TAAATGAACG TCCGAAAAAT TGATAGCCGT CCGTTGAGTA GCCTACTAAT TTTTGAAAAC 4260

TCCCTTGAAT AAGGGCTGGA AACTTCTCAT TTTGCGGTTG ATTTTGTCTT GATAAAACAA 4320

CAAGCTCTTC GCCCTCTTGC CAAATGTGAT GATCAAGATA CTGCGACACA TACGCTTCAT 4380

TTGACTGAAC GGCTCCTTTT GTCGCTAAGC CAATATCTTG ACCATAAACA ATATCTACTA 4440

AGCCAGTTCC AGTGACATTA ACCCGCCAGA ACCATTGATT GCTTGGCCCC AAACGAAAGT 4500

CAACTTGGTA ACGAACCGCT TCAAATTCAC CCCGCCATGT CAATTGATTA TCTGAAAGAT 4560

AAAAATCACT TTGACTGTTG GAACCAATCA TTGGTCGATA ATTAATTGTG TCCCCTGTAA 4620

TTATCCGCAA ATAGAGTTGG TTCATACTGC CGTCTAAAGG ATTACCTTCT AATTGATTAA 4680

TCATCATGCC ACAGCTATTC AATTCTTTTA AATCACCTGT CTCTAGAAAT GTAGCAACTG 4740

TGTCTGCTTG CTTAAGCTTC TTTATCTGCA TAACTTATCC TTCCCATCTT ATTCATTTAA 4800

CGTGACTGTT GCGGTTTGTA CATCGCGACT ATTTCCACCA ATCATTACGA TAAATTCACC 4860

TGCATCACTG GTTACCTGAT TGTCTGAATG AACATAGCGC ACCATATCCT CTGAAAGAAG 4920

AAACTGAACT TCTTTTTCTT CCTTTGGTTC CAGCCAAACT TTTTGGAAAT CTTTTAACTC 4980

TTTAATTGGG CGAACCACTT GTCCCACCAA ATCACGTACA TAAAATTGCA CGGTTTCCTC 5040

TCCAGCGATT GGACTGTTAT TTCTTACTAA AACAGAGACC GTGACTGGTT GTTCTTTGGT 5100

AAATGTTAAT GCACTTAGTT GAATTGGTTG ATAAGAGAAA TGACTATAAC TTAAGCCAAA 5160

TCCAAAGGGA TATTTAGCGT AATTAGAGAC ATCTAAATAT TTAGAAACAT ATTTTTCTTC 5220

TGGTGCTGAC TCATAAGGAC GCCCAGTATT GTCGACATTG TAAAAAATTG GCACTTGACC 5280

AACCGTTTCT GGAAACGACA TACTTAAACG ACCACTTGGA TTGTACTGTC CCCACAAAAT 5340

TTCTGCTAGC GCATTGCCCC CTTCTGTTCC AGGTTGCCAA GCTTGCACAA TTCCTTTTGC 5400

AACATCTATC CCTTGCAAAT CTAAGGGCCG GCCACTGTAG ACCGTGACAA CAACATTGGG 5460

ATTTTTCTCA ATAATTTGTG CAACTACCGC TAATTGTTTT TCGGGTAGAC GAATATCACT 5520

GCGACTGGCA GCTTCCCCAC TCATCCAATC TGTTTCGCCT AGTGCCACTA CTACTTTATC 5580

TGCTTGACTA GCTAACGTTA AAGCCTCTTC AATGGCCGCC GCGGTCGGTT CAAAGTAATC 5640

AAAGGGTTCT TGGGCAACTA GAAGATTTGT GGTTAATTTT TGGGCCCCAG CCACGAGGGA 5700

AATGGCTTCT TCTTTCTTTC CTTGCCATGA CCAAGCCCCT AAAATATCTT GTGATTGGGC 5760

CCCTGGGCCG ACAAGCGCGA TTTTTTCTTG CGGTGTAAAA GGCAAAATAT TTTCATtTTT 5820

TAAAAGAACC ATTGATTTTT GAGCAATCTT TTTAGCTATT TCTCGATGTT CTTTGCTTAA 5880

AATCACTTCT GCTTCCGCTT GTTCATCCGC ACCACGATAG GGATTTTCAA ATAAGCCTAG 5940

CTCATTTTTT AGTTTTAAAA TTCGCATAAC TGCTTCATCA ATAAGTGTTT CAGCAATCCG 6000

GCCTTCTTCA AGTAATTCTT TTAAATAATC TGGATAACAC GTGGTCATCA TTTCAATATC 6060

TACACCAGCT TTGATAGCCA ATTCTGCCGC TTGTTTTTCA TCTTTGGCAA CGCCATGAGG 6120

AATTAATTCT TTAATCGCTC CCCAGTCTGA AATCACAACC CCTTCAAATC CAAATTCATT 6180

TCGTAATACG TCTCTAAACA GCCATTGGTT GCCTGTAGCA GGAATCCCAT CCaCTGTATT 6240

GAAGGACGTC ATGACAAGTT TTGCACCCGC ATCTAATCCA GCTTTATAGC CAGGTAAATA 6300

CATTTCTCTT AGCTGCCGTT CAGACATATC TACCGTATTA TAATCCCGAC CAGCTAAGGC 6360

ACCACCATAT GCCGCAAAGT GCTTCACGCA CGCGGCTACT CGCCAATTGT CTTCTTTTAA 6420

ATTTTCTCCT TGGTAACCTC TAACCATTGC AGCAGCTAAG CGGCTGTTGA GTAATGGATC 6480

TTCACCTGTT GACTCCATGA CACGTCCCCA ACGTGGATCG CGCACTAAAT CAACCATTGG 6540

TGAAAAGGTT ACATGTAATC CTGAAACAGC AGCTTCTTTT GCTGAAATCT CCGCCATTTT 6600

TTCAACCGCT GCTTCATCCC AAGTGCTTCC TAAAGCCAAA GGAATCGGAA AGATGGTTCT 6660

GAAACCATGA ATAATATCCG CCATTAAAAG TGTTGGAATA TTTAATCGAT GATTTTGCAT 6720

ATATTCTTTT TGAACACGTT TTGCTTCGTT TGCACCGGAA ACACCCAGTG TTGTCCCCGC 6780

CGTATCAATC GTTTCTTGAC TTAGTCCTAG CTCTGACATA GGGCCGGTCT TTTCTTCTGC 6840

TTTATCTGAA TAAAACGCTG CTGCTAATTG TAATAATTGA TTAATTTTCT CATCAATTGT 6900

CATTTCTTCG ACTAATTGAA TCAGTTGATT TTCTTTCATA CTCTCACCTT TCTTCAATAC 6960

ATACTTCTGC TTCTAAAGGA AAATGGTCGG AGGGAAAACG TCCATCGATA CTATCTGTTA 7020

AACTTGCGGT TTGCTGTACT TGCCAACCTT TTACATAAAT ATAATCAATT TCTTCTAATT 7080

CCGCCCAAGG TTTTGTGTAA GTAAAATTTT GAAATGTTCC TCTTGGGCCA TAATGAGTAG 7140

CTAAATTTTT GCTATCTTGA AATTTTTTTT GAACAAGTTG ATGAACCTCT TCCCCCGACT 7200

CTGCATTAAA ATCGCCCATT AATAAAGTTG GATACTGTGC AATTTTGTCA TGTAACTCCT 7260

CTAAGATAAC GGTCATTCCT GCTAAACGAG CATGAGCTGA AATATGATcC AAATGAACAT 7320

TGATGACTAA AAAAGGCGTA TTTTGAGTCG TTTCCTTAAA TAGTCCCCAC AAAGCAATTC 7380

TTGGACAGCC AGCTTCAGGA TGGATACTTG GCTGCTGAGG TGTTTCAGAT AACCAAAAAT 7440

ACCCTGTATC AATAGCCTGA ACTTTTTGTT CGTTATAGAG AATCGCTAGT CCTTCTCCTT 7500

GACCATCCCC TTCTCTTTCC GCAGAAAGAC AGGTGAATGT TGTATACGCC TTTAAATCCC 7560

TGACTTGATT CGGTCGTACT TCTTGAATAC AACAAAGGCT CCAATCATGA AAATTGATTA 7620

ACTGACAAAC AGCCTCTTTA CGAAAAGACC ATTGCCAATC TTGGTCATAC TCTGTATCTA 7680

CCCGGACATT ATACGTAGCT ATTTTCATTT TTAACCTCCT GCAATAGTTG TTGAATTTGA 7740

CGCATCTCAC GACGGATCTG GGTTGTTTCA CCGCCACCAT GGCCTTTTTT CCAATACACT 7800

```
CTGACTGCTT TTCTTGCCTT TATCTGATGA TAAGAAGGGA AAAAATCTTT CATCGGACAT    7860

ATCGGATCGT GACTCCCAAC AGAAGCAAAG ACTGGATTTT TAATTTGCGA AACAAAGTGA    7920

CGACTATCAA AATAAGGAAG GACCGCGGCA ATTTTTTCAT GCGCTTGAGG ATGCTCTTGT    7980

AAGTACTGAT GAATAATTTG ATACGAACCC GTTCCTTCTG CTACTCTATT TTCTAAACAA    8040

GAATGACTTG GAACATCTGC AAAAGTAGCT AAAATCGGGA AGTCGACTAA ACTGTTCACC    8100

ATTAACACTA CGCCACCCCC TTGGGAACCA CCACGAAGAA TGATTTGGTC CTGATTAATT    8160

TCAGTAAAGG TCCGCACCAC CTCTACGCTA CGTAAATGGT CGGCAAACAA GCGGCGTTGG    8220

TACCAATCTT CTTTGTCTAG AAATCCTCGT CCCATTGGTG CTTCATGCCA AATTGTTTGA    8280

TATGGCACGC GATCTTTGGT TAATCCCCCT TGACCTCGAT TATCAATGAC AAAACAAGCA    8340

CAACCAATTT GTGTCCAATG AGCAAATTGC CAAGGTTCTT GaATATGATT CATATAGCCT    8400

AAACAGTCAA TCACTATTGG CTGTTTCCCC GCCGTTTTAG GaAGTAACAA ATAGCCAAAG    8460

ATGCGCTCAC CTAGGAAACT TAAAAAGTCG ACCTGTTCCA CTTGAACATT TTCCAATGGA    8520

TAGTCAATGA CCCGCCGATG GATATCTAGC GGTATTTTTT GAAGTTCTTC CTTGGTCTTT    8580

TGCCAAAACT CAAAATACTC ATTCATGTTC CACACCTAAC ACTTTCTGAA AAAACGCAAC    8640

TTGAAGAAAA GCCGTTGTCC CAAATCCTAA CGCAATATAT AAAAACAAAT TAAGCTTATA    8700

GAAAAAAACA AACGTTAAAA GAGGTAAAAT AAACAAAATA ATCGCATTTG CACTATTACT    8760

TAAGACCACA TAAACCGTCT CACCTAGTGT CTTTTTAACG GACCAATTAA AAACGGAAAA    8820

GAGAAAAAGA AAACTCGTCA AAAATAATGC AACAAAACTT AATAGAAAAA TAGCTAAGAA    8880

ACCTAAACTA GTGGCTATTT TAAATAAACT TGTTACGACA ATACCGCCCA GTAAAAAAGC    8940

AGCGAACAAG ACACCTGTTA AAATAAAACC TCGCATGATA TTATTTTTTA ACGCTGTCCA    9000

AAACGTTCGC AAAACAGGAA TATCTTTTTC ACGATACAAA CGATAGATCG TTCGATACAA    9060

CGCAATTAAG CCGACACCGC TAAAGATGCC TGTTATACAG GTAATTAAAA AGAGAAAATT    9120

AATAATGATC AAGTTGGCAA TAAAAGAAAA AATCCGATAA TACCAACCTT CATATGCAAA    9180

TAGCTTATTC ATATTAAAAG AGAAAAGCGG ATTTCTCGTT GCTTTATCCG CTTTTCTCTC    9240

CTCCTCTCTT ACTTAAGTTT GTCTCTGTTT TCCGCCATTT TCTCAGATTT TATTTTTTCA    9300

ATTGCATCCC ATTTATTTGA TTTTAAGAAT GCTTTGTAGT CTTCTAAAGA TTTGTCAAAG    9360

GCTTTATCAT CTTTCGCACG AATCATTGAA ATAACGGTTG TATTTAGTTT GGTTTCAATC    9420

GCTTCATTCG CACGAGCTTC CGGCGTTCCT GCATCTGGAT TAATATTTTC AATTACGAAA    9480

TGTGGGACTA ATTTTCCTTT GCCCCATTCT TGCATTTGCG TTAAAGCACT TTCCATTGGT    9540

ACTTTTAGTT TATTGACACG GTCGTTGTTA AAGTATAGGA AACGACTAAT GCCATATTTT    9600

TTGTTGTAGG CATCATTGTC TGTTTGTTCT AATTTTTTCA CTTCTGGTAA ATAATCAATT    9660

TTTCCTTGAT CATTGTACGC ATAAGTAACT CCTTCAACGC CATATTTTGT TAAAATTTGT    9720

CCCGGTTCAT CAATTAAATA TGTGAACAGT TGAGTGACTT TTGCTGGATC TTTCGCATCT    9780
```

TTCGTAATGT AATTACTTAA CCAACCTGAA ATGCCGGTTT GATTTAATGT TGGTTTTCGG 9840

CCAGAAGTGC TACTTGGTCC ATCAATGGCT ATATAACGTG TGCCAGATTT TTTCATAAAT 9900

TCTGTGAAGT TCCCACCTTG ACCACTGGTT CCAGCAACGA GCATGGTTGC ATAATTTCCT 9960

TGTTTCACTT TTTCATCAAA CGTTGCCCCA TCATCTGTGA AGCTATCATC ACTAATATTG 10020

CCTGCTCGGT AAACATCATT AAATGTTTTT AACCATTCTA AATATTCTTT ATCTAAATTT 10080

CGATCATAGT ATTTACCATT TTTATCCTCT AAAGGAACAC CTAAGAAGTC TTGTAATTTG 10140

TCTAAAAATG GTCCTGCACC ATCGCCCACT GTGGTGAAGC CCATTGGGGT CATCTCAGGA 10200

TATTTTTCTT TAATCTGTTG CATGACTTTT TCAAAATTTT CTGGTGTTGA AACGTCTGGC 10260

TTGCCTAATG CATTATAGAC ATCTTCACGA ATAACAAAAT TATCATTTAC TGGGATATTT 10320

CCACTTTCAT AATCCGCTTT TGTATTAGAG TAATTAGGGT AACCATATGT TTTTCCATCC 10380

TCTAAGGCAT ACCATTTAAA CGTATCTTGG TTAACAACTT TCATTAAATA GGGGTCATAT 10440

TTTTTAGCTA AATCGTTTAA AGAATAGGCC CAAGAATCTG CTTTCAATGC TGCTTGTCCA 10500

GTTTTTTCAG TTAATGTCAC GATATCAGGC ATATCTCCAC TTGAAATCAT GGCATTTAAT 10560

TTTGAATCAT CGCCGGAAAT AAATTCAATA TCCACATTTA AGTCTTTTTT AATCTGCGCA 10620

GTGGTCACGT CTTTTCCAAA AGGCAGCGCC GTCCAATCTG AGTTAATATA CCATTTAATC 10680

TTGGTCGGTT CTTTCTTCTT ATCTAACTTC CAAGCAGGCG TCTTTTCATC TAATTCATAC 10740

CGATCCTTAA TTGAGACATC TGAGCTAGCT TTCTTTTCCC CACCGCTACA GGCCGCTAAA 10800

CTAACGACCG CCGTTGCCAC CAATAAACCG CCTAAAACTT TTTTCATATT GTTCCCTACT 10860

TTCTTTATTC TTTTACTGCA CCGGCCATCA TACCGCCGAT AAAGTGTTTT TGTAGTAATG 10920

GATAAATGAT TAAAATTGGT ATCGTTGTAA TGACA 10955

(2) INFORMATION FOR SEQ ID NO: 237:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6604 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

GCTAGtCAAG GATTAAATGA ACGTGCTGAG TATTTAGACA TCAAACAATT ATCGTCTATG 60

TATTCAATGT CTGTTCCTGA AGTTGAACAA AATTTTGTAA AAGACAAACG TATGCAAATG 120

ATCGAAAAAA GAAAGCCTGG TACTAGCAAG GGAAAAAGAT aCTGGCCtGc TAAAGAAGCt 180

ATAAAAAtTT GtAATGACAT CmTGAATcAT TGGGATTaAA GGAAAGGrCg ATACAGAAAC 240

AwTGGACAAC TTaGTAATTA TGAAAAACCA ACAAGCAGTA ACAACTAGTT TGCAAGTCGC 300

AGAAACGTTT AATAAAAATC ATCGTGATGT TTTAGCAGCG ATTGATGATT TAmAAGAGGG 360

GGTTGCGGAA AATTACGCAG GGCTATTTTA CGAAGATAGC TATATTCATC CACAAAACAA 420

```
ACArAAATAT AGACAAaTCA TTATGAACCg TGACGGATTC mCGCTACtAG CaATGGGATT    480
CACTGGTCAA AAAGCTCTTA AGTTTAAACT GAAATACATT GAAGCTTTTA ACCAAATGGA    540
GGAACTGCTT AAAACTCAAT CAAACTTACC GATTAATAAC ACAGAATTGT TATTAGAAGC    600
TGCGTTAAAA CATGAACGTG GATTGACTCT TGTAAATCAA CGTTTAGATA AGCTAGAAAC    660
AGAAACTACA ATTAATAGAA GCCAACAACG AAAGATACAA GGGCTAGTTT CATCAACTGT    720
TATCAAAGTA TTAGGTGGCA AAAAAACATT GGCTTATCAG GATTCAAGTA TTAAGCAATC    780
AGCCTTTAGT AATTGTTATA AACAATTGAA AGCATTATTC GATGTAGCAT CTTATGTAGA    840
CATTCCAAAA GTTCGATATG AAGAGGCTGT AGCTCTAATT CCTAGATGGA AGCCTAACTT    900
AGAATTACAA GCAAGAATTG ATATGGcTAA TGaTAATGGa GATATGnTTT AAAGAAATAG    960
GATAGTAATG ATAAArGGAT AAGCCTTAmC TTATCCCsGG TTwAAAGTAG TGtCAmCAAT   1020
AATAGCtTTT TGATCCATAT CACTTTTGAT GGATTCAATA ATTTCTAAAA CAGAAGATTT   1080
ATAGTAATAA GTCTTGCTTG ATGCGATTAC TTCATTTTCT TCGGTTTTAA TTACAAAATA   1140
GTATTTTTTA TTAGAAGCTT TCTTTATTAC AAAAAACATT TTGACACCCC GCTTTCAACT   1200
GAATTATATC AAAAAAATTA TaGGAGGGTA AATATGAnAG CAaTwCGTGA aGCaCGATTG   1260
ATAGcGCtTA TTATTGATGA TTGCGCTAGG TGTGTTATTG AAAAGCCACT TTTCAGTTCC   1320
AGTGTTGGCA ACAATAAGTG TACCTCTTTT TATCCGTTGG TTTTTCAACT GGGATGAAGC   1380
GGAGTATCAG CATTTTCAAA AAAGACAAAA TAAAAAGCCC CAATCGTCTG CCAACGATTA   1440
GGACACATAC AAAATGAACT AAGGAGAGTA TACCAAAATG AACGATAAAA TTCAAAAATT   1500
GATTAAAAAA CTAGCAAAAG AATGCCAGAA AGAAGATGTA GCTTTATCTT TGGCAGCTAT   1560
CGATTTAGAA GGAGAGATGG CAATATCTCA AGTTGGAAAA GGCACGATAG TAGCCATTGC   1620
TGCACATAGC CAATATACAC TAACAAAAGA AAAACTGGAA CAATCAGATT GTGATTGTCC   1680
AAAACATCGT TTATTAAAAG AGATGTACGG TACTGAAACA GAAACTACGA CTAAAAATAC   1740
ACATACTTTT GTCACAGATG ATCCAAACGA TTTGATGGAT ATATTATCGA AGATTTTTCG   1800
AGGTGAGTTT AAATGACTAG AAAAGAAAAG TTAAACCAAG CAAAAAGATT AGCTGATTTA   1860
TGGTACAAAC AACAAAAAAG TCAACTATAC ATTGCGCAAC AAAAAGAGCA CAGAGGGATT   1920
GCATGATGAA AAAAGCGACT ACGCCGCCAA GCAAATAGTC GCATACAAAA TTATACTAGA   1980
AAAATTGTAA CACAGAAAAG AGGTTTTGTG AATGAATCGT AGAGAAGCAA ATGCACTAGA   2040
CAGATATTTA ACAGAGCCGA CTGAAAAACC ACATAAGGAA ACCTATGAAG ATGATCCaGT   2100
GGATACTACT GATTATTTTG GAAATGaAAT TGCTGATGAA GACGGTGTGT TTGAGATAAC   2160
TTTTGCAATG AAATGTCTTT ATACAGGACA ACCAGTACTC ACTTGTAAAA AAATTGCTAC   2220
ACAAGATACA ATCGTTGATT TGATAGAAGA ATTAGGCGAA GAAAACGTGT ATTTAATTGA   2280
ATATGTGAGT TCAGGAAAAA GATATAAGGA GGGCTTATTG AATGACTGAA GCAACCAAAA   2340
CAGATTTTTC TAAGTTGAAT GTTTATCAAA AATTAGCGTA TGTGAGACAA AAAGCACCAT   2400
```

ATATTCAAAA AAGTAAACGT GGCCAGCAGT ACAGCTATGT GGGGTCAAGC GATGTACTAT 2460

CCGCATTAAA TACAGTCATA AATCAAGTTG GATTAATTTT GAAGCCAGAA ATTGTTGCTC 2520

ATCAAGTTCG GGAATCACAA GATGAGGTAT GGAAAGCGGA TAAAGTAAAG AAAGAGCCTG 2580

TAGCCAAAAA ACGTACAACG TATTTTACAG AGCTAGAGTT AATGATGACA TGGATTAATA 2640

TTCATAATCC TTCTGAGATT GTTGCTTGTT CATGGTATAG CCAAGGAGTA GATATTGAAG 2700

GAGAAAAGGG GGTAGGAAAA GCACTAACGT ATGCGGAAAA ATATTTTTTA TTGAAATTTT 2760

TTAATATTGC AACAGATGAT GACGATCCTG ATAAATACCA AAAAGAACAG CTAAAAAATA 2820

CTACAATTAC TGAACGGCAA ATTGATATGT TGAACGCATC AATTAGTAGA GTAGCCGAGC 2880

TGGCAGGGCA AGAATTTGAA GCTGTGAAAT CGTTAGCTAT AAATGATTCT GATTTGAATC 2940

CTAAAAAAGC ATTTGAAGAA TATAGTGCTT ATGATTATGG TGTTATTTCT AAATTGCTTG 3000

CGAAGTGGAT AAATTTTTAT GAATCTAGGC AGAAGGTTCA AGAAGATAAG AAGTGATTGA 3060

ATGATTGGAA AAATCATAAA ACACAAAGGA AATAAATTGG CGATTGAGTT TGAGGATGAG 3120

ATCAACTCGA ATTTCCTCAA ACTCCTAGCT AACAACGATG ACAATTTAGT AAAAGTTGAA 3180

CTATTAGATA ATCGACAAAT GTCTCAAAAA CmGaATGCaC TTTTyCACGT TTTAATaGCk 3240

GATATAGCTC GTTGGAGTTA TGACGAACCA AATGGATAGA GGAAGTTTTA AAGTATTAcT 3300

ACGAAGCTAA aAGTGGCGTG TATTTTGAAC ATAGTAAGGC TACACGACAT GAAGCAACAG 3360

AATGGATTAG TTTTTTAATC GAATTCATTT TGAAAAATGA TGTACCACTA GAAAAGAGAT 3420

ACCAATACTT GCTAGAAAAT AACAAATGGT TTTATTACTG CCTTAAATAC CGTAAGTGTT 3480

GTATTTGTGG GAAACATGCC GATGTTTGTC ATATCGAGGT AGTCGGTATG GGGCGAAATC 3540

GTCAAAAGAT TAATCATGAA ACATTTACTT TTTACGCTGG TTGTCGTCAA CATCATCAAG 3600

AAGAACATCA GATAGGTACT AAGAACTTTT TAAACAAGTA TCAAATTAAA CCAGTAAAAC 3660

TAAACGTTGA AGAACGTAAG AAGTTGAATA TCGGAGGTTA ACAGTTTGGT AGAGGAATTA 3720

CTTGAAAAAT ACAGGCAATT AACATCGAGT CAAAAACTAT TTTTTGAACT ATTAGCATTT 3780

GTCTATATCG GTTCAAGAAA TGGTAAAGGA ATAGCTATTG AAGCACAAAC AATAAAAAAA 3840

GTCGTTAACG GAGAAATTAA GCATAAATAT GTTTATACGG TCGTTGTTGA TGAGGAGGAT 3900

AACTAGTGAA TGAACATAGA GGATATTACG CCATTATCCC AGCGATTGTT CGCTATGATA 3960

ACCACTTAAA TGGGAATGCA AAATTATTGT ATGGAGAGCT AACGGCATTA GCAAATGAAA 4020

AAGGCTATTG TTGGGCAACG AATCAATATT TTGCAAATCT GTATAACGTT AGCAAGCGAA 4080

CAATTATTTC ATGGCTGAAA CAGTTAGAAG AAAGAAACTA TATAAAAATA CAAATTTTTT 4140

ATAAGCCAAA CAGCAAAATG GTAGATCGGA GACATATCTA TATCTTACCG TATCCAACTG 4200

ATACAGAATT TTACACCCCT AGTGAAGAAA ATTTCATCAC CTATGGAAAA AATCATCAGG 4260

AGGGGGATGA AGAAAATTTC ACTACCCCTA GTGAAGAAAA CTTCACAGAG AATAATACAT 4320

TAATTAATAA TACmAAGAaT AATACAAAGA ATATATATAG TGTTGAACAA AGTTCAACCA 4380

```
TGTCTGAATT ATTTGAAAaG GTTtGGAAAa CATaTCCGAr GAAAaCCAwT AAGAAAAAaG    4440
CCAAaGaGCa ATTCTTAAAG AAAATTAAGT CAGATGAAGA CTTCGAACGG TTTAAAACAG    4500
GATACAAAGC TTATCTTAAG TATATCAAAT TAAACGACTG GTATCATCCA CAAGAATTGT    4560
TCCGCTGGAT ACGTGATGAA CGTTTTAACG ATGAATACGA CTTGTCTGAA ACGGCAACGC    4620
AAGTAAGATA TTCAAATAAT CCAGTTAGAC AAGAGAAGTT GCCAGATTGG GTAAATGAAC    4680
CAAAAAAAGA AGAGGAGAAA CTATCACCAG AAAAGCAAGC TGAACTTGAT AGGCAAATAA    4740
AAGAATACTT GGAGGGGAAA TGATGCGAAT TATCCTGCCA ATTGAACGAA AACTGCAAAG    4800
TCGCCCGAGG TTTGCAAGAC GTGGGAATTA TGTCCAAACC TATGAAGATA GCGCTATGGG    4860
TGCCTATAAA CAAAAGGCAA AACCAGAATT GATTGAAAAA GGGGCTATTT TTGCGCATAT    4920
TACGTTTTAC ATCGCTACCC CTAAAATATC TATTAAGTTC TAAAAAGAAA CGCTTAGAAG    4980
TGAAATTAGA GCGGAAATAT TGCGATAAGA AACCTGACTT GGATAATTAT TTCAAAGCAG    5040
TCACTGATGC TGCCGAAGtA TTTTATATAA AACcGATGgT CAAaTtGCTG TAaTGGTTtG    5100
yCAAAAGTTG TACAGTATGC GACCACGAAC AGAACTAGAA ATTACAAGTT TGGAGGAACA    5160
AGTGTAGTGG GTAAAACGGG TAAACAGATC AAAGAGAATT ATTCGGTTAG AAAGCGAGTA    5220
AAGAAGATGC TTAAATTTAA aGaATTTAAT ATTCAATTAT yCGACGTTCA CATTTATATA    5280
AAAmGGCTAA GTAGTTTTAA AACCAACAAA AAATAGTATA AAAAAACTAA CTAGTTTCCG    5340
CTAGCTAGTT ATAGATACAT TTTCATCTCA GATAGAGACA ATTAATAATA CATTACTATG    5400
AAATTTAAGG CAACAAAAAA AGCCAGCTGA CCACTAGCTG ACTTATGTGG TAGAAACAAC    5460
TGTTTTCCGC CAGTTGTTCT TTAGGTGTGA GTTAGCACTT TCCCCAAATA AAGTGCTATT    5520
AAAAAGCCGA TTAAAAATCG ACAATAGATC TCTTGCAGAG ACGTGTACTA CTATTAACTG    5580
TTTACCAGAA GTTAATGTAG AAAAGGAGGA AATTTATTTC ATAAATAAAA TCCCCAAGAA    5640
AGTTAAAGTG ATTATACCAT TGAAAAATAG ATTTGAAAAT ACTTTCTATG CCAAAAATAT    5700
TTATTAGATG GAAAATATCA AATGAATAAA GATGATGTGA AAGATGACGA ATGGAAGATT    5760
GTAAAATATT GTAGAATTAT TATTGTTTTC TTTATAGTCA ACCAATAATT TGTTGACTAC    5820
AAACATGTAT GAGTGATGAA TATTTTGATA AATAAATAGC CAGTCGGTTT CCGCCGACTG    5880
GCTGAGAAGT GAaTAGCTAT TGGAATAGTA TyCTaGTATA ATTTATaTCa TATGGAGTCG    5940
CTGATGAGCA AAGAATAAGC ATTTCTTATA TATGTTTGGA TAAATaAAAA AAGCTACTTA    6000
GTTTCCGCTA AGCAACTCTT AAATGATGAT ATGTTTATTA TAnACTATTA TACCATAAAA    6060
GGAGCGATTT CACTTGATTC TATTGTTAAA AGAAGTAGAT TTTCGACAAA CAAAAGCGAA    6120
TGCTAGAAAT GTGTTGAAGA GTTTTAGACG TTTAGAGCGA ATAGCTGGTC GCTCTTTGAT    6180
AGATTTAAAA TCACCAATTA TTACAGATAT GCCTAAAAGC CAAAGTCACG GGAATAAAGC    6240
AGAAGATGCG CTAGTACAAT TAGTAGATGC AGAAGCAGAA AGAGACGCAA TTGTATCTGC    6300
GCTTATGGCA TTGAGCCTAA CTAGCAGACA AATTTTGCAT TATAGTTTCT GTGTGCAGGA    6360
```

TCATTACTCT AATaCAAGtA GCTAGGGAAG TTGGATATTC CGAAAGAAGT ATTCAACGAA 6420

TGAAATCAGA GGCTTTAATT GAATTTGCCG AAGCGTATCG AAATGGAAAA ATAATTGCAT 6480

ATAAATAAAA TTTTTGGCGG TTTTTGGCGG GAAGTTGGCG GTTTTAACCA TATTAGGATG 6540

TTAATATGGn AGTGGCGAAG TAAGGAGACG AGGTAAGCCT GCATTACCTA CCTAGCTCCG 6600

TTCC 6604

(2) INFORMATION FOR SEQ ID NO: 238:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29729 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

AGTAACTTGT GACAAAGTCG TGCATGGATG CACTAAAAAG ACACCCTATT GTGGTAGGGT 60

GTCTTTTTTT GCTGTTTAGA CAGCAGCGGT TGGAAGCAAC TTGTTTATTC CTCGTTCTGT 120

TTTTCCAACC AAATGGTTAC AAGTTTCAGG ACAATCCCGA CAAAAATCGG CACAAGAATT 180

TTTGTGACAA TCTCGTACAT GTGATGCACC TCCTTTCGAG GCAAGTCGCC AATTTGAGTA 240

TAGCATAAAA AGTTGTTCAT TTATTATTGA AATTATTATA AATGGAATAA GTACGTGTTT 300

TCTGGTTTTC GTAACTAGAG AACACGTGCT TTTTTCTTTC AACTTTTGGC ACAATGGTTT 360

GTGCCAAAGT ACGACTTTTT CCAGATTTGG AAAGCGGATA CAATTTAGTC AACAAGAAAA 420

GTTACTTGAG AGAAGGGGAA TAAAAATGGA ACAAGTACCA GAAAAAAATA AAGTATCAAT 480

TAAAGCAAAA GTCCAAAGAA TTGGCAGTAC CTTATCTAGC ATGGTGATGC CCAATATCGG 540

GGCGTTGATT GCGTGGGGCG TTTTGACAGC GCTATTTATC CCAGATGGAT ATCTGCCAAA 600

TAAAGCCTTT GCCACGATGG TTGGCCCAAT GTTGACCTAT TTAATTCCAT TATTAATTGG 660

CTACACTGGT GGGAAAGTAA TAGCAGGTGA TCGCGGTGCC GTAGTTGGCG CAATTGCGAC 720

AATGGGCGTG ATTGTGGGAA CCGATATTCC GATGATGCTA GGTGCCATGA TTATGGGTCC 780

ACTTGGTGGC TTTACAATTA AGAAATTTGA TCAGTATTTC CAACATCGGA TTAAAGCTGG 840

TTTTGAAATG TTAGTAAACA ATTTTTCAGC TGGATTGATT GGTTTTGCGT TACTTTACTT 900

GGTTTTTCAG CGATTGGCCC AATTGTCGAT GCGCTGACTG CTGCAATGGC ACGTGGTGTG 960

GAAATCATTT TAAATGCTCA TTTGATTCCA TTAACGAGTA TTTTTATTGA ACCGGCAAAA 1020

GTGCTCTTTT TAAACAATGC GATTAATCAT GGCATTTTAA CACCCTTAGG GACAGAACAA 1080

GTCTTATCGG CTGGTAAATC AATTCTTTTC CTATTAGAAG CAAACCCTGG TCCTGGGTTA 1140

GGTGTCTTAT TAGCATTTAT GTTTTTTGGA AAAGGTGCGG CAAAATCTTC TGCGCCGGGA 1200

GCAATAATTA TTCATTTCCT TGGTGGTATC CACGAAATTT ACTTCCCATA TGTAATGATG 1260

AAACCGTTAT TATTCTTATC GGTTATTGCA GGTGGTGCGA CAGGTAGTTT GGTTTTCCAA 1320

ACGTTAGATG CTGGTTTACG TGCGCCCGCT TCACCAGGCT CCaTTATAGC TATTTTAGCG 1380

ATGACACCAA TGGGTtCTTA CTTACCAGTT AwTTTArGtG TTTTGGCAGC AACTGCTGTT 1440

TCCTTTGCAG TATCCGCAGT CATTTTAAAA GCAGATGCGA AAGAAGAAAG TGATGAATTT 1500

GCCCAAAAAG TGAAAGAAAC CCAArCTGCT AAAGCAGCAT CAAAAGGATT AACTGTCACA 1560

AATGGTAGTG CTTCATTGTC AGGAATTCAA CAAATTATTT TTGCTTGTGA TGCTGGGATG 1620

GGCTCTAGTG CAATGGGTGC GTCTATCTTA CGGAAAAAGG TCCAAGAGGC TGGTTTAGTA 1680

CAAACTGTGA CAAATCGGGC CATTAATAAT TTGACCGATG AGGGAAATAC ATTGATTGTC 1740

ACACAAGCGG AATTGCAAGA ACGGGCCAAA CAGAAAGCAC CTAATGCAAC CTTTGTAGCA 1800

GTTGAAAATT TCTTGAACTC ACCACGTTAT GATGAAATTG TGGGCTGAAC TATCAGGTAC 1860

CAAAGAAAAA AACGACGCAC AACCAACAGT CTCCAAAGCA ACACCTGCTT CTGAAATTGA 1920

GGGAGATTTA GCAAATGTCA ATGAGATTCT TTTGGTTCAC GATGATCGTG TCGGGTCAGC 1980

AACGATGGGA ATGAAAGTCT TAGAAGAAAT TTTAGATAAA GAGAAAATTT CAATGCCGAT 2040

TCGAAAAATT AATATTAATG AATTAACTCA ACAAACACAG GCTTTAATTG TCACAAAAGC 2100

TGaACTAACG GAACAAGCAC GTAAAAAAGC ACCGAAAGCG ACACACTTAT CAGTAAAAAG 2160

TATGGTTAAT CCCCAAAAAT ATGAAACAGT GGTTTCGCTC TTAAAAGAAA GTGCCTAGAG 2220

AGGAAGAAAA CAATGGAAAA TCTTACGAAT ATTTCAATTG AATTAAATCA ACAGTTTAAT 2280

ACAAAAGAAG AAGCTATTCG CTTTTGCGGC GAAAAACTAG TCGAGGCAGG CTGTGTTGAA 2340

CCCGCTTATA TCGAAGCAAT GATTGAAAGA GACCAATTGC TATCTGTTTA TATGGGGAAT 2400

TTTATTGCCA TTCCTCATGG AACAGAAGAA GCCAAAAAAT TAGTGAAAAA ATCAGGAATC 2460

TGTGTAGTGC AAGTTCCAGA GGGTGTTAAT TTTGGCACCG AAGAAGATGA AAAAATTGCT 2520

ACCGTATTAT TTGGGATTGC CGGAGTCGGT GAAGAACATT TGCAATTAGT CCAACAAATT 2580

GCACTTTATT GTAGTGATAT GGATAACGTG GTGCAACTTG CCGATGCATT AAGTAAAGAA 2640

GAAATAACAG AAAATTTAGC CATTGCTTAA AGGAGAGAAT AAGAATGAAC GCAGTACATT 2700

TTGGAGCAGG AAATATTGGA CGCGGCTTTA TCGGCGAAAT TTTAGCTAAA AACGGGTTTC 2760

ATATTACGTT TGTGGATGTT AATGAAACCA TCATTCAAGC GTTAAAAGAA CGTAAAAGTT 2820

ATACAATTGA ATTGGCCGAT GCCTCACATC AACAAATTAA CGTTGAAAAT GTGACCGGGT 2880

TAAATAACAT GACAGAACCA GAAAAAGTAG TAGAAGCAAT TGCGGAAGCC GATTTAGTCA 2940

CGACGGCAAT TGGTCCTAAT ATTTTACCAA GAATTGCTGA ATTAATTGCT CAAGGAATTG 3000

ATGCACGTGC CGAAGCAAAT TGTCAAAAGC CGCTGGATAT TATCGCTTGT GAAAATATGA 3060

TTGGTGGTTC AACCTTTTTA GCAGAAGAAG TGGCCAAATA TTTGAAAAAC CCAGCTTATG 3120

CTGAACAATG GATTGGTTTT CCTGATGCGG CGGTTGATCG GATTGTTCCA TTACAAAAAC 3180

ATGAAGACCC ACTTTTTGTT CAAGTTGAGC CTTTTTGTGA ATGGGTCATT GATGATACCA 3240

ACCGAAAAGC CAAAGAGATT CAGTTAGAAG GTGTTCATTA CGTTGCTGAT TTAGAGCCGT 3300

ATATTGAACG AAAATTATTT AGTGTAAACA CTGGCCATGC TACAGTTGCC TATACAGGGG 3360

CGTTGTTAGG CTATCAAACC ATTGACGAAG CGATGCAGGA CGCCTTAGTG GTAGCACAAC 3420

TCAAATCAGT TTTGCAGGAA ACCGGTAAAC TTTTAGTGGC TAAATGGAAT TTTGATGAAC 3480

AAGAACATGC AGCCTATATT GAAAAAATTA TTCAACGTTT CCAAAATAAA TATATTTCAG 3540

ATGCTATTAC ACGTGTAGCA CGGACACCAA TCAGAAAATT AGGTGCGCAA GAACGGTTTA 3600

TTCGACCAAT CCGTGAATTA CAGGAACGCA ATCTAGTGTC GCCGCATTTA TTAGCAATGA 3660

TTGGTATTGT CTTTAATTAT CATGATCCAG AAGATGAACA AAGCCGTCAA TTACAGGAAA 3720

TGCTTGACCA AGAAAGTGTT GATACAGTGA TCGCTGAAGT AACGGGCATT GAAGATCCAG 3780

AAACGGTTAA AAATATTAAA CAAAACGTAG AACGCTATGC GCGACCACAA GTAGCATAAT 3840

AACAAAATCC TTCTACCAAG ATACTTCACA TTTCTTAATT AAAGAAAAAA ACAACCGCTC 3900

ACTGATTAAA GTGAGGGGTT GTTTTTAACA TATTGACTAA GTGGTTTTTG ATAAAGCTCT 3960

CCTTTGAAGG CACCTGTATT GAAGTATTCA GCTAATTGGT AACCAATGAA TGGGCCAGTA 4020

GTGAGTCCTG ATGAACCAAG TCCGCTAGCA ACCACAAGGT GGGGCATTTC TGGCAAAGGG 4080

CCAAAAAACG GAGCGAAATC AGATGTATAA GCACGTGTAC CGACGCGGTA ATGCATGGGT 4140

TGCTTAAAAA GCTGCTCTGC TTCTTTTAGA AATGGTGCAG TCCCAGAAGT TAGTTGTTGA 4200

AAAGCGCTAA CAGTCTCTTC TAAATCCCAG GCTTGTTCAT TTTCGTGGGT GGCACCTAAT 4260

AAAATTTTTC CTTGATTAAA AGGAATTAAA tCnGctTCGC CATCTAACAT AGCAACAGGC 4320

CATTGCTGAC TGTTTGTAAA AGGGGTCTCG AAAACTAATA ATTGTCCTTT TTGGGGACGA 4380

ATGTCCACAG ATAAATTTAA AGTAGTGAGT AATTCCTTTA AATGAGGACC AGGTGTTAAA 4440

GCGAGAAAAT CAACGCTCTT TTTTTCTGAA TTTGTGACAA TTTCCCACCC TTGATTCAGT 4500

TGTCGAAAGT GCGCCCGTTG TGAACAAAAA GTGACCCCTT TGGCTTCGGC GCGTTTCTGA 4560

AGATGGTTTA GATAAGCTGG TCCATCTAGG CGACCGCCGC CAGAAATAAA AAGTGAGGGA 4620

GTTTCCTTAA GTAGCGGTAA GAACTTCGCT GTTTGAACCG CTGACAACAT CGTGATTTCT 4680

CCAATTTCAG GTGCAGTTTG CTTACGCTCT TCTGCCAAAT GGGCTAAGTC TGCCAAAGCG 4740

GCTGCCGGGC GCAAAATGAC TGTCCCTGAT TGACGATAAA TATCCTCCGT CAGCTGAAAA 4800

TCTTTGACTA GCTTTGGAAA TAGGGCAGCG CCCTCTCGCG CTAATTGGTA CCATTTCTTA 4860

TTCCGACGTT TGGATAGCCA AGGGGAAATA ATCCCAGCGC TCGCTTTAGT CGCTTGGCCG 4920

AGCCCTTCAT CATAAACGGT AATGTCAAAT TTTTGTGGAT CTAAATAATT AGCTAACGTC 4980

ATTCCGATAA TGCCGCCGCC AATAATGGCA ATTTTTTTCA TGCTTCTTCT CTCCTTCGAA 5040

CAGTTGTATT GTAACACAAA AAAACCGCCT CTAGACAGAG ACGATTGGGA GATAACTTGT 5100

TAACGGGTCC TTCTGTGTGT GTAAATAGAA TTGCCGCCTT CTAAAACGTG AACAATTTCT 5160

TGTGCCATTT CTGCCACAGG CACGGGATCA TAGCCAATAA ATAAGTTCTT TAAAAAGGGG 5220

ATTTTCTTAA AAAAATAAAT ACATTTACTA AAAAGTATGG TGCCGGTACG TTCTTGACCT 5280

ACAAGTAAAC TAGGATAAAA AATTACTGCT TCGGCGACTT CTTGGTGGAT TAGTTCCTCC 5340

GCTGCCAATT TGGCATCCAT ATATTTTCTC AAAGGGAAAG GAACAGTATT GGCTGAAATA 5400

AATAAAAACT TAGCTGGCTT TTTTTGTTGT GCTAAAAAAG AGCTGATTTT TTTCACTGGC 5460

GTCAGAATTA GCCGTTGATA GGTTTTCTTT TTTCGTGGAT TTTCAAATAA AATACCTACT 5520

GTATCAATGA CCCAGTCAGC ACGCTGGACT TGTTCTTGCC AATTCGTGTC ATGAGTAACA 5580

TCGGAGCATA CCCACTGCAC TTGGTGGCTC CAAGATGCGG AGAGCGTGGC AGGCTTTCCA 5640

TGACGTGAAA CACTGATGAC TTGATGTTGG CGTTGGACTA AAATTTTCGC GATTTGTTGA 5700

CCTAAAAAAC CGCTACCGCC AAAAATAATG ACTTTCATGG CAAGCCCTCA CTTTCAATTG 5760

TTTCTTCTAG TGTACCTTAA ATAAAAAAAG AAGCAAAAAA TCTCGATTGC GCATAGTTTA 5820

GTTATAAATA ACATAAAAAT AACCTTTTTA GATTGTTGAA ACACAAAGAC CCAACAATGG 5880

TTATTTATTG TCGTGAAATA AATAAAAAAA TGAACAAAAA TGAAATAAAT CAGACTAAAG 5940

GCCTAGAGAA ATGCCCTTGT TTCTTGAATT CAAGCATGGT TAAATAGTTA TAAGTGAACG 6000

ACCGAGCAAT TTAGTTTGTT GGGGAAAATA GACAGAATGA GGAGATGAAA CGAAGAAAAT 6060

GAAGAAGAAA ACGATAATTA TATTGGGGGC AGTTGCGGTA ATTGCGGTTG GGGGCATCGT 6120

AACTGTGAAT GCGTTAAATA AAAATGCACA ACAAGTAGCT GTCAAGCAAG CGCCTAAAGA 6180

TGACTGGGGA ATTGACTATT TTGACGTTCC CGACTTGCAA CAAATTTATA TTAACGGTGT 6240

CATCCAACCG GAACAAATGG AAGCCTTTGC GCGTGATCAA AAAATAACAA AGGATCCAGA 6300

GATTAAGGTG AAAAACGGCG ATGTCGTAGA TGCAGGCACA GAATTATTTA CTTATGAAGA 6360

TGAGGCGGTC ACAAAAGAAA TTGAGGCACA ACAAAATAGC TTAGCCAAAT TAGAAACGAA 6420

GCGGGCGAAT ATCTATAATA AGTGGAATCG GGCCATTGAT AAATTTAATA AAACTAAAGA 6480

AGAAGACCGC ACGATGTCTG GTGATGATTT AAATGAACAA TATCAAACAG AAGTCGATGC 6540

AGTAGATGAA GAGATTACCT TCACCAATGA AACCTTAGCG GATTTAGGAG CGAAgcAAtA 6600

TATTTCCACA AAGGCTAATT TCAAAGGTCG TGTATCAATT CCAGAAGTAA AAGATGCCAA 6660

TTCACCGATT TTACGGTTAA CTTCAGAAGA TCTTTATTTA GCTGGAAAAG TGAATGAAAA 6720

GGACTTGACT AAAATTAGTG TTGGGCAAAA AGCTAAACTA ACTTCTGTTT CCAACAATGT 6780

GGTTGTGGAT GGCTCAATTT CTTACATCGA TGATAATCCT CCTGAAGGCA ACAGCGATGC 6840

CGCGAGTGGC AATCCAGAGG GCGGCACAAC GATGTCTAGT TATAGCGTCA AAATTGCGTT 6900

GGCCAATTTA GACAAAGTCA AAAATGGCTA CCATATGCAA GCAACCATTG ATTTAGGCGA 6960

TTTAGGGGCG ATTGAGTTAC CGAAAAAAGC GATTCAAAAA GAGGGTGAAC AGGCCTACGT 7020

TTTAGTGAAT GATTTTGGAA CCATCATTCG TCGTGATGTC CAAGTCGGGC AAGAAAATGG 7080

CGACAAAATG GCGATTGAAT CTGGCTTAGA ATCAGCCGAC CGAGTGGTTA TTTCTTCAAA 7140

AAAACCAGTA AAAGTCGGTG ATATTGTTGA ATCAGATGCA GCGATTGCTT CTGATGAATC 7200

AGCAACCAAC GAATCAATGA CAGATGCGTC GAAATAGGAG GGCTTGATGG TGGAAGCAAA 7260

CAATACAAAT CAACCCCTCA TTGAATTGCG GGACATTAAC AAGTTCTATC CAGTAGGTAA 7320

GGAAAAACTA CATGTTTTAA AAGAATTGAA CTTGACCATC CATCAAGGAG AGTTCATTTT 7380

AATCATGGGG AAGTCCGGTA GCGGTAAGAC AACGCTGATG AACATTATTG GTTTCTTGGA 7440

TCGCTTAACA GACGGAAGTT ATCATTTCTC AGGAACAGAT GTTTCAAAAT TGTCTGAAAA 7500

TAAAAAATCA GCCTTTCGCA ATGAATATCT AGGATTTATT TTTCAACAAT TTTTCTTGAT 7560

TAATTCTTTA AATGTTAGTC AAAATGTTCA ATTGCCCTGT GTTTATGAAG GGAAAAAAAG 7620

CCGCGCTGAG AAAAAAGCGA TTGCTGAAAA ATACCTAAAA ATTGTCGGTT TGGAAACAAA 7680

AGCCAAATCA AAAGTTACTG AACTTTCTGG AGGACAGCAA CAAAGAGTAG CCATTGCTCG 7740

TTCCTTGGTG AATGATCCAT TATTGATTAT GGCTGATGAA CCGACTGGGG CATTGGATAG 7800

TGAAACGGGT ACAGAAATTA TGGAATTATT GAAAGAATTG AATGAACAGG GCAAAACAAT 7860

TGTCATGGTG ACACATGATG AAGATATGAA AAAATATGCC TCACGAGTGA TTCATATGAA 7920

AGATGGTCGC TTTTTGGAAG AAGAGGTGAT TAGATGATTG CTAATTTATT TGTTAGTACT 7980

TTTTTAAGCT TAAAGGCACA TAAGTTGCGG GTATTTTTAA CGATGGTCGG GATTATTATT 8040

GGGATTACAT CAGTCGTAAC GATTTCTGCG CTAGGTGAAG GGATGAAACG TCAAGTTGTG 8100

AAGGCATCCA GTGCAGTGAA CGCGGATGTT TTGAAGATTC ATTATACAAT GTCGGATGGT 8160

AGCAGTGACA ATTTTATGTC CTATGAAGAG CCAGATTATA CGTTTAGCCG AGTGGACTTG 8220

AAAAAATTAC AAGACATTCA AGGAATTGAA AGTATTTATC CTCAATATGG TGAATCAATG 8280

ATGGGGGGCG GAGACAATTT GTTTGTACCA ATGGATTATT TTGGGGCGCA AGCGAATTTA 8340

TCCATTACGT CGACTAAAGG GCAGAATGAT ATTTTATATG GTCGTGATTT TCAACCGAGC 8400

GATGCAAATA CAGATGCGAT TGTCTTAAAT CATGATATCT TTGAAGCGCA AATTCGGCTT 8460

GATGATCCTA GCCAACTCAT TGGAAAAGCT GTTTCAATTG GTGGTTATAT GTACAAGGTG 8520

ATTGGTATTT TAGCACCAAA AGATTTAGAT AGTTTGGGTA TGAACGATGA TTGGGCAACT 8580

GCTATGAGTA GTTTTGTTTC TCGAGAATCA TACAATAAAT TAGCGAAAAC AAAAGCAATT 8640

AGTGGCATCA ATATTAAGGT TCGTGAGGGA GCGGACCGTG AAGCAATTTT AGGACAGGCA 8700

ATCACCATTT TAAGTGAAAA TCATCCTGAA GTGAAAGGAA CGTTTAAAGA AAACGACCAA 8760

GACCAACAAC TCCAACAACA AATGGAAGAA ATGGTCACGG GTATGACAAT GTTTCTGATG 8820

GCAATTACTG CAATCTCCCT ATTAGTCGGG GGAATTGGCG TCATGAACAT CATGTATGTG 8880

TCTGTCACGG AACGGAAACG AGAAATTGGG ATTCGTCGGG CCATTGGTGC TAAACCACGA 8940

GTCATTTTAT TCCAATTTTT AATGGAAGCA GCGTTTATCA CATTGATTGG TGGTTTGATT 9000

GGTGTCGGCT GCGGCTACTT ATTGGCAACT GTGGTCGGTG GGTATATTTC GATTACGCCA 9060

ATTATTACGC CGTCCATCTT TGCTATTTCA ACATTGGTTT CGGTCTTCAC AGGTATCTTT 9120

TTCGGAATTA TTCCAGCAAT TGGTGCTTCG AGGATGGATC CGATTAAAGC GATTTATAAT 9180

TAAAAAGAAG AGCTTGGTAG TCACATCTTT TTTGATGATC CTATCAAGCT CGCTTTTTTA 9240

TTTCATGTCA CTAAGAATTT TCTGAATCAA TTCTTGTTGA AACTCTTTGG TATATTGTTC 9300

GGTGAGTTTC GCTTGTTCTT CTTGGGACAT TGATGGATTC GCCATCATAG CAGCCATCAT 9360

TTTTTTGCCA ACTGTTTCAG TCCCTTCTTT TTTGATATCA GCGGCATGTT GTTCCATATA 9420

TTCGCCTAAT TTTTTGGCTT GATCAGCACT TAAAACTAGC CAGTCACTAG CAACGTTGAA 9480

GGTGTTTTCT TGTTTTGCAA ATTGGTTATC GTTGTCAGCA AGGCCAAATA GTGTTTTGGC 9540

CCAAGACTCT TGATAGGACC ATTTCGTCGT TTTTTGGACT AATGTGGCTT TTTGAGCATC 9600

TTTTTGAACC GTGTTTGTGA CGTTTTCCGT TCCTGTTTTT TGTAGTTTCT TTTTATCTGG 9660

TGTGTTATAT AAATAAACTT TTTCGGAACC ATCGCCCAAC GGCTGGTACA ATAACATGTT 9720

GGGGCCATCT TTGCTTGAAC TACTGCTTTT TAGTTCAAaC GTTTTGtCCT GAGTAACTTC 9780

GGTCATACCA AAATGATCTT TATCGTTTAA AACGATAAGA CTTAACGAAG CTAAGAACCC 9840

TAAACCAAAA ATAACCATAC AGAATTTATG CCAGACTTTT CGGCCAAATA CATTCGTATA 9900

GGCAAAGCcT AACACGCAAA GGATTAAAaG GAAAaTAATC ATTATTTCAT TGCACCTACT 9960

TCCTTAACAG GCATCGCTTC TTTTTTATTT AAGAAGAAAG TAATCGCCAA GCCGAGAACG 10020

CCAAAGATTG TTGCCACGAA GAAGGCTGCG TGATAACCGT TTAGTGTTGC GTTTGTTGCT 10080

TGATCGCGAT AAGTGAGTGG GGCTGTTTTT AATAAGTCCG ATGCAGGTAA GCCGTCTTTG 10140

GTAACATTTG TTAAAACACT AATCAAAACA GCTGTTCCAA TTGAGCTGGC TACTTGTCGG 10200

AAGGTGTTGT TTACAGCAGT CCCGTGACTT AATAAATTCA TTGGTAAAGC ATTCATGCCA 10260

GAAGTTGTCA CAGGCATCAT GACCATCGAA ATACCAAACA TCCGAATCGC ATATAGGACA 10320

ATGATATAAA TAATAGGTGT TGCTTTTGTT AAGAAAGCAA ACGGTAACGT TGCAGCTGTT 10380

AAAATAAACA TCCCAGTAAT CGCAAGACGA CGGGCACCAA TTTTATCAAA AATGCGACCA 10440

GTTACTGGCA TCATAAGTCC CATGACTAAC GCTCCTGGTA AAAGCATTAA CCCTGAATGG 10500

AAGGCAGATT CGCCACGAAT ATTTTGGATA TACAAAGGTA AAACCATTTC GGCGCCAATC 10560

ATTGCCATGT TGGTTACGCC AGATAAAATG GCCGCAACAG TAAAGACTTT TGATTTAAAA 10620

ACGCGTAGTT CTAAGAAAGG TTGTTCTAAA TGCAATTGTC GATATGTAAA AAGACATAAA 10680

ACAATCGCAC CAACAATTAA GAATCCATAG ACTTGCGGGC TACCCCAACC TTTATCACCA 10740

ACGCTTGAGA AACCGTAAAG TAAGCTACCA AAGCCGAGAG TAGAAAGAAT GACCGATAAG 10800

ACATCGATAC TTGGATTAGA TAATTGAATG ACATTTTTCA TTAAAAATGA AGCTAAAATT 10860

AGGACTAAAA CGACAATTGG AATGACCATC CCAAAAAGGT CACGCCATGT ATAAGAATCA 10920

ATAATCCAAC CAGATAATGT TGGACCTAAA GCGGGTGCTA ATCCGATAAC AATTCCCGTT 10980

GTGCCCATGG CAGCGCCACG TTTTTCAGGC GGGAAAATTG AAAGCATAAT TGTTTGTAAA 11040

AGAGGCATAG ATACACCGAC CCcTGCGGCT TGAACAAGAC GTCCAGTCAA TAACATGCCG 11100

AAGTCTTGCG CTACAAAACA GATGATGGTT CCGATTAAGA ACGTTGACAT CGCTGTTATG 11160

TATAATTTTT TTGAACTGAA TTTGTTAATC AACCAAGCAC TGATGGGAAT CATAATGCCG 11220

TTCACAAGTA GGAATCCAGT GGTTAGCCAC TGTACACTTG AAGCTGAGAT ATCAAATTCT 11280

TTCATTAATG TAGGTAAGGC CGTCGTTAAA AGCGTTTGGT TTAAAATCGT ACAAAATGTT 11340

CCGATTAGTA AAACACCGAC TAGTAACGAC CGGTTATAAG GTTTTCCGTG AATATCCACA 11400

GGTTGTTGTG AAGTCACATA TAtCGCTCCT TTTTCTGTAA GTTTCAAATT AAATTGCATT 11460

AaCTACTTTA CmACTAATTA GAATCTtTGT CAACTTACTT GACAAAGGTG TAAAAAAAAC 11520

TACAATTAAG GAAAGGATAA GGGAAAGGAA GTTAAACTAA TGTGGTCAAA AATGATTCAG 11580

CGTTTGTTAA ATGAAGAAAA TTTGTTGATA GGAATTAGTG AACTAAGTGA AATTGCTGAA 11640

GTTTCGCCAC GTCAACTTCG TTATTGGGAA GAAAAAGGGT ATATTTCTTC TATTGCGAAA 11700

GACGCAAACG GTCCACGGAA ATATCGGTTA CATACAGTCG TCAAAGTGCA TTGGATCAAG 11760

CGTTTTTTAG ATGAAGGATA TACGTTACAA AGTGCGGTGG AAAAAGCAGA AGAGCAGCAT 11820

CGTAATGTGG GGATGACGAA AAAAATCTTT AGTCAAATGT TTCATGGTAT CTCAGAAGTT 11880

TGTCCCAATT ACGTAGCGAT TGATTTGGGC GATTTCGATG AAGAACAAAA ATTATATGTT 11940

GTATTTAATG AAACCACAGA AGTCAGTGAG TATCTTTTAA TTCCTAAGGA TGCAGATGTT 12000

TCTGATGAGT TAAAGAAAAG AAAATAAGAA TGAAATAGGA CTGGGAGAAA GACCTGAAGT 12060

GAAAGTAAAA CGTTACTGAA TGAATCTTCG AGTTCTTTTT TTCTTGCAAT TTCTTTATTT 12120

TTCTATATAA TTTCTTATAT GGAAAAGGAG GGTTTCGATG AAAGAAACAA AACCTTACGG 12180

AACAGTCTTG TTAAAAGCGG CTAAAATCAT GGACTGTCTA AGTTTACAAC CCGATCAAAC 12240

CTTACAGGAA ATTGCAAAGA ATACCGAAAT GACGGCATCA ACGGCAATCA AAATTTTAGA 12300

AACCTTAACA TTAATTGGCT ATGTTCAAAA AGACACGAAC AAGACGTACC GCTTAGGAAC 12360

AAAATTGATT CGCTATGCGA ATCAAAGTGT GGAGCAAATC GATTTAGCTG AAATTGCGGT 12420

ACCTTATTTA GAAAACCTTC AAAATACGGT CGATGAAACC ATTCATTTAG GGATTTTAAA 12480

TAATAATGAA ATTCTGTATA TCAATAAATT AGAACCGAAA AATCAAACGA TTCGCATGTC 12540

CTCGAAAGTG GGAATTACTC GACCTTTGTA TAATTCGGCC ATGGGTAAAG CCGTTTTAGC 12600

AGAATTTTCG GAAGAACAAG TGCAACAATA TTTAGACACG CAGACATTAA TTCCCTATAC 12660

CGAAAATACG ATTACCAATC CCTTGCGTTT GAAAAAAGAA TTAAAACAAG TGCAACAAAC 12720

GGGGGTTGCT TACGATGATG AAGAAATTGA ACAAGATATT TTCTGTAGCG GAGTATCTTT 12780

AATGAAAGAT GGCGAAATTG CGGGTGCTTT TAGTGTAAGC ATGCCGAAAT ATCGTTTGAC 12840

AGAGGAAAAT AAAACGACGA TTAATCAAGC CCTTTTGGCA ACAAAAGCAG CGATTGAAGC 12900

GAAATTATAG TAGAAAAAGA AAGACAAACA AAAGTCGATT GGGATTTTTG TTTGTCTTTC 12960

TTTTTTTATT TTTGTAAAAA CGTTTGCATT TTCTTTTTTT ATCTGTATAA TGAAGTTGTG 13020

AATTTCTTTA TAATAAATAT ATATTTCTAA ATAAGAAATG AGGGAGTCTG TTGAAGCGAG 13080

TATCAATTTT ACAAAAATTA GAAAATGCTG GCGTGATTGC TGTGGTTCGC GGAAAGACAA 13140

AAGAGGAAGC GCTAAAAGCt AGTCAAGCCA TTGTAGCAGG CGGTATGCGG GGGATTGaGT 13200

TAACCTTTAC TGTTCCTCAA GCAACAGAAG TGATTCAGGA ATTAGTTGCG TTATACCAAG 13260

AGAATCCTGA CGTAGTCATT GGTGCGGGGA CAGTTTTGGA TGCTACGACT GCTCGTTTAG 13320

CTATTGTAGC AGGGGCTGAA TATATTGTGA GTCCAAGCTT CGATGAAGAA ACCGCTAAAA 13380

TTTGCAATTT ATATCAAGTT CCTTACTTGC CAGGCTGTAT GACCATCACA GAAATGAAAG 13440

AAGCACTGAA AAGCGGAGCA GATATTATTA AACTATTTCC GGGAAGTGTT TATGGACCGA 13500

GTGTTATTTC AGCATTTAAA GCGCCATTGC CTCAGTTAAA TATTATGCCa ACGGGTGGTG 13560

TCAGTTTGGA AAATATGGCC GAGTGGTTCG CTGCAGGTGT GACGGCCGTC GGTGTTGGTG 13620

GCAATCTGTT AGCGCCAGCA GCAACTGGGG ATTTTGAAAA AGTTCGTGAG GTTGCCCAAG 13680

CCTATATGGA AAAATTCCAA GCAATCAAAG GGGTGTAGGC ACATGGGGAA AGTCGTAACA 13740

TTGGGAGAAA TCATGTTGCG TTTGTCTACG CCTGTCGGAA TACGCGTAgC ACAAACAGAA 13800

AATTTCGCTG TACATTACGG CGGTGGTGAA GCGAATGTAG CGATTTCGTT AGCGAACTAT 13860

GGCCACCAAG TGACTTTTGC TAGTAAAGTG CCCGATAATT CATTAGGAGA AGCCGCCAAA 13920

AAGCATTTAA GTCGTTATGG GGTTTCAACG GAATTTGTAC GAACTGGCGG AACACGTCTA 13980

GGAACGTATT ACTTGGAAAC TGGTACAgGA GAACGAGCAG CCAGTGTTGT CTATGATCGC 14040

GCCTATTCAA GCTTTGCTCA AATCGAAGAA ATGGAATGGG ACTTACATGA GCTGTTTTCA 14100

GGAGTAGAGA TCTTTCATAT TTCAGGGATT ACAGCGGCGT TATCAAAAAA TTGGCAAAGT 14160

ATGACGGTTG CGTTAGTTAA AGCAGCGAAg CAAGCAGGGT GCCAAGTAAG TTTTGACATT 14220

AATTATCGAG GGAAACTTTG GTCCCAGAAA GAAGCGAGCG TGGCATTGAA GGCCATTCTT 14280

CCATTTGTTG ATATTTGTTC AGCTGGGGTA TTAGATGCTC GTTATTTACT GGAGATTCCA 14340

GAACCAGAGG AAGAACTTGA AGACGCATTA ACGTGGTACT ACCGAAAAAT GCAAGAACGT 14400

TATCCCAACA TTCAAGTATT TTATTCAACC AAACGTCAAG TTCATTCAGC GACAGATAAT 14460

GAATTAATCG GAACGCTATG GTACAAAGGT GCGTATTATA CGTCCAAATG TCATCGATTA 14520

CAACCGATTG TTGACCGTGT CGGTGCCGGT GATGCGTTTG CTGGTGGGGT TTTACATGGC 14580

ATATTAGCTG CCTATGATCC ACAAACCTGC ATTGATTTTG CGACAGCGGC TAGTGCGTTA 14640

AAACATACTG TCCATGGTGA TTGTAATCAA TTTAGTCAAA AAGAAGTCCT TGCTTTTTTG 14700

GCAACGGGCT CAGGAAAAAT AAATAGATAG GAGATTTTAG TTATGCAAAA TATGGAAACA 14760

CGTTATACAC ATAGTCCGGC GGATATCCGT CACTATTCAA CAGAACAATT GAGAGATGAA 14820

TTTTTAGTAG AGAAAGTTTT TATTCCAGGT GCAATCAGCT TAACCTACAC ACATAATGAT 14880

CGGATGATTT TTGGTGGTGT GACACCGACC ACAGAAGAAT TAGAAATTAT TTTGGACAAA 14940

GAATTAGGAG TTGATTATTT CTTAGAACGT CGTGAACTAG GAGTTATTAA TATTGGGGGA 15000

CCAGGCTTTA TTGAAATTGA TGGCGCCAAA GAAACAATGA AAAAACAAGA TGGCTATTAT 15060

ATTGGGAAAG AAACAAAACA TGTTCGTTTT TCTTCAGAAA ACCCTGATAA TCCAGCGAAA 15120

TTTTACATTA GTTGTGTGCC AGCACATCAT AAATATCCCA ACGTAAAAAT TAGTATTGAT 15180

GAAATTACCC CAATGGAAAC GGGTGATCCA TTAACATTGA ACCAACGTAA AATTTACCAA 15240

TACATTCATC CGAACGTTTG TGAAAGTTGC CAGTTACAAA TGGGCTACAC CATTTTAGAA 15300

CCAGGTAGCG CTTGGAATAC AATGCCTTGT CATACACATG AACGTCGTAT GGAAGCGTAT 15360

GTTTATTTTG ATATGGAAGA AGATACGCGG ATTTTCCACA TGATGGGTAA ACCAGACGAA 15420

ACGAAACATT TAGTAATGAG TAATGAACAA GCAGCTATTT CACCAAGCTG GTCAATTCAT 15480

TCTGGGGTTG GTACAAGTAA TTATTCCTTT ATTTGGGCCA TGTGCGGAGA AAATATTACG 15540

TACACAGACA TGGATATGGT TGCAATGGAT CAATTAAAAT AAACCAAAGA AAGAGGAGAA 15600

AACATGGAAT TTAAAATGAA TGATTTTCAA TTAGATGGAA AAGTAGCGTT AATTACAGGT 15660

GCTGTTTATG GGATTGGCTT TGAGATTGCG AAATCCCTAG CTGAAGCAGG AGCAACAATT 15720

GTTTTTAATA ACTTATCTCC TGAATCAGTT GAACAAGGCT TGGAAAATTA TCGGGCCGCA 15780

GGGATTGAAG CACGTGGTTA TGTGTGTGAT GTCACTGATG AAGAGCAAGT TCAAGCGATG 15840

GTTGCTCAGA TTAAAGAAGA AGTCGGCTCT ATTGATATTT TAGTCAATAA TGCGGGAATT 15900

ATTAAACGAA TTCCAATGGT GGATATGTCA GCAGAAGAGT TCCGTCAAGT GATTGATGTT 15960

GACTTGAATG CGCCATTTAT TATGGCAAAA GCGGTTATTC CAGATATGAT TGAAAAGGGT 16020

CATGGCAAAA TTATCAATAT CTGTTCAATG ATGAGTGAAC TAGGTCGTGA AACAGTTAGT 16080

GCGTACGCGG CAGCAAAAGG CGGACTAAAA ATGTTAACGA AAAATATTGC TTCTGAATAT 16140

GGACAATATA ATATTCAATG TAATGGGATT GGACCTGGCT ACATTGAAAC GCCACAGACG 16200

GCGCCATTAA GAGAACCAGG GCATCCATTT AACGAATTCA TCTTGGGACG AACACCTGCT 16260

AATCGTTGGG GCACACCAGC AGACTTAGCT GGCCCAGCAG TTTTCTTAGC ATCTAACGCT 16320

TCTGACTTTG TTAACGGACA TATCTTATAT GTCGATGGTG GAATTTTAGC TTATATTGGA 16380

AAACAACCAT AAATAAGTAA CAAAAAGCGG ACATACCTAA AAAGGATGTC CGTTTTTTGT 16440

TATTTACTCA TTGATTGAAT TGGTAAAAAC CGTATCTAAT TTTTAAAAAA AGTCAATTTT 16500

TGACAGTAAT TCGTCAAGAT TAACTATATA ATTTGCTAAC ATTCCTTGCT AAACTAAAAT 16560

TACAAAAAAA TGAAAGCGTT TCATTTGCGC TGGTGGGATG GAAAAGGAGT TTTGATTGCC 16620

GATGATTATT AAACAGGAAC AATTCATACC AAAAGATACA GTAGAAACAA CGATTGACTT 16680

ATTAATTAGA AATTTAACAA CGATTAAAGA TAACACAGGT GAATTTCTGT TAGATTTTGA 16740

TGGTTTAAAA GTTGATGACA AAAGTTGGAC GATTTGGAAC TGGCCGCAAG GCGTTGGGTT 16800

GTATGGTATA TATAAGAATT ATCGAAATAC TAAAAGTGAA AAAGCTTTAC AAGTAGTTAA 16860

CGACTGGTTT GAAGGTCGGA TGCAAGAAGG CGCGCCGCCT AAGAATGTGA ATACGATGGC 16920

TCCGCTTTTA ACAATGGCTT ATCTATATGA AGACACTAAA GACTCTAAAT ATATACCTTA 16980

CTTAGAGCAA TGGGCGGAAT GGGTGATGGA AGAAATGCCC CGAACGAATG AAGGTGGCTT 17040

GCAACATGCG ACATATGGAC CAGAAAATAA AAATCAATTA TGGGATGATA CGTTAATGAT 17100

GACTGTCTTA CCTTTGGCAA AAATTGGTAA GTTGTTGAAT CGACTCGATT ATTTGGAAGA 17160

AGCAAAGCAT CAATTTTTAA TCCATATTAA ATATTTACAA GATAAAAAAA GTGGTCTATG 17220

GTATCACGGT TGGACTTTTG AGGGAAATCA TAACTATGCG GAAGCACTGT GGGCAAGAGG 17280

AAACTGCTGG ATTACGATTG CTATTCCAGA AATTATTGAA ATCTTAGAGT TGCCAAAAGG 17340

AGATAGTTTA CGGGAATTTT TATTGAGTAC CCTCAATGCG CAAGTGGCAG CTTTAGCGAA 17400

ATACCAAGAT GAATCTGGTT TATGGCATAC ATTAATTAAT GATTCAAATT CGTATTTAGA 17460

ATCTTCTGCT ACAGCGGGAT TCGCTTATGG GATTTTAAAA GCGGTTCATA AAAAATATAT 17520

TTCTTCTGAA TATGAAGAAG TGGCAAACAA AGCAATTGCT GGCTTACTAA ATGAGATTGA 17580

TGAAACGGGA GAAGTACAAC ATGTGTCAGT TGGTACAGGA ATGGGTGATA ATTTAGATTT 17640

TTATCGCACA ATTGGAATGA CAGCGATGCC TTATGGTCAA TCATTAACAA TCCTATGTTT 17700

GACTGAATTG CTTGTTTCTT ATTGCTAGAA AAGTAAGGAG AAAAGTTCAT GAAAAAAAGA 17760

AAGGTTTTAT TTACAGCAGT TATGGTATTG GCAGGATTAC AGTTGCTAAG TGGTTGCGGC 17820

AAAACAGAAG CTTCGGCAAA TGATACGGTA GTCTTGCGCT ATGCGTATGC TAGTAATAGC 17880

CAACCAGTTA TCGATTCTAT GAAGAAATTC GGTGAATTAG TAGAGGAAAA AACAGATGGT 17940

AAAGTTCAAA TTGAATATTT TCCAGATGGT CAATTAGGAG GAGAAACAGA ACTAATTGAA 18000

TTAACACAAA CAGGTGCAAT TGATTTTGCA AAGGTCAGTG GATCAGCATT AGAAAGTTTT 18060

TCTAAAGATT ATTCTGTATT TGCCATTCCG TATATTTTTG ATAATGAAAA ACATTTTTTT 18120

AAAGTAATGG ATAATCAAGC GCTAATGCAA CCAGTGTATG ATTCTACAAA AAAATTAGGA 18180

TTTGTTGGTT TAACTTATTA TGACTCTGGT CAACGAAGTT TTTATATGAG CAAAGGGCCT 18240

GTTACATCTC CAGATGATTT GAAAGGTAAA AAAATTCGGG TCATGCAAAG TGAAACCGCC 18300

ATCAAAATGG TAGAACTTTT AGGGGGTTCG CCAGTACCTA TGGGTAGTTC GGAAGTATAT 18360

ACTTCTCTAC AATCTAATCT AATCAACGGT GCAGAGAATA ATGAGTTCGT TTTATATACA 18420

GCTGGTCATG GTGGTGTGGC TAAGTATTAT TCTTATGATG AGCATACTCG AGTGCCAGAT 18480

ATTGTGATTA TGAACGAGGG AACAAAAGAA CGTTTGACAG CGAAACAAGA ACAAGCGATT 18540

GAAGAAGCAG CAAAAGAATC GACCGCTTTT GAAAAAACGG TCTTTAAAGA AGCGGTTGAA 18600

GAAGAAAAGA AAAAAGCACA AGCAGAATAT GGCGTTGTGT TCAATCAAGT AGACAGTGAA 18660

CCATTCCAAA AACTTGTTCA ACCGTTGCAT GAATCATTCA AAAATAGCTC AGAACATGGC 18720

GAACTGTATC AGGCTATTCG CCAGTTGGCG GACTAATTAA GGAGGTAATT ATGAAAGGTT 18780

TTAGAAAATT TTTAAATCGA CTTTTAGAAA TACTTGGATC AATCATTCTT GCTGCAATGG 18840

TTTTGGTCGT TTTGTACCAA GTGTTTGCCC GGACAGTGTT AAAAAATCCT AGTACCTTAA 18900

CAGAAGAATT TGTTCGCTTT GCATTGGTCT GGATGGCGAT GTTAGCTTCT GCTTATGTGG 18960

TTGGAAAAAA GGGACATTTA GCAGTAACTT TGTTAAGTGA AAAATTACAA GGGAAACAAA 19020

AACGCTTTTT AGAATTTATT GTTCAATTGT TGTTCCTACT TTTTGCAGGA GTGATTATGA 19080

TATTTGGCGG TTGGAATGGT GTAGTAATTA CGTTAGGACA AATTTCACCA TCTTTGGCAA 19140

TTCCTATGGG TTATGTCTAT TTAGCCGTTC CAGTAGCAGG TGTTTTGATG TTTATTTATA 19200

GTTTGATGAA TTTGCTAGAG TTATTCACGC AATCAAAAGA TTCTTAATGC GTTTATAGAA 19260

GAAAAAGAAT AGGAGTAAAA ATATGGCTTT AGAAGCAGGC ATTATCTTAT TTGTAGTTTT 19320

TGCAGTCTTA TTGGTAGTTG GTATGCCAAT TGGTATGAGC ATTGCGGCAT CTTCAATTGC 19380

AACATTGTTA TTAGTAATTC CTTTTGAAGC AGCTGCATTT ACCTCAGCTC AAAAAATGGT 19440

CTCTAGTTTA AATAGTTTCT CTCTCGTAGC AATTCCGTTT TTTATTTTTT CTGGTGTGAT 19500

TATGAATAAT GGGGGAATTG CCCAAAAGTT GGTAAATTTT TCTATGTTAT TTGTTGGTCG 19560

TGTTCCCGGC GCGCTAGCAC ATACTAATGT GTTAGGGAAT GCATTATTTG GTTCGATTTC 19620

TAGCTCTGCA ATTGCAGCAT CTACTGCTAT TGGTGGCGTA CTAATTCCGC AACAAGTAAA 19680

TGAGGGCTAT GATCGGAAAT TTGCAACGGC TGTAAATATT GCCTCTGCTC CTACAGGAAT 19740

GGTTATTCCA CCAAGTACAG CGTTTATTAT GTATTCATTG GTGGCTGGTG GGGCCTCTAT 19800

TTCTTCTTTA TTTTTAGGAG GTTATTTGGT AGGTTCTTTG TGGGCGTTAG GAATTATGGT 19860

GGTAGCTTAT GTTATTGCAA AACGTCATAA TTATCCTACT GTGGCAAAAG CAAAAAAAGG 19920

AGAAGTCAGC AAGGTTCTGC GAGAAGCGGT TCCCAGTGTA TTGTTAATTG TCATTATTAT 19980

TGGTGGTATT TTGACGGGGT TATTTACAGC GATTGAAGCA TCAGCAATTG CTGtTGCATA 20040

CTCATTATTG ATTTCAATGT TTTACTACAA AACAGTGAAA ATAAACGATT TACCTAAAAT 20100

GTTAAAAGAG GCTGTCTTGA TGAGTGGAAC CATTATGTTC TTGTTGGCAA CTTCTTCAAT 20160

GATGTCTTTT GCGATGGCTT TTACAGGGAT ACCAGAAGCA ATCAGTACAG CTATTTTAGG 20220

AATCACAGAC AATAAATACT TAATTCTACT ATTAGTAAAC GTTGTCCTAT TGTTTGTGGG 20280

AATGTTTATG GATGTTGGTC CAGCTATTTT GATTTTTACG CCAATCTTTT TACCAATTGT 20340

TATGAATGTT GGGGTGGATC CAGTTCATTT TGGTTTATTT GCGATTATGA ACTTGTGTGT 20400

TGGTTCAATT ACACCACCTG TTGGAACAGG ATTATATGTT GGTGCTAGTG TTGGGGGTGT 20460

TAAAGCTGAG CAAATGCTGA AACCACTAGT TCCATTTTAT CTGGTAATCT TAGCAATTTT 20520

ACTTTTAATC ACATATGTAC CTGAACTTGT CATGTGGTTA CCAAATCTAA TTTCTTAATT 20580

CAATACAAAT TGTCTAACAA CCCCAATTGA GCGAAATATA AGTTGTTCAA TCAACTAGAT 20640

TTGGTAATTT ACTATAAAAT ATTGAGTTTA GGAGGCATAG AAATAAAAAA TCTCTCTATG 20700

ACATTATGTG TCATAGAGAG ATTTTTTAGA AGCCACTTTT ACGGTATTGA CTAGGAGTTA 20760

ATCCAGTTCG CTTTTTAAAT AATCGACTAA AGTAGGCATA GTTTTCTAAA CCAACATCTA 20820

CAGCAATTTC TTTCACTGTT TGATTAGAAT AGCGTAGTAA ATCTTTTGAA TAAGAAATTT 20880

TTTTACTAAT TTGATATTTA ATTGGCGGTA ATCCAATGTA TTTGGAGAAC TCTTTGTTCA 20940

ATTGATATTT ATTTAAATGA AAGAGTTGAG CCAAGTCTTC TAAAGAAATT GTTTCTTTGA 21000

AATGTTGATC AATGTAAGCT TTCATTTCCA AAACCAATGA TGGAATCGTT TCGTAGGTAA 21060

AGTCTTGACG ATACTTTTGC AACAATAATT CGTTCAGTAA TTGATGAATT AAGACGGAAG 21120

ATTGGTAATC AGTTTGCGCA TGTTGTAAAT CTTGTAATTG TAATAATTGC TGAATGATTA 21180

```
AATGAATAGG ATTATCTTCT GGAAGCTGAT TAGTGGAGAA AACTGGTGAA CCATTTTTCA   21240
TAAATTCTTG GAAAAAAGCA CCAGAATTTC CTCCTTCAAA ATGAATCCAG TCCATTTCCC   21300
AAGGATCTTC GCTGGTTGCT TGATAGTATT GATAATCCCG ACAGTCAATA AAGAATAAAT   21360
CACCTCTCTG TAAGCTGTAT GTATGATAAC GATAGCGTAA TTCCCCAGTA CCATTTAGCG   21420
TGAATTTAAT TAAGTATGAT GGTAAATGTG CCCGCTCCGT ATAATAAGGG CGCGAGGCTT   21480
TAAACTGTCC AATTTCCTGT ATGTAATAAA AAAGTGTGCG ACTTTTGCTG CTAGGTGTCA   21540
AAATATGTCG CTTAGAATCG TCTGACCAAG AGTAGTCGGC TTTTTCCCaA TAATCCaTAA   21600
GAcCTCCTAA TTACCAAGAT AGTACAATTA TATCTCAAAA TTGTGAGAGG TGAAAGCGCT   21660
ATTTTTTGTT TATGATGAAT ACATAAAGTT AGTCTAACGT ACAAAAAATA AAAATAGTGA   21720
AAGGATTGAC AACTATCAAC AAAGTGAmAA AAGAACAAGT TGAAAAACGA ACCTATTTAG   21780
CGGTTGATAT TGGCGCTTCT AGCGGTCGCA TAATGAAAAG TCAACGCTTA GCGAATGGTC   21840
AGATAACAAT AGAAGAAATT CATCGTTTTA AAAATGGTTT TCATCAAAAA GATGGGTATC   21900
AGCGTTGGGA TATGGCTAGT TTAGTTCATG AGTTACTGCT TGGTCTGCAA AAAGTTAAAC   21960
AATTAGGTAT CAAGGAATGC TTTATTGGAA TTGATACTTG GGGTGTTGAC TATTGTCTGC   22020
TGGATCAATC TGGACAACTT TTGGACGAAC CGATTGCTTA TCGAGATGGA CGGACAGAAG   22080
CAGCTGTAAC AAATTTTTCA AAAGGCTATT CATTAGAAAA ATTGTACCAA CAAACGGGGA   22140
TTCAAGTTCA GCCTTTTAAT ACCATTTTTC AGTTATTTGT AGAGGAAAAG GAACGTTTGG   22200
CTGCAGCCAG CCAATTGTTG TTAATCCCTG ATTACTTAGG TTATGTCTTT ACTGGAAAAG   22260
CAGTCATCGA AGCAACCAAC GCTTCCACCA CGCAACTGTT GAATGCTGGA ACGAAACAAT   22320
GGGAAAGTGA ATTACTGGAT TTCCTAGGAA TTGATGAAAC ACTATTTCCA ACATTAGTAG   22380
AGCCGGGCAC GATTTTAGGC GACCTGCAAA CAGCGGCTTT TCCAGACTAT GATTTACCAA   22440
ATGCAACGTT GATTACGATA GCTAGTCATG ATACAGCTTC AGCAATTTTA GGTACGCCAG   22500
GCATTGGCGA TGACTGGGCG TACATTAGCA GTGGTACTTG GTCTTTATTA GGCATTGAAA   22560
CGACTGTTAC AACAATTTCT GCTGAAGCTT TCCAAGAAAA TTACACGAAT GAATGGGGCG   22620
CCCAGAACAC AATTCGTTTC TTAAAAAATA TTATGGGTAT GTGGCTAATT CAAGAAGTGG   22680
CACGGCACCA AAACTACCAA TATAGCTATG CTGAATTGGC CGCCTTAGCT GAAAAAGAAC   22740
CAGCATTTCA ACAATTTATT GATGTAAATG ATCCTCGTTT TTTAAATCCG GGAAATATGA   22800
TTACTGAATT ACAAGCCTAT TGCCGTGAAA CACAACAAAC AGTTCCTGAA AGCCCTGGTG   22860
AACTCGCTCG TTGTATTTAT GATAATTTAG CATTGTGTTA CAGCGTAGAA CTAGAAAAAT   22920
TAGCTCAATT AACAGGAATC GAACGGAAAA TTACAACTTT ACATGTTGTC GGTGGTGGTT   22980
CCAATAATCG TTTATTGAAC CAGCTAACTG CTGATGTTGC TAATGTTACA GTGAATGCGG   23040
GACCTGGGGA AGCAATTGCT TTAGGTAATC TTTTGATGCA GATGATTGCA ACAGGTGAAC   23100
TTAAAGATAT TCCAGCAGCT CGTACTTGTA TCCAAACATC TTTTCCCACG GAAATCTATC   23160
```

AAGCAAATCC AATTGATTCA ACAATAAAAA ATCGCTATCA AGCATTTATG AAAAGGAGCT 23220

CATTATGACA ACCATTACAC AAAAATATGA AGAAGCGAAA GAAAAATATG CCAGTATTGA 23280

CGTAGATACC GAAGCTGTAT TGGAAAAAAT GGCTGATGTC AAAATTTCAA TGCATGTGTG 23340

GCAAGGAGAT GATGTGAGAG GCTTCTTAAG TGAGGATGAA CTTTCTGGTG GTATTTCAGT 23400

TACAGGGAAT TATCCAGGCG TTGCTCGAAG TCCCCAGCAA CTACGTCAAG ATTTAGAAAA 23460

AGCGTTTTCT TTAATTCCAG GAAAACACAA ATTGAATTTG CATGCTATTT ATTTAGACAC 23520

AGAGGAAAGG GTGGATTTAA ACGAACTGGA ACCAAAGCAT TTTGAACCAT GGGTGACTTG 23580

GGCAAAAGAA AATGGCTTAG GATTGGACTT TAATCCAACC TTTTTCTCGC ATCCAATGTA 23640

CCGCGATGGC TTTACTTTAG CTCATCCGAA TCCGCAAGTT CGAGACTTTT GGATTGAGCA 23700

TGGGAAACGT TCACGAAGAA TTGCTGAATA TTTTGGAAGA GAATTAGGAC AAGTAGCTGT 23760

CAATAACTTT TGGGTGCCAG ATGGCTTTAA AGATAACCCA GTTGATCGTT TAACGCCGAG 23820

AAAGCGGCTA ATGgcTTCGC TAGATGAAAT TTTTTCAGAG GAAATTGATC CAGCTTATAC 23880

AGTCGATGCG ATGGAAAGCA AGCTCTTTGG GATTGGTTCA GAAGCCTATA CAGTTGGGTC 23940

TCATGAATTT TATATGGGCT ACGGTTTAAC GCGGAATAAA TTAATTTGTT TAGATGCTGG 24000

TCATTTTCAT CCAACGGAAG TGATTTCTAA TAAGTTATCT TCATTATCCT TATTTGGCGA 24060

AGGAATGCTA TTACATGTCA GTCGTCCAGT TCGCTGGGAT AGTGACCACG TTGTCATTAT 24120

GGATGATGAA TTGCAAGAAA TTGCCAAGGA GTTAGTTCGC AATGATTTAT TAGGAAAAAC 24180

ACACGTGGGC TTAGATTTCT TTGATGCCAC AATCAATCGT GTTGCCGCTT GGGTGATTGG 24240

CACCCGCAAT ACACAAAAAG CGTTAATGAA AGCGATGCTT GAGCCAACAA ACGTCTTAAA 24300

AGAAGCTGAA TTGATAGGAG ATTTTACCAC ACGTTTAGCC TTAACGGAAG AGTTAAAAGA 24360

CTTTCCATTT GCCGATATCT GGAACTACTA CTGTCAAGAA AATCACGTAC CTATCGGCTT 24420

AGATTGGTTA ACGGATGTTC AAGAATACGA AAAAGTGATT TTACCAACTC GTCAATTGCC 24480

AACAGGTAAA GATAGTTGTC GTTTTTCATA AAATTAACAA ACGAGAGGAA CTTCTAAAAT 24540

GAAAAAAATG AACGTGTTAC AAGCACCATT TGTGGAAGAA ATGGTCAAAA CGACCAAAAA 24600

TTTATATCGT TTAGGTTGGG ATGAACGAAA TGGCGGAAAT ATTTCGTATT TATTAAAAGA 24660

GGAGGAGATT TTACCTTTTC TAAATCCAAC ACAAGTGCTC CGAAAGATTC CAATGAAATT 24720

TGATGCAACG AAATTAGCAG GAAAATATTT TATCGTGACA GGCTCAGGAA AATATTTTAA 24780

AAATGTCTGT GATGCATCAA GCGAAAACTT AGGGATCTTA CGAGTTAGTG AAAATGGCCA 24840

GGAATTAGAA CTATTGTGGG GCTTAGAAGA TGAAGCTGTA CCAACGAGTG AATTACCTAG 24900

CCATTTTATG AGCCATATTG CTCGTTTAGC TGTTGATCCA GAAAATCGCA TTGTCATGCA 24960

TAATCATGCG AGTCATTTAT TAGCAATGAG TTTTACTCAC GAATTAGATG AAAAAGTCTT 25020

TACTAGAACA TTATGGCAAA TGTGTACCGA ATGTTTAGTA GTATTTCCAG ACGGTGTTGG 25080

TATTATTCCG TGGTTGGTTC CTGGAACGAA TGAAATTGGT GTAGCAACGG CTGAAAAAAT 25140

GAAAGAATCT CGTTTGGTTT TATGGCCGCA ACATGGTATT TATGGTACTG GCCGTGATAT 25200

GGATGAAGTA TTTGGTTTAA TTGAAACTGC CGAAAAAGCC GCTGAAGTGT ACACTTATGT 25260

TTGTGCGCAA GGAGGCGTTC GCCAAACAAT TTCAGATGCA GATTTATGGC GCTTAGCAGA 25320

AGCGTTTGGT GTGACACCGA AAGTTGGCTA TTTGGAAGAG AAAGTATCAA AACGGAGAAA 25380

GCTATGATAA AAAAAGCCTT TTGTATGCAA GTATATCCTG ATCAACATGC CGAATATCAA 25440

CGACGTCATG AGAAGTTATG GCCAGAAATG CGTCAAATGC TGAAAGAACA TGGGGCAATT 25500

AAGTATCAAA TTTTTTTAAA TCCTGAAACC AGTACACTTT TTGGTTATTT AGAAATAGAA 25560

GATGAAGCTC GCTGGGAGCA AATAGCACTG ACACCAATCA ATCAAAAATG GTGGAACTAT 25620

ATGGAAGATA TTATGGAAAC AAATCCTGAT TGCTCGCCGA TGACGGCTGA ATTAAAAAAA 25680

GTTTTTGAAC TATAATTTAA AATAGAACAG TTTGTTATAG TGAATCGTAT GATTGTACAC 25740

AAAACAACTT ACTTGAGTGG AGTTGATTTC AAGGGATTGT TCTGGAACTT ACAAAGAAAG 25800

CCTTTGTGGT AACGGCTGTT CCAGAAAGTA ATAAATTGTA CCAACATTAG CTGTAACTGT 25860

TTATTCAAAA AATAGTTAAC TTCTAAAAAA AGAGGTGAAT CTTGTTGAAG AGTGTAAAAG 25920

AAATCCGGAA ATCACAAGTA AATAGTATGA TTTCAGATCA ATATGAGATC TTTCATGTAA 25980

AAGATATGGT TAGTCCAGAA AAAACCATCT ATCATTACCA CGATTTTTAT GAAGTCCATT 26040

GTACGTTAAA AGGGGTAGCA ACGTTCTTTT TAGATGGACA TCAATTTGAT GTTGAAGCAG 26100

GGACCGTTTT ACTCATTCAT TATCATGATT TACATCGAAT TATTAAGCAG AGTACTGATG 26160

ATTTCGAACG GATGTATATT TTTTTAACAC CCGATTTTTT ACAACAACGA TCTAGTAAAC 26220

GGACAAACTT GTCTGCTTGT TTTCAACATT TTGGTCAGCG TCGTAGTAAA GTTATCAAAG 26280

TGGATGTAGC AAAATTAGCA AACTACTTGA CCCCGTTAGA TCATTCTCCT TTGCCaGAAG 26340

AATATGGTGC GGATGTTCGT TATGAACAAC AACTGTTAGA CTTTTTAATT TACTTAAATC 26400

AATTAGTTCT GAAAGAAGAG AATGAGAGCC AACCAAAACA AATGATTGAA AATGAACGAA 26460

TTGAAGCAAT GATTACCTAC ATTTCTCAAA ATTTGGATCA ACCATTAACC TTAGAACAGA 26520

TGGAAAAAAA TTTTTTTGTG ACTAAGTATT ATGTAACGCG GGAATTTAAG AAACATACAG 26580

GATTTACTTT CCATCAATTT GTGCTGAAAA AGAAACTACT TTACGCTAAA CAGTTATTAA 26640

AAGAATATCG CAGTGCTAGT GATGTTTACC TGAAATGTGG ATTTAAATCA TATCCACACT 26700

TTTTGAAATC CTTTAAAAAA GAGTTTAATA TGACACCAAA AGAATTTTTG GTGCAACATA 26760

AGAATAATCA GATCATTCAT TTTGATCATT ACGAAGAATC TATCAAGAAA GTGAGGCTTG 26820

AGTAAAATGT TGACACTTGC AGTACGTGCC CATGATGTTG GCACCAAAAT ATCTCCTGTG 26880

GAATTGGCTA AAAAAATTGC TGAAACGAAC ATAACTTCAA TTCAATTTGC CTTAGGTATT 26940

TCTTATCCAG AATATGCAAG TGCTGCTGTT CTATCTCCAG GACTAGGAAG CCAATTGAAG 27000

AATATTTTCG AAGATGAAAA GATTACGATT AGCGTCCTAA GTTGTTATAT TAATCTGATT 27060

CATCCAGATT TAGAAGAGCG AGAAAAACGA CTGCATCAAT TTGAAACATA CATTCGTTAC 27120

GATTCAATGT TTGGGGCAAA AATCGTTGCA ACTGAAACAG GCAGTATATA TGAAACGATT 27180

TATTACACTG AAGATAATTA TACAGAAGAA GCCTACGAAG AAGTTTTATT TTCTGTTAGA 27240

AGATTATGTC GAGAAGCTGA AAAGTATGGG ATAATTGTTG GGATTGAACC AGGAATCAAT 27300

CATCCAATTC ATACAATTGA AAAAATGCAA CGATTGATTG ACGAGGTTGC TTCAACTAAT 27360

TTAGGTATTA TTTTGGATCC TACCAACTTA ATTCGAGTAG ATATAGATAA AACGTATTTG 27420

GAAATAGTAG AAGAAGCGTT TGAGTGTTTT GGTGAAAAAA TCGTTGCTTT TCATTTGAAA 27480

GATTTCATTA TAAGAAATCA GCAGATATTT CCGGTAGCCA TTGGCGAAGG GCAGGTACCT 27540

CTTACAGAAA CCATTCAATT TTTGAATAAA CATAAGCCCG GACTGTTCAC TATTTTTGAG 27600

GAAACACCAT TTGAAAAAAT aGCAAGTGCC TaTCAAAAAG TTTCGcAATA TCATTCCATT 27660

TAGATaAAAA TGTGCAACAa AAAACATCTC GATTTTTTGA GATGTTTTTG TTGCTTAATA 27720

AGACAAAATA TAAAATACAC TTGCCTTAAA AACTCACATA AAAGAAAGAA GCAATCGCCA 27780

CAGCAAATGA AATAATAGCA AACACTAAAC TACTCCATAG TCCTAACAGA ATAAATGGTG 27840

AGGTTCTTTT ACCGCCCTCT TTCTTTAAGT AGGAAAATGT CCGTCCAGTA ACAATGAATT 27900

GCCAGATACT AATCAATAAA AAAATACCGC CAATTAAGTA ACGCATGTCC TCGTCCTCCT 27960

TCTATTTAGT ATACCACTTG ATGaAAACGT TATCaAAACG CTGTATAAAA AAGTTGCCTT 28020

TCAAAAAAGA GAAAGGCAAC TTTTTGACGG TTTAACGGAC GTCACCCATT AAGTTTTGGA 28080

TTGGTTTTTT CACAATGAAT AAGAGAATCC CAGCAACGAC AGCTACAAGT CCGACAATCC 28140

CAAAATATGC TGACTCTGTT CCAGGTGTAT AGAAACGAGC AATTTGAGCG TTAATGGCTT 28200

GTGAAGCGGC ATCTGCTAAT AACCAAATAG CAACGGTTTG AGCTTCGAAA GCTTTCGGTG 28260

CTAATTTTGT AGTAATCGAT AAGCCAACGG GAGATAAACA CATTTCTCCG ACAATCATAA 28320

TGAAGAAACT AAAGAATAAC CAAAGTGGAC TCACTCGTCC TTCGGTACCA TAAAGTAGAC 28380

CTGGCAACAT TAATAAGACG AAGGAAAGAC CAGCAAAGAC TAGCCCTAGT GAAAACTTCA 28440

CAACTGTTGA TGGTTGACGT TTACCTAGTT TTGTCCATAA TGTAACAAAA ATAGGGGTCA 28500

AAATAACTAC GAATACAGGA TTTAATGATT GGAACCAGCT AGCGGCAATT GGGAAGCCAA 28560

ACAAACTTGT TTGTGTCCGT TCATTTGcAA ATAGAGCmAG AATAGACGAA CCTTGTTCTT 28620

CCmATGgACC AGAAAACmAT TGCTGCTAAG AATwATGGcA gTAAGCCAGA ACTTTTGGnT 28680

TTTCtTCAGC AGTGACATCT TTAGACGTTA ACATTTTGCT GAAtAATAAA CTGGTAaCAG 28740

GATTCCTAaC ACGCTGATGG TATTAATAAA GAAATCAaTT GTTAAATGTC CTGTGACATA 28800

GGCGCCACCA AAAATAAGTA AAGCAACGAC AATTGCCAAC ATAAATGCTT TAGCAAATTT 28860

CTTTTGCTCT TCTTTAGACA TTGGATTAGT TGGGGCTTGA CCAATCCCAG CAAGAGATTT 28920

ACGTCCTTGG AAATAATATT GTAGTAGCCC GAAGAACATC CCTACTGCGG CAACGGAAAA 28980

GCCTAAATGA TAATTATATT CTTGTCCTAA AGTTCCGACA ACAAGGGGAG CGATTAAGGC 29040

TCCGATATTA ATCCCCATGT AAAAAATGGA GAAGCCTGCA TCTCTACGTA AGTCAGTTTT 29100

TGAATAAAGA TGGCCAACCA TCCCTGAAAC GTTTGGTTTt AACATCCCAG TCCCAAATAC 29160

AATCAACATA ATAGAAACGA ATAAGGCGGG GACACCAAAT GGGGTAGCGA GAACAATATG 29220

TCCAGCAATG ATTAGCAAAC CACCAATAAA AACAGTTTTC CTTGAACCTA AGACACGGTC 29280

AGAGACCCAA CCGCCAATGA TACTGGACAT GAAAACCATT GATCCATAAA TGGACATGAT 29340

GGCTAAGGCA GTTGCTTTGG GGAGACCTAA ACCGCCATTG GCTACGGTAT CGTAAATGTA 29400

ATACATTAGA ATAGCACGCA TGgCCGTAGy AACTGAAACG TTCCCACATT TCTGTGAAGA 29460

ACAGTGTTGC TAGTCCTTTC GGTTGTCCCA GAAAGGATTT GTCTAATTTC TCAAGTTGTT 29520

CTTTGTCACT CATAGATAAA ACTCCAATTC TTATAATATT CAGAATATTT GATAATAAAA 29580

TGACTTAAAT GTAGATTCAT ATATTTTATT ACTCATAGAT TGGAAAATAG AAAGAGAAAA 29640

AACTGATTAT GTTTTTAAAT AACAAGCGTT CGAGCGATTG TTTGTTTTTT ATTCGTGAAT 29700

TATGATAAAA ATGAATATTT ATTCTGCTT 29729

(2) INFORMATION FOR SEQ ID NO: 239:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7237 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

GGTTATGAGA GCAGGGATTG ATGTTGGGTC AACGACAGTA AAATTAGTAT TTTTAAATAA 60

GCAAAATCAG TCTATTTTTA CAAAATATGA ACGCCATTTT TCAGATGTCA AAGCAGCAAC 120

GGAACGCATT TTAAAAGAAG GGCTAGCAAG GATCGGTGCA GACCAACCTG TGACAATGAG 180

CATAACTGGC TCTGGTGGTA TGGGGCTGGC TGAGGTATTA GGGATTCCCT TTGTTCAAGA 240

GGTTATCGCG TGTACACGAA CTGTTGAAAC AATAATACCA GAAACGGATG TTGCTATTGA 300

ACTTGGCGGA GAAGATGCAA AAATCACTTT TTTTGATGGT GCTTTAGAGC AACGAATGAA 360

TGGTAGTTGT GCAGGTGGGA CAGGCGCCTT CATCGATCAA ATGGCTGTTT TACTTAAAAC 420

GGATGCCAAT GGAGTGAATG AATTGGCGAA GAATTATCAA ACTATTTATC CGATTGCTTC 480

CCGCTGTGGG GTTTTTGCCA AAACGGATGT TCAACCTTTA ATCAATGAAG GAGCAGCGAA 540

AGAAGATATC GCGGCGAGTA TTTTTCAAGC GGTAGTTAAT CAAACAATCG CCGGGTTAGC 600

TGCTGGTCGA AAAATTAAAG GTAAAGTAGC ATTTTTAGGG GGACCTTTAT TTTTCATGTC 660

AGAATTGAGA AAACGGTTTG TTGAAACCTT AGCAATTCAG ACAGAAGATG TTATTTTTCC 720

AGAACAACCG CAACTGTTTG TAGCAATGGG TGCAGCACTT TATTCAGAAG ATGCAGCCGA 780

ATGTACATTG GAAGAATTAA TTCAGCGCTT AGTGAACAGC CAGGCTACTG ATTTACAGCC 840

GAGCGACACG TTACCACCGT TATTTAACAG CGCAGAAGAT TTGGCAGATT TTCGAGAAAG 900

ACATGCGCAA gCCCAAGCAA CGGAGAAACC TTTATCTCAA CATACTGGCG TGACCTTTTT 960

```
AGGGATTGAT GCAGGCTCCA CAACGACAAA AGTCACATTG ATTGATGAAG ATGGTCATTT    1020
GTTGTTTTCT TTTTATGGTA ATAACCAAGG GCAACCTTTA GAAACCACCA TGACAGTACT    1080
GAAATCACTT TATCAACAAC TGCCTGAAGA AGTTTTTATC GGAAAAGCGG CTGTTACAGG    1140
aTATGGAGAA CAGCTAATAA AAAATGCGTT AAAAGTCGAT ATTGGTGAAG TTGAAACGAT    1200
GGCCCATTAC AAAGCAGCCA ACCATTTTCA GCCAGGTGTG GATTTTATTT TGGATATCGG    1260
AGGACAAGAT ATGAAAGCGA TGACCATTAA AAATGGCGTG CTTTCCTCGA TACAATTAAA    1320
TGAAGCTTGT TCTTCTGGGT GCGGTTCCTT TATCGAAACC TTTGCGCAAT CTTTAAAATA    1380
TAATGTGAAG GATTTtGCTT TGGCGGCTTT AGAATCGCAA GCGCCTGTTG ATTTAGGTTC    1440
TCGGTGTACT GTTTTTATGA ACTCAAAAGT GAAACAGGTT CAAAAAGAAG GGGCCTCCGT    1500
CGGCGATATT TCTGCTGGAC TATCTTATTC AGTCATCAAA AATGCCATTT ATAAAGTCAT    1560
TAAGGTTCGC CGACCTGAAg AGTTGGGAGA AAAGATTGTC TGCCAGGGAG GAACtTTTAC    1620
AATGAnGCTG TCTTGAGAGC TTTTGAnAAA ATTAGCGGTC GTGAAGTGGT TCGTCCTTCT    1680
ATTGCTGGTT TGATGGGGGC CTATGGAGCA GCGCTGATTG CTTTGGACGC TTATGAAATT    1740
GGGGAAGAAA CCACGCTTCT TTCTTTAGAA GAGTTAGCGG CATTTACCTC TGAAAAGGAA    1800
TTTACTCATT GCGGCTTGTG TGAAAACAAT TGTCAATTAA CGGTCACAGT GTTTTCAGAT    1860
GGGCGCCAAT TTATTACGGG AAATCGCTGT GAACGTGGTG CACGAATTAA AATTAAACGA    1920
GAAGAACGCA AAgTAAATTT GGTTGATTAT AAGTACCGCA AGTTATTTAA GTACCGACCT    1980
TTGCGGGAAA ATAAAGCGTT TCGTGGGCGC TTGGGTATCC CTAGAGTCCT AAACATGTAT    2040
GAAAATTATC CTTTGTGGCA TACATTTTTT ACTGACTTAG GCTTTCGGGT CGAATTATCG    2100
CCACGTTCTA ATAAGCAAAT CTATGAACAA GGCCTTGAAA CGATTCCCAG TGATACGGTT    2160
TGTTACCCAG CGAAAATGGC GCATGGACAT ATTCAAGCAT TAATTGATGC GCAAGTGCCA    2220
ATTATCTTTT ATCCTGGTGT TGTTTTTGAA CAGCAAGAAA CAGTGGAAGC AGATAATCAT    2280
TTTAATTGTC CAATTGTGCA AAGCTACCCA GATGTTATTC GAAATAATGT GGATGCCATC    2340
AGAGAGGGCC AAGTGGATTA TCGAAATCCC TATCTTAATT TAGCGAATGA AGCTGCAGTT    2400
GCCAAAGTTT TAGCTGAAAA TTTTGCTGAT TTAGGGATTT CCTTAGAAGA AATTCAAACA    2460
GCGCTGCATC ATGGTTATCA AGAATTAGCA GCTTTTAAAA AAGAAATTCA AGAAAAAGGC    2520
GAAGAAACGT TAGCCATGTT AACTGAAAAA GGCCAGAGAG GGATTGTCTT ATCGGGGCGA    2580
CCGTATCACT TGGATCCAGA GATTAATCAC GGCATTGCCG AAGTCATTAC ACAGGAAGGA    2640
TTCCATGTCT TGACAGAAGA CAGTATCTCG CATCTTGGAG ACGTGCAGAA TTTACGTGTG    2700
GTGAATCAAT GGGTCTATCA TTCTCGCTTA TATGCGGCGG CTAAAGTAGT GGCTAAGACG    2760
AAAAATTTAG AACTTGTCCA ACTGAATTCC TTTGGTTGCG GCTTAGATGC TGTGACAACC    2820
GATCAAGTAG AAGAAATCAT GGATCGGTCT GGCAAAATTT ATACAGTCCT CAAAATCGAT    2880
GAAGGTGCGA ATTTGGGCGC AATTCGTATT CGGTTGCGCT CCTTGAAAGC GGCTGTCAAT    2940
```

GAGCGGGAGA AGCAACAGTT TGTTCCGCAT ATGCAAATGG AAGAGCCAGA GAAAATTGTG 3000

TTTACAAAAC AAATGAAAAA AACACATACC TTACTGTTAC CAATGCTTAG CCcGATtCAT 3060

CAATCTGGAT TAGTCGATGT AGCATTACAA GCTTCTGGCT ATCATGTTGT CTGTTTGCCT 3120

GCGCAGGATA AAGAAGCCGT GAATGTCGGG TTAAAATATG TCAATAATGA TGCGTGTTAC 3180

CCAGCAATTA TTTCTATCGG ACAACTAGTC GAAGCTTTGG AAAGTGGTCA ATATGaTTTG 3240

GATAATGTCA GTGTCTTAAT GACGCAAACA GGGGGCGGTT GTCGCGCAAC GAATTACATT 3300

CCCTTATTAC GAAAAGCCTT AAATGATGCT GGTTTTTCTC AAGTTCCTGT AGTTTCAGTT 3360

TCGATGGGCA ATAAAGGGGT CGAATCCAaT CCAGGATTTA AATTTACGTT GCCAATGATC 3420

AAACGCTTAG TGGTTGCCTT TTtATATGGT GaCTTATTCG AACGAGTAGT TTATCGAACA 3480

AGACCTTATG aAACAGAAAA AGGAATGGTG GATCAACTGC ACCAAGATTG GCTAAAAAGA 3540

GTGGAAGCCA ATGTCCGGAA TGGTTCGCTG ACACAATTTA ATCACTTCAT GAAAAAAATT 3600

ATTCGGACGT TTGATGAAAT TCCTTTACAA GAGATAAAAA AACCAAAAGT CGGTGTGGTC 3660

GGGGAAATTT TAGTGAAATA TTCGCCGACA GCGAATAATG ATATTGTGCG TTTGCTAGAA 3720

GAAGAAGGTG CTGAAGCAGT TGTGCCAGAT ATTGTTGGCT TTATGAACTA TTCTTTGTAC 3780

AATCAAATTT GGAAATACGA AAATATGGGC ATGTCAAAAC AAAGCAAACG TTTAGCAGAA 3840

TTTGCGATAA AAATCATTGA ATTGGTAGAA AAACCAATGG ATAAGGCATT GAGAAAATCG 3900

GTCCGTTTTG ATGGTATTCA TTCCATTTAT GATATGGCAG CAGACGCAAG TAAAATTTTA 3960

TCAATTGGCA ATCATACGGG TGAAGGTTGG TTTTTAACAG CTGAAATGAT TGAGCTCTTA 4020

AAACACGAGG TGAATAACAT CGTTTGTATG CAGCCgTTTG GATGTTTACC AAACCATATT 4080

GTTGGAAAAG GGGTCGTGAA AGAGCTACGA CGTCAATATC CGCAAGCAAA TATTGCGGCG 4140

GTGGATTATG ATCCAGGGGT ATCTTTGGTT AATCAGTTGA ACCGAATCCG TTTGATGATG 4200

GCCACAGCGA ACAAGTTGTT GAAAGAAGAA AATGTAAAAA GATAATAACT AATAGAAATT 4260

CGTTTAAAAA CATTTGACTT GGTCACTTTT TACTGACTAA GCCAAATGTT TTTGTTATAA 4320

AAAGAAAGCA TGCAACTAGA CCAACTTTGA TAAAAATTCC AATTCCTCTT TTATTTATCA 4380

TTTTTTGGTT GTTTACACGG TCTATACCAG TTATAATAGA AGAGGAGATG AACGAAGAGG 4440

AGGAAGCGTA AATGGTTCAA AATATACCAA TTTATATTCA AATTCACGAT AAAATCAAAG 4500

AAGATATTGA AAAAGGTGTC TGGAGTATCG GGGATCGTTT GCCCTCTGAG CGAGAATTAG 4560

CCTTGAAATT TGATGTTAGT CGAATGACCT TACGCCAAGC CATCCAGACT TTAGCAGACG 4620

AAGGGATTTT AGAACGAAAA ATTGGTTCAG GAACGTATGT AGCACGTAAG AAAGTTCAAG 4680

AAACAATGAC AGGTACCACT AGTTTTACAG AAATTACCTT GTCTCAAAAT CGCGTTCCTT 4740

CTAGTCGGAC GGTGTCTTAT TTTGTAGCGA AACCTAGTTC TAGTGAGATG GAAAAATTAC 4800

AACTAGGTCC AGAAGATTCA ATTTTACGAA TGGAAAGAAT TCGTTTTGCG GATGACATTC 4860

CGATTTGTTT TGAGGTGGCG AGCATTCCTT ATTCATTGGT CTCGCAATAT GGCAAATCTG 4920

AAATCACGAA CTCTTTTTAT AAAACACTTG AAGCCAAATC TGGCCACAAA ATTGGCCATT 4980

CTAACCAAAC AATCTCTGCT GTTCaAGCAT CGGAACAGAT CGCAGAATAT TTAGAGATTA 5040

AACGAGGCGA TGCGATTCtG CGTGTGCGCC AAGTATCCTA TTTTGAAAAC GGACTACCTT 5100

TTGAATATGT TCGTACGCAG TATGCAGGAA GTCGATTTGA GTTTTATTTA GAGAAGTAGA 5160

GAAATGATAA GTATAGAAAA ATGCCAAGAA ATAGGCTGAG GTArGATATT TtAaCCTCCA 5220

AArGTkGTAT CTGAAAAaTA CAATTTTTGG AGGTTATTTT TTATGrACTT ATCAGGTAGA 5280

ATTCAAATTA AAATTAGTGA AAGCATCTCA ATTCAmATAA ATATTGAATA GAGTAGCAAG 5340

TCTGAATACC AAAAAAGTGA GATGGTTTGT CCCGCTATAA ATATTGTTAA AAATAAACTT 5400

ATAATAGCTG CTAGTTTATT TGCTTTTATT CTTTTTTTGG CTGTAGTTCA TTTTTGTCCA 5460

TTTTTTACAT TTCTTTTTTA AACTGTAACC TAATATAAGA TATACCCGTT CAACGCTTTA 5520

ATCAAATAGT TGCAAAATCA GCAGTTGTTT TGCAAATTGT TACATTGAAA TTGTTAGTTA 5580

TTACCAATCA ATAATACTTA TGATTGTCAA AGTTATTATA TAGATATACT GAAAGAGAAA 5640

TTTATTTTCT TTTTCAATAA AAAGTTTAGC GGAAGGAGTG ATGCGGAATG AAAAAAATTA 5700

ACGTACGTAA AGCCGATACA GTTGATGCAA CAGCATGGTG GATTATTATC TAATTTGATA 5760

ACCAAATAAC GGTTAGACGA GATAGAAATA TCTCGTCATT TTTTAATAAG GAGTGAAAAA 5820

AATGAAGTAT AAATATCCCA GAATAAAACC AATCTATCCA TTATATCAAT TAAATAAAGA 5880

GAGATTTAGA GTTGGGGCCC AATTAGGAAT CACTATTGAA TTTGATGATG ATGAAGGTCA 5940

ATTTTGGACA TTATCTAATT TACTCGATGG AGTTAGATCG TTTGATGAAG TTGTCACAGA 6000

AATGAAGAGG AAATATCCGG AATTAACAGT AAAAGATATC GAAGAAGGGA TAGATTTTTT 6060

AAATGATGAA GGTCTAATCG AGGAGACGTT TCCTGGTAGA ATGATTGAAG ATAGATACTT 6120

AGCTAATGTG AATTACTTTA GTAGGTACTG TAAAGCTGAT GATGATAGAT TTGAAATACA 6180

GGAAAAGATA AATAACTTGA AAATACTATT ATTAGGTCTT GGTGGAGGAG GATCAAATAT 6240

TCTAACGGTA CTAGCGGGTC TTGGCCCCAA GATGATtAGA ATGGTGGACT ATGACCGAGT 6300

tGAAGCAAGT AATTtAGGTC GTCAATTACT ATATAGAGAA GCAGATATTG GTGAGAAGAA 6360

AACTGTAGTT GCTAAAAGAG CCATTAATGA AATGAnACTC AAATATTAAC GTTGAAACTG 6420

TAGATAAAAA AATAATAGAT GTGAATGATG TTGTTGAATT GACAGAAGGA ATAGATATCA 6480

TTGTTTGCGC TATTGATGAA CCCCCTTTCC TTATACATAG AATTGtGAaT GAAGCTATTG 6540

TAAAAGTTGG TTTGCCGTGC GTGTTTGGTG CCTCACAGGT TAGCCGAGGA AGAGTATACA 6600

CTGTAATACC CCAGAAAAcA GGATGCTTTG aTtGCATGAA TTTGAACTtC AGTAAAAATG 6660

ACCCTAAATT TGTTGAACAA TTTGTAGGTT TTAGAAATAT TCAATTTGCT CCCCCATCCA 6720

TAGCTTATGG TCCTGGTATC TTTCAATTAA CAGCATCAAT TGTAGATGAG TTAATAAGAG 6780

TGGTAACAAG ATATGCTGAG CCAAAAAGTT TGGGAACGCA ATATGAAATT AACTATGAGG 6840

ACGGAAACTC GTTCACGCAT AAAACTTGGC CAAGGTTTGA GAGTGAATGT CCAACTTGTG 6900

GTAAAGGAGA TGTCTCCCAA TGGGAAATAT TTCAATATTA TCAAGAGAAG AAATAGGCCG 6960

AGACTATACC GAAGCTCGTA CGTTAAAACG TAAGCTAAAC AATGGCGATT GGGTTTATGT 7020

ATCAGAGGTA CTTAATGAAT TAGTTgAAAA TAAAATTATG ATAAAAAAAG CCGATCAGAG 7080

AGATGATTAT AAGCCGCTAT TTGTCGAGCC TAAGAATATC TTATTTAAAA GCTCCTCTTT 7140

CCTGATTGAT GTTAACGGTA TTGTTGCGTT TACGATTAGC ATATTTTTAT TCGTAATCAA 7200

GGGGTCTCCT TTTGGAAGTG CAGTAAATGT TTTGTTT 7237

(2) INFORMATION FOR SEQ ID NO: 240:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1989 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

GAGGATTTGA TTTGATCnTC TGAGTTCACG TgkCGCTAtC agcTCTGATA cTTAcaCTTT 60

AcGTCGGAcC CAgAAcAGGC TTATTAGAAA TGATTCaAAA gATGTtGgCT GTTCCCgArT 120

GTGTTTGTGT TTtATTTGTA TAATTAAGTT GaCTATTTCC tTGGAAAATw cAGGAGTCAG 180

ATGCTAAATA ATTTTTGCAT CTGACTCCTG TTTTCTTATT TTAAACGATG GATTGGTTCT 240

CCAATTCTTT AACGTATAAA TAAAAAGAGC CAAAGTCTAA GCGATTCATT TGTAAAGATT 300

CATGAATTTT TGTGAACCCT TTACTTTCGA AAAATCGAAT CATTCTATGG TTTGTATGAT 360

TAGTTTGAGC ATACATTTTA CGTCTTCCTT CCCAGACTGC CCGTTTCTCC AATTCATGGA 420

ACAATAAACT GCCATAACCC TTGGCAATAT AATTAGGCGA AGTTGCAAAG CGCTTTAACC 480

AAACAAAGTC TTGTTCCTGC TCCATACAGA ACGTTGCCAT TGAAAAAATC ATTTCTTCGT 540

GCACTAGCAA ATATAATCGT TTTTTTGTAA TGTCTTCTTG GATATCTCCT TGAGAAGGAT 600

AGTCCTTTGT CCACTGTGTA ATGCCCACAA CATCCATATA TGCCTTTACT TCTGTTAAAA 660

ATTGAAGAAT CATCGGTAAC TCTTTCTCAT CAGCCACTCT AATTTTCATT GTTTATTCTC 720

TTaATAAGGT AACAATAGAA TCGACAAGAC AATCACACTA ATAATGCCAA TGATGGTCGG 780

TACACTCGTC CGTTTTAAAA TTCTGATGGG ACTAACTCCT ACTGTTGATG CGACAATCAT 840

GACAACTGCT GCTACAGGAG AAATCGTTCT AGTTAAATTC GCAATCATCT GCATTGGCAA 900

GGAAATTAAG ATGCCATCAA TCCCAGCTTT TTCAGCAATC CCAGGGATTA GTTCGATAAC 960

CGCATAAAAC ATTGCCAGAC CACCACCGCT TAAAATTCCA AAAAGGGTCG TTGCACCGCT 1020

GAAAATTAAG GTCGTGACAA TTCCTGCTGA ACTTGATGCT TCGACAGATG CCATAATACT 1080

ATCAATAATT CCTAACGTTT GAAtCGCAGA CGTGAATAAT GAACCGCCTA CGACAAGCAT 1140

CACAACTTGA CTAAACCCTT GCCCCATTCC TTTAAACATC TCGACAGCTG TATCTTGAAC 1200

CTTTTTAAAA GATTTCAAGC GAACGyCTCA ACCAATACAG CGATGAAGAA AGAAATAAAT 1260

GTAAGAACAA AAATATCCAT CGTTATTCCT TTAATAAACA TTCCTGCGAT TCCCACTCCA 1320

ACAATTAATA ATAGCGGTAA TATTGGCAAT ATTGCATAAA ATTTAGGAAC ATCAAATTCT 1380

TTTTGTTTGC TTAAGCCCTC TTCATTTAAT GAGACATAGG CAGCCTCACC TTCTTTTTTA 1440

TCGCAGTATT TTTGCCAGAA GTACTGTGCA ACCGCAATGA TTAATAACGA TGGCAGCGAA 1500

ATTTTGGCAT TCCATACAAC ATAATTTAAC ACATCATAAC CAAGGGTATT TGCGGCGATT 1560

ACGTTATCCG CGCCTAAAGG TGTTGGCATG ATAGTTGCCG TCATTGCAAT CACACCTGCA 1620

GCAGTTAATG ATGAAATCCC CATGCTGGCT AGCATCGGAT ATAATATAGA CATTAAAATA 1680

ATCGCTAAAC TTGATGCACT AGGAACTACT AGTGACATTA AATTTCCAAT TAAAAAGACC 1740

ACAGGGACAA ATAACGATTT ACGTTTAATC TTCATTAATG GTTTCACCAA AAAGTTGACA 1800

GCCATCTGAT TGGCCCCGAT TGCATTCATA TAGCCTGAAT AACCAAAAAG AATCATAATT 1860

ACAATTCCTG CTGAACTCAC TTGTGCAATA ATTGTATCTT TAAACTTCAA GAAAATATCA 1920

AATAATCCAA AACCGGTTCC TTCTCCTGCA GGATAAATGG GATGCCCTAA CATCCATGCA 1980

CATGCAAGT 1989

(2) INFORMATION FOR SEQ ID NO: 241:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1814 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

GCTTCATTAC TAAGAAATAA GGTATGTAAA CAAAAACAAC TAAGATTCGG TATTTTTCCA 60

GAAAGAGTTA TTATTTAAAT GTTTTTAAAT AAAAAATATG GAGATAAATG GCAGATCGTT 120

TGTTGCTTTT TTTAATyTCG AATTATCAAT TAAGTACAAC ACTTTATCCT GTAAAAACTT 180

CATACGATGT TTTyCATTTT AAGCACTTTT GTGGTCTAGT CTATGCATAA AGGATTGAAT 240

ATAATTATCT GGAATTAAGA TTATATCTTT CTCTTTTGGT AGGTACTCTC TTAATAATCT 300

ATTTGTATTT TCATTTAGTC AAAAACTGGG TAAAGGAGTC TACAAAGATA AACCTAAAAA 360

ATAATTTTGC ACATATGCTT GATGaCCTTC CTGGAGTAGC TGATAATGCT AAAAAAATAT 420

TGTTTGATAT GAAATGkACT ACTGkAATAG ACGATTGTAA TGTtCATTTA CGAGATGTAC 480

TTAATTTATT AGAAGATGAA ATGCACAGCT TTAGaTTATC TTCTATTGaA GCtAATGATT 540

TAGaTAAACT AtATGATACT TTTTATTACT ATGGnAAAAA AGamCAATTA ATAGAATTTC 600

TGCTTGACCA GATGACTAAT GCGATAAATT TTAAACAAGC TGATAAGGAT TTTCTTATTT 660

ATTCATGGCT TGTAGAAAAA CTTACTGAAG AACAAATATG CGAAATGCTC GCGAGAATTA 720

ATGAAAATAC CCAATACCAT TGGAATAACA ACCTTAATTC ATTTGTTGAT GAGTTCGTTG 780

CTTATTATAA AGAAACAAAT GGAATTGATC TAAGAATAAA TTATTTGGGA GCAATTTATA 840

CTAATTTAAA AATTCTCAGC GATGAATATA CTCTCTCTAT CTCAGATTAC GAGCTTATTT 900

TCAAATTTTT TGAGGATATA ATAGAGAGTC ATAGCGAAAC AATATTTGAT TTTCAGCGTA 960

AATATAAAAA GGATATTGAA ACAAATTTTG GTAATAGTTT GGAAAACTTT CCTAAACTGA 1020

ATAAGTTaTT GTTCGAATAG ATTTTGAAAA GCTTACAATA AATAAAAGTA TTAATTTTTG 1080

TTACATCAAT TAATCTCTAA TTCATGAATT TCTTTTTTGA TTCTATATAT AGTTTGTCTC 1140

GTAATATTGT AATCTTTCGC AATCTTTGAG ATAGCAACCC CATTCTTAAG ATCCTGAACG 1200

ATGCTTTTAT AAACAAGGCG ACGTTGGGGA TCTTTTGTGT CTGCACTGTA TAACTTAGGA 1260

CGACCTTTAT AGACGCCGTT TGCTTTCGCA ATCTTTATCC CTTGTGCTTG GCGTCTTTTA 1320

GATTCCGTAC GTTCTTGCTC GGCAATCATG GCCAAAATCT GAATGATTAG GTCTTTGATA 1380

AATTTATCTA GGAGTGGATT TCCGATTGCT TCAGCCATGA TTGGTAAACT GGTAATAATT 1440

AGCCGCACAT TTTTCTTCTT TAGATAATTT ACTGAATCGA TAATCTCGTC ATAATTACGT 1500

CCAAGTCGAT CGATGGCCTC TACAATAAAA ATATCTTGGG CTCTAACAAA ATCTAGAGCT 1560

TCCTGAAAAA TCGGTCGATG AGTAATAGTC GCTCCGGATT GTTTCTCGAT AAAAATTTTT 1620

TCAGCACCAA ACTTGTTCAT TtCCTCAATT tGACGTtGTT CATTTTGATC GATTGATGAA 1680

ACTCGAATAT ATGCAATTTT CCAAATtTTT ACCTACTTTC TATAATTTTT AAAAtACACC 1740

CTAATCATAC ATTCAAAAAC CATATTTTTC AAkGTtCTGT TGGTGTCTAT TTAGGGGTCC 1800

TTAATTAGAC AATC 1814

(2) INFORMATION FOR SEQ ID NO: 242:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12445 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

TATGGTTCAA GGTCCAGTGT TTaATGAAAG ATATGTTTGG TCAGACTAAT CGAATTGTTG 60

TTCATGTTAT TTTTGAAGAA TGGTTGGAAC TGTTCAAAAC CGTTGATGAT ACAGAAGCTT 120

CTAAAATGGA GAACCGCTTT GTTGCATAGG AAGGAAGTGA GAATATGCAG TATCTAATTG 180

ACAGTTATAG TGCGTATACG GTTGATGGTG GTTTGTTGGA TAAAGGATGG CAAGTAGTGA 240

ATTGGTTCTT TGTAGATATA CCATTCTTTA TCTTACGAAT GTTTGTTTCT TTTTTTCTTT 300

TTTGTGAGAA CGTCTTAAAC CAATCTAGTT TCTTTGAAGG AAAACAAGAA ACAGTATTGA 360

ATATGTCAAA AAATGTTCTA AGTGGCATTG GTGGCAAGGG ATTTAGAGGT GGTTCTTTAC 420

TAGCATTAGC GATTCTTATT AGTGCATATT ATCTGCTTTA TCATTTCTTT TCCAGTCGTA 480

GAAGCTTTTC AAAAGTATTC CTTCATTATT TAGCAGTTGT TATTTTGTTT GGTTTTTGGT 540

TTGGCTCTGT ATCAACGTCA ACAGGCACAA CAAGTGGCTC AATGTTCTTG ATTCAATCAA 600

CTAACGCAAT TGCTAAAGGA GTACAAAGTA CGTTTACTAC AAGTGCCAAT TTAGGAACAG 660

AAGGTTCTAA AGATGGAGAA GTATACCAAA GTCCTATTTT TGATGCCACA GTAAAACAAA 720

CATTTAATTT TGTAAATAGT GGTTCACTTG ATGGGAAAAT GGAAAATGGT AAAAAGTTAG 780

ATGAAGGTAA ACTACTAGAA AAGCCTAAAT TGTCTAAAAA AGAAAAAGCA AAGTTTGAGG 840

ATGAAAGAAG TACCTATATA AAAAATAATG AGAAAGATAA TCCATACTTC TCACAGGACG 900

GAGCAAAAAC AATGGAAAAA GCTTTTGGCG TTGGTGTGGG TTTAGTCAAT TTAGCAGTCT 960

TAGCAGTGCC TGTCACGTAT ATTAATATTA TGCTAAATGT AATTCAAATC ATTGTTGATT 1020

TACTCATTCT TGTTTTTCCG ATTATTGCGT TGGTGTCTTT CTTCCCTCGG TGTCAAATGA 1080

TGATGTTTAA ATTTTTTAAA TCATTAATTG GCATTTTATT TTTACCTGTA GTTTTCGGTA 1140

TATTTTTGTC TGTGTTATTT TGGATCAATA AGTTAATTGA CCAAGCGTTT TTAGGATTGA 1200

TGGATAAAGT AAGTAGTAGT TTATTAATGG TGATGTCTGG TGGTGTTTTT CTTTTAGGCA 1260

CATTAATTGT TACTGCATTG GTGAAAATTA TTTTGTATCG GAAAATTTGG AAAAGCAGAT 1320

ATAAAATATT ATCATTCTTT TCTGATGGCC AAGTTCAACA ACCTTCTTTT GAATCAAAAA 1380

TGAATGAAAA AACAAAAGAA ACGGTTGAGC GTACAGGTGA AATTGGTGTT GGAGCAGTAA 1440

AAGCAGGAAT AGGTGCTkCC ACTGGAAATT TAGCGTTGGC TGCTGATGGT GCTAGTCATT 1500

TAATGCCAAA GCATGACAAG GCATTGAACT TAGCTAAAGA CCATTTTATT GACGATAATG 1560

GGCAATTTGC TGGTGTTAAA ACAGGATTGA ATTCATTGTT GAATCGTTCT ACTCAAGAAG 1620

AACAACAACC ATTAAAAGAT GATTTATTGC CAGAAGAAGA AATTGTGGAA GTAGAAGAGC 1680

CAACCAATGT TGATAATGAG GAACTAGAAG AAATAGAAGA TAAAGCTTTA ACTGATGAAG 1740

TGGATAATTT TGAAGATTTA ACAGATGATG AAATGGATAA TTCTTTAGAA GAAATTGATG 1800

TTCCTGTATT AGATGAAACT GTATCTGATT CTGATACTTC TTTAGCTGAT GAAAGTAATA 1860

TTGAGTTAGA GCCAATTATT GAAAATGAAC TTGATAGTTT AGATTCAAGT GTAGAATCAA 1920

GTGAAAATGA TAGAGAAGAG CCAATTCATG AAGTTGATGA GCAAGAAATA GCGGACAATT 1980

TGAATGTTAC GGTCGATAAT TTTGATGAGT TGGCATTTGC TCGTGAAGAA CGTGCATTTT 2040

TTGATGGTGG AGAGGATAGC GAGTTAATCA CAAACAATAA TGATTCATTA AATAACGTGT 2100

ATTCATTTTA CCAAACAGAA GAAAACTTTA ATAATTTAGA ACAAACAGAG GAACAATTTT 2160

TTGGTAGTGT GAATACGGAA GAATATAATT TTGATGTTGT AGAGGGCTGA TGAAATGTTT 2220

ATCAGTCTTC TTTTTATTGA AAGGAGAGAT CATGAAGAAA TTAGGCAAGG TTTTAATTGT 2280

TAGTTGTTTT ATTTTTATTC TTCCTTTTTT ATTATTTTTA GGTGTATTTT CTTCTAGTGA 2340

AAGCGGAGAT TCTTCCCAGT TTCAGCCCGC TACACCACAG GAAAAAGTAG CATTAGAAGT 2400

TTCTAACTAC GTGACGTCAC ATGGCGGAAC GTTGCAGTTT GCTTCCGCTT GGATTGGCAA 2460

TATGGAACAT GAAAGTGGAT TAAATCCTGC TAGAATTCAA AGTGATTTAT CGTTTAATTC 2520

AGCGATAGCT TTTAATCCTT CGTTAGGCGG TTATGGAATT GGGTTAGGAC AATGGGATTC 2580

AGGACGAAGA GTTAATTTAT TAAATTTTGC AAAAAGTCAA AAAAAGGAAT GGAAATCAGT 2640

AGCTTTACAA ATGGATTTTG CGTGGAATAA GGATGGTTCT GATAGTGACT TACTTAAAAG 2700

```
AATGTCTAAA TCAAAAGATG TGAATACACT TGCGGTAGAT AtTTTGAAGC TGTGGGAACG    2760

AGCTGGAACA AAAGATGATC CCGCAGaACA AGTAAAAAGA AAGGCTAGTG CTAATAATTG    2820

GTATAAACGA CTTTCTACAG GTTCCATGGG CGGAGGTTCA GCCAATGTTG GTGGAGGAAA    2880

AATTGATGCC TTGGAAAAAG TGATGGGGCA AACTATTAAT GGTGGTCAAT GTTATGGCTT    2940

ATCTGCTTTT TTTGTTGAAA AACAAGGAGG TCTACAAATG ATGGGTACGG GGCATATGTT    3000

TGCGAGTGAA ATTGGTAATG ATTATCCTTG GAGTTCAATT GGTTGGACAG TCATAAAGAA    3060

TCCAAATTAT TcAGATATTA AAGCAGGAGA TGTCATTAAT TTTGGTCAAG GTGGTGTGGC    3120

TACTAGTATT TATGGGCATA CTGGTGTAGT GGCAAGTGTT GAAGGTAAAA ACAAGTTTAC    3180

TACTTATGAG CAAAACGCTG AACAAGGTCA AATTGTTGCT AAGTATTTTC GGACTTGGGG    3240

ATTAGATTTT CCACATGTGA CCAGCATAGT AAGGAAATAG GAGGAATTAT GAATGACTTT    3300

GAGTAAGAAA AGTGGTTTAA CTAATCGTCA GATAAAAGAA ATCTGTTCGA AATTTAACTG    3360

TAGAAAAAAT GAATTTGCAG TTCAAAAGTA CCAAGATAGT TTTTTATTAG CTGTGAATAA    3420

TAAAGAGTAT CGTGTTAAAT TTTCTGAGGG GATTTTTTCA AAGATTATGT ATATTAAAGA    3480

AGTGCAACGC ATTTCAAAGA AAGTAGGTAG AAAATGAAAA AGTTAAGTTT AGTAGTTTTA    3540

GTATTAGGTG TTGGATTAAT ATTTTCTGCT TGTGGAGGAA ACACCAAAGG TCATGAAACG    3600

GCGGCTTCGT CAACAACTGA ATTGGTGACT ATGAAACCAG TTGATAATAT AAATACTACG    3660

AAAAGAAGTG AAGTAACTGA AAAaAGTAGT ACTAAACCTA TGCAGATTAA AGATAAACTA    3720

GCAGAAGAAA AAGTATCTGA TACTTACCAG TTGGTGGAGG ACTTTGGaAA TACTTATGCm    3780

AATTTTTCTA GTATCAATGA TCGTAACGrA AAATTAAAGr AATTAATGAC TGCTGAATGT    3840

GkAAAGAAAA ATGGTATTGA TGTGAAAACT GGCGCAAAAT TGGAAAGTAA AGGTGAAGTA    3900

AGCGCAATAT ATCAAAATGA CAAAAAGGAA TACGCAGTCT TACTTGATTG TGAACAAAAC    3960

GGAACAAAAA CAAGAGTACT TTTACTGGCA AAAGTGCAAG GAAATAAGGT AGCGGAAATG    4020

ACCTATAATT CAGTGAAACA AGAGTATTAA CCTTGATATA TATGGATTTG TACTTTAGTT    4080

AATGGTATAA TGTATGTATA AGTGAGGAGA TGAGAAAATG AAAAAGGTAG TTTCAATTTT    4140

GTTGATGGTT GTTGCAGTCT TCACATTAAC TGCATGTAAT GGTTCTAAAT TAGATAAAAC    4200

AGGTGAAGAA TTTAAAAATT CTATAATGAA AGATTCTTCA TATGGTGATG AATATTCAGA    4260

AGATGGTTTT AGTTTTTTAA TATATAAAGA TAAAGACACT AATCGTTATT TGGCTGATGT    4320

TTGGGTTCCT GTTAAAGATG AAACTAGCGC ATTGGAGTAT TTTTATTATT ATGATGAAGA    4380

TAAGCGATTA GATAGTACTA AAAGTAAAGT AACCTTTGAT GATATGAAAG CTAGTGGAAA    4440

CTATGAAGTA GTGTATAAAT CaGGGAAATT TAAATAACAT AGCCTGTCAC GTAATTTGTT    4500

ATTTTTAAGT GGCTAAAATT TGGCTAAAAC ACTTTGAAAT GTATGGAAAT GCTGGATTAT    4560

TTGTAAAAAT AATGTTGATA TTAAAGCATT TGTGTGGAAA ATACATTTTA TAATTTACTC    4620

GTAACGCGTA GGTCACAGGT TCGATCCCTG CAGCTGGCaT CTAAAAAAAC AGACAGGcAC    4680
```

TACCATCTGG TAGTGCCTGT CTGTTTTTTT AgmTAAAATT TTGGaTTAGT GCAAATGaAG 4740

CAATGCCTAC CAGTACAATA AGCAATAAGT TTTTAGTAAG TATTGCGGCT AAAATGGTAG 4800

GAAATGAAGC GAATAAATaA GGAAGATTGA GGCTTGGTAA ATGCCCAATT CTTTGTTCAA 4860

ATAAGCTACT AAACCATAAG GCGGACATAA TCACAATTGG AACAAAACTT AGATATTCTA 4920

ATAATTTCTG TGGCaGACTG ATTTTTTTAA AAtAAAAATG GCAGTATTCT TGAAAGCCAG 4980

GTTGCAACGG TAcAGCCTAA AATTGTTAAA AAGAGATACT CaGAAGAAGG CATGTTTTAT 5040

CATCACTCCT ATTCCACACC CGATTAaCGT GmCAaCTArG ACAATTaAAT TACTGGGAAT 5100

AAAAATTAGA CCAATAGACA TCAAaCCAAA AGTAATCAGG ACCATGATGA GTTGTAAAGC 5160

GATTTTCATG GAACGaTCAC TaATTATCTG CAGATAGAGT AAACCAATGA ACATAGCCAC 5220

GACCGCAAAC TCAAGACCTA ATGCTTGAGG GTTTGCAATA AAGCGCCCTA GATAGGCACC 5280

GATGATCGTC GAAAAAACCC AAGTCGCATA GGAAATCAAA TTGGCAGCAT TAAACCACTC 5340

AAAACTTAAA CGTCCTTTGG TGTGGTTTAA TTTATTCATT CCTAGTGCAA AACTTTCATC 5400

CGTCAGTAAG GTTCCCAACC AGAGATTTTT ACCGAGAGAT TCGGCTTTAA AGTAAGGCGC 5460

AATGGTCATT CCCATAAGAA TCATACGCGC ATTTACTAAG AAAGTCGCTA ACACAATAGA 5520

TAAGATAGGG CTGCCGCCAG TCAACATACT AACTGTTACA AACTGAGCAG ATCCTGCAAA 5580

AATAAAAAAC GACATCGCAC TAACAATAAG AGGGGAAAAA CCTGCTGCTT GTCCTACGAT 5640

ACCAAAAGCA ATGCCAATAC CAATATAGCC AAAAACCGTT GGCAAGGCTT CACGAATGCC 5700

AGTGCGTGCA TCTAATTGGT GGTTCATTCT CTTTTTTTCC TTCTTTCTGT ATAAAATTTA 5760

AACACGTGTG TATGAAAATC TAGACCTTCG ATAATGAGAC ACAACGAGCG AATGAGCGTT 5820

TTCATTATTA ACCACTCCTT TTGAACAACT AAGATATATA TTGTATAATA CAATAAAGAA 5880

TACTTTTGTT CAATATATTA AACTTTTGGA TGATAAAAAG AAACAGAAGG AGCAACTATT 5940

TTGGAGATTA ATACAATTAT TGGCAATAAT TTAAAaAAAA TTAGACAAGA AAAAAAACTT 6000

ACGTTAGATG AACTAGCTGG GATAACCGGA GTGAGTAAAG GGATGCTTTC GCAGATTGAA 6060

AAAGGAACCA CCAATCCTAC CATAAATACA ATTTGGAAAA TAAGTAATGG ACTCAATGTT 6120

CCATATACGT ATTTATTAGA AGATGACTAT GTTAAAGAAA ATCCCGTGGT TAAAAAAACT 6180

GAGACGGTAA CACAACATTC TGAAGATCAC TCCTATCGAG TCTTTAATTA CTACTCGGAC 6240

AACCAGGAAA GAAACTTTGA ACTCTTTCAG TTTGAATTAG ATCCAGGGAA AAAGTATCAA 6300

GCTGTTGGGC ATTCAAAACG TTCGACAGAA TATGTCATGG TTATTCAAGG AGCATTAGCT 6360

CTTGAAGTGG GCGGTCAGAG ATACCTTCTT CCAAAAGACG CCGCTATGTA TTTTGATGCT 6420

AATCAAGTAC ACACGTATAT AAATGACGGA ACAGAGAAAG TCCAAGCAGT CGTCATTAAC 6480

TATTATACGT AACAGTGGAA AATAAACATT AGAAAATTGC TGCATCTGTT ATTTTATGGA 6540

CAAAAAAAGT AAACAATGCT ACCCTAAAAA TAATAGAAAA TGAAAGATAA AAGAAGAGGT 6600

CAAAATGAAT AAAAAAATTG AAAGTCTTGG TAATACAGCA TTAAGTTTAC TAAGTAATAC 6660

GATTATTGTT GTGCCTTATT TACTTTACAT TCAATTGAGT GATGGACACA ACCTTCTAAC 6720

AATTTTACCA TTTGTCTTAT TTTATACATT GCGTATGACA GGCATTTTTC TGGTTCGAGG 6780

AATTAATTTG TCTTTAACTA GTCTAGGTCT ACTAAAGATT TCTCTATTGT TAGGCTTGGC 6840

AGGCTCCTTA ATTGGGATTG CCGGCGCTTA TTCTTTCCCG CTTTATAGTA TTTCAGGAAT 6900

GTTGTTAGGA CTTAGTGGAG CGTTGTTGCC ACCAAGTAAT CAATCAATTC GCTTTTTCTT 6960

AAAAGAAGAA GGTCAAAAAA TCAgTCmTAG CAATCyTTTA ACgACAGcTT AwTGaCCgCT 7020

GtTTTAwTTG GCGCTTTATT TTTTAAAGCG ACTGAGATTG GACTGGCGTT ATTGGTGTAC 7080

GCGCTCTATT TTCTGGTTGC TTTAGTAGCA ATCAAGAGCT ATCCTGTTTC ATTAAACGAG 7140

CTAGAAGAAG AATCAGCAGA AGTTTCACGA AAAGAATTAA TTCTGTTTGT TATCTTTTTT 7200

GTGTTATTAC TTTTATTAAG AAGTGGTCGC TTATTAACGA ATGCGGTGGA ATTTGATTAT 7260

GCGATTGTTG GCGCCTGCTG TTTCTTTGTG GTGGCTGTCC TTTTCGTTAG TCATTATGGC 7320

AAACGCGGTG CGTATAAAGT ATCATTAGAA ATGAAcTTGC GACCTTGATT AACGGGATTG 7380

TGGGGAACTA TTTATTnTAT TGGTTCCATT TATGTTGCAG GGtTaCGGCA GAGmtCATAT 7440

GGGCAtCTAT CTCwATTtAC CTtATGTTTt AGGGAtGAwT TTtGCAAtGA TTGCAGCTAG 7500

TAAAcGAAcA ATGGTCAAAC AaCATACCAa TGATTGGTCT GGCGGGAAGT TTGGGAATAT 7560

TACTtCTTAC TTCGTGGACT ATTTAGGCC TGTTTCTCTT GaGTTTTTTC AAAAGCACAT 7620

TGAATAGCTT TTTAGCTAGG CGATATTATC AAGAGACAGA TTTGCCAAAA GATAGCCGTA 7680

TCCTTATCAA ATACACAACA CAAACAAAAG GGAGTTTATT CCATCAATTT ATTTTGATGG 7740

CGCTTATGTT AGTCACTGTA GTAGGCAAAG GTGGCACGAC ACAATTTTTA TTAGAATTGA 7800

CTGGCGGCAA AGGCATTTCA ATGCAAACAA GTCAATTTTT AACTTTCATT AAAAATAGTA 7860

ACTCGCTCTT AGTGTTATTG TTTATTGCAG TTTATTTCGT GGTCGTTTTT CGTCAAAAGA 7920

AACGATAATT TACAAATATT AATAAAAAAA TACTTGGAAA AACAGTGATT GTGGCTGATT 7980

TTCCAAGTAT TTTTTTATGA AAAGATGTTT TTTTTTTTCG AAGATTCAAA AAAAATTAAT 8040

CAATTGTTTC ATTTTTAACA AGCTATTATT AGGTTATATG AATTTTATTG AAAGGGGATT 8100

AGTATGAAGA AAAAAACTTT TTCTTTTGTG ATGTTGAGTA TACTTCTCGC ACAAAATTTC 8160

GGGTTTGCCG TAAATGCCTA TGCTGTAACA ACGACAGAAG CACAAACAGA GACCACTGAT 8220

ACAGCAAAAA AAGAGGCAGA GTTATCGAAC TCAACACCAT CTTTACCTTT AGCAACAACG 8280

ACTACTTCAG AAATGAATCA ACCAACTGCA ACAACTGAAT CGCAAACCAC AGAGGCGAGC 8340

ACAACAGCTT CCAGTGATGC TGCTACACCA TCTGAACAAC AAACAACGGA GGACAAGGAC 8400

ACCTCACTTA ATGAAAAAGC CCTGCCAGAT GTTCAAGCGC CAATTACAGA TGAACTACTT 8460

GACAGTATGA GTCTTGCGCC GATTGGTGGA ACAGAATACA GCCAAACAGA GGTTCACCGC 8520

GAATTAAATA CAACACCGGT AACCGCTACG TTCCAATTTG CTGTTGGAAA CACAGGTTAT 8580

GCACCTGGAT CAGTTTATAC AGTTCAATTA CCAGAACATT TAGGTTATTC AACTGTCAGC 8640

GGAGAAGTGA CAGGCATTGG CGCAACTTGG GCAGTCGATG CGGCGACCAA AACATTAAGT 8700

ATTACGTTTA ATCAACGAGT TTCAGATACT TCCTTTAAAG TAGAACTAAA AAGTTATCTA 8760

ACAACAGAGG CGGAACCATT AATCAAAATT GAAACTCCAG GAAAAAATAA AAAAACCTAC 8820

TCGTTTGATT TATATGAACA AGTGGAACCA ATTCAATATA ACGAACGAAC CAGAACGACG 8880

GGGTTAGATG GCGAAATTTT TTATAATTTA GACCGGACGT TAACTGGCAA TCAAACATTA 8940

GAATTATTAA CAACAGAGAC GCCAGGCGCT GTCTTTGGAA AACAAGATAA CTTGGAACCT 9000

CAAGTTTTCA GTTACGATGT CGACATTAAT GGTCAAATTT TACCAGAAAC GCAAACCTTG 9060

TTAACACCTG GCAAAGATTA TACATTAAGC GATAATTCAC TCGGGCGGAT TGCTGTAACT 9120

GTTCCAAACA TGAATCAACA AAAAGCCTAT TCCTTATCGA TTAATCGGAC AATTTATTTA 9180

GAGAGTGCTT CGGACTATAA CTACTTATAT TCGCAGCAGT ATCCAACAAC AAAAATTGGG 9240

TCAATTTCTT TGAAAAGTAC GACAGGAACT AAACAAACAA CCGATTTTAC TGCTAAGACG 9300

AGTCAAACAA GTAAAGTAAT TGCTGATCGT GAAATGCGTA GTATGTCCTA TATCAGTTTT 9360

CAAAGCAAAG GGAAATATTA TGTAACAATT TATGGCACGT TAACAGAAAC AAAAGTGGGT 9420

CAACAAATCG TATTAGAGAG TACAAACGGT CAAGAAATTA AGAATCCTAA ATTTACGGCG 9480

TATGGTCCTT TATATGAAAA TGTAAAATTG GAAGACTATT TTGATATTAA AACTGAAGGT 9540

GGCAAGCTCA CTTTAACGGC CACAAAAGAT AGCTATTTAA GAATAAATAT TTCTGATTTA 9600

ACAATGGATT TTGACAAGAA GGACATTAAT CTATCATTAA GTACACCTGT AATTGGTCCT 9660

AATAAAGCCA TTCAATTAGT ATCCGATCAA TATATTGAAC CAATTAGTGT TGTTAATCCT 9720

TTGAATGCTG AAACTGCTTG GGGTAATTAT GATCAAAATG GTGCCTATTC ATCAAGAACA 9780

ACTGTCTCAG TTATGGGAAG CAAAGAGAAA CCGATTCAAA ATTTAGAAAT TAAAGTAAAG 9840

CATCCTAATT ATCTTTCATT ACGAGCTACA AAAGAAATTT ATTTTTATTA CAAGTTAGGA 9900

ACGGATTATA CAGTAACGCC AACGTCAGAT GGTTCAGTTA TTAAGTTCAC TACGCCAATA 9960

ACCAACGAAA TCCAAATTCC AATTGGTTTT AATTATGTGC CAGATAGTTT GCCAAAAGAT 10020

AAAAGTATCC CAGTCGATAC GATACCGATA ACAATGAGTG CTGAAGGTTT AACTCCAGTT 10080

GATACGACAG TAACTACTAA TAGTAAGCGT GGTTCTGAAC GAACACTTCA AAGTAGTAAA 10140

AATCAATTCC TTGTCAATGC ACGAAATGAT TCTTTTGACT CACTAAGCGT CCGTACAAAA 10200

ATTcCAGcTG GCGCCGATGT TCTTTTTGAC ATTTATGATG TTTCAAACGA TCAGGTAGAT 10260

TCAATTTATC CACAATACTG GGACCGCGGT CAATACTTTG ATAAACCAAT GACGCCAAAC 10320

AGCCCTGGAT ATCCaACGAT TACTTTTGAC GAAAATACCA ATAGTTACAC GTTTGATTTT 10380

GGAAAAACCA ACAAACGTTA CATTATTGAG TATAAAAACG CCAATGGCTG GATCGACGTG 10440

CCAACTCTTT ATATAACAGG GACAGCGAAA GAACCACAAT CGAATAATAA TGAAGGCTCT 10500

GCTTCGGTTT CTGTTCAAAA TGAAGCGTTA GACATTTTGA GTGCAACACA AGCGGCGAAT 10560

CCAACATTAA AAAATGTAAC AAAAACGACA GTAACAACAA AAAATATTGA TAATAAAACA 10620

CATCGTGTGA AAAATCCAAC GATTGAATTA ACACCAAAAG GCACAACCAA TGCTCAAATC 10680

GATTTGAATT CTATTACCGT GAAAGGCGTG CCAGAAGATG CTTATTCATT AGAGAAGACT 10740

ACAAACGGTG CGAAAGTCAT TTTTAAAGAC TATACATTGA CAGAAAACAT TACGATTGAA 10800

TACAATACGG TCTCTGCAAA CGCTGGCCAA ATCTATACAG AAACAACAAT CGACTCTGAA 10860

ACATTGAACC AGATGTCTGC TAGCAAGAAA AAAGTCACCA CTGCGCCAAT CACATTGAAA 10920

TTCTCAGAAG GTGATGCGGA AGGTATTGTT TATTTAGCAA CTGCCACATT CTACACGCAT 10980

AACGTAGAGG ATGAAAACCA AGCAATTGCG aAGGTTTCTT TTGAACTAAT TGATAATGTC 11040

ACGCATACAG CAACCGAATT TACAACAGAT GAAAAAGGTC AATACTCCTT TGATGCCATC 11100

ATGACAGGTG ATTATACTTT GCGAGTAACG AATGTACCGC AGgAATATTC CGTGGATGAA 11160

GAGTATTTGA CAGGAAAAGC CATTAAGCTG GTCAAAGGAG ACAACCAACT AAAAATTCCA 11220

TTAACGAAAA CAATTGATcA CAGTCGTTTA CAAGTCAAAG ATTCAACGAT TTATGTCGGC 11280

GATTCATGGA AACCAGAAGA GAACTTTGTT TCAGCAACAG ATAAAACAGG TCAAGACGTT 11340

CCcTTCGAAA AAATCACTGT TTCAGGTCAA GTTGATAaCa sCAAAGCAGG CGTTTATCCA 11400

ATTATTTACA GTGACGAAGG TAAaGAAGAA ACAGCcTAtG TGACCGTCAA ACCCGACCAA 11460

TCTAAGTTAG AGGTCAAAGA TACAACGATT TATGTTGGTG ATTCGTGGAA ACCAGAAGAT 11520

AATTTCGTTT CAGCGACAGA CAAAACAGGT CAAGACGTyC CGTTTGAAAA AATTGATGTT 11580

CAGGGAACAG TGAATGTTGA TAAAATAGGC GATTATGAAA TTGTCTATAA AAATGGCAmA 11640

AAAGAAGCGA AAGCAATCGT TCATGTCCGT GATGACAGTC AGTTAGAGGT TAAAGATACA 11700

ACGATTTATG TTGGTGATTC GTGGAAACCA GAAGATAATT TCGTTTCAGC AACAGACAAA 11760

ACAGGCCAAG ACGTTCCGTT TGAAAAAATC ACTGTTTCAG GTCAAGTTGA TACTAGCAAA 11820

GCAGGCGTTT ATCCAATCGT TTACAGTTAC GAAGGTAAAG AAGAAACAGC TAATGTGACT 11880

GTCAAACCCG ACCAATCTAA GTTAGAGGTT AAAGATACAA CGATTTATGT GGGCGATAAA 11940

TGGGAACCAG AAGATAATTT CGTTTCAGCA ACAGACAAAA CAGGTCAAGA TGTCCCGTTT 12000

GAAAAAATTG ACGTTCAGGG AACAGTGAAT GTTGATAAAA TAGGCGATTA TGAAATTGTC 12060

TATAAAAATG GCACAAAAGA AGCGAAAGCA ATCGTTCATG TCCGTGATGA CAGTCAGTTA 12120

GAGGTCAAAG ATACAACAAT TTATGTGGGT GATAAATGGG AAGCAGAAGA TAACTTCGTT 12180

TCCGCGACAG ACAAAACAGG TCAAGACGTT CCGTTTGAAA AAATTGATGT TCAGGGAACA 12240

GTGAATGTTG ATAAAATAGG CGATTATGAA ATTGTCTATA AAAATGGCAC AAAAGAAGCG 12300

AAAGCAATCG TTCATGTCCG TGATGATAGT CGTTTACAAG TCAAGGATAC AACGATTTAT 12360

GTCGGCGATT CrTGGArACC AGAAGrGAAC TTTGTTTCAG CrACAGATAA AACAGGTCAA 12420

GATGTCCCAT TCGAAAAAAT CACTG 12445

(2) INFORMATION FOR SEQ ID NO: 243:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2546 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

```
GAACACAAGA AAATGTTTTT TGAATAAACT CTTCTTGTTT TGAGGTTGGA TAAAGTCTGT     60

ATTTATAAGC TTTCAGAGCC TTCATCTCAT AACCACCCTA CCAAGATAGT TAGAAAAGAA    120

GCTGGTTATA AACCAGCTTT TCTTTTCTTC TCTTCAATTT TTTCACGGTA CATTGCAACT    180

TCGTCTTCAG CACAATAAAC ATAGTGTCCA TTGAGAATTT CACGGAGCTC ACGATCTTTC    240

CCATCTGCCG CTGCAGGCTG GTAAACGATC CGTTTACGTG TGCGTTCATA GTCTGGATCT    300

GGTAATGGAA TCGCAGAAAG CAAACTTTCT GTATAAGGAT GAACGCCGAA ATTATAGACA    360

TCATCACTTG TGCCCATTTC CAATAATTTC CCACTGTGCA TTACGCCAAT TCGGTCACTG    420

ATATGTTTTA CCATCGATAA GTCATGTGCG ATAAACAAGT AAGTTAAGCC TTGTTCTTTT    480

TGTAAATCTT GTAATAAGTT TACTACTTGG GCTTGAATCG ACACATCCAA AGCCGAAATT    540

GGCTCATCAC AGATAATGAA ACGTGGATTA ACCGCTAAAG CACGAGCAAT CCCGATCCGT    600

TGACGTTGCC CACCAGAAAA CTCATGAGGA TAACGCGTAC TGTGACTTGG ATTTAATCCC    660

ACCGTTTTTA AAAGTTCATT GACTTTTTCT TCCCGCTCTT TTGGTGTTTT AGCTAATCCA    720

TTGACATCGA TTCCTTCAGC AATAATATCA TTGACTTTCA TTCTTGGATT CAATGAAGCA    780

TATGGATCTT GGAAGATCAT TTGAACATCG CGACGGAATG CTTGCATGTC CGCTTTTGAT    840

TTAATTTTGC TAATGTCTTT TCCATCGAAA AGAATTTCAC CAGCAGTTGG ATCATATAAA    900

CGAATGACAC TGCGACCAGT TGTTGATTTT CCACTTCCTG ATTCGCCAAC TAATCCAAAG    960

GTTTCGCCTT CATAAATATG AAAACTAATG TCATCAATGG CTTTTACTTC GTCTTTACGA   1020

CCAACATTGA AATACTGTTT CAAACCAGTG ACTTCTAATA GGACTTTTTT CTCTTCTGCC   1080

ATTAATTACT ACCTCCTGGA TTTTTTTGAC GCTCTGCATA AATTGCCCAA CGACGAACAA   1140

TCTCTGCTGG AGGCGTAACT TTTGGTGCTT GTGGTGCTAA TAACCACGTT GCTGCTTTGT   1200

GGGTTGGTGA TACTTCAAAA TAAGGTGGTT CTTTTTCTGT ATCAATTTTT AAAGCAAACT   1260

CATTCCGTGG ATAGAAGGCA TCTCCAGTTG GTGGATCTAA TAAATCTGGA GGTGAACCTG   1320

GAATCGCATA AAGTTTGTCT TCTGTTCCTT CCATCGTTGG CATTGACCCT AAAAGTCCCC   1380

AAGTATATGG ATGTTGTGGA TTATAGAAAA TTTCTTCTGC TGTTCCCACT TCAACCAGAC   1440

GTCCACCGTA CATGACAGCG ACACGGTCAG CCACGTTTGC TACTACCCCT AAATCATGGG   1500

TAATGAAAAT AATCGAAGTG CTAATTTTTT CTTGTAAATC TTTTAATAAC TCTAAAATTT   1560

GTGCTTGGAT GGTTACATCC AACGCTGTAG TTGGTTCATC TGCAATTAGA ACTTCGGGAT   1620

AACAAATTAG AGCAATTGCA ATGACAATCC GTTGACGTTG CCCACCAGAA AATTGGTGTG   1680

GATAGTTTTT CAAACGTTTT TCAGCATTAG GAATGCCGAC TAATTTTAGT AATTCTAATG   1740
```

ctTGCGCTAA GCCTTCTTTT TTTGAGACTT TATTATGTTT GATTAATGAT TCCGCCACTT 1800

GTTTGCCAAT TGGCATCGTT GGGTCTAATG AAGTCATCGG ATCTTGGAAA ATCATTGCGA 1860

TTTCTTTACC ACGAATGGCT TGCATTTGTT TTTCTGTTTT ATGAACAATA TCTTGTCCTT 1920

TAAACAAAAT TTCCCCGTTA TCAATATTGG CATTACTGCT CAACAAACGC ATAATACTGC 1980

GGGTTGTAAC AGATTTACCA CTCCCTGATT CACCTACGAT GGCTAACGTT TCACCTTTAT 2040

AAAGATCAAA ACTCACACCA CGAATGGCGT TGACTTTACC TGCAAATGTA TCAAAGGAGA 2100

TTCTTAAGTC TTTAACTTCT AATACTTTTT CCATAGTCAT TCACCACTCT TACTCTTTCA 2160

TCTTAGGATC AAATGCGTCA CGTAAACCAT CAGCTAACAA GTTGAAACAA ATCATGATAA 2220

CTGAAAGTGT TGCCGCTGGA ATCCACATTA AATATGGAAG GAAACGGAAA GTTTTGTATC 2280

CTTCGTTTAA CAAGGTTCCT AAAGATGCGG TTGGTGGACG TAAGCCTAAC CCAATAAAAC 2340

TTAGGAACGC TTCAAAGAAA ATTGCCGATG GGATACTGAA CATCATTTGG ACAATAATGA 2400

CACTTGAGAT ATTCGGTAAA ATATGTTTAA TAGCmATTTT CCAACGAGAT TCGCCCAnTG 2460

TTTCTGCTGC CATGACAAAC TCTTGGTCTT TTAGTTTCAt GGTCkGGGcT CGCACAATCc 2520

TGGCCATkGA tATCCAGTTT GTAATA 2546

(2) INFORMATION FOR SEQ ID NO: 244:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8102 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

AACTTACCAA GAATCGCTTT AGAAGCGCAA GGAAGTCAAG AAAAGTTCTG GCAAATTTTA 60

AATGAACGTT TAGCGATCGT TAAAGATGCT TTGGTTTATC GGGTTGAACG TGTTAAAGAA 120

GCTAAACCAG CGAATGCACC GATTTTATAC ATGTATGGTG CATTTGGAAA ACGTTTAGCT 180

ACACAGGATG CCGTAGATGA ACTGTTTAAA AACAAACGTG CCACGGTTTC ACTTGGCTAT 240

ATTGGCTTGT ATGAAGTGGC TTCTGCCTTT TACGGCGGTG CTTGGGAGGA CAATCAAGAG 300

GCCAAAAACT TTACACTGGA TATTTTAAAA GAACTGAAAA AAAATGCCGA CAACTGGGGC 360

AACGAATACG GCTATCATTT TAGTGTCTAC TCTACACCGA GCGAAAGCTT GACTGACCGT 420

TTTTGCCGTT TAGATACTGA AAAATTTGGC ATCGTCGAAA ATATTACAGA TAAAGAATAT 480

TACACAAATA GTTTCCATTA TGATGTACGC AAAAATCCAA CGCCTTTTGA AAAACTTGAT 540

TTTGAAAAGG ATTATCCAAA ATACTGTTCT GGTGGTTTTA TCCACTACTG TGAATACCCC 600

ATGTTACAGC AAAATCCCAA AGCACTTGAA GCTGTCTGGG ATTACGCTTA TGACaAAGTT 660

GGCTATTTAG GAACCAATAC GCCAATTGAT CATTGTTATG CCTGTGGTTT TGAAGGGGAC 720

TTCCATCCAA CTGAACGTGG CTTTGAATGT CCAGAATGTG GCAATCATGA TCCTAAGACT 780

TGTGATGTCG TTAAACGGAC TTGTGGTTAT TTAGGTAACC CGCAAGCTCG CCCAATGGTA 840

CATGGTCGTC ACAAAGAAAT TTCCTCTCGT GTGAAACATT TAAAAAATTA AACTCGATTA 900

TTCTATAGAA AGAAGTTTTG CTGATGCGTA ATCCGAAACC ACAAGAATGG AAAACCGAAG 960

ACTACAGCCA AAAAAAGATT GCCGATTACA AAGCGTTCAA CTTTGTTGAT GGCGAAGGTG 1020

TACGCAATAG TTTATATGTG AGTGGCTGTC TCTTTGCTTG TGAAGGTTGT TTTAACAAAG 1080

CTGTCCAAAA TTTCAATTAC GGTACGCCGT TTACAGAATC TTTAATGAAT CAAATTATTG 1140

AAGATTTGTC CCATGACTAT GTCCAAGGTT TGACTCTTTT GGGTGGCGAA CCTTTTCTGA 1200

ATACAGATGT TTGTCTTAGT GTGGTCAAAC GTGTGCGGGA AACATTTGGC TCAGCCAAAG 1260

ATATCTGGTC TTGGTCTGGG TATACCTTTG AAGAACTACT TTTAGAAACG CCTGATAAGC 1320

TTGAACTACT CCATTCAATT GACATTTTAG TTGATGGGCG CTTTGAACTG GCAAAACGTA 1380

ATTTAAATTT ACAATTTCGT GGCAGTAGTA ATCAACGGAT TATTGATGTA CCCAAATCAC 1440

TTGCTGCTGG CAAAGCTGTT ATTTGGGAAA AATGTCATGA CGCCGAAACG TCTTATGAAC 1500

AAATCAAAAA ATCAATTTAA AAAAGTTAGC CCTTTAGCCA TCTCTTGAGG ATGTGTCTAA 1560

AGGGCTAACT TTTTTCCCAT AAAGGTAAAA CAATGTTTTC ATACGCTAGT GGATCTAAAT 1620

TCGTTAAAGT AATGCCTAAT AAACGGATTC CTTGTTCTAA ACCTAAAATA TCTTCCCAAA 1680

TTAAATTCGC TTGATAGAAA AGTTCTTCTT TTTTAGAAAT ATATTCAGGT AATGTCACAC 1740

GTTTCGTAAT CGTCGTATAG TCAGTATAAC GAACTTTCAG GACTACTGTT TTGCCATGCT 1800

TTTGCACTCG TTGCAATGCA CGCTCAACCC CTGCTGCTAA TTGTCTCAAT TGAGCCGTAA 1860

CCTGCTCTTC CGTCGTTAAA GGGGTGCCGT AAGTATGTTC CTTGCCTACC GACTTACGAT 1920

CTCTAGTCAC TTGAACTGGC GCATCATGGA TACCGCGGAC TTTACGATAC AATGAATAAC 1980

CCATTTTACC GAATAACCGA ATCAACGTCA TCTCATCTTG CTTGTATAAA TCTGCTCCAG 2040

TAAAAATCCC TAGTTCATGC ATTTTAGGCA CTGTTTTTTT CCCGATTCCG TGGAAATCTT 2100

CAATCGGTAA TGCTTTCAAA AAATCTTCGG CCTCTTCTGG TAAGACAACC GTCAGGCCCT 2160

TTGGTTTCTG AAAATCTGAT GCTAACTTGG CTAAAAACTT ATTGTAACTA ACACCTGCCG 2220

AGCAAGTTAA CTGCAATTCA TGCCAAATAT CATATTGGAT CATTTTCGCA ATTTTAATAG 2280

CCGACTTACT GTTGACCTTA TTTTCAGTTA CATCTAAATA CGCTTCGTCA ATAGATACCG 2340

GTTCAATCAC ATCTGTATAG CGGCGAAAAA TTTCACGGAT TTCTTGAGAA ATTTCACTAT 2400

ATTTCTGATA ATTACCTGGC TTAAAAATGG CACTTGGACA TAGTTCATAC GCCTTTTGAG 2460

CACTCATAGC CGAGTGAATA CCGTACTGTC GAGCAATATA ATTCGCTGTG GTCACCACTC 2520

CTTTGCCCCC AGTATCACTT GGATGGCGCG CAATAACTAA CGGATGTCCT ACTAACTCTG 2580

GATGATCTCG TTCTTCAACT GAAGCAAAAA ACGCGTCCAT ATCTATATGT ATAATTTTTC 2640

TCGACGTATC TTTTTTTAAG GGAAAACGTA ATTCGCTCAT CATTTTGATT CGCCTCCTCC 2700

CTCCATTATA TACGAACGTC CGTTCTTTTT CAAATTGAAA CTTCTTTATT TCAACCAGAA 2760

AAAGCTGAAA ATCCGTGAAG TTATAAACAC AGATTTCCAG CTTTTTAATT AAACGACCAA 2820

TTGTTTTTGT TGATTATAGT AGCCTTTATA GGCAAGATAA GTTGCAATCA CTGAACCAAT 2880

ACTTGTTGCT GATAGCAACA TAAAAGTCAC CATAATCTGA TATTTTATCG CATGGACAGG 2940

ATCAACGCCC GCAAAAATTA AGCCTGACAT CATCCCAGGT AAACTAACAA TTCCGACAGT 3000

TTTGGCAGAA TCAATCGTCG GCGCCATCCC GGTTTTCACA CTTTGTTGGA TGATTGGTAA 3060

AGAGGCGAGT TTGATAGTTG CACCTAATGC TAGTTTCTCC ATTACCGCTT GCCGTTGGTT 3120

TTGAAAAAGA GTATCTAAAT TGCGATAGCA TAAACCAATT GCCACCATCG AATTGCTGGC 3180

AATCATCCCA CTGATAGGAA TGATTTGAGA AGGAATAAAT TTAATAGAAC CAGAAAAAAT 3240

CAAAATACCA ATGGTAATTC CTGTACTACA ACCAATTGCA ATGAAAGAAA TCAAAAAGCC 3300

CTTTTGTAAT TGCTGGCTGC GTTTTTTTGC ATTAAAGGCT GCGTTGGAAA CAATCACCAA 3360

AACCATGATG AGCGTTAAAA TAAGGTTATT GACTTGAAAA ACATATTTTA ATAAATAGCC 3420

GACTGCGACT AATTGAATGA CTGCTCGAAT CACACTAATA ATAATATCTT TGGTAAAACC 3480

TAATTTTTCT TTAGTACTAA TAATTAATGC GACCACCACT AGCATGGCTG AAAAAAATAA 3540

GGATAAATTA TTAACTGCTA AATCCATCTG TCGGCGCCAC CTTTCCTGCT ACAATACGAA 3600

TCACTTGCTG TGCTTGCTGA ATTTCTTCGG TATCATGCGT GACACGAACC AGCGTGACTC 3660

CTTGCTCTTT GTTTAATTGG TTTAACAATT GATTGACAAT CTGTTTGCTT TCTTCATCTA 3720

ATCCCGTTGT AACTTCGTCT AATAATAAAA CATCTGGTAC AAAAATAATG TTTCGTAGCA 3780

AAGCAACCCG TTGTCGCTCA CCACCAGAGA GTTCGGCTAT TTTCTTTTCA AGATAAGCAG 3840

CGGGCAATTT CACTTGTTGG AGTAATGCCA CGACTTTTTC CTGATTAAAC GCTTCTTGTC 3900

TGACGGTAAA TGGAAATAAC AAATTATCAT ACACGGTTTC ACCAAATAAA GTTGGCTGCT 3960

GAAAACAATA AGAAACCTTT TGGCGGTATG TTTCGATTGG CATTGTAGTA ATTGGCGCTT 4020

CTTGATAAAA AATTTCGCCT TCTGTAGGAC TTAATAATGA AGATAAAATT TTTAAAAATG 4080

TACTTTTCCC GCTACCTGAA GGACCAACAA TTGTTGTATT GCTGCCTGTT TCAAATGACA 4140

AATCAATATC AGAAAGAATG CTACGACCAT TTACTTGATA AGACACGCCT TCTGCTCGCA 4200

ATAAATTACT CATGTTTCTC TCCTTTTTTA AATGGAATGA TATTCTTTAT ATACTTCTTG 4260

TTTTTTTACA TTTCTCAATT TAGCCACTTC TTTGATTGCC TCTTTTGAAC TCCGTCCCTC 4320

TTCCTCCATT AAAACTTGCA CATGTTCCTT GAGGGAAATA GCTGGCATTG CTGCTATTGT 4380

CTCTTTCTCA CCAGTAAACC CGCTAACTAA TAAACAACAC TCGCCTTTTA GCGTATTTTC 4440

TGCTAAATAT TCTGTTAGTT CTCCTAAGGT ACCACGCAAA TACTCTTCAT GAAGTTTCGT 4500

TAATTCCCGA CAAATCACAG CGGGGCGTTC TTGACCATAG ACTTCAGCAA ACGTTGCTAC 4560

CGTTTTAGCA ATCCGATAGG GGGACTCATA AAAAATCTGT GTCGGCCGTT CTTCTTTTAA 4620

TGCTGATAAC ACGTCTTTCT GTTCTTTTTT CTTTCTTGGT AAAAAACCAT AAAACGTAAA 4680

TGGCTGTGGT AAAAGACCCG AGGCAATCAA AGCGGTCATT CCTGCCGTAG gTCCTGGCAA 4740

GGCAATCACT GCCAGTTCTT CTTCAAGACA TGCTGTAACC AATTCATGGC CAGGATCACT 4800

AATTGAAGGC ATACCTGCAT CGCTAACTTG GGCAATCGTT TCACCATTCA GCAATCGTGT 4860

AATCAGTTGA GGAATTCGCT CTTTGTAATT ATGTTCGTGT AAACTAATTT GTGGTGTTGT 4920

AATCTCAAAA TGATTTAGCA ATTTTTGTGT GTTTCTTGTG TCCTCACTAG CAATCACCGT 4980

TGCTTCTTTC AATATGTTTA AACAACGAAT ACTCATATCT TCTAAATTGC CGATCGGTGT 5040

TGGGACAAGA TACAATTTTC CTTTTGTTGT TTGCTTATCA AAACTTTTCT GTTTTTGCAT 5100

AGCGCATCCC TCATTTTCCT TTTTCGTTCT ATAAATAGAT AAAGGTTTGA TTTGACAATC 5160

GATTGACTGT CAAACCAAAC CTTTCCTTTT TATTTGCAGT GTCCACTAGA CTTAACGCGT 5220

ACGTTCCCCA TAAATAACAT CAAGACAAAA GGCACATTCT TCATCATTTT CACGTCTTGA 5280

ACCATATAAA ATATTGCAGA CATGGAAGCC CTCTTCATAA AGTTTTTCCA AATTCATACG 5340

TGATTTTGAT AGCTCTTGTT TTTCAGTTTC AGTTGTATTT CCTGCTAACT TATTCAGTTC 5400

TTGGAGATGC TCTCGTAAGC GTTGGTTTTC GATTTCCAAC GTGGTATTTT TTTCTACCAA 5460

CTCATGGAGT GCTGCTTTAA TTTCTCTTAA TTGTGTCACT GAACTGTCTA AATCAGTTTC 5520

CAACGAATTT AATCCATCAT AGAGCGAACG CTTATCCATT TTGATCACCT TTTTCAGCAG 5580

CAACTTTCGC AGTAGCTTCT TTGATTTCTT CGTAATCATA ATCCGCTGCG ATTTCACGAC 5640

CATGCAGACG AACTTTGACA ATCCGACTTA ATAGATTCAA CCCGACGACT TTGCCTTTGC 5700

CATCTGGTGT TATTACTTCT TTTCCATAAT CTGGTAATTC TTTTTTAGCG GCTTCATATT 5760

CGTCGTTTTC GTATTTCAAA CAACACATTA AGCGGCCACA TAATCCCGAA ATTTTGACAG 5820

GATTTAAAGA TAACCCTTGA TCTTTAGCCA TTTTGATTGA AACTGGCATA AAATCTCCTA 5880

GAAATGTTGA GCAACATAGT TGTCTGCCAC AAGGGCCAAT GCCACCTAAT ATTTTCGCTT 5940

CATCTCGGAC ACCAATTTGa CGTAACTCAA TTCGCGTCCG GAAAATAGCC GCTAAGTCTT 6000

TGACTAATTc ACGAAAATCA ATTCGCCCAT CTGCCGTAAA GTAAAAAATC ATTTTGCTAC 6060

GATCGAAGGT ATATTCTACT CGCACTAATT TCATTTTTAA GTCATGAGCT CGAATTTTTT 6120

CATTGGCAAT GCTTTTGGCA GCTTCTGCAT CAGCCAAATT TTTTTGTTCT TTTTCTAAAT 6180

CATTGGCTGT TGCTTTATTT AAAATGGGTT TTAGGTCCTC TGGTAAATCG TCTGAATCGA 6240

CTGTTTTTTT AGGAATAGCA ACAGTAGCTA ATTGTTTTGA CTGTTGAGAT TCAACGAGTA 6300

CTTTCTCATT ATAAATATAC TCAGATTTTC CAGGAGCAAA ATAATAGATA TGACCGGCTT 6360

CACGGAAGCG AACTCCTACT ACTTCTACCA TTTTATTCCT CCTAATCTAG TTCAAGTGAA 6420

CGTCGCTTCA ATCGGTTAAG GAACTTGCTG AAACAACTAA GTTCCTACTA TATTATGAAA 6480

CTGAATGCCA CTTGGCACTT TTTTCCTTTA TGATTTAGGG TGAATCATTT GGATAACTAA 6540

TTGTTCACAA ACATTTTGCC AACTAACATT GGCAGTCCAT TTTTGGCGTG CTTTCAAAAT 6600

TAGCGCCAAC CGTTCCGCCT GCTCTTCCGT TACTTTTTTG GCTTGCTGTG TCGCAACACT 6660

TTCTTCCAAT AATTGACGGT AATAAACCAT GAGCAAGTCA AAGCTAAGCG CTTGTTGTTC 6720

TTTTTCCTTA AATACTTTGA CCATTTTCTT CTGAACGTAG ATAAATGCCT GTAAATCATT 6780

ACTTTTTAGA TAATTAAACC ATTGCAAAAT GATTTCCCTA GCTTCATTAA ACCATTCATC 6840

TTGAGAGATT TCaACTGCTT TCTCAAAACT ATTTGTCAGT TCAGCTAAAA GGGTTGCAGT 6900

CTTTTCACCA ATCCCCTGTT TGATTAAGCG ATCAATTAAT GTTTTTTTGA CTAATGGTTG 6960

AAAATGTAAG GTTTGGCATC GTGATTGAAT CGTTGGTAAA ATTCGAGAAA GCGAAGTGGT 7020

TTCTAAAATA GCTAAAATTT GTCCTTCTGG TTCTTCTAAA AATTTTAAGA GACTATTaGc 7080

tGCGCCGGTA CTCATTTTAT CTGCTTCTTG AATTAAGAAA ACTTTTTTAG CAGTCTCGAC 7140

CCCACTTTTA GAAAACTCCG CTTTTAATTC ACGGATTTGG TTCACTTTGA TGGTTTGCCC 7200

ATCTGGCGCA ATTCTTAAAA CATCTGGATG TTCATTTTCA TTAATCCGCA CACAATTATG 7260

GCATTCGTTA CAAGGCTGTT GATTTACTAA ATTCGTACAA AAGACATGTT TCGCCATCCA 7320

TAAGCCAAAT TCTTGTTTTC CAGTTCCTGT ATCTCCTTCA AAAAGATAAG CATGGGCAAG 7380

ACGACCATGC TCAAAACTTT TTTGGAGTTG CTTGTACAGC AAAGGTTGCA TTTGCTGTAG 7440

CTGTTGTGCT TCATTCATCT TAATATTGAT GGAATCCTTC AACTGGTAAG ACGAAGCAAG 7500

TAGCGCCGCC TACTTCAACT TCCACAGGAT AAGGAATTTG GCCATCCATT GTGATATCTA 7560

AAGTCACAGG TGTTGAAACA TATTGTTTTC TTGATTGACA TGTTTCTTTA ATTAAAGCTA 7620

ATGTTTCGTC GACACGTTCA TCATCAATCC CAATAATAAA TGTGCTGTTT CCCGCTTTTA 7680

AGAACCCACC TGTTGAGGAT AATTTTGTAG CACGAATATT GGCATCAATA AATTCGTTGG 7740

CTAATCGGTT ACTATCTTtG TCtTGTACAA TGGCTAAAAT AATCTtCATG GTCTACACCT 7800

TCCTATAATT AAAAGTTTTC TGGATAACGT TCAATAATCG CCTGATACGT TGCTTCTACG 7860

ACAAGTTCTA AACTCATCCG TGCATCAATC TTTTTGATAC GTTGTGGATT TTCTTCTGcT 7920

AATTTCAAAT ATTCATGACG AACmCGTTGA TGAAATTCTA ATCCTTCTGA ATCTAAGCGA 7980

TCAATTTsCT GTGTTCGATT TTCTTGAATC CGCCGTAAGC CAGTATCTGA GTCAACGTCT 8040

AAATAAATTG TAAAATCTGG TGAGACACCT TCTGTCGCAA ACGCATTAAT TGAAGCAATG 8100

Gn 8102

(2) INFORMATION FOR SEQ ID NO: 245:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1670 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

AATATTTATT TTATAGGAGG AGAAAGTTAT GGGGCATCAA TTGAGCACTC ATTTTTTTAA 60

CAAGAAAACA GGGATTATTT CAACTTTTTT AAGTTTGATA CTTTTAGTCG GTGGGTTACT 120

GTTTGCACTA CCTGCATTTG AAGTACAAGC AGCATCAACT GTTACACCTA AAACAGTCAT 180

GTATGTAGAA GTAAATAACC ACGATTTTAA CAATGTGGGa AAATATACTT TAGCTGGAAC 240

CAATCAACCT GCTTTTGATA TGGGGATTAT TTTTGCAGCC AATATTAATT ATGATACGGT 300

CAATAAAAAA CCTTACtTAT ATCtAAATGA ACGGGTCCAA CAAAcTTTAA ACGAAGCTGA 360

AACACAAATT CGTCCaGTTC AAGCACGCGG CACTAAaGTC CTATTGTCTA TTTTAGGCAA 420

CCATGAAGGC GCTGGCTTTG CTAACTTCCC AACTTACGAA AGTGCGGATG CCTTTGCAGC 480

TCAATTAGAA CAAGTCGTTA ATACGTATCA TTTAGATGGT ATTGATTTTG ATGATGAGTA 540

TGCTGAATAC GGAAAAAATG GGACGCCTCA ACCAAATAAT TCTTCGTTCA TTTGGTTACT 600

TCAAGCCTTA AGAAATCGCT TAGGTAATGA TAAATTAATT ACTTTTTATA ATATTGGACC 660

AGCGGCTGCG AACTCTTCAG CTAATCCTCA AATGArTTCT TTAATTGATT ATGCTTGGAA 720

CCCTTATTAT AGTACCTGGA ATCCGCCACA AATTGCGGGG ATGCCTGCTT CTCGTTTAAG 780

TGCTTCTGCC GTTGAAATTG GTGTCAATCA AAATCTAgCT GCGCAATATG CTAAACGTwC 840

AAAAGCGGAC AATATGGTAT TTATCTAATG TATAACCTCC CAGGAAAAGA TTCATCTGCG 900

TATATCTCGG CGGCGACACA AGAACTTTAT GGTCGAAAAA CGAACTATTC ACCAACTGTT 960

CCAACACCTT AATATTATCT CGTGTCTACA AGCGCCAACA TGTAAAGAGT TAAAAGCCAG 1020

CTTACTACAA GCTGGCTTTT TGGTATCTTG AGCGAATAGT TTTACACAAA AAAAGACCCG 1080

CCCTTTGGGT AAGGACGGGA AAGGAGTTAA AAATGAAAAA GTGTTTTGTG TTGTTAAGGT 1140

TGTTTGTTGA TATGCTTATA TCTTACAAAC TAAATGTGAA GATTTTGTGA TCATTTCATT 1200

CTTATTTTGC TACTTTTATA GGAATTAAAT AGATTGCCAC GAGACTATCT TTAGTTGCCT 1260

CGTGCATCCG TTATTCTGAC TGCTCCTTGT GGTGATACTC TGAAAAAACC AACAGTCCCA 1320

GTCCCTCCAT TTGCAGCCAT CTGTTTCGAT TGGGCTCTAA AAAAGAAGTC ATTTCCATTC 1380

ATACCCAAAA ATCCAAGGTT CGTATCATTA TTTTCTGCAT CAGTCAGTGT CTTTCTAATA 1440

TATGTCTCTG CTTTTTCTTG TGTATCGATG ATTTTTGCCT CTTCATTTTG GTCTTGATTT 1500

TCCTCTTCAG AAGAACGATA GTAATACTCT TCCGCaGGAT AATTTCCACT ATCTTGAGTA 1560

ATTAACAATC GCGGTTTCGT CTTATCCGAT TGGTCGCCGT TAGGATTTTC AACACCAATT 1620

TTATATAATA ATAGCGCAGC ACCAGTAGGC CCCCACCTAA TACTTAAGCT 1670

(2) INFORMATION FOR SEQ ID NO: 246:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1453 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

TACCGAgGTG TATCAgCGAT TCtaGTAAtC CACAaGTTCC ACTACAGAAT CAGAAtCTTA 60

AAAGGTCTGG CGGAAGAAAA AACGCTAAAA GAAAAAATCC CGTCCTTAAA TCAAACAACG 120

ATTAAGAAAG TCACTGATAA TGAGTATAAA CGCATTCTTG TGaTGGAAGA AAAGGCTACA 180

GGAACGACCT ATAAAAGTAT TTTGATAAAA AAACAAAATC GCTTAAAAAT TGTCAACGAA 240

ACAACCGATG AATTGTTATA CAATGGAACC CTATAAGAAC GAAACCATTC AATAAACTCA 300

AAAGTATAGC CATTGTTTTC AAAGAAAAAT TTCTTTGGAA ACAATGGTTT TTTTATTGAC 360

AAAAGAATTT CTAGGATATA TAGTTAGTTA CATAAGTAGT TAACCGTTTA AGCGAGGTGT 420

AAAAATGGAA TTTAATTTTT CGGGGGAAAA GCCTTTATTT CAGCAAGTAG CCGATCAGAT 480

TGCCGAAGaA TTTTTAATGG CGCTTACTTA GAAGGCGAAC AAATTCCCTC AACCACAGAA 540

ATATCAAAAA GTTATCAAAT CAATCCAGCC ACCGTTTTAA AAGGAATGAA TTTATTAGTA 600

GAACGACAGT TGATTGAAAA GAAACGGGGG ATTGGCATGT TTGTTTTACC AGGCGCTCAA 660

GAGAGAGTAA GAAGTGCACG GAAAGAAGAA TTTTTAAATA AAGAAGTCTT AGAAGTTGTC 720

GCAGAAGCTA AAAAATTAGG GATTACAGCT GAACAATTAA AACAACTAAT TGAAAGAGGG 780

TATGACGCAT GAGTTTGCGG GTGGAATCAG TAAGTAAAAA ATATCATCAA AAACAAGCAC 840

TAGATAATAT TTCaATTACT TTTGAAAAAG AAACGATTTA TGGGCTCTTA GGCAGAAACG 900

GGGCTGGTAA AAGTACGTTA TTAAATATCA TAAATAATCG AAGTTTTGCC ACTTCAGGTT 960

CCGTTAAATT GGCGGGAGAA ACAGTGACCG ATAATGAGGC TGCTTTGACT CATATTTACT 1020

TAATGAGTGA AGATAATTTG TTTCCTCCTC AGTTAAAAAT TAAAGATATT TTTAAAACGA 1080

CTGAAGGTTT TTATGGATCC TTTGACTGGT CTTTAGCCGA ACAAATGTTA AGCGATTTTG 1140

ATTTAGATGG TAAAAAGACA TTCAAAAAAT TATCAACTGG TTATCGGAGT ATTGCTAAAC 1200

TAATAGTGGC GCTATCCGTC CCCTGTGAGT ATATTTTTTT AGATGAGCCA GTCTTAGGTT 1260

TAGACGCCAA TCATCGGGAA TTATTTTATA CTTATTTAAT TGAGACATAC CAAGAACGTC 1320

CCCGTACGTT TGTGgATTTC TACCCATTTG gATtGGAAGA AATtGCGAAT TtATTAGAAG 1380

ATATTATCay CATTGAkCmA gGGAAAATTG ATCCGCGCAG AATCTATTGA AACCAATTTG 1440

AAAAACCGGT CGC 1453

(2) INFORMATION FOR SEQ ID NO: 247:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1820 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

TAGGAGTGAC AACAAATGAC AATTGATTGG ACAAAAGAAG TCGAAGCACG TAAAGACGAC 60

TTATTAGAAG ATTTACAAAA CTTATTACGG ATCAACAGTG AACGTGATGA TGCACAAGCA 120

ACACCTGAAG CGCCCTTTGG ACCTGGGCCA GTTGCTGGCT TAAAACACAT GTTAGCGTAT 180

GGTGAACGTG ATGGCTTTAC GGTGAAAAAT GTCGACAATT ACGCTGGACA CATTGAATAT 240

GGTGAAGGCG ACGAAACGTT AGGAATTTTT GGTCATATGG ACGTTGTTCC TGCTGGTGAT 300

GGTTGGGAAA CAGATCCTTA TGAACCAGTC ATTAAAGATG GCAAAATTTA TGCACGTGGT 360

GCAAGCGATG ATAAAGGCCC AAGTATGGCT GCTTATTACG CAATGAAAAT TATCAAAGAA 420

TTAGGCTTAC CCGTGTCTAA AAAAATTCGT TTCGTAGTAG GAAGTGACGA AGAAAGCGGC 480

TGGGGCGATA TGGATTACTA TTTCCAACAT GAAGAAGCGC CAGACTTTGG TTTCTCACCA 540

GATGCTGAGT TCCCAATTAT CAATGGGGAA AAAGGAAATG TGACTATTCG TTTAACTTTC 600

CGAGGCGGTA ACGGCGCTGA TTATAAATTA GAAAGCTTCA AATCTGGCTT ACGTGAAAAC 660

ATGGTTCCGG GAACAGCAGA TGCTGTTGTA ACGGCTGCTT CAGCAGACGA AGCGGCTTCA 720

CTAGCTGCTA GCTTTGAAAC ATTCATCAAA CAAGAAGCAA AAATTTCAGG TAACGCAGAA 780

CTTTCAGATA AAACAGTGAC TTTCCATGTA GTTGGTAAAG GTGCTCACGG TGCTAGCCCA 840

CAATCTGGAA TTAACGCAGC GACTTTCTTG GCAACTTTCT TAAATGATTA CAGCTTTGCA 900

GAAGGCGCAT ATAGCTTTAT TAATACCATT GCTGAATTTA TCCATGAAGA TTTTTATGGT 960

GAAAAATTAG GCGTGGCTTT TGAAGACGAA AAAATGGGTA AATTGACAAT GAACGCTGGG 1020

ATTGTTAACT TTGATCCAGA AAACCCTGAA AATAGCTTAG TAACATTAAA CTTCCGCTAT 1080

CCAAAAGGAA CTTCAGCAGA AGAATTACAA GCAAAAGTTC AAACAACAGT CGGTGAAACG 1140

GTCACTGCAA CACAAGGAGA TCGCAACCAA GAACCGCATT ATGTACCAGT TGATGATCCA 1200

TtGGTGGCAA CGTTACTACA AGTTTACGAA GACCATACAG GCGAAAAAGG CCAAGAACAA 1260

ATCATTGGTG GCGGTACTTA TGGCCGCTTA TTGAAACGTG GTGTTGCCTA TGGCGCAATG 1320

TTCCCTGGCT ACACAGACAC AATGCACCAA GCCAACGAAT TTATGGAATT AGATGACTTA 1380

TTCCGAGCAG CTGCTATTTA TGCAGATGCG ATTTATCGTT TAGCGAAGTA AAGTAAAAAA 1440

GCyGTTTTCG ATTTTCGAAA ACAGCTTTTT TTTGCAATTT TTGCCTTTTA AAAAcCCaAA 1500

ATCAGAACCG TCAATTTACT GAAAATTTCT TGCAAATCAA CGGATCCCAC GATACGATAA 1560

ATAGGAGAAA ACGCAATGAG TGAAACACTT ATTGTTAACT ATTTAATCAA AGGACGTGTG 1620

AAGATGAACG TAGAAGAATT AGTTCnAAAA GCmAAAGATT TAATTGGAGA AGGCmACGTG 1680

GAACATGCGC AACAATTCAT TGAAGAACAT AAAGATGAaT TaGGCGrACA CTATCmTAAA 1740

ATTACmGCGA TGTTAGCTGA TmCAaCGACC GATGGTTtGT TTGaTAAAAT TAAAGgATtG 1800

TTcAATAGTT GAAAAAAATC 1820

(2) INFORMATION FOR SEQ ID NO: 248:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 931 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

GACATTTAAT TGAACAATTA GAACAACACT ATCACGGAAT GATAATAGTA CTTTATACAA 60

AAAATTGTCA ATCATACGCA GCATCCTTAA ATCAAAAAGG CATAAAGACG ATACTTTTTG 120

ATAAGAAAAC TTTGTTGATT ACAACGGCGA TTTCATACTT GAAGTCAGCG CCGTTAATTT 180

TTGTAGATAA TTATTTTCCA GAATTGGCAA TTTTAAAAAA AAAGCAAGTA GTTATCCAAT 240

TATGGCATGC TAATGGGGGC TTTAAAACcA GTTTGGCTGG GGTGCTTATC ACACGAAACA 300

TAGAAGTAGC AAGGATCAAA TCAGGTTTCA AAAaGTTtAC GmTAAAATGA ATTATTTTAT 360

CGTTAATTCA CTAAAAATGG AAGCAATTTT TAGAAACAAT TATCGTCTCG AGAATGCTCA 420

ATTTTGTCAT TTTGGTTCGC CGCGTCTTGC TTATCTTGAA AAATTAGAGG TAGGAGAAAC 480

AAATCAGAAT AATAAAGTTT TATATGCACC AACTTACCGA GAgGGAATGA ATGAAATGAT 540

GCTTGTGATA AATCAAGCAA TAAAGGCTTT TTCTGCTATG CCAAATTTTC ATTTTTATAT 600

GAAATTACAT CCATCAATCC AGTTAGATGC CATTGAACTA CCGAGAAATG TTTCTATTTG 660

GGAAAAGCAT ATTTTTGAAA GCTTTAGTGA AATAGGgCTA CTTGATTACT GATTATTCGA 720

GTGTGGTCTT TGAATACATG CATGTTAAAG AGCAACCGAA CATTTaTTTT TgTCCAGATT 780

aGAAAAtACG CATTAAATCC TGGAATkGAG CCTGATTTTT TTGAATATTT ACCAGGACCG 840

TTAGTTGaAA ATGaAAAACA GTgGATgTGG GCATGtCAAA TTTTcCAtAA ATAGTTAtAA 900

rGAkAATaTT ACtrGAtTaA CGAGGAATnG G 931

(2) INFORMATION FOR SEQ ID NO: 249:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1948 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

CCTATGTAAA TAATGAAAAT CCCATATTTT CCAGAAGGGA TAAATGTGAG TTTTGTTGAA 60

ATTCTTGGCA AAAATCAACT ATTTGTTCGA ACGTATGAAC GTGGTGTAGG TTTTACAAGT 120

GCTTGTGGGA CAGCTATGTG CGCAAGTAGT TTACTTTATA CGTTATTGAA AGATGGTGTT 180

TTTTATGAAG AAATAACCGT CAAAAATACT GGGGGTATGG TGAAAACAGT CGTCCACGAA 240

ACGTCAGATG GCAGTTATTG GATGGAATTA ATTGGGAATG CCACGATTAC GCATTTGATT 300

GAAGGATCAC TGACTGATTT ACTAAATGGT GCATTTGAGA AAATAACTAT CACTGAAACG 360

AATGAACAGA AACATTATCA AGAATTTCTG CAAACACTTT CTCAAAAATA ACAATTGTTC 420

CTTGATAAAG CGCTCTTTTA TTTTACAATT GTGTTAAGTT CTTGTTCCAA AAAATAAAAA 480

ATATTGCTTG AATAAGCAAA CTATGAGAAA ATAAAGGGCA TGAATTAAGA ATTATTCTTA 540

GATAAACGAG GAAGTGTCTT ATGAAACTAT GGAAAAAAGT AGTACTGGGC TTCTTGAGCG 600

TCTTGCTAGT GGCAGTAGCT GGGGTATGTG CTTACGGGAT TAAGATGTAT TCAGATGCCA 660

ATTCAACAAT TAATGGGATC TATCAATCAG TTAACCGTAA GTCCAATAAA GGAGCCACGG 720

CCAATATTGA TGCGCAAGAA CCATTCTCTG TTCTTTTAAT GGGGATTGAC ACAGGCGATT 780

TAGGTCGAAC AGAACAAGGT CGTTCAGATA CTACAATGGT TGTAACCATC AACCCGAAAG 840

AAAACAAGTC CACAATGATC AGTTTGGATA GAGATATTCT AACAGATATT GTTGGAAATG 900

ACACACAAGA CAAATTGAAC CATGCCTATG CATTTGGTGG CGCCGAAATG GCGATTAATA 960

CAGTCCAAGA ATTACTGGAT ATTCCCATCC ATCACTATGT ATCCATCAAT ATGAAAGGGT 1020

TAAAAGACCT GATTGATGCG GTAGGCGGTA TTGAAGTAGA TAATACGATT GGTGAATTTA 1080

CGTTGGACGG TATTACTGTT CCAGCAGGTA AAATCmAGTT AGATGGCACG ACAGGTTTGG 1140

CTTATGCGCG GATGCtCaCG AAGATCCTGA AGGAGATGTG GGTCGTCAAC GCCGCCAACG 1200

AGAAGTTGTT GAGAAAATTG TGCGCAAAGT CATGAGTTTT GACGGTGTTT CAAAATATCG 1260

AAAAATTCTT GATGCTGTTG AAGCCAATGT AAAAACAGAC TTAACTTGGG ATGACATGAT 1320

GGACATCCAA TCAAAATATT TATCTGCCTT TAAAACAATT GACTCTGAAC AATTACAAGG 1380

ATACAGTGCA ACCATTGATG ATATTTATTA TCAAGTCCTA GATCCAAATA GCTTATACAA 1440

AACACAAACA ACACTCAGAA AACAACTTGG TTTAAAAGAA CATGCTTCTG AACGTGAAAA 1500

AGATTTAGCC TTTTATAATC AGTTTAGCTA TGCTGTTACA GATACTGCAC TGATTGGTGA 1560

ATCTGGTGAT TCAAGTGCCA ATAGCTCTAC TAGTTCGAGT AGCACAGCGG ACACCACAGC 1620

CGGACAAACA GAATACAACG GcAATACGCA ACAAGCAGCA GAAGGAACCT ACGATCCGAA 1680

TACGTACGTT GACCCAAATG CGTACACAGA TCCAAACGCT TATGCTGATC CAAATACGTA 1740

TGTGGATCCC AACGCACCAC AATATTAATC TGTTAAACAA GCTCGAAAAC TTCTATTTCT 1800

GATGGGAAAT AGAAGTTTTT TTCTTTTATT CTAGAAATTT TTTTCAGAAG GTGTTATAAT 1860

TGGACTATAC CAnAATAAAC CGTAAGGAGC GTGTCAAAAT GCAAATTATC CGTGTAGCAA 1920

ATGCAGAAGA AGGCGGCAAA AAAGCGTT 1948

(2) INFORMATION FOR SEQ ID NO: 250:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4765 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

TCAAGCCAGC AAGTAAGGAA rGArGTGTrG ATTTCCyGtC CyGaCTTtCC TACTATCGCA 60

TAAAACTTCC CAAGTTCAAA TTTTTGATTT ACTCCTGACA AAACTTTTTC TTTAGAATTT 120

GCATAACTAT ATGCTACATT CTTTATTTCT AATATATCCA TTATTTTCTC CTAACTTATT 180

TTTGATAAGA TTTCTTTAGG CTTCTTGATT AACATTAATG AAGAAGCAAA TACAACTGAT 240

AAAACAATAA TACTTATTAA TATTAAATAA CTTTGTCCAA GCGTTGTTAT ATTTGACATA 300

AAACTACTAC TATTTATTnA ACTTCCaCCG GAAATCATTG rATCCTCTGA GTTAATGAAT 360

CCATCTACAA TTACTTTTAG TAATACATTT CCTAAAAATA AGGAAGATAT TATACTTGGT 420

ATCGAyATGA ATATTAACTC GAATATAAAT TGCATTATAA TTTGTATCTT AGATGTTCCA 480

ATAGATAAAA ATATACCTAT TTCATAAATT CTTTCCCTTA ACCATAGAAT CAAGATTAAT 540

GAAAGAACAA CCATCCCAAC TAACATAATA GAATAAGTCA TTACTTTTAT TATATATTTT 600

ATTCCACTCA CTGACTCTAA AGACTCTTCA AACGCATTAG AATCTTTTTC AACAAAATAC 660

TTTGACTCAT CAATTTTAAG CTCTTTCAAT TTGTTTAAGG CTAAGTCTGT AGATTCTGCA 720

CTACCAGAAT ACATTAAAAT TTTATTTGCA ATTTCATTAT TTTCTGAATT GTTTAATACC 780

TCTTGGCTTG TTGAATAATC TACAAACACC ATATTTTCAC TAAAATCAGA AGATAATCCT 840

GTATATGTTT CCTGTTTTTT ACCAGAAAAG ATCCCTATAA TTTTAAATTT ATGACTTTTT 900

ATTTTTCCAC TTTTTTCAAT ATCTAGTAAT TCAAGATCAA CTTCATCACC TAATTTTAGA 960

TTGTTTTGTT TAGCAAATTC TTCATGAACA ATAATTGAAT TCTTATCATT TTCTTCTATA 1020

TTTTCACCTT CTTTAATTGT AAATACTCCA CTACTAAATA AAATATTTCT TTTAGTATTA 1080

TTTGTAGCTT CGAGTGAAAC AACATTCTTA AATTCGTCAG ATAAATCTTC TCTATTTATT 1140

CTTTGTTCTC CGCTAACTAC TTTAGCATCT TTTAGTTTTG CTAATCCATC ATATTGAATT 1200

ATTTTTTCTT CAATTTCTTT TAATTTTTCA ATATCTTTAA ATTGATTAAC ATTAAAATAT 1260

TTACCATCTT TTTTTGTTAT TGATATTGAA GAATTAGAAC TTTCATATAA AGCCTTTTCT 1320

ATTTCATTGC TTGATTTCAT TATTGTTAAA CAAGAATACA AGCAAGAAAG AACTATAGTT 1380

AATATGATAA AAATTATTAA AGTTCTATTT TTCTTTCTTG TAATATATGC TATAGCATTT 1440

TTTATCACTT TCCTAACTAC TCCTCTCTAT TTTCTTACTC CATAACTAAT TGCATCCACT 1500

GGACAAGCAT TTTTACATCG TCCACATCTG ATACACTCCA GATGATTGCA ATTCTCAACT 1560

GAATCTTTAT TCATTTGGCA AACTTTTTTA CATTTTCCAC AATTAATACA CTTATCCTTA 1620

TCCAATCTAT ATCTAAATAT TGAAATAGGG TTAAAAATTG AGTAAATAGC TCCAAGTGGA 1680

CATATATATT TaCAAAAAGG TCTATATATA ATTATAGAAA GTAACAAAGT AACTATTAGA 1740

ATGATTCCTT TCCAATAATA CAAGAAACCT ATCGTACTTC TCATACTTTT ATTTAATAAA 1800

ACAAGTGGTA ATCCTCCTTC CAGCATACCA ATCGGACAAA TAAGTTTACA AAAATACGGA 1860

GAACCCTGAC CTAAAATATC AACCAAAAAT AAAGGTAATA TTAATGAAAA ATACAAGTAA 1920

TATTATATAT TTCAATTTTC TCAAATGTTT ATCAAACTTA AATGTTTCAA CTTTTTTAAA 1980

AAAGGGAATC TTGTGTAGCA AATCTTGAAT CAATCCAAAT GGACACAGCC ATCCACAAAC 2040

AAATCTTCCC ACCAATGCTC CTATAAAAAA CAAAATCCCA ACTATATAAT AAAATAACTA 2100

ATCTTATATT TTGAATTTCC AATTACTGCC TGTAGTGATC CTATTGGACA AGAACCTTTT 2160

GCTCCCGGGc ATGAATAACA ATTTATACCC GGAACACAAA TTTTCTTTAA ATTCCCCCCG 2220

TAGATTTTAC CAGTCTTAAA CCCTTCaAAG TAAGAATTAG TAATTAAAAA CCAAACGGrT 2280

TGAAATAAAT GCTTAAAACG ATCTCTTATT TTATATTTAT CTTTTTTTGt ACTACCATTT 2340

ATAATATTtT TAATTCTATC CaATTCCAAT ACACTCCATA CAAATTCTTA TCGCCTTATC 2400

CmAGaCmACy TTAGTTTCtC CcATTACTAT CAACAAAGAA AGTCTCTGGC ATAGCCTGTA 2460

TACCATTCAA TCTACCATTA AAGTTTGTCT TATCTGGCAT TAAGAAAGGA TATGAAGCCT 2520

TTGTCTTTTC ATGTATTAAC TTTGACTTTT CAATTGCTTC CTTGTTTTCA CCGTTATCAT 2580

CAACAGGATC TGTAACAACT CCAACGATAT TTACACCCTT GCTTTTCATT TCATTCTGAA 2640

CCTCAACCAA ATCAGGAATT TCCTTTACAC AAGCTGTACA CCAAGTTGCA AATACATTTA 2700

CCATAGTAAG ATCATACTTC TCAAAATCTT TACTTGTAAA ATCCTTGCCA TTTATGTCCT 2760

TTGTTGATAG TTTACGTAAA TCTTTAACCG ATTCTTTGTT AAATGCTTCT GTATTTTCTA 2820

AATCAGTTTT TTCTGATAGC ACAAAACCAT TTTTAGGACG TTCTTTTTTA TCTATAATTT 2880

GAATCTCAGT TCTTTTTAAT TCATCCAGAA GATTACTTTC AGCTCCGCTA TTTGTGCTTA 2940

AATAGCAATC GTATTTTCCA TCACTTGAAA CTCCTATTTT AGTATGAGTA TCACATTTTA 3000

TAATTTTAGA TTTTTTTCTT CTGATGTATT TTTTTCAAAT ATACCAATAG TCCCaATTCT 3060

TTCAAGCTCA TTTTGCCAAT TTTTATATCC ATCGCCCATC TTTTCTATAA CgCATyCTTT 3120

TGgTtCTtCT GTCATTTTTt CAAATGTCAA AATAGCATAC TTTAGTTCTt TAwCyATAGG 3180

ACTTnGAnCA TCTAACAnCG CTAnCTTTTa TCAGCTATAT AATTTCTAAA TTTGTCTGAG 3240

AGTTTAAATT TAAGCCCTAA ATATTCATAA ACATATTCTT TCTTtGCCTC CaATGTATAA 3300

TCTGaTGGCT TAAAAGTTTC ATTTTTGcCA TTTGCATTAA TTGACATGCC TTCCTCTTCA 3360

GAACTTATGG ATTGTCCTTC TTTTATTTTT ACTTTTTTGA CACCCAACTG CGCTTAAGCT 3420

CAAACATAAT GCAAGTAATA GGCATGCTCC TCTTGGTACT ATTTTTTCAT AAATTATTCC 3480

TCCTTTATTT TTTACACAAT AGCATTTTAT AGCTACTTTC AATAACATTC TACACACCTT 3540

TTATGAAACG TTGATGAAAT ATAATTTTTT ATTTTTCCAA ACTATGGTAT AACAATATAA 3600

GTATTATGCT TAAAGTTGTT CTTACACGAT TAGTATATGA TCGATTAATA TATGAGCAAG 3660

CACAGTAAGT GTACGGACAC TTACGAAAAA AATGATGAGG CAGCCCTTTG GGATCTGCTT 3720

CTTTTTTTAT CAATTTTTAA CTTGTTAGGG TGTACCCTAA GACCCCGAAA TACATTCAAA 3780

GGAGGATTAC CTATAGCAAA TAGAATACGA AATGAAAGGC TTGAAATTAA ATTAACAGAA 3840

GAAGAAAAGG CTCTTTTTGA GGAGAAAAGA AAACTTGCGA AATGTAAAAA TATGAGCCAT 3900

TTCATCCGCA AGTGTGTTTT AGAAAAAGAA ATATATTAAG TGGATTTAGA GCCGTTTAGA 3960

GATTTACAGG GTTTCCTTTC CAATGCTACC AGCAACATAA ATCAGATTGC AAAGCGTGTG 4020

AATTCCACCG GCATAATCTA TAAGGACGAT ATAAATGATA TGAAAAAGCA GATTGAGTAT 4080

TTCTCAAAAG AGTTGTGGCA AATACATTCT CTGCTTCTTA ACAGAACTTC CGGAGTGTTA 4140

AATGAGAGTG TAAAATATTT TGTGTAAATA GAAAAAAGAA AGTCCCTTCT GTAGAATAGA 4200

GTTACCACAA CACATTCACA GAAGAGAGGA CTTCCCTATG AACGATTTTA CTACAGAAAT 4260

TCTAAAGACT CTAGCGAACA AAGGCGATTT GAATGAATTA TTCCGTGTCC ATTTGGAGAA 4320

AGCAGTCAAT ACGCTTCTCA AAACGGAGTT AACGGCTTTC CTCGATTACG AAAAGTATGA 4380

TCGCATTGGT TTTAACACGG GTAATTCTCG TAACGGCTCC TATGACCGTA CGGTCAAGAC 4440

CGAGTATGGG GAACTTCATC TCCAGATTCC GCGCGACCGC AATGGCGAGT TCAAGCAACA 4500

GACTGTTCCT GCTTATAGAC GGACGAATGA CACGTTAGAG GAGACCGTCA TTCACCTCTT 4560

CCGAAAAGGT ATTACCATGT CGGAAATCGC AGACTTGATT GAGAAAATGT ATGGGcATCA 4620

CTACACGCCC CAAACCATGT CCAATATAAC AAAATCATTT aCAGAAGAGG TAACGGCGTT 4680

TaAaGGGcGG GAcTTTCATG GACCGTTATG GCGCTATTTA TTTGGGCCnC ACGTTTTTCn 4740

CGTAAAAGGG AAAAnCGCCA CGAAG 4765

(2) INFORMATION FOR SEQ ID NO: 251:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4054 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

GTGTGTCTAA CGGTTCTAAT GGTTTATAAA CGACCGCATA TTTAATCCCA TATTCTTTTA 60

ATTTTTCTGG AGATAAACGA GCAmTAATCG CATCTTTTTG CATCAACTCT AAAACAGGGG 120

TCTGATTTCC TACTTCAACA TTGATGTGCG CATAGCGATT ATAAATAAAT GTGTCTTTTC 180

CTTTTGGATC TAATGGTTTC CAAGCCTCTA AGTCAGGTGT AAAAGCAACA CCGTTAAATG 240

TATGCACACC TAACATTGGC AAATAGCCAT ACATCAAGCG TTCGCCCACC CACAATTGGT 300

TGGGGTCTTT TTTGTTAATT TCTTCCACTT CTACCGCTAA TTTTTkACCA TAGACAGCAG 360

AAACGCCACG AGAAATTGGA TTTACAAAGG CTCCAGAAAC CATAATGATT CCTAATAAGA 420

CAAGGACAAA TAACGTGCGC CAACGTTTAA ATAAAGCAAC AATTAAAACG GTAGCGACAA 480

CTAAGACACC AATGATGTCC ATTCGTGTAA CGTATAACCG TAAATTGCCC GTATATAAAA 540

CAAAGAAATA CAGAACAACA TTTAATCCTA GTGCCCCTAA CCAGGCCATC ATTGGAATTT 600

TTGCTTGGCG ATTCCATAAw TAATCGATGA mCCwTAAaCT TAACAACACA GCGGCAAaCC 660

CGAAGaTAAC ATCGCCCGTT GTTCTGGCAC ATAAGACCAC AATGTCAACT TAGCAAAAAT 720

TGTTGGAAAT TGCACGCTCA TCCaAACTAA GTTAAATAAA CAATACACGA ATAAGACAAA 780

ACCGTAAATG tTTTCTTTGA TTTTTTTATA AAAAATAAAT GGTGAGAGCG GCAAAATAAC 840

AAAAAATAAT TGATAAAAAC TACTGATTTC CGAATTATTG GAATATGGAA CATCTTGAAA 900

CGGCATTTTC CAATTTGTTA AAAATAACAA AATATCGCGT TTAGGCATAT CGCCACCTAA 960

ACTTACCCGT TTTCCTGGAT AAGCGGTATC CATTACTGCT TTTAAGGAAT CCAATGAGCC 1020

ATATAGCGAT CCGCCAACAA TAATCGCCGT CATCAAGAGT GCCCCrCCTA TCAGGACCCC 1080

ATCCCACTTA TCCAACTTAA TTTTTTTACG AAATTCTAAA AAGAATAAAA TTAAAAATAA 1140

TAAAATCAAA TAACCAAGTG GTACTTGTAA CGCAGGATAA ATAACTAAAA CAAATCCTGA 1200

GGCCATAATA ACAGCCGTTA TCGCACAAAT GACACGACGT CGAACACTAC TATGATAATA 1260

GAAATAATTA TAAATTCCCA CTAAAAAACc TAAGCCAAAA AATACTAAAT CTCCAACTGG 1320

CGAGACAAAC CACCATTGAA GGGCCGAAGA AAACGGAATC CAAACACTTG CTAACAAAGC 1380

TAAATATTTA TTCCGCCGTG TGACGATCAA GCCAATTTCA AAAGAAAATA AAATTAAGCC 1440

AATTAATTTA AGATTCCAAT ACCATGATAA GCCATATTCT TTGCCTAATA GTACATAGCC 1500

CCACGTAAAA GGTTTCGCTA ACGTGGCCAA ATTCCACGCT GGTGAATTAT AGCCTACCAC 1560

CATGTCTTGC CCATTTAACG TAATCACATC ATTATGTGTT TTCAAACCTG TTTGTGTTTG 1620

TGAGAGGCTA AATGGGGTTT GAACGAGCCA TTCATCCGAG CGAACACCTC TTGGTTCACC 1680

AGCAATTAAT GAACTTTTTT GACCATCCGC ATATTCAGAA ACATACGCAT CCCACATGTT 1740

TAATGAACTA CCATGGACTT TAAAGGCGAC CAGTAAAATA AAAATGACTG CCGCTAGCAG 1800

ATAGCGCCAs GTrAACAAAT GATCAACCAA GCGAGCACTC TGCTCAGCTT GGTTGCGTAA 1860

TGGTTGTTTT TCTTTTCTTG TAGTTAGTTG TTGTTCCATC TTTTCGCTCC ATTTCGTCAT 1920

CCGCAATTGC ATCTAGTCAA TTCGTAATGA AAAGTTGGGA ATATCCCTTG TCAGATTAGG 1980

ATTATAAAAT GGATCATTTT CAATTAAAGG TCCCCATTTT TCTTCCATCA TCACTTTTTC 2040

TTGTTCAAAA CGTTTTTTCT TTTTGCCTTT ATCATCATAT CCACGAGTTT GAGATTCAAA 2100

ATGATAAAGC TCTGCTTCAT GAAGCCAGAC ATTGTCACGT CCTAAAGCTT GGACTTTCAA 2160

GCATAAATCA ACATCATTGA AAGCTACGGT AAACGCTTCT TCAAAACCAC CAACTGCATC 2220

AAAATCTGCT TTTTTCATAA GTAAGCATGC GGCAGTTACT GCTGAATAAT TAACATTAAT 2280

AGCTAACCGA CCAAAATAGC CTAAGTCACC ATGGGGATAA CCATAATGAC CATGTCCAGC 2340

GACACCACCT AATCCTAAAA TAACGCCTGC GTGTTGGACC GTATTATTTG GaTACAATAA 2400

TTTCGCACCT ACACAGCCAA TCCGTTCTTG CTGAGCmAAT GAAACCATCA AGGTTAACCA 2460

ATTCTCTGTA ATCACTTCGG TATCATTATT TAAAAATAAT AAATACTCTC CGTGCGCTTT 2520

TTTCGCTGCA CGATTGTTAA TAGTTGAGAA ATTGAATGGA ATATCGATTG ATTCAACGAA 2580

AAAGCGACCA GGTAATTGTT GTTCAAATTC TGCATATAAT TCGTGCATTT TTGGATCTGT 2640

ACTTCCATTG TCAGCCATAA TAATCTCGTA ATTTTGGTAG GTTGTTTTTT CAATAATCGA 2700

TGAAACACAA CGTTGTACGT CTTTATAGCC ATTTTTAGTT GGAATAATAA TTGAGACTAA 2760

CTTTTCAGAT TCAATGTCAT AATAAACATC GTATAAGCCA TTCGCCGCCC CATGCGTTGC 2820

ATGACCATTA ATTCCTCGAC GAACGAGCGC ATCTTGCACG GCTCTTAGCC CCGCTTCAAA 2880

AGCATAACCT TTGGAGCCTT GATCCACTGC TGTTGAGGTT GGCAGCATCC GCCAATAGTA 2940

AAGAACTTTT GGAATATGTT TGATGCGTTC TTTTGTCGTT TTTTCAGTAA ATCTTAAGAC 3000

TAAATCGTAG TCTTGAGAAC CTTCGTAACC TTTTCGAAAG CCGCCAATTT CTTCTAGAAT 3060

ACTCCGACGA TACACGCCTA AATGCGAAAT GTAATTCGTT CCTAAAAGTA AATCAGGTGA 3120

CCAATCTGGT TTAAAGGCTG GATCAGAACG ATTTCCATCC ATGTCAATCT TATCTTCATC 3180

ACTATAAATT AAATCCAGTT CTGGGTTTTC ATTCAGCACT TTAACCACTT CATAAAAAGC 3240

ATTGATCGCT AATTCATCAT CGTTGTCTAG TAAGGCGACA AATTCTCCGG TAGCAATCGC 3300

TAAAGCAGAG TTGGTTGCTT CGGAAATATG ACCATTTTGT TCACGAAAGA CGACCCGAAT 3360

TCGCTCATCC AATTGCTGGT ACTCTGTTAA AATCTTTTTG ACATTAGGAT CTGTTGAGGC 3420

GTCATCGGCC ATACATAATT CCCAGTTAGT ATAGACTTGA TTCAAAATAG AATCAATGCA 3480

CAACCGAAGC CATTTTTCTT CAACATTATA CACTGGCATC GCAATGGAAA TTTTAGGCTG 3540

ATAATGAAAT GTTGCGATTT CTTGCGTCAT CGCTTCTATG TCTAGTACTT CATTTCTAGC 3600

TAACCAGTTA GGATACGAAG CTTGATTTCT TAGTTTCTCA ATTTTCGCTC GTTGAATGGT 3660

ATGAGAAATA CCATTTTTAG CCATATACTT AAATCCTTTT TTGGCTTTTA TCCAAAGACG 3720

AGTTACTTTA TCTTCTGTCC CTGGAATCAC TGGATATTTT TTATTCAAGG GAAAATCCAC 3780

CGTTATAACA TGTGCTGATG ATTGAAAATG AAAAGGGAGT ACTTTTTTTT GTTTAATTCC 3840

TTCAAGTGTA ACAACAAAGC CAGCTTCAAT CGCAGGCTCT GTTTGATAAA TTTGATTAAC 3900

ATCTTCACGT AAAACGCGCT GAATGTTATA AGCGGACACT TGGTTTtCAT TATTAATTGT 3960

GAAAGTTGGG CTTTCTTTAG TGATtGTAtC TAACGCCCAA CCTGTAAtGG TTAAaTtATT 4020

tGTTGTTTTG TCGCGATAGA TACTATCAAA GATC 4054

(2) INFORMATION FOR SEQ ID NO: 252:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3074 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

AGCTCATAGT TTTTCATAAA AGAGTTCATC CGCTTAGCAG TAGCTTTAGA TACACCATTA 60

CCAGATTTAG TAGTATTATC AAACGGCTTT TCGAATAAAT CCTTTGCTGA AGCTTTGATT 120

TCTTTTATTC CTGATTTGCT TTTTTTAACT GCTTTATCAA TACCTTGAAC TAGAGGACCC 180

AATAAAAAAT TATCCTTTGA TGATTCCCAT AATTCACTAA TTTTTTTAGA GATATCGGGC 240

CATTTTTCTT GCAGTTTCTT TCCAAATTTA CTTCCAAAAA TTGATCCTGC AGCTGTTCCT 300

AAAACGCCAC CAATTAGAGT ACCTAACGGA CCAAATGTTG TCCCTGCTGC TGCACCCAAT 360

TTTGCGCCAG CTAATGTTGC TGCTTTAGCG CCTAATGCTC CACCAACGAT AGAACCTAAC 420

GAATCACCAA TTTTTTCAAA CTTATTATTT TTATTAATTT GAGTTAAAGT TAAACCTGCC 480

GAAATATAAG ACAATAAAGG AACCGATTTG GCAAACCTTC TCACTCCGCT AAAAAGTTTA 540

GTACCTTTTC CTATTTTGCC AACAGATCCA ACCGGTAAGG ATTCCTGTAT ATTTGGACTA 600

GCAGTAGTAT GCGTTGGTGC GCTAGCTTTT TGAGCTGCAG AATCAACATC TTCTAAAGCC 660

TTTCTTGTTT TAAAAGCTTC TCTCTGAGCA GTGTTCCCAA GCTTTTCTAG TTTACCAATA 720

ATACCTAAAA TTCCCTTTGA GACTCCCAAG AGCGGTTTTC CTACTAGCTT AAACCCTAGC 780

AAACCGCCAA TACCATACGT GGCCCATTTG CCAACTTGTT TCATAGCAAC AGGGTTTTCA 840

TCAGTAAATT CTTTGAGTAT ACGAATGCTT GGCTTAATAA TATATTTAAA CCCTTTATCC 900

```
ATTACGTTAA ACATATTTTT AGCACCTTCT GAAATGCCAG GTCCTATAGC TTTGCCTACC     960
TTTTTAGCTG TAGGAATTAT TTCATCAACA AAATAATCTT TTGCTTGTCC AAAGTGTTTC    1020
ATTACCTGTT TTGCTGTTGG CTCAACTTTT GCGCCAACAT TTtCTAACAT ATCTGGTAtT    1080
TCTTCACCTT TTTTTACAAA ATCTGTCGCT TTGATAATTA GTGGCGTGAA TATTGGTAAC    1140
AAACCTTGCC CTAACTTCGC CGATGTTTCT TTTATTGATT CTGTAAAAAC ACGGGTGGCA    1200
TTTGCAGCTT GATCACTTGT TCTTGCAAAA TCGCCTTGGG AGTTCTTAGT TTTAGACATT    1260
ACGTAGTTAT AACGTAAGCG TACTAATTCT TCCTGACTTA GAGAATCTAA TTTTGCTTGT    1320
TGAACCTCTT CACCTTTAGA TTCTGCTTCT GTTAATTTTA ACTGCGCATC TCTAGCTTCA    1380
ATTGAATTTT CACCATGTTT TTTAATGGCT TTATTTAAAC GATCTTGAGC TTTTTCTCTT    1440
GCCATGGCAT TCTTTGCAGC CTTAGAATTA TCTACTGACG ACTGCAACGC TCCACTGGCC    1500
ATCGCAAACT GTTCCAAGTT GGTTTGAGTC ATAACAATAC CCAAACCTTT TAACGCCTCA    1560
GTTTCACCAG TGAATACACC ATTTAATGCA GTATTTACAC GGTCAATTCC TATATTTTTA    1620
AATGAAGCAA GATCTCCTGC CAAGTCAACT AATGAAGTAG ACATTTTTGC AGCTTCTTCT    1680
GTGCCAATAC CCATTGAAGT TGACATATCT CCATATGTAG CTGCTAAGTC TAACGCTGTA    1740
CCTTGTGCCA AACCAATATT AGTCAGTGTG GACTTCGACC AATCTTCTAC AACTTTATTG    1800
TwATCACCAA AAGCTACTTC TACTTTGTTT AAAGCTTCAT TAGTATCAGA TGCATAGTCA    1860
AAGGCCTTTT gCCTGcTCCT GCGATTGCTG CACCAGCTAT CAtTGCACCA TTTCTTATCT    1920
TTGTAAAAGC ATTCCAAGCA ACATCTGAAG CCTTTGTTGC AACAGTACCT ATATGGCCAA    1980
TACTTGTTTG TACACCCTTC GCAGCTTTTC CAACTACTTT TAGAGATGTT GCACTAGATT    2040
GAGCGAAAAC ACTCGCAGAA TTTTTACTTT TAwCaAACTG aTTTCCaAcA TCTTTTAcTT    2100
GtTTTTCaGT TTGTTTAgCa GAGTTkGTTG TCtTTTGCaT AGAATCTTTT GcTCGaTTTC    2160
CgAACTGGAC AACTTTATCT GTATTTTGkT TTAAAcCTyC ACTATTTCTT TTTAAAgAGG    2220
TTGTGGCGCC ATCTATGGAT TTTGCTGAAT TTTGATAAGT ACCTTCCATT CGTGCAGCTT    2280
TAGCAAGAAT TTTATCAGTT TCTTCATTGG CACGCTCTAA CGATCTATTA TTTATTTTCC    2340
AATCGAGTTC GATTACTGAA CTACGTAACG CATCTGCCAT TATTAGATGA CACCTCCTCT    2400
TGATTTAGTT AGATTTATTT TTTGATATGC tAcTTCGTTC CAAACTGCTA GCTCCTCTGC    2460
CGTGGCAATA TCGACTTCAT CTTTAGTGGC AATGCCTGCT ATTACAGGCA GCCACCTAAA    2520
CATATCTTTT TGAACTTCTC TTTCCGTAAT TCTCGGATTA GGGCTAGTTA AGCATTCGAC    2580
CAAGAAAGTT GTCGGCTAAT GCCATAACCT CACGATATCC TTCGTGTTCA TCCCAATAGT    2640
CCCAGTTCAA TTTTGGTGTA ACAATAACTG AGTCCATAAG TTGAGAGTGG TAAGCAACGT    2700
CAGAAAAAGT ATTTGACGGT CCTTTTGAGT TATCTAAAAT CGCTTGAGCT GCACGTGTTC    2760
CAGGGAATTG AAATGTATAT TCCACATCTT CAACAGTATG TTTTTCTTGT TTTCCGAACT    2820
TATTAAATGG TTTTTTTTCT GCTTCTGGAA TGTTATTTTT CACTTCTAAT TCTTTTACTT    2880
```

```
CATCyTTTTT CGTCATGATA AATTCCTCCT AAAGTTAAAT AAAAAGCACt TAACGAGATG     2940

TTaAGCGCyT AGtTATATTC GTGTTTGTAA TCTAATGCTT TAATCGTATA AGAACGAGTC     3000

GGAACGCCTT TACCAAACGA TCCATCAGGT GTTTTTTCAA TATAGGCTTT TGATGCCCAT     3060

GCTTTTTCAG TTGA                                                       3074
```

(2) INFORMATION FOR SEQ ID NO: 253:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1828 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

```
CTTGTGATTG TAGATGAACC GCAGATTCCT ACGAATCCAG TGCCCTTTTA TCCTTATGTA       60

ACGAATCCTA ATCTCGTGCT CTTCCGGTTT CACGGACGTA ATGCGGCTGG CTGGTTAGCC      120

AATGATGCAG AGTGGCGAAA AAAAAGAACC CTTTATCATT ATAATACGCA GGAAATCGCT      180

GATTTAAGCG AAGCTGTTTT AAAAATGTCG CAAGAAGCTA AAGAAGTGGG CGTAATTTTT      240

AATAATAATT CTGGCGGCGA TGCCGCAGAA AACGCTCTAC AGATGCAAAA AGTCTTGAAT      300

TTGAGTTACG ATGATTTGAA TCCAAAACAA TTAGATTTAT TTTGAGAAAG AAGGTAGCTG      360

AATGTCTGAA AAGCAATCCC CTTATGAAAT GGCTTGTGCA TTCCACGAGG TTTTTAATCC      420

ACAACAACCA GAGCAACCAA CTGCTTTTTC GCCAGAGCTA GCCAGTCGAC GAGCAGAGTT      480

TAAAGTCGAA GAACTTGTAG AATTTTTATA TAGTGCTGCT GATAATGACC AAGTGCTTTT      540

TCAACAATTG GTCAACGGAT TAAAAAAATC TGTCGATAAA GCAGAGCAGA AAGTCTTGGC      600

GAAAGGCAAA CCAGTATCTG ATCCTTTAGT GGAGGAAGTG GATGCACTAG TTGATTTATT      660

GTATTTTACT TATGGTTCGT TTTGCTTAAT GGGCGTCAAT CCAACAAAAA TTATTGAAAT      720

CGTTCATCAA GCCAATATGG GGAAATTATT TCCAGATGGG CAGCCACATT ATGATCCAGA      780

AACAAATAAA GTATTGAAAC CTGAAAATTG GGAACAAGAT TTTGCGCCAG AAGCTAAAAT      840

TAAAGCGGAA ATTCAACGCC AATTGAAAAT TACTGaAACC AATGAAAAAG ACGAATGAGA      900

AAACTCATTC GTCTTTTTCA TTAATCATGG AAGAACGTCC GATATAAGGC TTTGATGGCT      960

TTTTCCTCTT GTTCTTTTTT GATACCAAAC ATAATACTTA CTTCCGAAGA TCCTTGGTTA     1020

ATCATTTCTA AGTTGATTTT TTCTTTGcTA ACGCCATTGT ACTTTCAGCA GTAACCCCAA     1080

TTCGTTGGCG CATGCCTTCC CCGACAATCA TAATCATAGA AAGATGATGC GTGATGCGTA     1140

ACTCATCAGG CGCTAAAACT TCTTGGAGTT GAGCCATCAG TCGATTTTCA ATTTCAGTTG     1200

TTAACTGGTC TTCTCTTAAA ACAATCGTGA TGTCATCAAT TCCTGAGGGC ATATGCTCAT     1260

AGCTCAAGGC TAATTTTTCT AGTACTTcTA AGAGCCGTCG CCCmAAACCt AATTCTCTAT     1320

TCATTAAATA CTTACTAATG TAAATACTTG CGAmCCCTTG ATCACTAGCA ATTCCTACAA     1380
```

CAGGAGCATG CTTCACTTTG CGAGAGGTGG TAATTAGGGT GCCAGGATGA TGTGGGTTGT 1440

TGGTATTTTT GATAACGACT GGAATATTTG CGCGATAGGC AGGCATTAAG GCTTCGTCGT 1500

GGAGGACGGC AAAGCCTGCA TAAGCTAACT CCCGCATTTC TCTGTAGGTC AATTCTTGAA 1560

TCGTTTGAGG CTGTTCAATA ATTCCAGGAT GGGCGACATA AATTCCGTCC ACATCGGTAA 1620

AGTTTTCATA CAAATCTGCT TGAACACTGC GGCAACAATG GACCCTGTGA TATCGGAACC 1680

GCCACGAGAA AAAGTACAAA TTTCGCCGGC TTCTGTATAA CCAAAGAAAC CTGGAATAAC 1740

CAAGATTTGT TGAGTCTCAC GATAGACACT AATTTTCTCT AAAGCTTTTG GTAAAATGCG 1800

GGCATTGCCT GGCTCAGGTG TCACTAGG 1828

(2) INFORMATION FOR SEQ ID NO: 254:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1623 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

GGCCGCCACG GTTTTGGTTG TTCCCGCCAT TGTTATTTTG GCTGGTCCGA TTTTGATTCG 60

TTGTAGAACC GTTATTGTTT TGGGTATTGC GGTTTTGATT GTTTGTATTT GTTnGATTGC 120

TACGGTTTTG TGTGTTGCTT TGGCCGTGAC GATTTTGGTT GTtGttGTTt CTTkGTtGyC 180

CATtCGTkGA TkGATtGTTt GGACGATTTk GGTwATTtCT TkGTtCACCA GGTTTATtGT 240

yCATkGGTTy CTTATTGGCT GGTTTATTTk GTGTkGTTkG TkGCTTtTGG TTGTTTGTTG 300

CTTTTGATTG AGTATTTACG GTACTTCCTC CtCCGAATGC TTGACGAAGC TTTGTTTCTT 360

GCTCGGAACT AATTGCGCCC aTATGGTTTT TAACATCCAT ACCTAATTGA TGCGCTTTAT 420

CGACAACGTC TTTGCTTGCT TTATTCATTT CTTTTGCTAG TTCATAAATT CTTTTTTTTC 480

CCATGCAATC ACCTTCCTAA CCTTTTAGTA ATTCCTTGAT TTTGGTGGCA AATCCTGCAT 540

CAGTTACTCC GATAACCATT CGTGGTCGCC CAATCGCTTG GCTGAGCTCG AATTGACTGA 600

ACAGTTCGCT AACAGGAACA TTATAATAAG AACTTTTGTC CTTGATTTTC TTTCTAGTAT 660

TCTCACTGGC ATCTTGCGCA ACAAACACAA ATTTCGCTTT GTTTGCACGA ATATCTTTAA 720

GGGTTAATTC CTCGCCTGTG ACTAATTTAC CGGCGCGCAT TGCTAAACCT AAAAGGTTCA 780

AAATTTTGGT TTTATTTTCC ATCGCCAAAC AACTCTTTGC GGGCTTTTTG ATGTGTTACA 840

TAATCCAATA GTTCTTGATA AAATTCATCA CTTAATTTTG TTTCTAAGAC ACGATCAAGA 900

ATTTTTTTAT CCCAAGCTTT TTGGACTTCC GCAGGATCCA AATCAACGTA AGCGCCACGT 960

CCAGGTAATT TACCAGTTGG ATCTAAGGCC ACCTCGCCTT CTTTGGAGCG AGTAATTCTG 1020

ACTAATTCTT TTTTCGGTTT CATTTCGCCT GAAACAACCG ATTTACGCAT TGGAATTTTT 1080

CTCTTTTTCA TTTTTTCACC CTCCAAAAAC GTTATTCTTG TTCAGGTAAT TCTTCTGTAA 1140

TTTCTGCTTC AACAACTGCT TCTTCAGCAG CCAAGTTTTC GTAATCATCT GCCGTCATAT 1200

CTGAAtGGAT AATTGCTTCA TCCATTTCTT CTGCCACTTC GGCAACTTCT TGATTGGCTT 1260

GTTGTTCATA AAATTCAGCC ATATCAGATT CTGATTTGAT ATCAATTTTA TGGCCCGTTA 1320

ATTTAGCAGC TAAGCGCGCA TTTTGACCAC GtTTACCAAT TGCTAATGAC AATTGATAAT 1380

CAGGTACAAC AACTGTACAT GCTTTGCTAw TTtCTGGATC AAAAaTAACA TCTAATACTt 1440

GAGCTGGAtT CaAGGCATTA CTAATATAAA TAGCTGGATC TTCGTTCCAT TCCACAATAT 1500

CCATATTTTC GCCTTTTAAT TCATTCACAA TCGCTTGGAC ACGTTGGCCT TTTGGACCAA 1560

CACAGGTACC CACTGGGTCA ATATTAGGAT CTGTTGAACG AACGGCTACT TTTGAACGAT 1620

CGC 1623

(2) INFORMATION FOR SEQ ID NO: 255:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3601 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

ATGCCTGCCG GTCGACTCTA GAGGACCCCC GCATGAACCA GACGTGACGG CTTTGATTCG 60

TCTAAGTGAT GTCTATGAAA TTCCTTTAGC GACCAATATT GGCAGTGCTG AAATTCTTTT 120

GAGAGGTGTC GAGGCGGGGT TTGCGGACTT CCGTGAGGTT ATCCaTGAAG GGGATCGTCG 180

ACCGTTAGCT TTTTAGAGAA CCTTAATGGT TATTTTGACT ATTTTGTGGT ATATTTTATG 240

AGGACTTTTG TCCAAATAAA ATAACAGAGT AGGAGACAAA GCGTTAAGTG CTAAAAGGAT 300

GGGATGTTGC CTTTTGGACG AAGAACATAA ACAGTGTGTT GCTAGAAAGC AAAGGTTCTT 360

ACTTTGCTTA TCATCAGAAC AACACGTTCA GCTTTTTAAT GTTCGCGGTT TTGTTTTCGC 420

ATTCGCTGCA TGGTTCGTCA TGTAGCAGCT CTTAGAGGAG ATGCTAAAAA TGACAAAGAA 480

AAAATTTGGC ACAAAATCCA TTGCGTTAAT GGGCGTACTC ATTGCGGTGG TCGTTGTCTT 540

TTCACGCTTT TTTGCGTATG AAACAACGTT TTTAAAAATC AGTTTCACGT TTATCCCAGA 600

ATCATTGATT GGGATGATTT TCGGACCTTT TTGGGCTGGA ATCGGCACGG CGGTGGCTGA 660

TGTGGTAGGA ATGTTGCTTT TTCCAAAGGC AGGCTATTTT CCTGGGTTTA CTTTAAATGC 720

CTTTTTGGCT GGTGCTATTT ATGGTTATTT CTATTATAAG AAAGAAATGA CATGGCAACG 780

AGTGATTTTG GCGACATTGT TGGTCACCGT TTTAATTAAT ATCATTTTAA CTCCTTTGTG 840

GTTAAGTTTA ATGTATGGCG TCAATCTAGC AAATTTTGCT TGGTGGGTAC CACGCTTAAT 900

TAAAACAGTT ATTTTCTTTC CTATTCAAGT AATTGCCACT TATTATTTAG GCAATAAAAT 960

TCCCTTTAAA CGTTTATTTG GAAAGCCATT ATCTGAATTA GATCAATAAA TTTAAAATTA 1020

AAAGAGTCAC TTTTCATCAT CCCAGAAAAG TGACTCTTTT TTATTGTGCT TTTTGATTAA 1080

ACTGTGCTTC GTAAAGGTCA CGATAATAAC CCTGTTCTTT TTGCAATAAT TCGTCATGGG 1140

TGCCAATTTC AACCACTTGA CCATTATCCA TAACTAGAAT TTTTTCAGCA CTTTTGATTG 1200

TTGCTAAGCG ATGGGCAATA ACAAAACTGG TCCGGCCAGT CATCATTTGT AAAAAAGCAT 1260

CCTGAATTTT TTGCTCGGTC AATGTATCAA CTGAACTGGT CGCTTCATCT AAAATAAGCA 1320

TCGGCGGATT ACTAATCATC GTGCGTGCAA TAGTCATTAA TTGTCGTTGT CCTTCAGAAA 1380

TTTTAACGCC TTGGGAACCA ATCTCGGTAT CTAGTCCTTG AGGTAAACGC ATGACGAAAT 1440

CAAAAATATG GGCTTTTTTC ATGGCTTCTA TAATTTCTTC GTCAGAGGCT TCAGGATTTC 1500

CGTACGTTAA ATTTTCCCGA ATGGAGCTAT CGAATAGCCA AGTATCTTGG AGAACCATAC 1560

CAAAACTTTT TCGCAAGCTG TCCCGAGTAA TTTGAGTAAT GTCGTGAGTA TCAACCGTAA 1620

TTTGGCCACG GTCGACCTCG TAAAAGCGCA TTAACAAATT CACTAATGTT GATTTACCCG 1680

CACCAGTTCG CCCAACAATG GCAATTGTTT CTCCTGGTTC AGCAATTAAA TTAAAGTTTT 1740

CGATTAAAGG TCGCGTCGGC GTGTAAGAAA AATCAACGTC TTTAAATTCA ACACGGCCTC 1800

GCGCTTCTTT CAAAACAATT GCGTTCGAGG CATCAGGACG TTCAACTGGT TGATCCATCA 1860

TATCAAAGGT ACGTTCCAAT CCAGCTAAGG CGGCTTGGAT TTGGTTGGTG ATTCCTGAGA 1920

GTTCAATAAA AGGTTTAGAA AATTGACTAG AGTAAATGGT GAAACTGGAA ATAACACCAA 1980

CTGTGACTAA AGGATTGCCA CTTAAAACTA ACAAGCCaCC GACTAATCCA ATCGATAAAT 2040

AAGATAAATG GTCGATAAAT CGAGAAAGTG GGTTGGTTAA GGAGGAAGAA AACTGAGCCt 2100

TCTGTCCTTT TTCATACAAC GTTTGATTCA ATGCTTCAAA TGTTTCTTGC GTAGCTTGTT 2160

CTTGTTGAAA GGCTTTAACA ATTTTTTGGT TGCCAACCAT TTCTGAAACG AAACCGGAAA 2220

TTTCACCAAC AATCGTTTGT TGTGCAGCAA AGTTTTTTTG TGAAGCTTTT GCAACAAGCC 2280

AGTTTACAAT AAAAATAATC GGCGTACTGA TAAGAACCAC TAATGTGAGC AGGGGACTTA 2340

GCCGAATCAT AAAGAGCAAA GCAACAACGA CGGTGGCCAT GCCAGAAAAT AATTGGTTGA 2400

AGACAGCCGA GCAAGCCATT GAAATATTGT CAATATCGTT GGTAAAGCGA CTGACGATAT 2460

TTCCGTGTGA CGTTTGGTCA TAATAACTTA ACGGTAATTG ATTTAAATGG GCAAAGGCAT 2520

CTTTTCTCAG CTGTGTGGTC GATAAATAAG ACACGCGATT ACCGAGACGT TGAATCAGCC 2580

ATTGACTTAG AACGGTTACG AGTAAAATCC CTGCTAATAA ACCTAAAATT TTGATGAGTT 2640

GCGCAnCATT tGACTTGTCC TTTACCCACC ATTGTATCAA CGGATTTACC TATATAATAA 2700

GTCATGAGGA CAGTCGAAAA ACCACTGAGA ATTCCTAGGA TTAAGGCGCC AATAATTTCT 2760

TTTGGATAGC GTAGTAAATA CGGCTTAAAA CGTGAGAAGG CACCAAAATG AATTTTTTTA 2820

GTTTGCATGG GTATCCTCCT CCTGTGACGC AACGATTTCT TGATACTCTT TAGAAGAAGT 2880

AAGTAATTCT TCGTGGGTGC CAAGGCCAGC GACTTTTCCA CTATCTAGGA CTAAAATTTG 2940

ATTCGCCGTT GCAATCGAAC TCACGCGTTG TGAAATTAAA ATAACGGTTG TCTCTGCTCG 3000

TTCTTTTTGT AAAGCACGAC GCAGATTCAA ATCTGTTTGA TAGTCGAGGG CACTCAAGGA 3060

ATCATCTAAA ATAAGTAAAT GCGGTTTCCT AATCAAGGCA CGGGCAATGG TCAACCGTTG 3120

TCTTTGTCCA CCAGAAAAGT TTTTTCCGCC TTCGTTAACA GGCGTGTCTA ACCCTTGGTC 3180

TAATTGTTCA ACAAATTCTT TACATTGGGC GATAGCTAAT GCTTCCCAAC AGTCTTCATC 3240

AGTTGCATTT GGTTTGCCCC ATTGTAAGTT TTCTCGAATA GTCCCTGTAA ATAAGACGGC 3300

AGTTTGTGGA ACACTTGCAA CTTGTTGGCG TAAAGTAAAG AGCGGCCAAT CGCGCACATT 3360

TACACCATTA ATAAAAAGGT TTCCTGCACT CACATCATAA AATCTGGGAA TGAGTTGGGT 3420

TAAGGTACTT TTCCCGCTGC CGGTAGGTCC TGTGATGCCG AGAATTGAGC CTTTAGGAAT 3480

CGTAAAATTG ATATTTTCCA AAGCCAAACC GGCCTCTGGT TCATAGCGAA AATCTACATG 3540

GTCAAAGGTA ATATCACCAA ATTGTGGAGA TGTTTTTGCT GATTCTGGTG TATCTGTTAA 3600

C 3601

(2) INFORMATION FOR SEQ ID NO: 256:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9589 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

GGATGATTTT CAAGAGGAGA ATCCTCAATT AGTCTTGATG GATATTAATT TACCTGTTTA 60

CGATGGGTAC TATTGGTGTC AAAAAATTCG TGAAGTTTCT AAAGTACCAA TTATTTTTAT 120

TTCGAGTCGT AGTACCAATA TGGATATGAT TATGGCGATG AACATGGGCG CGGATGATTT 180

TGTCACAAAA CCGTTCCAGA TAGATGTTTT AATTGCTAAA ATCAATGCTT TGTTGCGCCG 240

CTCTTATAAC TATTCCGATA CGGATAGTGA AGTGCTTTCA CATAATGGCA TTACTTTAAA 300

TGTGGACAAT GGACGAATGG AAATCaGAGG TGAAATGATC GATTTAAGTA AAAATGAATA 360

CCGATTACTG TATCTTTTAA TGAAAAAGCA CGGAAAAATT TTAACACGTG AAAAACTATT 420

GCGCGCTTTA TGGGACGATG AACGTTTTGT GGATGACAAT ACGCTGACGG TAAATATCAA 480

TCGTTTACGA AAAAAAATTG AGCAAGCAGG GATTGCTGGC TATATTGAAA CCAAAGTGGG 540

CGTCGGTTAT ATGGTGCCCT AGATAGGAGT AGACAATGAC CATTCTAAAG TATTTAAAAG 600

ATCGCTGGTT ATTACTAATT GGTTGGTTAT TTTTTCTGTT CTTAACTTGT TTTATTTTAT 660

GGTTGGCTCC TAATGTCCGC TTAGATTGGA CAGTGGTTGG CTATATCTTT TTACTGCAAA 720

GTGTTTTTTT ATCGCTATTT TTAACAATTG ATTATTATTT AAAACGTAAG TGGTGGCTTT 780

CGTTGGCAAC GGAAAAAGAG CCGCCTTCTT TACAAGAATA TTTGAATACA GCAGAAAAAG 840

AAGAAGAACT ACTTGTCCAA ACGTATATTA ACGGTTTATT GCAAGAGCAC CAACAAACCA 900

TGCAGCAAGC AATTAACAAC CAACAGGACC AAAAAGATTA CATTGACTCT TGGGTTCATG 960

AAATCAAAGT CCCCCTAGCG GCTATCACGT TACTGGTACA ATCTGTCGAA GATGATATTC 1020

CAGAGAAAAA ATATTATCTA TTGGAAAATG AGTTAGGTAA AATTGATGAA TATGTCGAGC 1080

AAGTCCTATA TTATGCTAGA CTGGACAGCT TTTCTCGAGA TTATTTACTC CAAGAATATT 1140

CTTTAAAAGA AATTGTGCAA TCTGTTGTTC GTACGCAAGC AAATTATTTT ATCCAAAAAA 1200

GATTGCAATT CTCAATTGAA GGGGAAGATG AAGCGGTTTT AACCGACCGA AAATGGGTCA 1260

TTTTTATTTT CCGCCAATTA TTAAGTAATG CAGTCAAGTA TACGCCAGAA GGCGGCACGA 1320

TTACGGTCCT AATTTCCAAA AATAAACAAG GTATTTATTT ATCTTTAAAA GATTCTGGGA 1380

TTGGCATTCC GACACAAGAT CAACGACGCA TTTTCGATAA AGGATTTACT GGTGAAAACG 1440

GGCGAAAAAG CGAACAGCAT TCGACTGGTA TTGGTCTGTA TTTGGCTCAT AGCTTAGCGA 1500

AAAAGTTAGG TCATGATTTA ACAGTCGAAT CGACAGAAGG GCAAGGTACG ACGATGACAT 1560

TGTTTTTCCC TTCCCTCAGT TATTATAATG AAGTGAAATA AAAAGTAGGT CTATCAGCAA 1620

ACTAAAGAGG TGATAGGCCT TTTTTTATTT ATTAAAACTT TCTTGCTAAC TTCTCTTACT 1680

CTTCTTAGTT GTTTAGAAAA CGTTTTTATT GTTAGAATGA AATGAAATAA ATTTGTTAGG 1740

AGAATATTAT GAGTTTACTT ATCGCTTTAG TGCCCATGAT TGCGTGGGGC AGTATTGGTT 1800

TAGTCAGCGG TAAAATTGGT GGAAGCGCCA ATCAACAAAC GTTAGGTATG ACAATCGGCG 1860

CATTACTATT TTCAATTGTT GTCTTTTTTG TCATCCAGCy TACGTTGACG ACAGCCACAT 1920

TAATTGTTGG ATTTATATCT GGTTTATTCT GGAGTTTGGG ACAGAATCAA CAATTTCATA 1980

GCATGAAATA TATGGGCGTT TCAGTTGGCT TACCAATTTC AACTGGGATG CAGTTAGTTG 2040

TTAATACCGT TGCAGGGGCT GTTTTTTTCC ATGAATGGAC AAAAACAAAA GACTTTGTTG 2100

TGGGCTTTAT CGCTTTGGCC TTTTTAGTTT TTGGGGTGTA TTTAACAGCT CGACAAGATG 2160

ATGATAGTCA ACCGAAAACC TCAAACAGTA TGTTAGATTT TAATAAAGGG ATTCGGGCGT 2220

TGATTTTCtC AACAGTTGGG TATGGCGTCT ATACCATCAT TATTAATGCC ACTGGTTTAG 2280

ATCCGTGGGG AATTATTTTG CCACAAAGTA TTGGTATGTT AGTGGGGGCT AGTTTCTTTG 2340

CCTTCAAAAA AGTCAAAGTT GATCGTTTTG TTTGGATGAA TATGACGACG GGTCTGCTTT 2400

GGGGCTTAGG AAATATTTGT ATGTTATTAA CAATGCGTGA AATTGGCTTA GCTATTAGCT 2460

TCTCTTTATC ACAAATGGGG aTTATTATTT CGACTCTTGG GGGCATCTTC CTTTTAGGCG 2520

AAAGAAAATC AAAAAAAGAA ATGTTTTATG TTaTTTnTGG TTGCATTTTT GTTATTCTAG 2580

GTGGTATTCT TTTAGGGTAT ATGAAAGCGT AAATAGTTTA TTTACGCTTA TTTTTTTTAC 2640

GGGGGTAAaC ACATgTCTAT TTTCAGAAAA AAAGAAATCA CTGTGCCTGT CAATGACAGT 2700

GGTGGGATGA AAAAAAATCT TAAGACGATG GACTTAATTT TTCTAGGAAT TGGGGCTGTA 2760

GTAGGAACAG GGATTTTTGT TGTGACAGGT GTCGCAGCGG AACGCTATGC AGGTCCAGGG 2820

CTCGTCTTAT CATTTCTGGT AGCAGCAGCC GCGATTATTT TATCTGGTTT ATGCTATGCA 2880

GAATTTGCAT CACGTATTCC GGTCATTGGT GGGCCGTATG CCTACATGTA TGTGGTCTTT 2940

GGAGAAATTG TGGCTTGGAT GACTGGTTGG ATGATTATTT GTGAGTTTTT CTTGGCGGTT 3000

TCATCCGTTG CTTCAGGTTG GTCAGGCTAT GTCCATGGCT TTTTGGATAG CTTAGGTTTT 3060

TCTTTGCCGC AAGCGTTAAG TGGCGCTTAT AATCCAACAA ATGGTACATA TATTGATTTG 3120

ATTGCCATGT TGGTCGTTGT AGCCGTAACT TTTTGGGTAT CTTTAGAAGC GAAAACAGCG 3180

TTACGTTTAA ATAACTTAAT GGTTTTCGTG AAGTTTGGGA TTATTTTATT ATTTGTTCTC 3240

GTGGGAATCT TTTATGTGAA ACCAACGAAC TGGCAACCAT TTATTCCTTA TGGTTTTTCA 3300

GGCGTTTTCA GTGGTGCTGC TTTAGTTTTC TTTGCCTTTC TAGGGTTTGA TGCGGTCAGC 3360

ATGGCCGCGG AGGAAGTGAA GAATCCGAAA AAAGATATTC CAAAAGGGAT TATTGGTTCC 3420

ATTATTATTT CAACATTGTT ATATATTGTT GTTACGTTAG TTTTAACTGG GATTGTTCCA 3480

TTTACGGATT TAGGGGTCAA GGATCCGGTG GCTTTTGCGA TGCGCTTCAT TAATCATGGC 3540

GCAATTGCGA CAATCATTTC AGTTGGTGCT ATTTTAACCT TATTAACAGT GACGATTGCG 3600

ATGATGTACA GTTTAGCCCG TGTGATTTAT GCGATTAGTA AAGATGGATT ATTGCCGCAG 3660

TTTATGAGCA AAATCGATGA GAAGCGGCAC ACACCTAAGA ATGCAACCTA TGTAGCTGGA 3720

TTTTTAGCAA TGGTTTTTGC TGGGATTGTT CCAATGGAGA TGTTGGCTGA ACTTACAAAT 3780

ATCGTGACAC TCTTTTATCT GATGTTTCTC GCTTTAGGAA TTATCAAATT ACGGACAATG 3840

AAAGAGAACC GCAAGCAGtG rATTtAAAGT ACCACTgTGC CGGTTCTACC ATTAATTTCA 3900

ATCGTGGTTT GCGCCGCTCT GATGTTCCAA TTGTCTTTAG CTACTTGGCA AGTATTTGGC 3960

GTAgcgTaGT GATTGGTTTG GCTATTTATT TCTTTTATGG ACGTCACCAT AGTATCGTTC 4020

GTTTCGAGAC AAAAGAATAA AAATGAACGA AAAAAGAGAG GAATACTTGT CAAATAGACA 4080

GGTGTTTCTC TCTTTTCTTT CTGGCATTTT CTTCATTTTC AAGAGAAAAC CGACAAGAAA 4140

TATAGTTTAA TAATGTCTAA TTATGGTATA CTGTTTAGGA GTAAATCCTG AGGGTTTAGG 4200

AAAATGACTA AAGGAGGTTT TTTCTGTATG TCACAGAAAG AAACATTTTA TATTACCACC 4260

CCAATTTATT ACCCAAGTGG GAAGTTACAT ATTGGGAATT CATATACAAC AATCGCCTGC 4320

GATGCAATGG CCCGTTACAA GCGTTTAATG GGTTTTGATG TATTTTATTT AACAGGTGTG 4380

GATGAACACG GCCAAAAAAT TGAAAAAAAA GCTGCTGAGT TGAATGTGAC ACCAAAAGAG 4440

TACGTTGATA AAATGGCAGC AGATGTTCAA AAATTATGGA AAACATTAGA TATCAGCTAT 4500

GATAAATTTA TTCGGACAAC GGATGATTAT CACATGGCAG CCGTGCAACA AATTTTTGAT 4560

CGTTTAGTAG AACAAGGAGA TATCTACCTT GGAGAATATG AAGGTTGGTA TTCAGTTTCT 4620

GATGAAGAAT TTTTCACTGA AACACAATTA GCTGAAGTAT ATCGTGACGA TGAAGGCAAT 4680

GTCATTGGTG GGaAAGCACC AAGTGGTCAC GAAGTCGAAT TAGTTAAAGA AGAATCCTAT 4740

TTCTTCCGCA TGAGTAAATA TGCGGATCGC TTGGnACAAT ATTACGAAGA ACATCCAGAA 4800

TTTATTCAAC CAGAATCTCG TAAAAATGAG ATGCTAAATA ACTTCATCAA ACCAGGATTG 4860

GAAGATTTAG CCGTTTCAAG AACGACTTTC TCTTGGGGAA TTCCATTAAA AAATGATCCA 4920

AAACATGTGG TTTACGTATG GATTGATGCG TTATCAAACT ATATTACAGC ACTAGGTTAT 4980

GGcTCAGAAG ATGATTCATT ATTCCAAAAA TATTGGCCTG CGAACGTGCA AATGGTTGGG 5040

AAAGAAaTCG TGCGTTTCCA TACGATTTAT TGGCCAATTA TGTTAATGGC CTTGGATTTA 5100

CCATTACCGA AAAAAGTCTT TGGACATGGT TGGTTATTGA TGAAAGACGG AAAAATGTCT 5160

AAATCTAAAG GAAATGTTGT TTACCCTGaA ATGTTAGTAG AACGCTACGG tCTAGATGCG 5220

TTACGTTATT ATTTATTACG AGCAATTCCA TTTGGTAGTG ATGGCGTCTT TACACCAGAA 5280

GATTTCGTTT CTCGCTTAAA TTATGATTTA GCTAATGACT TAGGAAACTT GTTGAATCGA 5340

ACAATTGCGA TGATTAATAA ATATTGTGAT GGCAAAGTAC CAGCCTATGC TTCAAAAGTT 5400

ACCCCATTTG ATAGTGAACT ATCGACAACT GCGGCGAATG TTATTGGAAA ATATCATGAA 5460

GCAATGGAAA AAATGGAATT TAACACGGCC ATTGCAGAAA TCTGGACGCT AGTTTCTCGT 5520

GCCAATAAAT ATATTGATGA AACAGCGCCT TGGGTATTGG CTAAAGAAGA AGAGAAACGC 5580

AACGAATTAG AAAGTGTCAT GATCCATTTA GCGGAAAGCT TACGAATTGT GGCAATCTTA 5640

TTGCAACCAG TCATGACAGA AACACCAGGG AAAATTTTCG AACAATTAGG CTTAGATCCA 5700

GAAACAATGA ACATGGAAAA TATTCATTTT GGTGAGTTCC CAACAGACGT AACAGTAACA 5760

AGTAAAGGAA CACCTATTTT CCCACGTTTA GAAATTGAAA CAGAAGTAAC GTATATTCAA 5820

AAGAAAATGT CCCAAAGCGA AAGTGCTACA GAAGAGGATA TTAAATGGAA TCCAGAAGAA 5880

ACAACGTTGG TTTCAACGAA AGAAAAACAA ATTAAGTACG ATGATTTCGA TAAAGTGGAA 5940

TTAAAAGTTG CCGAAGTTAT TGACTGTAAA AAAGTCAAAG GAGCAGATAA ATTATTGCAA 6000

TTCCGTTTAG ATGCTGGAGA TGAAAACCAT CGTCAAATTC TTTCAGGTAT TGCTGAATTT 6060

TATCCAGATC CAGCTGCGCT AATTGGCAAA AAAGTAGTCA TTGTGGCTAA CTTAAAACCA 6120

AGAAAAATGC GTGGTCAAAT TAGCCAAGGA ATGATTCTTT CAGCTGAATC ACCAGAAGGG 6180

AAACTTCAAA TTGTTGAAGC ACCAAAAGAA ATGCCAAATG GTGCAGGAAT CGCATAAAAG 6240

CATAAGTGGT AAAGATAAAA GGTGAATTGT AACCTTTTAT CTTTACTACT TTTTTTGTAT 6300

ACTAGAAGAA AAGTGATTTG ATGTGGGGGA AGTTTGATGG CTTATCATAA ATTTAAATTT 6360

aTGAGCmACG AAAAAATGAy CCTAAATGGC GTCCGTATTA TTTAGaAGCt ATGAGAGTAA 6420

GGGAACAAGC ACGGCTAACA GTACAACGTA GAAAAAGAAG GGCAAGAATT AGAACCTTTT 6480

tAGTGGTTGC GATACTGGTT ATTTTAGTTT TTCTGTTATT GGCTCATTGG ACATATTCAC 6540

AAACAAATGA ACTCTCTTTC TCCAGTTTTA TCGACGTCCT AAAATCGAAA ATCAACGAAT 6600

TTTTAAAtTA AAAGAaTCGA AGAAAAGCCT AGAAAAAATC ATTTGCGATT TTTCCTAGGc 6660

TTTTCTATTT TAGAtTTCAC TTTTTTAGGT ATTCgTkTTT TACCAAAACG GCCTTTCGTT 6720

AGGCGACTTG GTAAAATGAT CAGTAAGAGT GAAAAAGGGA CACTGATAAT TTGCCACGTG 6780

AGTAAACGGC CTAAGCTATT ATTTTTATAG AGCCACTCTA ATATAAAATG CAGCCCTAAA 6840

ATACTTAACA AAGCCAAGCC TGTCAGGAAA TAAGGCGTTG ATACCCAGTT GGTCGTTTTC 6900

AAGAGACTAA AACCTAAAGG AACTAGATAA AAGTAAGCAC TTGTTAAAGT AATACGCCAA 6960

CTTTTCATAA CGGTGCCTAA ATTAATTATA AAAAAAAGTA AAGATAGGAC TAAAGCGAAA 7020

GGTGGAAACA AAGCCACTAA CGCCGAATAG ATGGTTCCCC AGACGAGCAT TAAGGGACCT 7080

TTAACAATAG CGGCAATAAG TGTAAACAAG CCAATTAGGA GTAACTGCTG TGTCAGTGAG 7140

CCAAAGGTTA GCAAGCCAAA TAAAATAAGC AGTAGCCAGT GCCaAATGGT TTGGTTTTGT 7200

TGTTTTCGCT CAATAGTTAC ATTTTTTTGG ATAGAGGTTT TAAAAAAATT TTCTAAATCA 7260

GGTGATTGTT TCACCTTCTC ACGCTCCTCT CAACTTAAGA GACAACAAAC GAGTAGGTCT 7320

GCTAATTAAA AAGCAGTGCC TACTCGTCTA TTATAAGTGG TTTTTCTAAA AATAAAATCC 7380

GCCTTAAGTC TCAGTTACTG GGCTGCCAAG TAAGGTGTAA ACCATTCGGA AAGTGTGCCG 7440

GTATAAGGTA AAACGACCCG TTTCATTCCA CGAGAAATAG CAGTATACmG GATTTTTTTA 7500

TCTCGTTTAT TGTCGGAATA ACGATTAACA TTTGGGTTAT GAAGAATTAC ATTATCAAAT 7560

TCTAGACCTT TTGCCATATC AATTGATAAA ACTTGCAGTG ACGGATGCTC TTCCAATGAA 7620

AGAGAAGCAC AGAGTTCTTT TGTTAATTGT TCTTTTTCTG CAAAAGTTTT AGTAATAATG 7680

GCGGTGCTTT CTTTCGAATC AAAGGTTGTA ATGGAAGTGA CGATTGCTTT AAGATAGTCA 7740

GACGCAGAGT CACAAGGAAT AAATGTCGGC TTTTCTCCAT CATGCCGGAT GGGCACAATG 7800

GTTAATTGCT CTGGTTGAGT GCCTAATGTT TGGAAGAGTC GTGTAATTTC TTTACTTGAA 7860

CGATAGCTGT TTAATAATTG ATAGGTTCCA ACAGAACGGC TGGTTTCGTT AACCAATAAA 7920

TCATGGGcTT CAGAAAATTC AATGGCTGTA TTAAAAATGG CTTGATTTTC ATCACCAGCT 7980

AAAGTAAAAC TAGCTTTAGG GAAAAGTGTC AGCAGTAATA ATAATTGGGC TTCTGTATAA 8040

TCTTGGACTT CATCGACTAG GATAAAAGCC ATTTGCTGGT TGGCGAGGTT TTCAATAAAT 8100

GCATGTTTCA CCAAAATAAG GGCCACAATT TCATCGGCTG TGAAAACAGC TGTCTCTTTT 8160

TGATACGATT TTTCAGTGAA TAAAGAATAA AACGTTTCAA AAAAGGCCCA TTGATCTAGC 8220

CAAGCATCAT TTTTAATAGC GGCAGTGATA GAGCGATATT TTTTCTTCAA ACGTTTTAAA 8280

GCAAAATCAG CCAGATGTTT TTCATTATCT TCGGTAATCA CTGTGCCAAA ATAACGAACT 8340

TGTTCAGCCT CTGTTAAATC TTGGaATTGA TCCAACATTT TTTtAGAATG GGCTTGTTTT 8400

AATAAGTACC GTGTCCAAAG ACTGGTTAAT TTTTCtTTTG TGGCTGAAAT tCGTkGGCGA 8460

nTACtCArGT CaGCTGGTGk yTCyTGaTaA AGTTTaAAAA TgTTTCTGAA GTAAATAAGG 8520

TCTGCTTyTT AAAAGTtAAA GGmCGGAwAA AATGTGGCTG CACAACCGCT TTTTCAGGAT 8580

TGCTTAAAAG AAACGTTGCA AATTCTTTTG AACGAATAAC GTGCTCTTGC TCCGTTGCCT 8640

CAGAGCGAGT AATGCGTTCG AAATAGTCCG TTTCTTCTTC CaTTGGCCAT GATTCTTTGG 8700

TAGTGAAGCG CAAAAATTGT AAAAGAGTCA AATTCAAGGG ATTTCGTTCG CCTAAATTCG 8760

GTAAAACTTG AGAAATATAA TCAATAAATG TTGAATTTGG TGAAAGTAAT AAAATATTAT 8820

CCGCTGTAAT TTCTGCTCGA TGTTGATAAA GTAAATAAGC AATCCGTTGC ATAATCGCAG 8880

ACGTTTTCCC ACTACCAGCA ATCCCGTTGA CTAGTAGGAG TGGCTGCGTT TCGTCACGAA 8940

TTATTTGATT TTGTTCTTGT TGAATGGTTG TCGTAATGTC CTTCATATAA GGAGAAGCAT 9000

CTTCTTCTAA AGAATGCAAT AACACATCAT CTTGAATCGC AATGGTTGTG TCAAAATAGT 9060

TGATTAATTG ATCTTTTTCA ATAATTAACT GTCTTTTCAA ATCAAGTTGT ACAGGAATGG 9120

TTTGTTGATG TGCCGTGTAG GCTGTTTTCC CTAAAACATT GTCATAAAAT AACGAAGCAA 9180

TCGGTGAGCG CCAATCGTGA ATGCGCGTTT CTTCTTCTAA AGAAGTAAAA TCGTTTACGC 9240

CAATATAAAA GTTTTCTGGC TCTTCAGGGC TGTCTTCAAA AGTCACGCCA ATTTTGCCAA 9300

AATAAGGAAC TTGTAATAGg CGTTTCACTT TGTCTAAATT TTTAGCCGTT GCTTCATTTT 9360

TTAAGTTCCA CTGATCAATT TCACGGTTCT TCATTTCCAA CATGGCAAAC GTTTCTAGAT 9420

TATCGGCATA ACTATCAAAA TTCAGCTTAG AATCGCGACC AAATTCCTCT AGCATATGTT 9480

TGCCATCTTT CATTGTGTTC GCCAATTGCT CTTCGTAAAC TTTTTTAGTy CGtTGCAGCT 9540

CATgATAAAC CGTTGTTAAA TGCTGTTGTT CATAAATTTT TTCATCGGT 9589

(2) INFORMATION FOR SEQ ID NO: 257:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7992 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

GGTCCTTATC ACCTTTACCT TCCAAATAAA TCAAAATAGC GTACGcAwAw CTTgGTCGAC 60

GTTGTTCAGT AACTTCAAAC AGTTCkTCTk TCAAATTTTT TAAtACGTCG ATCTGCTCaT 120

TAACCCcGAT TGAAATCATT AGTATCTATT AAACCgTTAT CCATTTCAGA TTGCCTTTTT 180

TCAACCAATG TTTTTAAATG TAGTTCTATA GAAGCTGTTT GAACCATGCT TCTAATTAAA 240

TCTTCTGCTA ACCCTTCACT AGCTTGTTTA TTTTCTGCCA TTGTAGCTAC CCCCTAATTT 300

CTATGTGAAA ACTTGAAAAA TTCTGCAACA GTTGCTTTAA ACACTCCATT GCCATATATT 360

TTTATTTCTT TTAACCCACT CGACAAAATA ACGTTTTGTT GCCATGGCTG TAAATGTGGC 420

AATATACCTA AAAGATTATA ATGCAATAAT CCATCGTCAC TTTTAGGAAA GACACCATTA 480

CCATCTATCC AATTTGATGG ATACTCTAGT TTTATTGCAT CTGTGATATC CACTCCACTG 540

ATTTCCAATC GGAATTCGGT GGCTGTTGAA GGAAAAGTTT GAATGCCAGC ACTCAGACTG 600

TGACTATGAG GATTTGGATT TAATTCATGA GTATGTGGAT TTGGGCTTAA TGAATGGCTA 660

TGGTTATTTG GTGGAATATT CGTGCTACCT ATTTCTAATC TGGTATTGGT TAATTTTAAT 720

GATGCATTTG TAGTTTCTCC ACCTTTAACA GGCATCTCAA ATCCAGAAAT ACGAATTTGA 780

AGCTCCATAA CTTCGTGATA TACAAGGTCC TTACTTACAG GCAAATAAAC AATGGCTGGT 840

GTATTTGAAT CCCCTTGAGC AGTATATTCT TGACCATAGA TTTCAGTGAC CTGATGTCTA 900

CGTTGCATCG TATCTTGTCG ATTTTGTTTA CGATTGTTTT GAGTTCTTTC TGCTACTTTA 960

ATCAATGCAT CATCATAATT CATCATCGTT CATCTGCCTT CCGCTCAATT GATAAATTTT 1020

TACTCAAAGT CAGTCTATTT GTTTCTAAtC CTGTTTCATC AAAATTGTAA TCAATTTTCA 1080

AAATATAGTA CCAATCACTC ATCTTcATGA TtTCTTTTTG ATAATCAGAA CATCCTTCTG 1140

TTATTAATTT ATTGTTATCA TATAGCAAAC GAATTTGATC TCCGACATTA ATGTCAGACG 1200

GTAATyCAGT TGTTGTGATA TCaATGTAGT ATTTTCTTCT TGCTTGTTTT AATCTTTTAA 1260

CCGCTGTTTC ATATGCAGTT CTAGTTGCTT TATTTCTATC TTCATCACTA ATTTTTTTAT 1320

CATTTACTGC AAATGGTGCT AAATCATTAA ACGAATAACT CGCTTCAATA GATATATTAC 1380

TTTCATCTCT TACACTTTGT TCATCAATTA CTGTGTATTC AACaTTATTA TTAGAAGCAA 1440

TCTTTGCAAG ATTGATATAA TCATAACCAC GTTCGTTATT GATACCTTTT CGAAGAACAC 1500

GAACTGGAAA TCCTTTTATC TGTGTATGTG GTTCTAAATA AACTTCACGC AAACTCATAC 1560

TAGACATACC ACTATCTGAT TTTTCTCCAT ATACAGTAGC CATATTTATT ACTTGATCAA 1620

ACTGATAGTC AATTTTAGGT TCGCTAATTA AACGATATGG CGCATTTCTA ACGGATGATA 1680

TCTGAACAGG TTTAGTTTCT CCAAAAGTTC CAAATTCTAA TTTTCTACCA AAATTAAAAC 1740

CAACTCTCCA CCAAATATCA TCTGTTAATT CACATGTTTT TGTTAATGCC TCAAGTTTGT 1800

TTTGTCTACT ATATACATAG TCTATTCTTT TTTGAGAAGA GTTTTGTAAA TAATCTAAAT 1860

GCCATTTATT TGAATATCGA AAATCAAGAG TACTAAAAAT ATCATTAATT GTTCTATTCT 1920

TACATGCTAA ATTGGTGGAT ACTTGTCGGA ATTCCCATTC GCTGATTATA TGATCAAGTT 1980

GTATAGTTAT TGTTTCAAAT TCCTTATCGA CATCATAATC ACTTGTAATC CCATGAAAAA 2040

CTTTATTGTT AATATATATT TTCACTTCTT CTCGACCTTT AAAATATTCA CGATAAGTGA 2100

TTGGCAAAAC GAGTGAAGTT GTAGGCACAT ACATCAACTC ATTACTCCAA GATAATGTAT 2160

TTAAAATCTT ATTACCTCTT GCAACAAGTT TTCCTTTTTC AAAGATTTCG AAGAAAAATT 2220

TTCTATTATC TTCAGCGTAA TTATCTACTG GGGCATTTTC AACTGGTTCA AACATTAAAA 2280

GGTtGTCAAA ATACAACTCC TCTTTTTCTT GCCCGTCAGT TCGAATACCT TTGATAAATA 2340

TCGCTTCaTT TGCTTTCaGC GTAAACGATA ATTCAAATCC TGAcTTTTCG ACATCTTCAr 2400

CATTAGGATA TTTTTCTTTT ATATCATTAC GAATTTTTAT AGGAATGTTA TTGAATTGGT 2460

AAATTAATTT CGCTGTTTCA GCATTATAGA TTTCTATATT TTCCATTGGT AAGTCACTCG 2520

AAAAATGCCA ACCTTTTACA AAAACTTTTC CATTTTTAAT ACCTAATTGA TCAATTTCAC 2580

CAACATTTCG TAATTTTTCC TTAGTATTTT CTGGAGTTGG ATCATCATAA TCCCCTTCTT 2640

TATAATTCAG TAAGAAGTCG AACCATTTTT TTGCAGCTGT TATACGTGCT TGTGTTTTCT 2700

CAACTCCTGC ACGTTCAAAG TTTTTCAGAA ATGCTATTGT TAATTGATCG ACATTTGTAC 2760

CCTTAATAAA ATCTTTTGCA TCAGTCGGAC TAACAACTCC AATCCATTGA CCATGAAACA 2820

TACCATAGTG AATTAATTTA GCTTGTGTTG ACGCCATTTT ATAATCACCA TCAATGCCCG 2880

CTTCTCTAAG TAAACGTTGA ACATAATTCG CTCCGCTTTC CCATGCATTC GGACTTGTCC 2940

ACTGAACTAC TCCATATCCA CCACCGCCGC CAATTTCAGA AATATCGGGG CGCATTCCAC 3000

ATTCAACATC AGCATTACCA CATATAGCAG CAATAGCTTG TATTGTGTAG CCTTCTTTTG 3060

AAAGCACATT TCTAATATCT ACAGCAACTT GTTGCTCTTT ACTTAAAACA CGTTTTGGTG 3120

GAGGGGTGGG ATTATTTCCG CCATCATTTT TTGGATCTGT ATAAATAACA CTTGGCATTC 3180

CACCATTATT TTTTAAAATA AACTGATAAT TATCTATGGA AATTCCGTTA GCTCCGTAAT 3240

TACAATGAAT CACTGACGAA CTGTCAATAA ACATGCCAGT GTGTCCTCCT GCACCATCGC 3300

TAGCTCCACG AACACCCCAA ATAAATATAT CTCCACGTTT GGGATTCGGC GTCTGTTTCC 3360

AACCAATACT TTCTAATGAT CCAAACAAGG TTTCTGTATT GCCCATTGTA CCACTTGGCA 3420

AAAAACCGCC TGCGATCAAT GATAAGTAAA CTGCGCTTGA GCAGTCGTAA CTATTCGGGC 3480

CAAGACGAGC ACTCATAGAG TAAACAACTT TTCCCTTTCT CGCTTCAAAC CACTGAATTG 3540

ATGCATCATG ACTAGGCATT AATTACCACC TTCCTTTGGC CTTTCCTAAT GATTATTAAA 3600

TATAACGAGG TGGTTTAGGT TTCCCATTAT TTTTAAATTC AATTTGTAAA TTGGCTCTAT 3660

CATCAGCAAC AGGGCCATCA TTGAAGAATT CTAAAAATTG TAACCAACCA CTAGAACTAG 3720

AAATCCCCAC ATTTCCATTA TTATTTAATT TTAGTGTTTT ATTAATATTG TCATTAAAGG 3780

ATTTAGTAGT ATCATGTGAT TTTGTATATA TCTCTTGTTC TCCAGCACGA GTTAAATAAA 3840

GGTGTAAAGT AAATGGTACT GACCTAATGT TATTTGTTTT ATATGATAAA TTTTGGATTA 3900

CTACATCACG AACTTGCCAA TCAAAGCTAG CATTTGATGA TGGTCTCATT CCAAAATTAA 3960

CAGTCCCTCT TATTTCTCCT TCTCCTAAAT ATTGGGTCTG TTCAGCGTCT CCCCAATCTT 4020

TCCATCTAAA AATGAAATCA CCATTACCTC GATTAATTGA AACACTCATC CCTTTTGTTG 4080

ATTCTTGAAT CCCATATCTA GTGTTAATTT CCCAATTCTG TGAAGCTATA TTATCGATAT 4140

CATCTTCCAT TTTTGCAATT TTTTTCCAAA TTTTTTCAAT CTGTTCTTCA ATCCAATGAA 4200

TCTTTTTCCA AATTCCACAA ATTGCACAAA TTGTCATCTT CTGTAGATTA TAAAGTGTAG 4260

ATACTAATTT TTTTAAATAT CCTTTCCAAT CACAATCATC ATGAATATCG AGAGCACTGT 4320

CGGCTCTACC TAGTAAACAA TCATTTAAAT CATGCATATC TTGACAATCG TTTTTCTTGC 4380

CATTAAATCC TTTATCATCA CCTAAGCTAG CACATTCTTT GTCGCCAATT TCAAAACAAC 4440

AGTCAACATC TATGCCCCTA TTCTTTAAGT CCTTACAAGA TTTGCAATTT TTAGACATGT 4500

AAATCATCCT TTCTAGTACG TAATTTTATC CTCTTTGATA AATACACAAA CTGGCTCACA 4560

GCAATCGTTT GTTTCTACTA GTAAGCTGTT GTACCCATGC TTTATGGACC ATCCAAAAGT 4620

GTTACAATTC ATCAATTCAA GATTTTCAAT CGGTAACTTA TCAAAATCGC AGCAGTCATC 4680

TTCACGATAG AAAACATCTC CATTTGATTC AATAAGTAAC TCTCCATCGT ATATACCTTT 4740

TAAACGCATT GCGTTACCAT TTAAATGAAT AACTGGATCT TCTACTCTGC CAATAATAGT 4800

CATTGTGAAT ACATCAGTCT CCAAAACGGT ATTACTATAA AACTTTCCAG CAACATAGTT 4860

ATTGCAAAGC TCTTTCTTAC AAATTTTCGT TCCGAGCATC TTATTATATC CCCATAGATG 4920

ACGACCTTTA TTACAGTTAT AAATTATTTG ATAACCACCA CTACAATATT TCCAGTAATC 4980

TTTTTCTATA TCTTTTTGTT TAACACACAG ACTGTCTTCC TTAGTAAGAA AATCACATTC 5040

```
ACAAGAACAT TTTTGACAAT CTTTATGTTT AATATTATTT TTTTGACAAC TAATACAGCA    5100
TCCGTAGATA TCGTCCTGAC AAGTGTCTAT TTCTTCAAAT TCTAGACAAT CAGCGAAAGA    5160
ACACGCATCG TAAGGTTTCA AAAATGTTTT CTTTTCATCT GCATAGTGCC AAATTCCTTC    5220
ATATAATATA AAGTTCAAAT CGACAGAAAA AGTGTTTTTT TCAATGCTGT AAGGTTCTGA    5280
AAAACTTTCT ACAAAAGCAT ATGTCCAAAT TATTCGACGA TTTTCAATTG CCCATAGCTT    5340
ACCAATCTTA GAAAGCTCAC TATATAAGAA TTCTTTATAA AACTGTTTTT GGTCCATAGA    5400
AAACATAGAA TAATTTATCT TTAGTGTTAA TGCTAACTTT TGTTCAGTTG AATACTGTTG    5460
TATTTGTTTT GCATTAGTAT AACTACCATG ACCAAAACTA AATGCTTGAG TGGTAGTTTT    5520
GAATCCTCCA CTCAAATCTG CATCTTGAAT CATATCGTAT GCACCAATTA CAACATCTCT    5580
GAATTGTAAG TACCTTTTTT GAGGATAAAT TGTTCTGCAA TCATACATCA ATTTACACTC    5640
CTCTCAAAAA TCTATTTGCA CGCAAATTGG CAAAATTAGG ATTTCCAGTC ATACTAATAT    5700
TTTGGTTAAT TGTATTATTG TTATTTCTTG TATTTTTTGT CGTATTATAA TATCTATTCA    5760
TTACATTTGA GACAGGTTGA CTATTTCCAA ATCGATTAGT GATACTAGCT AGTGCTCCAC    5820
TTAGATCTAA GTTATTGATC CGTTCCATAA AGTCAATTCC GAAATGATCT ACTGCTTTTT    5880
TACGTTGAAC ATATTCTCCA GCAGTTAACA TTGCTGGAAC ACGATCTGTA CCTTTGGATT    5940
TAAAAATATG TTTCATAATG CTTCCACCGT TTGCTCGATA CTGAGGTATA ATGCCTCCAT    6000
TTGCACGTGT AACTCTTCGA GTCTTCACAT TTTCTGTTAC ATCAACAGTA AGACTAGGCG    6060
CTGTTAATGA ATTTAGATTG GTTTGAATAG AAGAAATTTT GTTATTAATA GATGTAACAT    6120
AACTATCAAT ATTACTATCT AAACCTTGTA GTGCATTTGT AAATCCTGAT TTCAGACTAT    6180
CCCCAAATCC TTTACCAACA TTATTAAACT CGGTATCTTT ATTTCTTAAA TTAGTTATTA    6240
AATCATCAAT AACTTTTTTG ATTTTAGAAG GAACATCAGC ATTTTtGAAT CCATTAATAA    6300
CTTGCTGTCC ATATGATTTC CCAATTGGTT CAAATTTAGA TTCTAGACCT TGTAACGTAA    6360
TAAGTAAGTT TTTAAAAGAT TCGACTAGCT CGGGTATACT ATCCAGTCCT TTTGCACTTG    6420
GAAATGTGTT CATTAGTTCA ATGATTGATT TCATTCGGTT AACGTTGTTT AATACACTCG    6480
TCCAATCAAA TGGTAATGCG TTTAATACAC TTAATTTATT TGATATCGTT CCTAACTTAC    6540
TAACAGATGA AACAATCATA CCTAACGTTT CGCTGCTTAT AACGCCTTCA GATATCCCTT    6600
GCCACTTACT AATGGCCATC TCTGTTGCGA CATATTTAAT ACTTTTTATA TTATTACTCA    6660
ATTCATCAAC ACCCAATGTT AATTGAGGAA TCTGATTAAT TCTATTTGCT AATGTTATCA    6720
GTCTATAAcT GTAGTGTCGA TTTTATCAAT GACATCTTCA CCAATCATCG TGTTTGAATC    6780
AGGAAAAGTA TCTATCGTTA ACCTATTAAT AACACCTTTA ACAGCGGTAA TTGTTGmATG    6840
GACTTTATCA TTTTCTAGTG TGATACTAGC TAATCCaTTA ACTCTATTTG CTcAATTGTA    6900
TAAAGTTTAT CAAAACTAAA TCCAAAGCTT TTATAAAGTC AGCTTTGACT ATTCCTTCAT    6960
TGTAATCTCG CCAAGAGCTA GCCTGAAACA TTCTTAATGC CCCTTTGGTA CTTGCAATAG    7020
```

TGGAATTTAT TTTGtCAGTT TCGATATTTA TTCCCgAAAG ACCTGTTAAA CGATTTGCTA 7080

CTTTAGTGAC CCATAATACA GCTACATCTA ATCCGCTAAG CATTTCCCTA TCAACAAATC 7140

CATCATTAAA ATCACTCCAT GaATCGrTTC TTAACATTTC GAATAGAnCC nTTTGGATAT 7200

TTGGAATAAC AGATGGTGCC TTACTGnTTC CTAATGGTAA AnCGAAAGTT CATTTAATTT 7260

ACTGCCCACT TTATTCAACC AGTAAATTGA TGTGTCTAAC TGGTCCAwTA ATCTTTTACC 7320

AACAATTCCT TCATTAAAAT CGGTCCAACT ATCTGCTTGC AATAATCTTA AAGCACCTTT 7380

GATATTTGCT ATTATTCCTT GTAACTTGCC ACCATCAAAT TCAAAATTAG AAAGTGTTCC 7440

AAATTTTTTA GCTATACTAC TTAACCAATA AACTGCAGTG TCTAGTTTAT CTAATGTTCC 7500

ACGTTCAACT ATTCCCCCAT CGAAGTCGCC CCAAGATTTA ATTTGCAACA TACGTAATGA 7560

GGATTTTAGA TTTGCTATCA TGGTTTGAGT TTTTCCTAAG TCAAAAGAAG TTTCTGATAG 7620

TTTAGACAGC TTTTTGTTAA TGCTATTCAA ACCATCCACA CTTTGGCCCG TCATATCTAT 7680

AAAATCTTTT GGCACAACTC CACCAATGAA ATCACTCCAA CTTTCATTAG TTAGCATACG 7740

TAATGCACCT TTAATATTTT GTAGTACATT TTGCACTGAA CCCGTTTCTA ACTTACTTTC 7800

TGaTAATGCT GcTAACTTTT TGcTTAATTT ATTTAACCAG AATATAGTAG TATCTGCTTG 7860

ATCAATAATT TTAGATGAAA CgACTCCaGA TTTCGCAATA TCCCaATCaG CAGGrtCTAA 7920

CTTtTCyAAw AATATCTGnn ATAACTGTAA CTTTACTTGG TAAATATCAG GATATTGGAA 7980

AATTTAATnT TC 7992

(2) INFORMATION FOR SEQ ID NO: 258:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8991 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

AAGAAAACCT TGAATTTGTT CAAGGGCAAT TGCAACAAAA TTTACCAGGT TTAACAATCA 60

AATTAAAATC ATTGCCTTnG CAAAATCGTT TGGATTTACA AACAGCAGGT AATTACGATT 120

TAGCCTTTGG AACATGGACG CCTGATTATG CTGACCCCAT TAATTTTTTA GAATTTTATG 180

ACTCGAAAAG TGGCTTAAAT ACTTCTGGTT ATAACGATTC GGCGTATGAT GCGGGTCTCC 240

AAAAAGTTAG AAAAGATTAC GCCAATGAAC CTGAAAAAACG CTGGAATGAA TTACTAAGTT 300

TAGAAAAAAC ATTAATTGAA AAAGATGCTG GTGTTTTACC CTTATTCCAA GGAGCCATTG 360

GTTACTTGAA ATCAGATCGC TTACAAGGCT TGCAAGTATT TTCTTTTGGT CGCACAGTAT 420

CCTATCGTTT AGCTTATGTA GAAGAATAAA TTGACTTCTG AAAAGTCCTA TCTTTTGAAA 480

AAGTCTGCTT TCTGTTTTAT CCTTAAAGAA GGAAAATTAA CTAAAAAAAT CAACTATGAA 540

AGGTAGGTTT TTCAAATGTT TAGCTTTCTA TGGGCATTAA TTGTCGGAGG AATTATTGGC 600

```
GCAATTGCTG GCGCAATTTT AGGAAAAGAC GTGCCAGGTG GCATTATTGG AAATATTATC     660
GTGGGCTTTT TAGGAAGTTG GGTAGCATCA CTCTTGTTGC CTTCTTTAGG ACCAGTCATC     720
GGCGGGTTTC CAATTATCTC AGCGTTGCTA GGAGCGATTA TCTGTATCGC TATTTACTCT     780
TTTATTGTAA ACCGTAAGgC GTAAAATTGT ATCAAAATGA TATCTTGAAA AGCATCATAA     840
AACTTGTGAA GGTTGATCAA ACCATTACTT TTGATGTCTT TTCTCTCATA TTTTAAGTCT     900
ATGACAAAAA TCTCTTTGGA TTTTTGTCAT AGACTCTTTT TTTAAAAATA TTTATTATGA     960
GAATTTCCgC GTTAGACTGC TTATTTTTCA TTTATTATTC ATTTTTGATA AAGGAAGTTT    1020
TGTTGGATAA AGTGAATAGC TcAAGGCTCA ACAAAAATAA rGArAAAATT AATATGTTAA    1080
GTTCAAGGTG AGTGAAATTC TGCTCTCATA ATACGCCGTT TTATTCATTA AACGATAGTT    1140
GAAAAAAATA AGTGCCCTTG TTATTCTTTG AATTAAACTA ATGGGGGCGA AAAGCATGGA    1200
ATCGTTAAAA AGTAATCTTT TTTTGACCAG AGAAGCaTAT TTGGAAGCTT TTGAAACAAT    1260
CcATGACACC TTTAAAAACG GCTATTTTAA CATCAGAAAA ACCTGGCGTG AATTTTGGGT    1320
CCAGTGGTGC TGTCTATGAT CAGCAACATG CTGAAATGGA AGCGCTCATT CGACCATTAT    1380
GGGGAATTGG ACCGTATTGG GTTACTTGTA AAGATGATGT TTTAAGAGAT GCTTATGTGG    1440
AAAAATTAAT CGCAGGAACT GATCCGTCAT CCCCAGATTA TTGGGGCGTT ATTGTTGATT    1500
ATGATCAATA TATTGTGGAA GCAGCAGCAT TATCTTTAAC TTTATTGCTT CATAAGACAT    1560
ATTTTTGGTC CTGTTTTTCT GAGAAACAGC AAGAAAATAT CATGGTTTGG TTAAATAAAG    1620
CACTTGGTTG TAAAATTCCT AAAAATAATT GGACGTTTTT CAAAGTGCTC ATTCGTCTGG    1680
CACTCGAGGG ATGTGGACAA GTGATTAACC AAGCGGAACT AGAGGAAGAA TTGGCGTTGA    1740
TTGAATGCAT GTATTTAGGT GATGGTTGGT ATATGGATGG TAAAACTACG CAAAGAGATT    1800
ATTATATTTC GTTTGCTTTT CATTATTATA GTTTGATTTA TGTCAAATTT ATGCGAGAAA    1860
GAGATCCTAT TCGTGCGAAA CGTTTTACAG AACGAGCCGT TATGTTCGCG CAGGACTATC    1920
TGTACTATTT TGATGAGGAG GGCGAAgcGT TACCTTACGG ACGTAGTCAA ACGTATCGAT    1980
TTGCTCAAGG AGCATTTTTC GCAGCTTTGA TATTTGCAGA AGTAGAGGCT TTACCTTGGG    2040
GACAAATAAA AACGCTGTTA GCAAAACATT TACACTGTTG GATGCAACAG GCTATTTTCA    2100
CTTATGATGG TCGGTTGTCC ATTGGCTATC ATTACGAAAA TTTAGTAATG GCAGAAGGTT    2160
ATAACGCACC AGGCTCTCCG TATTGGGCTT TTAAAACATT TTTGTTGCTC GCAGTGGATC    2220
AAGACCATCC GTTTTGGCGT GCTGAACCGC TGGCGATAAA ACGAGAAGAA AAGCGTTTGA    2280
TAGAAAAGGG AAACATGTTA TTAATCCATC CTAAAATGAA AACGCATGTC CTCGGCTATC    2340
CTGCAGGACT ATGGCTTGAA AATCAAGCAC ATGCTGCGGC TAAATATAGT AAATTTGTCT    2400
ATTCTAGCCG CTTTGGTTTT AGCGTTCCCA AAGCTGGTGC GCGTTATGAG GAGGGGGCTT    2460
TTGATAACGT ATTGGCTGTT TCTCGTGATG GTGAATATTT TCGTGTTAAA GGGAAAGTTC    2520
ATGATTTTTA TTTAACAGAG GAGCAAGTTT ATTATAAATG GCTACCTTTT CAAGAAGTAG    2580
```

AAATTGAAAC AACCATTTAT CCTTTTGGAG AATGGCATGT TAGAGTACAT GAAATAGTGA 2640

CATCGGTTCC ATTAAAAGTG CGCGAAGGAG GATTCAGTCT ACCGTTACTT GGAAAATTTT 2700

CAGAAAAAGA ACTTGGTAAA GAGTGGGCTA GAGTCGATTC AGGAGAATTG TGTTCGCAAA 2760

TAGTTGCAAT TGAAGGGTAT GACCACGCAG GTATTCTTCA ACCTGAACCC AATACATCGT 2820

TGTTTTTCCC AAGAACAAGT TTACCTTATT TGCAAAAAGA GCTACCAATA GGAACGCATC 2880

GCCTGATTTG TTTGGTAGGT GGTGTGACAA GGGAGGAATA AAAAATGATA AAAATCGAAT 2940

TGAAAAATAA AGACAATCAG ATTATTACTG GCCGAATCGA TCACGAGAAG CCAGAAGCGC 3000

CATTCGTGGC GGAGGGAGAC ACGTTAGCGG TATTAGCAAT TAAACAGTAC CTTTATCAAG 3060

TTGGCGATAA AATTGTAGTG ACTTATACAG GGGCTACCCC TTATTTGATG GTTCAATTAG 3120

ATGAAACTTT GGCTCCTTCA TTGATTTATC TAAAAGAAAA AGTTTGGGAA TATCAAATTC 3180

CATTAGAAGA ATCTGCTCGA AAAGCTTTTG TTGATACGGC CTTTCATTCA GCAAGGCACC 3240

AAATAATGGT TCGACAAGCG CACACGTTTG AAAAAGACAA TTATCAAAAT CTTTCTATGA 3300

ATACACATGA CCAAAAAGAA TATTCAGGAG CCTATCCTCA TGCGTATGCA AACGTTGAAA 3360

CGAGAGACGA TGCTGTTTTC TTTGCGAAAA ATGCGATTGA TTCTAAGTTT GGTAACCTTT 3420

CTCATGGCTC GTATCCATTT GCTTCTTGGG GAATTAACCA AAAAAAGGAG GCACAGTTAA 3480

CGATTGACTT TGGCAGACCT GTAGAAACGG ACTGGATCCG CTTGCTATTT CGGGCGGACT 3540

ACCCACATGA TAGCTATTGG GAGGAAGTCA CTTTAGCTTT TTCAGATGGT GAGGAACGAG 3600

TAGTTGCCAC AACGAATACC ACCTGTTTTC AAGAAATAAA GTTCCCAATG AAAACAACAC 3660

AAACTTTGAC ATTAAAATGC TTACGAAAAG CGCTGGACTC TTCTCCATTT CCAGCGTTGA 3720

CCCAAATTGA AGTATTCGGT CGGAATAAAT AACAAGAGGT CGGGACAGAA CCGTTTAGCT 3780

CCGAGAAATA AGAAGAAATT TCCGAAAATT GATCTTTGAT TTTTGGAGAC CAAGTAGGTA 3840

TTGTTCATCA TCGAAtCACT ACGTTCTTCG TGATTCTCAA CAATTTCGGC TTATTTCCGA 3900

AGGACCTTCA ACTCTTAGAG CTTTAGCTCT TAGAAGTTGT TGAGATCGGA GCAACGCGTA 3960

GTGTTGGTTC TGTTCCCGCC GTTTATCAGT TTTTGAGCGT GGAGCAAAAA TCCAAAGTGA 4020

TTTTTGTCCC ACGCTCTTGT TTTTAAGAGA AATAaTCAAG AAAGGTCGCT TTTTATTTTC 4080

TTTTaCGTTA TAATACAATA AGTTAATATG TTAACAAGTT AAGGAGTGAT GGAGATGTGT 4140

AGTCAAGaAA ATAaACAAAG ACTGATCTCT CTAGGAGAAA ATGTCACAGA TGAATGGATT 4200

CAAACGCAAT TAAATGAATG CGTTTCAAAA ATTGCTAAAA ATATGCAGCG TTTCGGAACA 4260

CAGTTTCCTT CAGCTTGTGC TACGAATGGT CGATATCGTC TCAAAGCTAA CGATGATTGG 4320

ACCAATGGCT TTTGGACAGG CATGCTTTGG CTGGCATATG AGTGGTCTCA CCAAGAAAAA 4380

TTTCGTTTAC TGGCAATGGA AAATATTGAC AGCTTTCAAC AACGATTAGA GGAGCATTTT 4440

GTTTTGGACC ACCATGACAT TGGTTTTTTA TATAGTCTAT CGGCGGGTGC TGGGTATAAA 4500

ATTACCGGTG ATGAGCGCTG TAAACAAGAA GTGCTAGCTG CCGCTGAAGT ATTGTCAGCA 4560

CGATTCCAAG AAAATGGCTC CTTCATTCAA GCGTGGGGCC ATTATGGTGA TCCCAAAGAA 4620

TATCGATTAA TTATCGACTC ATTAATGAAT TTACCGCTTT TATTTGAAGC AACTAAATTA 4680

TCTGGAAATA GGGTTTATGA GGAAATTGCC TCACGTCACT ATCATACGTT AATGAAAACG 4740

GTGATTAAAT CCGATGCGAC AACTTATCAT ACCTATTATT TTGATCCAGA AACAGGCAAA 4800

CCAAAACATG GAgcTACGCA TCAAGGACAT AGTGATACTT CCATTTGGGC ACGAGGCCAA 4860

AGTTGGGCAA TTTTAGGGAT TCCCCTAAAC GAAACATTTC TCCATTCCCA ACCATTTCCA 4920

GAAATGTATC CACAAATAGT GGAGGTTTTT CTAGCTCATT TGCCAGCAGA TTTAATTCCC 4980

TATTGGGATT TTGATTTTAA TGACCAAATG CCATCTGACA AAGATAGCTC GGCGCTGGCA 5040

ATTGCGGCAT GTGGTTTATT GGAAGCAGAC AAACTACAGG CTTTTCCTCA AGCGAAAGAG 5100

CTTGCCAAAG GAATGATTTA TCAATTAGGA GAATATTATC GTACGCAAAA TGATTCCGAA 5160

AATGAAGGAC TGCTACTCCA TGGTGTTTAC GCACATGCAG AAGGAAAAGG GATTGATGAA 5220

CCCAATTTAT GGGGCGATTA TTTTTACATG GAAGCCTTAA TGCGTTTAGC AAAACCATCA 5280

TGGCAAAGAT ACTGGTAAAG GGGAAAAATG ATGCAAACGT TTGAGATTAA AGAAGACTTT 5340

TTATTAAATG GACAACCAAT AAAAATTATT AGCGGGGCGA TTCATTACTT TCGAATGACA 5400

CCTAGCCAAT GGGAAGATAG CCTCTATAAT CTGAAAGCTT TAGGAGCTAA TACAGTAGAA 5460

ACCTACATTC CTTGGAATAT TCATGAACCA GAAGAAGGCG TATACGATTT TGAAGGAATG 5520

AAAAACATTG AAGCATTTGT CCGCTTGGCA GAAAAACTGA ATTTACTAGT TATTCTGCGT 5580

CCTTCGGCTT ATATTTGTGC AGAATGGGAG TTTGGTGGCT TACCAGCATG GCTTTTGAAA 5640

GAAAAAGGTG TGCGATTACG TTCGACAGAT CCTATTTTTA TGACTAAAGT TCGTAACTAT 5700

TTTCAAGTAT TATTGCCTAA ACTAGCACCG CTGCAGATAA CACAAGGTGG ACCAGTGATC 5760

ATGATGCAGG TTGAAAATGA ATATGGGTCC TATGGCATGG AGAAAGCGTA CTTGCGACAA 5820

ACCAAGCAGA TAATGGAAGA ACTGGGGATT GAGGTTCCTC TTTTTACATC TGATGGCGCT 5880

TGGGAAGAGG TTTTGGATGC GGGTACGTTA ATTGAAGAGG ATGTTTTTGT GACAGGAAAT 5940

TTTGGTAGTC ATTCGAAAGA AAATGCTGCC GTTTTGAAAA AATTTATGAC TCGTCATGGT 6000

AAAAAATGGC CGTTGATGTG CATGGAATAC TGGGATGGGT GGTTCAATCG TTGGGGCGAA 6060

CCAGTGATTC AGCGGGAAGG TACCGACTTG GCGAAAGAAG TAAAAGATAT GCTGGCAGTG 6120

GGTTCTTTAA ATCTCTATAT GTTTCATGGA GGAACAAATT TCGGCTTTTA TAACGGGTGT 6180

TCGGCGCGAG GTGCAAAAGA TTTACCTCAA GTCACTAGCT ATGATTATGA TGCTTTATTA 6240

ACAGAAGCAG GAGAACCAAC AGAAAAATAT TATGCCGTTC AAAAAGCAAT TAAAGAAGTA 6300

TGTCCAGAAG TTTGGCAAGC ACAACCACGC ACGAAAAAAC TGGGGAATTT AGGAAGCTTT 6360

CCAGTAACTG CTAGCGTATC CTTATTCGCA GTAAAAGATC AAATGATGAC ACCGAAGACC 6420

ACCACGTATC CATTAAGTAT GGAAGAAGCT GGCTCAGGCT ATGGCTACTT GCTTTATTCG 6480

TTTGATTTAA AAAATTATCA TCATGAAAAT AAACTmmAAG TGGTCGAGGC CAGTGATCGT 6540

TTACACATTT ATGTAGATGG AGATTTAGCT GCTACCCAAT ATCAAGAGAC AGTGGGAGAA 6600

GAGCTACTCA TTTCAGGGCA AACTGAAAAA GACACACTTG CGTTAGATAT CCTAGTGGAA 6660

AACTTAGGTC GCGTTAATTA TGGCTTTAAA TTAAATAATC CCACACAAAG TAAAGGAATT 6720

CGAGGTGGAG TGATGCAAGA CATTCATTTT CACCAAGGCT ATCAACACTA TCCCTTAACT 6780

TTTTCACAAG AGCAATTAGC CAAGATTGAT TACACAGCTG GAAAAAATCC ATTGCAGCCT 6840

TCGTTTTATC AAGTAACATT TGAATTAGAA CAGCTAGCAG ACACGTATAT TGACTGTCGA 6900

GGCTACGGCA AAGGATTCGT TGTGGTCAAT GGCCATCATC TAGGTCGTTA TTGGGaAATA 6960

GGACCCATTC ATTCGCTGTA TTGTCCCAAA GAATTTTTAC AGCAAGGACA AAATGAAGTA 7020

GTAATTTTTG AAACAGAGGG AATTGAAATT GAGTACTTGA AATTTACGAA TCAAGTAATC 7080

ATAGATTAAA AAAATAGAAG TAGAAGCTGA AATTTATTCG GCTTCTACTT CTATTTTGGT 7140

TTTGAAATAT TTCCAATGTT TATAACTAAT AATATATTCA GCAACACTAC CATCTGCGAG 7200

ATAGGAATGT TTAACTTGTT TAACAGCGGG TTCACGAAAA CTTAATTGAA GTAAATTCAA 7260

TAACTCAGCA TCTTCGGGAA AGACAATTTC ATTTGTTTCT ACAGAAGCCA ACGAGAAGAG 7320

GTCTACTGAA AAATCTTTAC GGACTCGTTC ATACACACTT GCATAAGCTG TGAGGTCCTT 7380

TGAGAGCGGC TCTTTAACAA GCTTTTTGGG AAGATGTGTG ATATGTACGA GAAAGGGTGT 7440

ATCTTCAGAA TAGCGTACAC GCTTGATTCT ATAATAAGAA GCGTTAGCTG GTAAGCCTAA 7500

TTCCTTTAAA ATTTCGGCTT GATTATCTTC TTCAATTGAA AGGACTTTAA CTTTTTCAGA 7560

GTCTAAAGAA TGCAACTCGA TATCAGAAAA TTTTACATTT TGAGAAACTT TTGATTTCGA 7620

AACAAAGGTC CCTTTTCCTT GGATACGATA CAAGTAACCT GCACTGGTCA GCTCATTCAG 7680

TGCTTTGACA GCTGTGATAG AACTAACAGC ATATTTTCGT TTAATATCGG CTTCAGAATA 7740

AAATTTATCA CCAGGAATAA AGGTATGATT TTTTATTTCA CGTAAAAGGT CTTGTTTAAT 7800

TTGTTCATAT TTAGGTAACA TGCGATTGCC TCCATTCTAG TAATCGAATG AATTAAGCAT 7860

AACATGTTAA CTAGCCTTTA GCAAaCACTT TGAAAAAGAG GAAAGACAAT TTTTATTTAA 7920

gtTCAAAAAA aAATaATATA TTAAGAaTaG GTTTACGGAT AACTTAGTAT GTGTTATAGT 7980

AAGTTCGTAG AGAACATAAC ATGTTAAGTA GGGTCACCTA CATCTAGGGA GGGAAATTCG 8040

TGGAAAACAT TGTACTAATT AGTCATGGTA GCATGGCTGA AGGCGTCAAA GTTAGCTTAG 8100

AAATGATCGT TGGTCGGCAA GAACACGTGC ATACAGTCTC ATTACGACCA GATAGCGACA 8160

ATCTTCAATT TGAGAAAGAG CTAAATGAAA AAATGAAAGC CCTTAACGGT ACGACACTGA 8220

TTATTGCTGA TTTACTAGGA GGAACGCCAT GTAACGTGGC AACTAAGAAT TATTTGAACG 8280

TTGACGGAGT GGAAATTATT GCAGGGATGA CGCTCTCTGT TGTGATTGAA GCAGTGGTTA 8340

ATCAACAGGC TTCGATTAAA GAATTAGTTT GTTTAGCACA AGAAAATATT GTTGATGTAA 8400

AGGCTGGAAT GAATCAAGCG GAACaAGAGA TTTCAGAAGC CAGCAAAGAA AAAGAGTTAA 8460

GTAACTATAG TCAATATGCT GGAAAAGAAA ATATTGTAAA CACACGGATT GATGAACGGT 8520

TAATTCATGG ACAAGTAGCG GGAATTTGGT CGACAAGTCT TAGTACACAG CGAATAATTG 8580

TAGCCAATGA TGAGGCGGCT ACAGATCCTT TACAAAAATC ATCTTTGAGA ATGGCAGCGC 8640

CTTCTTCGAT GCGTTTATCT GTTCTTGGAG TAGAGGCTGC CGCTAAAAAT ATTCaATCaG 8700

GAAAATATGG TAAACaGCGT TTATTTTTAT TATTTAAAAA TCCGAAAGAT GTGCTCCGCT 8760

TTATCGAAGC GCAAGGACCT ATCAAAACTG TAAATGTTGG GAATATGAGT TACAAAGAAG 8820

GAGCACGCGA AGTAACAAAA AGTATTCAAG TTTTACCAGA AGAAGAACAA ATTTTTGAAA 8880

CAATCGCTTC AAAAGGCGTC ACTGTGAcAG CCCAATTAGT TCCAAACGAT CCAGTCGTTG 8940

ATTTnATGAA AAAATTACGC GGCTAAAACT GTCAGACTAA GAAAGAATAT n 8991

(2) INFORMATION FOR SEQ ID NO: 259:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4245 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

ATAnAAGTAn ATAmAATGGC GATTATTGCT GTCTGAAAAG ATAAAAGAAT TGGTCGAACA 60

TCCATTTTCC CACCTACTTT GTTAACTTAA AGCCAATATT yTCAAAATAT TTTCTGCATT 120

GTTGACTCTG TAAAAAATTT AAAAAAGCAT CTGCTGATTT TTGTTTCTTA GAGGCGGCAA 180

CTTTACCAAC TGGATAGATA ATTGGCTTTT TCAAAACAGC TTCAGGCATG GCCGCAACAA 240

TCGCTACTTT TGAATTGGTT GCTGCATCTG TCGCATAAAC TAAGCCAGCT TCTGCACTTG 300

CATTAGCTAC CCATTCAAGG ACTTCTGTTA CATTCGTGCC AAAGCTTGCG TGTTTTTCTA 360

CATAAGACCA AGCGCCTAAA GCTTTTAAGC CTTCTTCGGC ATATTGACCA GCTGGAACAC 420

TTGCAGGATC ACCAATTGCT AtCATTTGGG CTTTTTTTAA ATCAGAAAAA TCATGCCACT 480

TTGCTTGATC TTGGTTAGGC ACAATAAGAA CGAGCTGGTT TTCCAATAAA GGAACGACAC 540

TTTTTTTATT AATTAGTTTT TCTGCAACCA ATGCATTCAT TTGTTTTGTC GAAGCTGAGA 600

AAAATACATC GGCTTTTAGG CCTTTTTCAA TTTGCATCTG TAATTTTCCA GAACTATCAT 660

AGGTTCCAGT TACCTGAATA TCTGGATGCT CTTTTTCAAA GGCTGGAATA ATTTTCTTCT 720

CCATGACTGA TTCTAAGCTA GCTGCTGCCG CAAGGGTTAT TTCTTCTTTT TTTGTTGAAA 780

CTGTTTCTTT TTGCGGCTGA TTTGTACATG CAGCAATTAG GGCAACTAAA CTGAATAAAC 840

TAAGCAGGAA CACAGCCTTT TTTTGTTTAG ACATTTGAAC GCCTCTTTTC TTTTATAAAT 900

ACCCGCCATA ACTCAAGTCA ATAATTTCAG ATTTTTTAAT CACTTCTTTG GCCATTAATC 960

TTGGTAACAA CAAATCAAAA ACCGTTTTTT CTGAAAATAA AACACCACCT GGTAAGCCAA 1020

TGATCGTCTG CTCTCCTTTA TAAGCCAGCA AAAACATTGA GCCTGGTAAG ACTGGTGTCC 1080

CATGGGTTAC GATTTCAGCA CCTGTCTGCT TAATTGCTAA CGGTGTCAAA TCATCAGGAT 1140

CTACACTCAT ACCACCTGTA CAAACAATCA GATCAACCCC TTGTGCTAAC ATCTTTTTGA 1200

```
TCGCAACAGT AATTTTTTGC GGCTGATCAT CGACAATTTC TTGCTTGACA ATCGTAACTA    1260
AAGGATACGC GGAGAACTTG GCTTTTAACA CAGGGTAAAA AGCATCCTCA ACCAATCCTG    1320
TGTAAACTTC ACTGCCAGTA GTTACAATCC CGACCCGTAT TTTTTTGAAA GGTTTCACAG    1380
AGATTAGTGG CTCAGACGTG CTGAGCTGGC GTGCCTTTTC TAATTGGCTT TTAGAAATTG    1440
TTAGTGGAAT CACTCGAAAG GCTGCAACTT TTTCATCTTT CTGAACTTCT ATATTGGTAA    1500
CTTTAGTTGA CAAGGCGAGA CCTTCAATTT CATTTAAAGC AATTAATTTT TCTCGATTGA    1560
CTTTTAAAAG ACCTCTCGTA CTAGCTTTAG CAACAATTTT TCCTTCATGG ACTTCACTCG    1620
GCTGACATGA GTCATCTGTT AACAGGTTAT ACAAAAATGT CGCkGCCTCT TCTTCATGTA    1680
TTtCTGCTGG ATCCAAGGAA ACAAACAGAT GTTTTTTCCC TAATGATAAA AGAATTGGAA    1740
TATCTTCTTG AACAACTATA TGTCCTTTTT TAAAACACAC TTCCkgGTCT TCCCATAGTC    1800
AATTCGCGTA AtATCATGAC TCAATGGAAG GCCAACTGCC tCTTCTGTTC GAATTGTTTT    1860
CATGGTTTAT TCCTCATCAT GTATGATTTT TTTATTTGTG ATTAAATCCA AAGCGTGGGG    1920
CAAAACAGGC AACAAAAAAT GCAAATGTTC TGTTACACTT TTCGGGCTAC CAGGCAAATT    1980
AACAATGAGC GTGCGTCCTC TTTGAACAGA AACGCCTCGA CTTAACATAG CAAATGGTGT    2040
TTTTTCCAAG CTAAGTTGAC GCAAGCCCTC ACTGATCCCA GGAATGATTC GCTCTGCTAC    2100
ACGTAAGGTA GCTTCTGGGG TCCGATCCCG TTCACTCAAC CCTGTGCCGC CTGTCGTTAA    2160
AATCAAATCA CAGGTCTTAG ACCACTCTTG CAATAGCTCT ACTAATTGAT TTTCTTCATC    2220
TGGTAAAATG GTTTGTTTGG TTACCTGGTA GAGCGCAGGT AACTGTTCAC GAATCAAGGG    2280
CCCCGATTGG TCGACAGCAA GTCCTTGAAA GCAACGATCA CTTAATGTTA CAATGCCAAC    2340
TTTAAACAAC GTCTAACTGG TCTCCTTTTT TAATATGACC TTCTGTCACA ACAACTGCAA    2400
AGATTCCTTC TCTAGGCATG ATGCAATCGC CCACTCGCGC GTAAATCTGA CAGTGTTTAT    2460
GGCACTCTTT CCCAATTTGA GTAACCACCA ATTCGGAATC GCCTATTTTC ATTTTCGTTC    2520
CAACAGGTAA TTGACGTAAA TCAATCCCTG AGACGATTAT GTTTTCCCCA AAAGCACCAT    2580
TGCCAACCCG AGCTCCTTTT TCATTGAAAG CCGTTATTTC TTCAAAAGAT AACAAACTGA    2640
CTTGTCGATG CCAATTACCG CCGTGCGCAT CATTTTCTAA TCCGAAACCT TTAATTAAGT    2700
TGACTTCAGG AATTTCCTTT TTCTGTGTGC CTCTTCGCGG GCTAATATTT ATTGCAATAA    2760
TTTCTCCTGT TTTCATACTT ACCCTCCTAT TTGTGACATA AATCGTTGTT CTCCTTGTTG    2820
TTCCAAGAAA AGATGTTTTT CAGGCTTACA GGCAATGCCA CTCTTGATAA TCCCTAACAA    2880
TTCTGCTGTT TGTTCCATAG CTAAAGCTTC CTTTAACGAG TAGCCGTCTG CAGAATGCAG    2940
GCACGTTTTT AAATGGCCGT CTGCAGTAAT TcGGATTCGA TTGCACGTCT CGCAAAAACA    3000
ATGCCCTAAT GCACTAATAA AACCTATTTT CCCTTTAAAA TCTtTTAAGG CGTAATAACT    3060
AGCAGGGCCA TTGCCCAATG GTTTTTGGTA TGGTTGTAAC GTTCCATAAT ATTTGGTTAA    3120
AATGGCTTCT ACTTCTTCTT GTTGCTTACC TGGATGCTCC TTCCCTAATC CAATTGGCAT    3180
```

CATTTCAATG AAACGAAGAT GCACAGTCTC TTTCTTmGCG ATTTCAGCCA GTyCTACAAT 3240

yGTTGCTTCA CTTAGCTCTT TCCTAGCTAC AGTATTAATT TTGATATTAG GTAATCCCAC 3300

ATCAATCGCT TGTTCTAAGC CCGCCAGAAC ATTCCTTAAC TGACCAACTC TTGTTATTTC 3360

ACGAAATTCT TCTGGATCCA AAGTGTCTAA GCTAATATTA ATGCCGTCCA AGCCCGCCTC 3420

TTTTAATCCT TGAGCCTCAC GAGCTAACTT CATGCCATTA GTCGTTAAGG TTACTTTTTC 3480

GATGCCTGAG ATTTGCTTTA TCCTTTTAAT TAAAGAGAGT AAATTCGGTC TAACAAGCGG 3540

TTCACCACCT GTTAATTTAA CTTTTTTGAT GTCTTCTTTC GCTAATATCC TCAATAAAAA 3600

AATAATCTCA TCATCTGTTA GAAGCTGCTC TTTTTTTAAA AAGCAAAGAC CTGTTGCTGG 3660

CATGCAATAC GTGCAAgTAA GTCACAACGA TCTGTTAAAG ACAAACGAAC ATAATCAATT 3720

TCACGGTTAA ATGCATCTTT CATGCTGTTC ACTCATCACT TTCCTTTGAT TTAATTGAAG 3780

AACGAATGAA ATGACCACTA CGTCCACCAT TTTTCTCTAA TAAACCAATG TTTGTCATTG 3840

TCATTCCTCG ATCAATTGCT TTACACATGT CATAAATTGT TAAGAGTGCT GCCTGCACAC 3900

CTGTTAGCGC TTCCATTTCA ACGCCAGTTG AACCCACTGT TTTTGTAAAG CAACGTACTT 3960

CTAAACAAGT TTGATTGCGC CAATCAAACT GAATTTCACA CTTTGTCAAA GCAACCAAAT 4020

GACACAAAGG AATCCATTCA AACGTTCGTT TTGCGGCGGT AATCCCAGCT ACTCTTGCTA 4080

CCTGTAATAC TTCACCTTTT TTAATCATTT GGGTCTGGAT CGCTTCTGCT GTCGCTTCCA 4140

ACATGTGAAT TTCTCCATAA GCGAGGGCTG TTCTTTCAGT GTCATTTTTT TCCGAAATAT 4200

CAACCATGTG CACTTCTCCt TGTGTATTCA AGTGCGTAAA TTCCT 4245

(2) INFORMATION FOR SEQ ID NO: 260:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4013 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

TGAAATCAGC ATGGTACGTT TAAACCAAGA TTGTATCCGC TTTAAAAAAC AAGAATGATT 60

GACAAAAATT AACTTTACAC TATAATATTT TTAATAGATA ATGTTGAAAA TTGAAAAAGG 120

CAATGGTAAA AGGGGCTGTT AAAATGTCAA AAAAAACAAT TATGTTAGTT TGCTCCGCAG 180

aATGAGTACT AGTTTACTTG TTACAAAAAT GCAAAAAGCA GCATTTAATC AAGGCTTAGA 240

TGCCCATATA TTTGCGGCTT CTGCATCGGA AGCAGAAACA AATTTAGAAA ATAATCCAGT 300

CGATGTGTTA TTGTTAGGTC CACAAGTTCG TTATATGAAG AGTGATTTTG AAAAGCGTCT 360

ATCACCAAAA GGTATCCCAA CAGATGTCAT TAGCATGTCT GATTACGGCT TAATGAACGG 420

GGAAAAAGTG CTGGAACAAG CCTTGGGTTT GTTAGCACAA GCAGATAAAA AAGAGTAAAA 480

ACAAGTTGTT TTTTCGAGAA TCTTCGAAAA AGCAACTTGT CTTTTTTTAT GACTTTAGTT 540

AATTGGAATT GAGCTGAAGC AGTCATTCTG GTTTATCAGT TCCATAAATT ATAGAGCCGA 600

AATTAGTCCT GTGAGCTAGA GAAAGACAAC TGTCTTTAAT TTTTTACTTT AATCTAGATA 660

CTAAAAAAAC ATACTTAACA ATCTATTTTC TGAATAATGA ACGCCTTTTT CCGTAGTCTG 720

AAAGTAGTAA TTCAAAGAAT GATTAACAAT ATTTTATATA CGAGTCATTA TAGTATGGCA 780

TAAGCATTTC GAATAACAGA CTTATTCAGC AAATAAATGA GAGAGAATTC TGTTTTTATT 840

GTTTAAAAAT AATGAGGTTA TTTCATAAGG CATACTTTCT TCATACGTGA TTTCTAAAGG 900

ACTATCTTTA TCAAAAGAAA TAATTTGTGA GTCAGGCATT GCAAGTAATA TAGGTGAATG 960

AGAGACAATA ATAAACTGAG ACCCCTCATC AATTAGTTTT TTTATAGATA TCATTAAGGC 1020

TAGTTGGCGT TGAGGTGAGA GAGCCGCTTC AGGTTCATCT AACAGGTACA GACCATTTCC 1080

TCTAAACCGA TGGTTCACCA GATTCAAAAA ACTTTCGCCG TGTGATTGAT GATGTAGGGA 1140

TTTATCGCCG TATTCAGAAA AATCTAAGTC TATATTTTCT GCATATGTAG CAACATTGTA 1200

GAAACTTTCA GCTCTTAAAA AGAAACCATC CGTAGCACGC TTTACACTTT TTCCAACAGT 1260

TAAATAATTA TGTAAATCCA AGTGAGTATC TTTTGTAGAG AAATTAAAAT TAATACTACC 1320

GCCTTCAGGA TTAAATCCCC AGTTGACTGC GATTGCTTCT AATAAAGTGG gATTTTCCaG 1380

TTcCATTTtc ACCTGaAAAA AAAGTAATAG GttTTTAAAA TTTAGkGaAT nAAATTTATT 1440

CmGcCCTTCA ATTTTTGAAA AAACAGAATC TTCTTCAAAA GCTTTGAATT CTAGTCTTTT 1500

TAAAAAAATA TTTTCCATGA TAAAGCTTGC TCCTTTCTAG TCTTTTCAAA TTAAAGCTTG 1560

GATTACTAGA TATACTTTAA CATTAATTTT TCTTTAAAGG TAGATAAATT TGTTATAAAA 1620

AATCTGCATA TAGGTAAGTC GAAAGTGGAA GATTCGTGTA AACAATAAAG CGTATCAAGA 1680

GTTTGTTTAG TAGGTAACGA CTAGATAAGA ATAGAAAGAT AACGGTGGGA GATTTCCTTT 1740

TTCGCAAAAA AGCTAGAGAA ATGCGTATTT TATAATTGAG GTGGTTAACA TGGAAGATTT 1800

GCATGGACGT AAAATAATTA TTGAAGAAAC AGAACGTTCA ATACAGACAA AACTGGTTTA 1860

TTTACCAACA ATGTTAGAGC GTTCCCCCCT AACAATTCAA TGTCAGCGGT GCGGGGAAGT 1920

CGTATCCAAA AAAGAAAACA GACTGGCGAC AAATGTATAC TACTGTCATG CGTGCATTCA 1980

ATTAGGCCGT GTGACTTCTT GTCAAAAATT TTGTCATTTA CCAGAAAGGC CCAATAGCCC 2040

AAGAACTGTC TTTTTTGAGT GGTCTGGACA ATTAACAAAG GGACAGCAAG CGATCTCGGT 2100

TGAACTTTGT GAGACAGCAA AAGCAAGAGA GAATCGCTTA GTGTGGGCAG TTACAGGTGC 2160

AGGCAAAACA GAAATGtTAT TTGCTGTTTT ACACCAAACC TTACAAGAAG GGGGGCGGAT 2220

TGCGCTGsTT CGCCGCGTGT CGATGTTTGC TTAGAACTGT TTCCCAGAAT CCAAGCGGTT 2280

TTTCCTCATG AAGCAATTGC TTTACTCCAC GGAAACAATC AAGAATCGTA TCGTTATACG 2340

AAACTAGTGA TTTGCACCAC CCATCAATTA TTGAAATTCC ATCAAGCGTT TGATTTACTT 2400

ATCGTTGATG AAGTCGACTC TTTTCCTTTT GTTAATAACG AACATCTTTA TTATGGAGTG 2460

CGTAACGCAC GAAAAAAAAG AAGTAGTTTG GTGTACTTAA CAGCAACGCC TACGAAAGAA 2520

TTGCGGAAGC TTTTACAGCA AAAGAAGTTA GCAGCCAGTA TTTTGCCAGC AAGATATCAT 2580

CGTAGGCCAT TACCTGTGCC ACGCCGATTT TGGGTTCAGT CGTGGCAACG TCTTTCTAGC 2640

AAAAATATTA AAATCATTGC ACATCATTTG AAACGTTTAT TGGTGCGTAA TTCAGTTTTA 2700

CTATTTTGCC CAAGCATTTC ATTAATTCAC CAATTGCATG CATTAATAGC TGAGCGATTT 2760

TCCGACAAAG AAGTTGTCGC CGTTTATGGT AGTGATGAGc TTCGCCTGGA AAAAGTTCAA 2820

GCAATGCGTC AgcAGAAAAC GGATATTTTG ATCACTACAA CCATTTTAGA ACGTGGTGTG 2880

ACTTTTGATG CAGTCTCAGT TATTGTTTAT GGTGCCAATC ATCGAGTCTT CACTTCTTCA 2940

ACACTTGTCC AAATTGCGGG CCGAGTCGAT CGAAGACAAG AATTCAATTA CGGCGAAGTT 3000

TTATTTTTAC ATGATGGGGA AACAAGAGAC ATGAAAGAGG CAATTCGGCA AATAAAGCAA 3060

ATGAATCGCT TGGCTAGTAA ACGaGGGATG TTAGATGGAT tGTAAAAATT GCGCAAgCCC 3120

mTTCGTTTGG AGATAACATT AAAAATGCTC TTACGTTTTA ACAAATGGAT GAGCCCTTCa 3180

CTTTGCCcAG CACTGCCAAG CAAAGTTCCa GAAACTGCCA ATGACAGGCA CTTGTTTTGG 3240

CTGTAGTCGC GTCAGCCAAG AGCGCTATTG TTTTGATTGT CAAAGGTGGC AGCTGCTTTA 3300

TCCAGAGTAT TCTTTTCATA ATGCAGCGTT GTTTCATTAT GATGAAGGAA TGCAAGAGTG 3360

GATGGAACGC TATAAATTTC AAGGAGACTA TCGCTTAAGA ATGTGTTTTA ACGAGGAAAT 3420

CAACTTTTAC TTTCAGCAAC AGTCTGCGGA CTATATTATT CCGGTGCCCT TATCCGAAAA 3480

ACGAATGCAA GAACGAGGCT TCAACCAAGT AATTGGCTTA CTAGAAGCGG CTGATGTTCC 3540

GTACAGCCCT TTTTTAATTA GAAAAGAGGA AAATGTTCCG CAATCTAAGA AAACTAGAAA 3600

GGAAAGGATG CAGTTGCAAC AACCATTTGC CATCCAAAAA GAAAACCAAA AAAAATTAAA 3660

AAAATGTTCT GTTATTTTGG TGGATGACAT CTATACCACT GGACGAACTC TTTTCCACGC 3720

AGCTGCTGTT ATCAATGACT GTTATCCTAA AAGTTTGAAT ACGTTTACTC TCGCGCGATA 3780

AAAAGAAAAG AATTCATAAA TTTGATGACT ATTTTGTTGG CAAAACCCAG ATTATTCGAT 3840

ATAATGAAAT TAAGAAGAAG GGAATTAAAG AGTGGTCCTT TAATCTTTTC TTTACTGGTA 3900

GATGAAAGGG GTAATTGTTA TGTTTAGATA TAATGTACGT GGAGAAAATa TCGAGGTAAC 3960

CGAAGCGATT CGAGACTATG TTgAGAAAAA AGTCGGTAAT TAGAAGATAT TTC 4013

(2) INFORMATION FOR SEQ ID NO: 261:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4951 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

TCAGTGAAGA TTGTTTTATT TTCGTTATAG ACGATGGTAA AGATGCTTTT tTATCCACCC 60

ATGATACAGG TTACTTTACG TCAGAAATGT TTGAATATCT TGAAAAGAGC CACTTACTTC 120

TTAGTATTGT TTCTTTAGAT TGTACTAGTC AAACAAACGA AACTGGCAAT AGTCATATGA 180

ATTGGGAAGA AAATTTAAAA TTAATTGCAG AATTAAAAGA AAGAAAACTC GTCACAGATA 240

AAACGATTTA CGTTGCCAAC CATTTTTCTC ACAATGGCGG CTTGTCTTAT GCGGAGATGG 300

CGGCTTTATC TCAAAAGCAT GAAATCATTA CGAGTTATGA TGGTTTAGAA ATACTAACTT 360

AATCGGATTA TTTGCCTTCT TTTCAGCACT ATTATATGCT TTTAACCGAT AATTACAAGG 420

TACTTTTATC GTAACTTACA CTATCTTAAA AAACAAAAAT ACTCGAGATT TCCCGTGATT 480

CTGGAAATTC CGAGTATTTT TATGGTTCTT TACGCTGCTG GcACAGCGGa TAAACGTTTA 540

TTTTCTTCTT CATTCACTCG ATTAACAATA AAGTTATAAT TCACTTCATT TAGTCTTTGA 600

CCATGATTTC CTGTTTGCTT AATGCCGCCA TTGGAATGAA CACCAACTAC TTCAACCTGA 660

TCATTATAGA TTGGTGAACC aGATTGaCCG CCGGTTGTAT CGATATCATA AAATAGTAAT 720

GGATTTTCTA AGTCTGTAAA GTTAGATGTA AACAAATCAT TTTCATGAGA CCATTGTGTG 780

TGGTTTTTCT CACCTGGATA GCCTGATATT GTGACATGGG TATCTGAAGA TTCAAACTTT 840

TTCAAAACAA ATGGCGTTAA AATTTCTCCC AACTCTGGGC CATCTGGACG ATCGTTTTGT 900

TTGCCGACAG TCACTACCGC AATATCCGCA TTCGGGGAAA AAGCTACATC AATCACTTTG 960

AATTTGCCaA ATGGTGTCGC ACTGCCATCT CGACCTGGAT AAAACCAAGC ATCATCTTTG 1020

GCATTCGGAT TTAATACTTT GGCATTCTTA AAACTTTCAG CCACATGATT ATTGGTGACA 1080

ATTGTATTGG TTCCAACAAC AAAGCCTGTT CCTAAAGAAA TATAGCCTGG TTTACTGGCA 1140

GGGGAAATGA TTCTTCCGAT TGACGCAAAA GGCGCTTCGG TTGTATCTGC CACTTCTTGT 1200

CTTCTGTCCT CAGGGTCCAG TAACGATCTT TTTTGTCGTG AATGACTTTC GGCAGGAACG 1260

ATATACTCTT CTGCATTTGC AGACAAGCTA AAACCGGCGA TCAAAGTTAC TAAAATCAGA 1320

AACAAAAAAC CAGCACTAAT TTTTCGTATG GAGAACTTTT TCATAAGATT ATGCCACTCC 1380

TTATCCATTT TTATTTAATT GGGGAATTTT TTTCATTCAT TGACCAGAAC AGATTCACTT 1440

GGTTGGTTTA CCTGAATGTC CTCTTTAGCT CCGATTCCAG CAGAGTTAAA GGCTGCTGAT 1500

ACCACTGAAG CTGCTTCATC GCCATATTGA ACTTTTGCAG CAGCAAGCAT CGCATCACGA 1560

GCATCACTGA ATTGTGCTTT AGGTGTTAAG TAATTTACTA ACGAGCTGTA GAAAATAGTC 1620

TGTGCTTTTT CAATGCCTAA GTTCTGGATA ATGGTGTAAC CAATCCGATT AATAATTCCA 1680

CTATTATAAT GAACACCGCC TTGATCATAA TAAGGCGTTC CTTTATACCG TGCTCGATCG 1740

TCGTATTGAG CCATGGTTTC TGGTTGTCCG TGTTTACTTG GCGTTTGTAA ATTTCGAATA 1800

CCTGTTTTTC GGTCAACACT CTGAGTATCC GCACCAATTT CTGGATTAGA TGCACCCGAA 1860

ATAATATAAC CCATCAAATC AGAATAAGAT TCATTCAACG CACCTGATTG TCCTAAATAT 1920

TCTAAACCGG CAGTATGTTC CGTTACACCA TGTGTCATTT CATGACCAAC TACATCTAAA 1980

GAGGAAGCAT ACGTTTTTCC TGTTGGTGTA CTTGTTTCAC CATAACGCAT TGCTTTTCCA 2040

TCCCAAAAAG CATTGTCATA AGCATCTGGA TGTTGTTCAT CAACAACTGA CAAGATGGGC 2100

ATCCCTCGAT CATCAATACT GTGTCGTTGG AAATGATCTT CATAATATGT TTTcACAAAT 2160

TTTCCATGCG TATAGGCGTC AACAGCATTC TGATTTAATG CTACTAGTGA TGGTGCTTGA 2220

ATAATAGAAT AATTATTTTT GCCATCAACT ACTGCATGGT AAACCCCTGT GTTATCCGTT 2280

CCGTGTAAAG CAATTCCCGT ATTGCTTTTT TCAACTGGCA CATTAAACGC TACTTGAAAA 2340

GAGTTTTTTA ACGTTACTTC ACTACCGACA TGTTCCGTTA AATCTTGTTT TTCCATAATT 2400

GTTCCATCTG TAGCGTTCAC TTTATAAACC AATGATACGG GGTCACCAGT TGGCGAAgAA 2460

tTTCTCTnAC TTTATAAAAT AATTGCCCCT CATTATTTAC ACGTTTGTCA ATAGCTAAAA 2520

AGACAAGATC TyTAGTTTCC GGATTATCAG AAGCAATAAT TTTTTTAACT CCTTCTTCAC 2580

cGATTGCATC arCCGATTGA ATTTTCACTA CTTGATCAAC TGGATTTCCt GTAACACTTG 2640

TTACTGCACC ATTTCTATCT AAATGAATAA TCACATTATC TGAATCAACA GTAACGCCTT 2700

CCGTTTGCTT TTCTAACTTA AGATGCTTCA TCCCCCACGC ATCTTCTTTT CGATCTACAA 2760

CCGTATATTG TTCAGCAGCT GCCTCAGAAA TTGCCTCCTT TTCCaACTTC GATAAAACTG 2820

TTGAAACTTC ACTTTCTGAA TAAACTTGTT CTTCTGCGGC TACAAAAAGT GAAGAAAATA 2880

GACATGAACT TCCAACAAAG ATGCCTGTAC CTAAAATGTA TAAAATTTTA TTTCCCTTCA 2940

TCAAACAATT AACTCCTTCC CCAGTTTCCT TTTATTTCTT TTCCCAACAA AAGTATTGCT 3000

TTATCCTCCC TAATTTTTcC CGTTAAACGA CCGCCCCTTT CCCTGCGTAA ATATTTGTTG 3060

CTATTTTCAT TATAGACAAC TACCTAACCG GCACTATCTA AATTCCTTAA TCAGCCGACA 3120

TTTTTCCCTG TATTGCCCCT CATAACTCAA TACCATAGCA AAAAAGTTGT TAACAAATTC 3180

ATTCGTTAAC AACTTTTTTA CTGTTATGAT TTTTTTCAAT AATTAAATGT TGAGTGACCG 3240

CTCCTTCATT GAAGGACGTT TGCAAATCAA ATTGATCTGA AAAATCAACA ATTTCTTTTA 3300

CATATAACAA TCCCCAACCG TGCTCTTCTG GATTACTTTT AGAACTAAAT TTGGTTTTCT 3360

TCATAACCTT AAAATCAAAT TCGGCTTGGA CACTATTTGT TATCACAAAC TCTTGTGTTT 3420

CATTTTTATT GAAAATTGAG ATAGCGAACA TTTTCAATTC TGaTTTTGCA CTGTTTTCAA 3480

TCGCGTTATC TAATAGAATA GCCAACAAAC GAATCACAAC CACTAAGTCG ATTGATGTAT 3540

CAAGAATAAA TACTTCTGGC ACATCAACAA TCACTTTTAG TTTTGCTTGC TTTGCGGTAT 3600

TAATTTTCAT ACTTATGAGT GCCTTAAGTT CCATATTTTT CAAACGATTT AACTTCATAA 3660

GTTCAAATTC TTCATTATCA ATAATTTTTT TGGTTGGTGC AATCGTTTCT TCATAAATTC 3720

TTTTGATTTC CAGAATATCT TsGkATGAAA TAGCAATTTG TAAACTATAC AATAAATTTT 3780

TATAATCATG ACGAAACATC GCTAGCTCTT CGTTAATGGA TTCTATTTTT TCAGTATAAA 3840

TCGTTAAATT ATTCAATAAT TGAGTGTTGA TACGTTGTTC TTCCAAATAA GCTTCAATGA 3900

CTAAGCCAAT CAGAATTAAA AATAAACCTA CACTTATTAA TAACTCCAAT TTAGAACTAT 3960

TTACCGTGAC AAATCCTCGC ATAAAGTCAG GTGAAATCTC TGGCGAAATC AAAAACACAA 4020

AAAGTAACAA CAGAATCAAC AATAAAAGTA AACAACCAAA AAAAATTCGG CTATTTTCTA 4080

ACACACGATT TAATCTTAAA AAAACACTAT TGGGAATTTT TAAAGCAATC CAGATATTTA 4140

TAAGCAACAG AAGAACATTT GAGGcAaTCC ATAATAATTT ATTTTGATAA ATCTGTTGAA 4200

AAAAAACACT TAAAATCGTT TGCTCCGTGT AATTGATAAn CGAAaCCATA AGCAaTGTAA 4260

AAnCAGACAA AAATATCGCT TTAAAAATTG AGAACCTCTT TTTGTGAATG AGTAAAAATC 4320

CTGTAAAAAT AACGACTGTC AGCGCATAAA TCAACCAAGA GTAGCCCAAC ATTAATGCCG 4380

CAAACGAACA AAGATTAATG ACAATAAGCA ATGAAAGTAA CGGTATTTTA CATTCAATCT 4440

TGATCAATGT TACTCGTAAA AAGACAAAAA CGATAAAGCT AGATACAAGC AAAACGTTAG 4500

TAGCTAATAA CGACAAAATC ATTTTTCAAT ATTCTTTTTA GGTTTTTCAG TTTGTCCCAT 4560

CCATTGTCCA AATATATTGG GCGAATTTTG ATTTAATCTA GGCTTAGTTC CCACACCATC 4620

ACTAACTTTT GCAACTCTTT TTTCAATAAC ATTTTTAATT ATTTTTTTAC CAAATTTCAT 4680

AAAAACACTT CCTTCAATTA AATTTTTTGT TACAGGGAGT ATCATCAGAC CTTGGATGAC 4740

GAGACCGTAG AGTATTACTG AAGCAATCTT TTCTTTGACA AACAGACTAC ATAAAATAAG 4800

TAAACACACC CTTATAAGGC TTTGCTTTTT GAGCACGTTT AAATCAGATG GCTGAACAGG 4860

TTCAATCGCT GTTCCTTGTG GCGCATACCA ATAAAATAAT ACGACTAAGA GGCACATCAA 4920

TAGAATTTTA TAAATAAACG GTAAGCTCAC A 4951

(2) INFORMATION FOR SEQ ID NO: 262:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5277 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

TTAATTGATA ATCAAAGTAT AACATTTTTT AATGAATTTA CATAATAATA CAAGTGGAGG 60

TCTGTTCGTG AAGGTAGCAG GAGACAAAAT AAAACAAGCG AGAAAACTTA AAAAAATCAC 120

ACAATCAGAG CTGGCTAACG GCATCTGTAC GCAAGCAACA ATTAGTAATA TTGAAAACCG 180

TAATCTATGT GACAGTTTAG ATATTTTCTC ATCTATTTGC CTCCGATTAG ATTTAGAAGT 240

AGAAGAATGC ATGATGATTT CTGAAGAAAA GAAAATTGAA AATCTACTAA ACAAAGTCGA 300

GGATTTATCT TTACGGTTAT TTCACTTAGA GGCCTACAAT TTATTACAAG AAGTGCCTGA 360

AGGTCTTATC TTATCTAATA ATGAACTAAC GACTAAATTG TTATATTATA AGGGrATTAC 420

TTGTTTATTA GGAAAAAAAG ATAAAGCAGA GGCCTTGTTT TACCTATACC GGGGCGCGGA 480

AATTGACCGT AAAGTCAATA TTTACAACAT TTTAAGCCTG AATGCGCTAG GTACGCTATA 540

CGAACTGGAA AAAGATATGA GAAAAGCGCA AGTGTATTAC GAAAAATCAT TACAAGAATT 600

GGAACAATTT AAATTAGAAT GTTCCTTGGA GCGTTGTAGA ATTTATTATA ATTCTGCTAA 660

ATTCTACTCG GAAATGAAAG ACTACCAAAA AAGTGTCATT TTAAGCGAAA AAGGGATTCA 720

GATTTGTCGT GACAAACACT CCATTTATTT GCTAGATTAT CTTTTATATG AAAAAGCCTT 780

TAACAAACAA ATGCTCGGGG AAGACACAGC CGATGACTAT CGCCAAGCCT ATTATTTTAC 840

ACAATTTTTT GGCAATACGG AAGTCCTGCA ATATATTGAG AAAGATATGA AAGCTTTTAA 900

TATTTCCTAT TAATTTAATC AAAAAGCCGA TAAAAGCTGA AAACTCAGTT TTTACCGGCT 960

TTTTGAAAAA TATAGGCAAG TTGCTTTTAA AAATCAGCAG TCACGGTTAC GATAAGCAAG 1020

ACGAAgTATT TAGGAGGATT TAAAAATGAA AAGAGTAATA TGGTTTAGAC GTGATTTACG 1080

ATTACAGGAT AATAAAGCAT TAGCACACGC GTTACAAAAT TCTGCAGCTG ATGAATTGAT 1140

TTTATTATTC CAAATGAATC CTCAACAATT TATTCAAGAA AGTGCTAATC ATAACGCTTT 1200

TTTTGCAAGC TTAGCCTCGT TCAAAGAACG AATCGATCAA GAGGCACATT TACAAATCAT 1260

GGTCGGCGAA CCATTAGATT TATTTTCACG TTTGAAACGC AAATTACCCG ATTGGCAGGC 1320

CATTTATTTT AATGAAGATA CTTGTGGCTT TGGGGCAAAG CGGGACCAGC AAGCTATGCG 1380

CTTTTTTGAA GAAAATAATA TTCAGTCTTT CTCTTTTCAA GATGCCTATT TGCATGGCTC 1440

TGAAGAAATT AAGAAGAACG ATGGCAGCAA GTACCAAGTG TTTACGCCCT ATTACAATAA 1500

ATGGAAAGAG GCGCCTAAAG AAACACCGAT TCCTGTTTCC TATACAGCTG mAAAAATTTT 1560

TAGTGCGTGT CTTTTTCCAG AAGAGGAAGC AGCTTATCGT GAACAGATTG CGAGGATtCC 1620

TTTAACACAC TATAGTGTCG GCGAAGAAAC AGCCAGAAGG CGCTTAAATA CTTTTATTGA 1680

TCAAAAACTT CAATCCTATG AAAATAAGCG TGATTTTCCT TATCAGGATC AAACGAGTCA 1740

TCTGTCTACT TTTTTAAGAA CGGGAGAACT TTCGATTCGC ACCATTTGGC AAGAGCTTGC 1800

ATCTGTGCCT TCTAGCTTAA GTAAAGAAAC CTTCAAAAAA GAATTAGCTT GGCGCGACTT 1860

TTACAATATG ATCTATAGTG CGTTTCCACA ACAAAAAGAG GAAGCTATTC AAGAAAAATT 1920

TCGTTATATT CAATGGACAA ATGACCCAGA AATGTTTGTC AAGTGGCAAA AAGGGGAGAC 1980

GGGGTACCCT ATAATTGATG CCGCAATGCG ACAACTGAAT CAAACTGGTT GGATGCACAA 2040

TCGCTTAAGA ATGATTACTG CCTCTTTTTT AGTTAAAAAT TTACACATCG ATTGGCGTTG 2100

GGGTGAAAAA TACTTTCAAA AAATGTTGAT TGACTATGAT GCTGCCAATA ATATCGGTGG 2160

CTGGCAATGG GCTGCTTCAA CAGGAACGGA CGCTGTCCCT TATTTTCGGA TTTTTAATCC 2220

AATTATCCAG TCAAAAAAAT TTGATAATGA CGGCCAGTTC ATCAAAAAAT ATGTTCCAGA 2280

ACTTAAGCAA GTGCCACAAA AGTATATTCA TCAACCAAAT CTAATGAACG AaGCCTTACA 2340

AACGCAATAT CATGTACATT TAGGAGAAAA TTATCCAAAA CCCATTGTCG ATTATGCATC 2400

AAGTAAAAAA CAAACATTGT TTCTATAksA AGCGAGCAAA GAAATtCATC AAGAAATGAA 2460

CAATCCAAGG TTTCAATAAA CAGTAAACCC AACTAGCTTA GCAAACAACT TGTAAGGTTA 2520

GTTGGGCAAA TTAATTAGTC GAAAGAGAAG TGCnATTTAT CGGCTAAAGT TTTTTATTCT 2580

ATTCTAGTTA ACATAATATA CATTATACAA AGTAGAGTAA AAAGCATTGA AAAGCAAACA 2640

AAACCAGTCT TTAGTTTATC TAGACTGGTT TyGTCACGTA CGTTATATAA ATTATGCTTG 2700

TTGCTTGATG GCTTGTGCGA CACGTGCTCC ATATTCTGGA TTtACTCTAG TGAACAAATC 2760

```
AATTTGACGC GCGATGaTTt CTTGaTTTTt CACTTgACCT AAAgATGCCG CTATATTGTT     2820
AATTAAGTTT TCTTTTTCTT CGCTTGGCAG CAAATTATAC AAAGCGTTTG CTTGTGTAAA     2880
GTGATCTTGA TTATAGCTAT AATTACCAAC ATTTCCTTCA ACTTCAAAGG AACTAATTTT     2940
CGCTGTAGGA TCTTCTTTTG GTGTTTCAGT ATAACTATTT GGTTCATAAT TAATTTCGCT     3000
ATTGCCATTG TTAAAACGCA TATTGCCGTC TTTTTGGTAG TTATTTACAG GCGCTTTTGC     3060
TTGGTTGATT GGCAATTGGT GACTATTAGC GCCCACGCGA TGACGATGTG CGTCCCCATA     3120
AGCGAACAAA CGTCCTTGCA ATAATTtATC CGGAGAAGCT TCAATACCAG GAACGAAATT     3180
CCCTGGTGAA AAGGTTACTT GTkCAACCTC TGCAAAATAA kTCkCTGGAT TTCTATTTAA     3240
AGTCATCGTG CCAACkTCAA TCAGCGGATA CTCTTTTTGA GAAACCGTTT TTGTTACATC     3300
AAAAAGTGTT TCTTTCATTG TTAACGCATC TGCATACGGG aTAATTTGCc ACAGATAATG     3360
TCCAAGAAGG AAATTCtTGA TTTtCAATtG CATTATGCAA ATCTyCAATA TGGrAATCTG     3420
GGtTTtCCCa GCAATTTCyT CTGCTAATTG ACTCTCTAGA TTTTTGATTC CTTGATTCGT     3480
CTTGAAGTGA TATTTAACAA AAAATACTTC TCCAGCAGCA TTGACCCATT TGAATGTATG     3540
GCTACCAAAA CCATGCATAT GACGGAACGA CAACGGAATG CCACGATCAC TCATTAAAAT     3600
TGTTACTTGA TGAAGACTTT CCGGTGAATG AGACCAAAAA TCCCAAACAG CTTCTGGACT     3660
TTTCAAATGC GTCCGCGGAT TTCTTTTTTG GCTATGAATA AAATCTGGAA ACTTAATTGC     3720
ATCACGAATG AAAAAAATGG GCGTATTATT GCCAACTAAG TCGTAATTTC CTTCGTCTGT     3780
ATAGAATTTA AGCGCAAAAC CGCGAGGATC ACGCAATGTA TCTGATGAAC CTAACTCTCC     3840
AGCTACAGTT GAAAAACGAG CGAATAAAGG TGTTTCTTTG CCAACTTCAG ATAAAAAATC     3900
GGCTTTAGTA TATTGTGCCA TTGATTGGCT CACCTTGAAA ATCCCATGAG CACCAGCACC     3960
TTTGGCATGA ACAACCCGTT CTGGTACGCG CTCACGATTA AAATGAGCTA ATTTTTCTAA     4020
TAAATGAACG TCTTGGATTA GGACAGGACC AAATTCTCCT GCAGTTAGCG AATTTTGATT     4080
ATCGCCAACT GGACTTCCTT GTGACGTAGT TAAATGTTGA TTTTTCACCT AAAACACTCC     4140
TTCATCTAAA ATTATACTCA TTGTATAAAA AAACCACTCT GCATAACAAA AATAATGCTT     4200
AAAAACACTC GAAAACATTT GTTATTTAAT AGAATGATTA AAAATAGCAG TAGAATTTAA     4260
ATCAAAAACA AATGTATTAA CTTTAAACTG TTTACACCAC CAAAAATAAC AACCAATAAT     4320
TCAAAGAAAA CTAAAATGCT GTTTTTTTAT TCAAAAAAGA CCTAGCATCC CAATCATGCT     4380
AGGTCTTTCG TTACTTTTTT TGAGCGGCCG TTTTTATTGG TCATTCAGCG TTATTGATTG     4440
ATCGCAACAG TCGTGTCTTC AGTCGATAGA CCTGTCATAG AGGTGCTATC GGAAAATATT     4500
TTTTCAACTT CTTGTTCTGA TAAAATATTy TCTTTCTTAG GCATkGAAAT CGTCGCTTTT     4560
TTGGCTTTTT GTGTAATGCT AATTTTACTA GATAATGATT CAAGTCCCAT ATCAACGTTC     4620
TTGTCTTGCT GCGGTTTCAC ATTAATCAAC ATATTTGTTT TATTTTCTTT TGTATTAACA     4680
GAAACTTTGG CAGAAACATC TTTGACTTTT TCGAAAGGAT CTTGTTCTTT ACCTTTTTCA     4740
```

GATGTTTCTG TCGTAATTTT GATTAACTTT TGTAATTCTT TTTTTGTAAA CGTATGGGTA 4800

ATGGTGTCCC CTTTTTTCTC GAACGTATCC TTGTCTAATC CTTCTAAAAA CTCTTTGTAT 4860

TCCTTAGCAG ACTGTTTTCG TTCTTTTTCC GATTCTTGAT ACTCTTTGAT TGTATCTTGC 4920

CATTCTTTCG TACTTTCCTC GTTCATGGCT AAAAGATCAA TGTATTTTCC TTTTAGTTTT 4980

TCTGTATCTA ACTGAGAACT ATCGATAGCA CCGTCAGTCA TTGAGTCTGC GACAGCGACA 5040

ATATATTCCA TCATATCTGT TGCTAAATAG AGTTTTGGCT CTTTGCCCTC GTTATCTAAC 5100

GATCCAACCA ATGAAATCGG TACATCCATT CCCATCGCTT TTAATTTCAT CTCTAAGTTG 5160

AATGCTTTTT CTTTTTTGGC ATCTACTTGA ATCTTCCCAG AAAGCGATGC GTCTTTGATT 5220

TGCGTGATGA GCATCCCAAT CATAGGATTC GTTTGTGmAC CATCTTCTTG TGAAAAT 5277

(2) INFORMATION FOR SEQ ID NO: 263:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 891 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

TCACAATAAA TAAAGAAAAG TAAGAGGATA AATAACAAAA ATCTCATCAA ATTGAAAAGA 60

AGAAGTGGTT GTTTTCACAA GTGTGAAGAG GGTTAATTAA ATACAAACAA GATTTCAATT 120

ACATTACAGA AATTTTAAAA CTCAAAATTT AATACGTTTA TTTCTTGTTT CAATAGGAAA 180

TTAACGGTAA AATAGAAAAG TGAGACAGTG AGGAGGAAAA ATGAATGAGC GAACAGTTAA 240

CAAGAAGTAG TAAACATATT CAAAAAGCTC TTAAAAAGCA AAAAGCTTAT AAACGTGCTA 300

CCGCCGTTGC GGGGACTTCA ATGATTTTAG CACCTGTGGT TGGTGCTGCG GTACCTGCCC 360

AAGCAGAATC AAGCCAGCAA GCATTTATTA ATGAAATCGG AAATTCAGCA GCAGCCGTTG 420

CCAATAGCAA TGATATGTAT GCTTCCGTCA TGATTGCTCA AGCGTTGCTA GAAAGTAGTT 480

ATGGTAGTTC TGGGTTAGCA TCTGCACCGA ACTATAACTT GTTTGGTGTA AAAGGAAGTT 540

ATAATGGACA ATCTGTTTAT ATGCCAACAA AAGAATATTT AGACGGACAA TGGGTGACCG 600

TCACAGCwGC ATTTAGAAGT TATAATTCTT ATGCGGAATC GTTCCAAGAT CATGCCAATG 660

TTATTCGTTC CACTGCTTTT GGGGATACAT ACCATTATTC AGGCGTTTGG AAGAGTAATA 720

CTAGCAGTTa TCGCGATGCA ACAGCCGCTT TAGCTGGAAG TTATGCAACT GATCCAGGCT 780

ATGCCGAAAA ATTGAATTGG TTAATTGAAG CATATAATTT AACTCAATAT GATTGGGGAG 840

CACCAGTCGC TCAAACCACT AGTTATTCAG GACTCACCGG GAACCGTTGA T 891

(2) INFORMATION FOR SEQ ID NO: 264:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4030 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

AATTCTCTAA TTCGGATAGT TTGGCTGGCT TAAAGTCCAC ATAATCCACC TCTGATTATT 60

TTTTCTTCGT ATTTATCGTT CCATTGGCCA TTGCTTTCTC TTGTAAACGA CGCTTTTTCT 120

TTTTAATTTT TGATTTTTTC TTACCCATAA ATTTTCCTCT TTTCATTAAT TAATATCTTA 180

TTATTTCTTT TTTTCAGTCG CAAATCCCAT GATGaTTTGa TTCTTTAGAT ATCGTAATTG 240

CTGTCTGGCA TATGCTTCCT CTCTACAGCA AGATAATCTA TAGTTAATAT CCTTTAAAAC 300

CGCTAAATCA AAATTGTATT TGCCCAGTAA TTCATTAATT TCTTCTTCAG CTGTTTTCAT 360

TTAATTCTCC TTTAGCGTGA TCGGTCTGCC ATACTTTAAA ATTTCCCAAG CGCCATCTTT 420

CATATTGGTT TTATTCATAT GATTTCTTTC ATCACGAGCA ATTGTATAAT CGAAAAATAA 480

ATCAGCTTGC TCTGATCCAT GCAGGTACTC AACATACACG CCGTCTACTT GACGGCCCAG 540

TATATATAAT TCTGGATAAC TTAGCATTGT TGTTCTCCTT CAAGAGATAT AGTGGCCCCA 600

GTGaTTTTTA AGCCAATTAG TTTACTTAAA ACCATAATTG CTTCATCCAC TAGCAATCCA 660

TCTCGTTGAA TAGAACTGGC ACATGTAAAG CGGTCCATGC TTACAATCTC ATCCACCGTA 720

ACAGGATATG GAATAGTaAC GCCTGCTTTT TTGGCAATTT GTTCAATAGC TCGGACATGC 780

GATTTATTCG ATGCTAAAAT ATACTGACCT GTTCTAGCTG ATTCTTTAAC TAGCTCTGTT 840

GTTTTTCCAG TCCCTCGATC TCTTGCTATA ATTTTCATCT ATTTGGCCTC CTCACTTAAT 900

CCCCAAAAAG GCTTTACCTT GTAAAACTTC TCAACTTCTC GATTTAAAGA TCGCACCATA 960

CTTTCTAAGA CAGTTGCTCG TGTTCCTAAC GTGACATTTT GCTTTTTCGC TTTTTTAACA 1020

CTTGTAATAC CTAAGTTGTG CCTTAGTTCA GAAAAAATTA GTATTCCCTG ATTGCTATAT 1080

TTCCATTTTA AATCTGGATT GGCTTGAATC TTTTCCCATA ATTCGTTAGT GACCACTAAA 1140

TAGTTATAAT CACCTAAAAA CGTCTGTTTC GCAGAACTTT TCAAGTCTGC CATTGTAACT 1200

TTAATTTCAT AGCATCTAAT GGTGTTGTCA GTAGAATAAG TCATAAAGTC GACTCTTTCT 1260

TTGCCAAACC ACCCAATAGT AACTTCAAAA CAGCCGAACA CGCCCATTTT GTTGGTATAG 1320

TGCCACAAGC ATTTCTCAGC TTGTCTGGTT AAATCAGTTT TCATTCACCA CCATCCGCTT 1380

TCACAGCAAA CGGCCAATAG CGCTCATCAA TTGCTTTGAT TTCTTGTTCT GTGAATATTT 1440

CTTGAACCAA CCCAATCAAA TTATCTGACC AGTCAACAAT ATTATATTTA CCATTTTCAT 1500

GATCAAAAAC CAGTCGCTTA TGCACATCAT CGTTTTTTAT GAAATCAACA TAATACAATT 1560

GCTTTTTCTC TACTTCGTAG CCGTTATCAA TCGCATTCAA TAGTTTTACT TTATTTTCTG 1620

GATGTAAAAG CCAACCCTCA ACATCGTCAG TTATCAATGT ATATTCGTTA TAATATAAGC 1680

ATTCTAACTG TTGACCAACG AATACCATAA ATGCAGCACC AGAACTTAGT TTTTTATCAC 1740

TAATTTTCCC TTTTAACCCA TCAAGCCAAT CAGCCACAAA TTTCTCTTCA TGCGAAAATA 1800

TAACTTTTCC CTGTACTTTG AGCTCTTTTA TGACGTCAAT TGCTGCATTT ACTCCGTCGT 1860

CATAACsTCT TGCCCATTCA TCTTCTGCCT CGCAGCCTCC GATATCATAG AGCGCTTGAA 1920

TTAAATCTTC TTTTTCTTGT TTATTCACCG CTGATCCTCC ACTCTAAAAA GTTCGTCTTC 1980

TAGTCCAGTT TTTTGATACT TCAAAAATTC AGTAATTATG ATCGGTTGTT TATTACGCCA 2040

TGAAATTTCA GTAATATCTG TATTTTTAAG AATTACAGTT GATGGCAATG CACCTTGTGC 2100

ATTTTCATTA ACATAACCCC ACGAACCTAG GTCCATATAT TTTTTAGGCT TCAAGGAAAA 2160

AAAGGAAAGA TATTCGCTGT TGCAGTCTCT GACCACATAT TTAAACCCTC TTTTAAAAGC 2220

TTCTTCCATA ACCTCTAGCG CTTCAGCTCT TGGTCTAATC TCCATTATCT TTTTTATTCG 2280

ATCTAAATAA TCACTCATTC CGCTTCCTCC TGTTCAATGG CCCATCGGCT AAACGCTTGT 2340

AAGACTTGTA ATTGCCCAGT TTTTGACATA TATCTGTAAC TTCTATAGAC AGGTCTATCA 2400

CGATAGTCTG GTTTTATAGA ATTAACTCTC AGTCTCCAAA ATAATTCTAT AGGTTCAATA 2460

TTTGTGACTG TATATTTTTC TTTCAACCAA TCCAGAACAA TCTGCTGATT TTCGTTTAGA 2520

TATGGTCTTT TAAAACCTTT TACGATGAAT AAAACATCTT CtGCTGACAT ATTTCCTTCT 2580

TCAACACGAT CCATCTCAAG TTGACGTTCA ATTTCTTTTA TTAGTTCATG CATTTAATTT 2640

CCCTCCAACA GTTCTGGGTT TTCATGGATG TTACCGATGA CTAAATAGCC ATCGCGATAC 2700

AAAATATTAG ATTGACAAAA TACAAGAtCT AAAAATTCCC AGAAAAAGCG CCCATCTGCA 2760

TAAGCAATTT CGTGCATACT TCTTCGGTGG GCGTATTTTT TTGGTGCATC ATAAACTAAA 2820

ATGTCCCCCT CAAAAATTTC AACGCCGTTC TTGTCTTTCA AGCCTGTTGA TTGCATGATG 2880

ACATAAGGAA ATTCATCCCA ATTAAAAGTT GTATCATAGC TTCCTATCTC TATTTCAGAC 2940

GTGCCAGGAC TTCCCTTAAT CTGAAAATCC GTAATTATGG CTATCATGCT ATTCGAGTCA 3000

TCAAACAAAG GTGTAACATT TTCTAGCATT TCTTTCTCGT AGGTATCCCA CGCTCTAAAT 3060

TTTGGAATCA TCTTCTTCAC TCGCTTTCTC CAACAGTTCT GGATTTTCGT GAATATTTCC 3120

GATAATACAT AAATTCCTGT AATATGATAG GAAGTTGCTA GCATCGTTCT TAAAACAAAA 3180

TTGCGCATAG TCTACATCGT AAACAATTTC GCAAGCTTCA AACGTTTCCC ACGTGCCATT 3240

TATATTGTCA TCATGTTCGA CTAAATCCCC CTCAAAAATT TCAACGCCGT TCTTGTCTTT 3300

TAAGCCTGTT GATTGCATGA GAACATATTT ATCAATCATT CCCCACATGC CATTTTCTAG 3360

ATTGATAAGA GGCGCTATAA ATCCTGTATC ATCATCAATA GTCCATTCTA CATTTTTATC 3420

TTCATCAGGA TAATACATTA TGTTTTCTTC TACTGAATAA GCTCTAAACT TCGGAATCAT 3480

CTTCTTCACT CGCTTTCTAA TATTTCCACT TCTTCCGCTG AAAGAAAACT AGCAAGACAA 3540

TTTGCATCAC AAAACATATT GTCTGTTCCA TCTTCAAACT GGAAAAAATT AACTATCATA 3600

AAATTGTCTC TAATTGTTAA ATATGaTTGA TTAGTAAATt CTCCATCATT AAACAATGCG 3660

TTACCACAAT TGGcACAGGT AGAGCTATCT TTCTCAAAAT TCGGATAAGT cACAATTTTT 3720

CACTACCcTTC TGATTATTTA ATTAATACAT AAAATCCATT TTTCTTTGCA ACGTCTCCTC 3780

GAATTCCCAA ATGGCCTTcC AACTCTTCAA AGGTCCTTTT TGTAATTTTG GAAAGTTGCG 3840

TATCATAGCC CAAATTTCTA AGTGTCACGT ATTCATCTGG TGTTAACTGG TCTAAATCTA 3900

TcgCaCAAGT GGCGGAAAGC GTTCAACCGA CGACTCAAAT ATCACACTAG TTAATTTCAT 3960

TTTTtCACCC TCAACAATTT AGAATTAATT TTTTACGGCT TTCGTTAAAT CAAAACCAAG 4020

AGCAGTTGGA 4030

(2) INFORMATION FOR SEQ ID NO: 265:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4147 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

TAATGCAATG CTTGCAGCAG ATTACTGTAT TATTGTTCTT CAAACACAAG AGTTATCTTT 60

AGACGGTGCT CAAACATATA TCGCTTATAT GCAATACTTA GCTGACACAT ACGATAATGA 120

TTTACAGGTC TTAGGTATAA TTCCTTGTAT GTTGCGTCCT GGGGGTAGGG TAGACAATAA 180

AGTATTAGAA CAAGCAAAAG AATTATATGG AGGAAATGTA TTAAACACTG TTGTAAAATA 240

TCAAGAACGT TTAAAAGTCT ACGATGTTGA GGGAATTAAA ATTAGTCATA ATTACACAGG 300

TAAAGCAGAC GGTTGGGACG TTAAAGCGCA TAAGGTATTT ATTGACGTAT TAAATGAATT 360

AGATCAACAT CAACTAGTAC TAGAAGAATT GGAGGGATAA AATATGAGTA ATAGTTtCGA 420

ACCTATTATC AAAaCTAAGA AGAAAAAATC TTCTTTAGAA GAAAAGCATG TTAGTGTTAA 480

AGGAAATGAT AATTTTTCAA GAGATACTAC AAATAATTTT GCTACAAAAC CTACATCAGT 540

ATCTAAAAAA ACAAGCACAG ATAATAGTGA TGTTAAGCCT TCTACAGTAG ATCGACGATA 600

TAAAGCTACT ATGACTCAAA AAATATCTCC TTCAGTCAAT CTTAAAATAA ACACTTTAAA 660

ACCCTTTCTT GGTGATTTAG AAAACATGCC CAAAGCCACA GTAAATGATA TTTTGAACAT 720

CTTGCTAGAT AATTATGTAA ACACAAAATT TTCTACTCGC CAACAAGAAG CATTCAATTC 780

TATGTATAAC ACACAAATTG AACTTTTAAA GAAGTAGTAA GTTTAACTTA GGTATACCAG 840

TATACTGGTA TACCTAAAAA AGAGAGTATT TAATATACTC TCTTTTTTTA TATTATAATT 900

TATGATTCCG CATTATAGAA TTCTTCTTAA TCGTGTATAT TCGTTTACAT AACGAGCATT 960

TATACTTGTC TAAATCACTC GACACTCTAC TTGCAAGACA GCCACACGAG CACTGCACAT 1020

AGGCAATACT AGTTGTAGTA GTATCTTCTT TGATTAGTTT TCCTTGGTAA TCATATGTTT 1080

TCATAAAAAT CTCCATCACT GTAATAGTCG TTAATGAAAG CATCAATTGT TTTACTTTCT 1140

TCTTCGGTTA ATTCTGAAAA TTCTACATCT ATAAAAGGCT CTTCTACAGT TCCTACTGCA 1200

GACCATTTCA AAAAGTTATG AATTTTAATG TGCCTTATGT CGTAGAAATC GATTTGAAAA 1260

TCACTAATTA AAACAACATC AAACTCAGCC ATTCCACGGA AAACACCAAA CACATGCGGT 1320

TTTACACGAT CATATTCATC CAATGAGTTC AACTGGATTT CAAGCACTTT ATTTTGTTTA 1380

ATTGAGCGTT CTAATAAGTA TTCTATTTGT ACTTGAGATT GCTGTGGCAG TCTTTCAATC 1440

TCTCGTGCAT TAAATTCATC CGTTTCTTTG ATTGCTGCTT GCAACTCTCC TAAAGGGAAA 1500

GCAGTTGGCC ACTTCAAATC AAAAGGGCGG TCTACATAAT CGTTATAAGG TGTAAATTCC 1560

TTTTTAGTTT TCTTCATTTG TTAACCTCCA AGACCGACGT TGCCACCTGC ATGACCACCA 1620

ACTAAGCCCG AACGATTCAC AGcTGTCGCC CCtTCAAGAA GTGAGGACGC ATGGATCAAT 1680

GCTTTAAAAC CAAATTTCTG ACGAATTTTA TCAATTAGAA AGTCTAACTG ATTGTTTAtT 1740

ATTTGTTCTT CAGGAGGGGA GAAGAGATCT ATTTGAAGTG TAGTATCCCA AACCAGCTTT 1800

CCAAAGGAGA CTCCTAAATT TCGAATATCC ATGCCAGGAA CATAGTTTTC TCGAAAAAGT 1860

TTCAAAACAT GCTCGGTTAA CACTTTTGTG TTGTTGCTAC GAGCGATTTT AAGCTGTTTT 1920

CGCCAATGAG ACCTGCCCAG TCCATCAACT TGTCCTTTAG AATAACCAAC AAAAATTGAA 1980

ACACATTCAG TTTGACAATT TGAATTTCTC AACCTGGTTG CTACCTGGTC TGAAAGTTCC 2040

TTCAGCACTA ATTCAATTTG ATCCTTTCGA GCATAGTCTC TAGGTAATAC TTGACTATTA 2100

CCGAATGATT TTTCTTTGGC CACCTGGCTT TTTTCTCCTA AAAAACTACG ATCAATGCCC 2160

CAAGCATGTG CGAACAGTTG TAATCCCATG ACACCAAAGC GGTGTTTTAA TAGATAAGGA 2220

TCTGCTTGTG CTAAATCATA AATCGACTCA ATTCCTGACA TCTTTAAACG AGCAGCCATA 2280

CGCTTGCCAA TCCCGCAAAA TTCAGTGATA GGGGAGATAG ACCAAACTTT TTCTTGTACA 2340

TCTTCATATC TCCATTCTGC AACAAAACCT GGTGTATTTT TTGCTTCGTT ATCTAGGGCC 2400

AGCTTTGCTA ACAATGGGTT ATCTCCAATG CCAATCGTGA CGTATAGCCC CATATGGTTA 2460

TAAATGACAC GTTGAATAAT TTTGGCCAAT CGATAAGCAG TATCACAATT AAAATACTTC 2520

AGAGAAGCAG TTATATCAAT GAAGCTTTCA TCTACACTAA AAACAGAATG GTCTTCATCT 2580

GCCACAAACC GTTTGAGCAG ATTATTAATT TCACTATTTC GCTTCATATA TAAATTCATT 2640

CTTGGTGGAA CAATAAGCAA TTCTGGGTGA TCAGGCACTT CATCTTTACG AGTCACATTA 2700

GAAATTCCCA GGACTTTTTT TGCCATTGGC GAAgcAGcAA GyACwArTCC GCCTGCATTA 2760

TCCGCACCAC TCATyACTAT AAGCATTGTT TTrAGGGGAT CTAAGCCACG TTCTACACAT 2820

tCTaCGCTCG CGTAAAaGGA TTTtACaTCG aTrCAAggAC ATCTCTCCGg GTTCTTgGAA 2880

TAATCAAtTT CATGCAAAAA CCCGTCCTTT CGAGTTACAT TATACGAACT AACGTTCTTT 2940

TTTGCAACAA CAAAATTTCA GTTTTCCGAA TTATTTTAAA TTTAAAGTAG CTCAGAAATC 3000

TACAGTCATT ATTTTTTTAG TTTGTTTTTA ACGCGCAAGT AGATATCAGC CAGCCAGTAA 3060

ATAATTGTAA TTAATACTAT TACTTTTAGA AAGTTCACCA TGAATACTAG ATTGAATGCG 3120

TATTTTAACC ACAACAGCCA TAATATAATC GCGATAACTA CTTTTCCAAA GGTATTcaAT 3180

TTCCtAACCA ATAACCAATG TAAAAAATCT TCTCCTTTAT TTTTAGTAGC TCTATATTCT 3240

TCTAGAATAT TTTTTTTCAT GTTTAAACCt CCTTTTTTCT CrTCTTmTAT TmACATATAT 3300

ATAATTACAT GTTAATATAm ATAwATGAAA AAGAAAGGAG GTAGTTTTCT TATGAAAAAA 3360

TTAGCACTAG GTTCATTGCT TTTAGTTGCT CTAGGTGTTA TCGGGACATC TTTTGCTTTA 3420

ACTACTGAAG TAAAACAAGA GGCTGGATAT GGACATCCTC TTATTGATCA CAAAATTTAG 3480

TCTTAAAAAG GACATAGCCT TGAATATTTG TAAAGGTTAT GTCCTTTTTC ATTTTTATCT 3540

CTTATTACTA TTTTCGGGCC GTCGTATAGA TTCCCCCAAC CTAAGAACTT TTCATTTGGT 3600

TGAGCTGACT AAACGAAgGG AAACCAAACT AAAACGTTGA AGGCCAGTTG AGCAAAGCGA 3660

AACTAATTAT tCGTCTTTTC GGGTTtGACC AtGTtCCTGG TAtaAAAAAT TATTCGTAtT 3720

TAAATAGAAA TATTTTTATT ATTTCCGCCT GCAGATCACC ATTTAATCAT GTAAAAGCGA 3780

ACGTTAAAAC ACCTTCATTG CATAACTACC GATCAATATT TGTGAAAATG GCTTAGAATT 3840

GCTCTCTCAA ATCTAAATTC TAGTTTCTTA TTTTACTTTT GGTACTAATT CATTTATTTT 3900

TGAAATATTT AAGGCAATCA TTAACTCACT GATTGTTTGT ACACCTACTT TTGTATCATC 3960

AAATAGTCCT TCATTGTACA TTCCATCTAG TAATACTGGG TCTATTTGTA CATTTCTTTT 4020

CAGTTCGTCA ATTGTCATTC CTTGTTTATC CATGATTTCC TTtAACTTGA TkTCCATTTG 4080

ATTATTCTTC ATAACTTTTT CCTTCTTTCT TTTTTATCCC TTATTACTAT TTTCAAnGTG 4140

GCACATA 4147

(2) INFORMATION FOR SEQ ID NO: 266:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2705 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

GTTCTTCTTC GAATTCATCT CCAATTACCT CGACTTCCTC ACCTTnAATA TCTGGTnGGT 60

CAAAACCAAA CTCTGTCATC TCAAAATCTA AmTCTAAATC TGCTAATTCA TCTAACTCCG 120

CATTTAATAA TTCTTCATCC CATGTAGCAA TTTCTCCGAC CTTATTATCA GCTAAACGAA 180

AAGCTTTGAT TTGCTCTGGA GTCAAGTCAT CAGCAATAAT TGTAGGTACT TCTTTTAGTC 240

CCAAATACTT AGCAGCTTTT AATCTTGTAT GCCCATTTAC AATGACATTA GAAGCATCTA 300

CCACAATAGG AACTTTAAAT CCAAACTTTT CGATACTTTT AGCAACTGCA GTGATAGCAC 360

CTTCATTATG TCTGGGATTT TTTTCATAAG GTATCAAATC TTCTGTTTTT TGTACTTTTA 420

TATCCATATA TGTAGAACCT CCTTGTATTT AAGTATCTAG TGGGATTCGA ACCCACGAAT 480

GCTAGTTTTG CAAACTAGAG CGTTGCCACT TCGCCATAGA TACATGAAGC AAACTTATAG 540

AGCAGTATAA GCTTGCTTGA CTAAACGAAA CATTTATTGA CATACATTTT TTATTTTGTG 600

AATCACTAAA ATTCACAATG GACCAGATAG GATTTGAACC TATATTCAGC ATCTGCCTAC 660

TTCCTTTATA TGCACTTAGA TATCATTACG CATAACTGAT CCATAAAATA TGCAACAAAA 720

AAACCGACTG CATAATTTGC AATCGGTTTT GTGTTTTACA TATTTTTCTA TACTAGTATA 780

ATAACATCAT TTTTATGACT TTATCCGCCA TAAAACCGCA AAAAAACCAC AATTTCAGTC 840

AGTCGCACTT TTATAGCATT GAAGCAGACC AGATTCATAT GCTTCAGCAA ATTCAAGAAG 900

TGCGACaCaC TTTAACGTTT CCACGTTGCT AACAGAATAA CCTAGTTCAG AAGCAATCCA 960

AACATTAGTT TTCTTTTCAA CACTGCAGTA AGTATAATAC AATAGTTGTC TATTTATCAA 1020

TGACAATCGG GCCAATGCAT TATCTATTTn TAATTTTTCT TTTTCCGCAT CAATTTTTCT 1080

AATAAAATGC TGTTGAGATC GATGTGTATT TGTGTGCTGA ATAGGCACTT CTTTAATTAT 1140

CGGCGAAGTT ATATTAAATT GAGGTGTTCC TGCAATTCTT TGCAGCCGTC TGTATTTTTT 1200

TAAAAGATTT CGTGCATTCT GTTTTGTTTT TTTGTCATCA ATATCAGGAA ATAGCTGCAT 1260

TTTTTATCCC TCCGTGATAC AATATCTATG CAGCGTTTAT TTCGGGAACG CTAAATCACA 1320

CATCCGAATT ATATTTGAAG TCGGTTTTTG TTTAATATTT TTTCAAGATG CCAACCACTC 1380

GCCGTCTAAC GAGTGGTTGT TTTTTTTATT GTCTATCATC TTTTAATTTT TCCAGTTCTT 1440

CTTCCAATTG TATTTTTTGT TTATATAGCG AACTTACGCA CAACGCTAAC AAAATAATTA 1500

CAAGCCATAA GAATTCATTT AGTGTTATCA TATCTATTCT CCAGTTAATT AAATATTTTt 1560

ATTCCTCCAA TGCCCAGCCC ATTAATTTAA CTATTTCAAA TACTGCATTC CGCCTTACAC 1620

CAGTTGAGAC ACCtAGGTwA TTAAACTTCA AAACATCCCT GTCTAgATTA TTTGTAACTG 1680

TTTCAAAATT CCTAACATCT TCarasTTCA TTGTTCCGCC AGTCGGTAGC CATATAATCa 1740

AACTTTTTGA TTTTTCCATT TAtTGATCCT CCTCACTTAA TCCCTAACAC TACATATCCA 1800

TCTTGTTGGG CATAATCTGT AATGTAAGTT ATTTCTGCAA CATGGACATC ACCTGTATAT 1860

TGTCCGTCTT GATATTCGTT TAAGCGTAAT ATATCGCCGT TTTGATAGTT TCGGTCATTC 1920

TTACGTATTT CAAAACGCTT ATCACCTGAA ACAACCGCTT CAAAGTATTC TGGTAAGATT 1980

TTTAGTTCGT GAATGGTTGG TTCTTTCTCA ATTCGTATCC GTTAGCTAAT GCATCAATCA 2040

ATAAATCTCC ATTTGTCTTA AACCACTTAG AAAATTCAGC AGTTGGCATT GCACTTGAGA 2100

AAGAAACTGc CGAACCGATa ATGTCTACTT TATcAGAACC CATTGaTAAA CCTTCTTTTA 2160

TAAATTCATC GGCACTTCTA GGCAATGCAG CTTTTTTCGA TTCTtTAATT GTTTTGCTAA 2220

GCTAATTGCT CTTTCGTTGG CATAGTCAGC ACCTTTCAAA TAATCAAGGC TATCTGTAGA 2280

AACTTCTAAG CATTCTAACT CTTCAATCAA TTCTTGTTTT TCCATTAGAG TCCACCTTCC 2340

AAAACTTTAA ATGTACTTTT ACGCTCATCG AATAACGCTT TATTTATTTC CTGTTCAGAC 2400

CATGGAACAA TCATTTGTGC TGTGTATTCA ACAACCAGAT CAACAAACTC ATAGTTAATA 2460

CTATTAATTG AAATCACTTC ATACGCCATT TTTTTATCAC TAGAAAATAT GATATTTCCT 2520

ATTTTAGCTA ATTTTGGCTA ATCTTAGCGT GTCATCTTCT TTACTTTGAT GTAACGTTCA 2580

GTGAATGTAA AATTCATTTA TATTATGATA ATCGTTATTC GCCTACTTTT TGGACCAATA 2640

TTTTACTCTA ATCTTACAGG TTACTAAAAT ACAATTTTTT CAATCAGTAT GCGGGATAAC 2700

AATAT 2705

(2) INFORMATION FOR SEQ ID NO: 267:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1302 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

```
GGTAAGTAGA TCCCATGGTT GATTCACCTT CCAAAAACnA GCACnCATAG AAGAGCTTCT     60

TAAATCTTTC CCACTGTGCA ACATAATTAT AAATAAATTT CTTTGGTGTC ACCTCGCCAA    120

AGGCAAAATC CATATCATAG TATGCGGGAA TTATTATCTC TTTTTTCGAA TTTTCAGGAT    180

ATAAAATATT TTTCTCCAAA CGCGCCTTGG CAGCTTTAAA TTCAACTGAT TGTTCTCCTA    240

AAACAATTGC CTTAGAGGAA ATACCGCCAT TATAAAACAT AGCTACATCT TCCTTATAGC    300

GATTTGCCAT GtCAACTwGa GCCAAGCGGG CACTAAGTTC TtTCCCCGCA TTACGTAATT    360

CCCaTAAAAA ATACAACaTA ACAATAATAA AAACTATAAT TAGCCAAGGC TTTAACGACC    420

TTCCTATTCC TAATAAATAT AACCAAATTT TCTGCAAAAA ACTAACCTTA GGTAGCGAGG    480

TCGTTTTGGA CATAGTATCA GTAGTAAGTA GTTCCTTAAT CCGAGTGGTG ACAATATGTA    540

TTCCTGCATT AAAATCCTTT AATTTATAAA AATCTTTCAC GGTACTATCT GTTATTTCAT    600

CGGTTAAACT ATCTGGGATT ATCTCTTCAA GCCCATAGCC TACTTCAAGT CTTGCTTGCC    660

TGTCTTTGAC AGCTACTAAG TAAATCACTC CATTATTATA AAGTTTATTT CCGATACCTA    720

ATTTATTACC TATTTGATTG GTAAAGCTTT CAATCGTTTC GTTTTCAGGA ACTkGCTCAA    780

CTGTGATAAC CATTAGTTGT GGTTTCGTnT CCgTCTTACT AArTTCTTGa TTAATTTTTk    840

GTACATTATT AAkGACTTCy TCGTTCAAAA TTTGTGCATT AyCTGAAACA AAAAGATGAT    900

TTyGaTTAAT TTTTACTACG ACATCGTCTG TTGGAGCTGA ATAAGTCCTT TCAAATCGTT    960

TTAAGAAACT CATTCCCAAA TAGATCCCTA TTAAAAGAAC AGCTAGATAA ACTAAATTCT   1020

TCATTACAAT ACTTTTTACT TTTATTTTTC GATACTCCCC CACGCTCATT TGTAATTCTA   1080

TCCCTACTTT ATTTTTCCCA TACACTGATT TTTTCTCCCT TCCACTCCTA CCGAATGTTC   1140

TAGCTCTGaT TTTCCACAGA GCTAAATAAA TTTTTTTGTA TTAAAAATTT ATTtTCTTAA   1200

TAGTTAGTAT ACCGAATATA GACGTTCAAA TCTCTATCTT TTTAACTAAT TTCTGTTAAC   1260

ACGCATAAAA AGAACGTAAC GCCTAGTGGA TAAGCGGTAC Gn                      1302
```

(2) INFORMATION FOR SEQ ID NO: 268:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2311 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

```
AAACTTGCAA AAGAGAACGC AAGCAAGATA ATAGAAGACG GTTTTATATA GATAGCTACC     60
```

GnAAAGGGTA CCAGAAACCA AATTAAAAGC CTAGAAACGT TGGAAAATCA ACGTCTAGGC 120

TTTGTTTTTA GACTATTCCC ACTCTGTAGT AGAAAGAATA GTTTGACTTA CTTTTGTTTG 180

TATATCAAGG TTTAAGTTTG GTTGAATTGC TAAAATATGG CTTAAAATGA TATATACTGT 240

GATAAAACTG TGATAAACTT TTTTGATGAA AAAATAGATT TTAGTAATTA AAAATTTGTC 300

TGAATTTATT TATATGCTTA ATTTGAATAA TGTTCTTTTT TAGTTTTATT GATAATTTAA 360

AATTTACTAT GTGTTTTATt GTAGTAATAA TTtAAATATA CAGAAATTAT GGkTTTGTAT 420

AGkAGTTTTT ATTATAAAAG TATTGAAAAT ACAGGTTTTT AATGTCTTTG TTTGCATAGT 480

AAATAAGTTA AAAAATAACT TTTGTGTTAA CACTTATTGC kTTTTGAATT TATTATTTTA 540

AATTAATCTG GAATGATATt GTTATGTATA AAAAAACTAT GTTGTTTAAA CTTATCGACG 600

TGGTATAATT AATATAGGTA TTTTGAAAAT AAAAAGAAGG GGTATTTTAT GTTTACTAAC 660

AATAAAAATA AAAGCTTATA TTTTGAAAGA ATGCAAATGA TTGATATGAT AAGAAATTTG 720

GACAATATTA GTTACAAAGA GTTTCCTAAG TTTATnGGGG AGCTGTACAA TCTTATGGGG 780

TATAAGGTAG AaCTTAATGA GAGAtGAATA ATATTGATTT ATACGCAAGT AAGTTTTCAG 840

AAAAGTTAGC AATACAGGCA AAAGCTTACA GCTTAAmCTA TAATAAGTCG AAaGCAATTA 900

ACmAAAGAAG aGTAGAGTCA TTTGCTTTGC AAGCAAGAGA AAAAAGAAAA AAACCTATAT 960

ATATAACTAC AGGAGTTTAT ACAAATCCAG CATATATAGA AGCTAAAAAA ATTGGTGTTA 1020

CTTTATTTGA CAGAAAAAAT ATTTTTGATT TAATTTCAAA GGCTGATCCT AGCTTACTGG 1080

TCGAGGTTAG TTATTTGGAA TCAGTTGAAA AACACGATTT AAAAAAATGT AAGCATTGTA 1140

AAATTGGGCA TGCTACGAAA GTTTACTCAG AAGATACGAG TAAGTTTTAT TTTATTTGTA 1200

TGGATTGTAA GAAACATACA TAGGGAGAAG TAAAAGTCTA GTAACGTTTA GTTTCTAGAC 1260

TTTTTTATTT GAGTAAAATC TCTTTAAGTT GTGTTTGTTG TTGAAAATAG TAACTATTAT 1320

AAATAAGATT GGCTTGTTTA TAAGATAAGT TGTGGGAATA GATTAATTCG TCAAGAATTC 1380

TTATCTTGAG AACTATATCG TTTACGTTGA AAAAACTAGC CATTTTTTCA AGATCATCAT 1440

AAGTACTAAA TGCATACTTT AAGGCTTTTT TATTAATAAG TAACATTCCG GCACCCTTAT 1500

TTGCTCTGAA TTCTTTAGGA TAATCTTCAG GAAGATAGCC CTTTTTTATA AGTAGTGTAT 1560

CAAATGATTT TTCATTGTTA GGCATATCTG TCGTATAGTC AAAATAATGA TGAATTATTT 1620

CGTGCATTTT CGTAAAATTT TTAATTTCTA TTTTCAAGTG TTGGGCAGTC GCTACCATAG 1680

TTACATCACC AATTTTTGTT ACTGATCCCC AGATTAATTT TTTAGaTTGT TTTTTGAATG 1740

GATAATCAAA AAAATCAATG TTTTCGGTTT CTTTTACGTA GGCTTCAATA TCATAACAAT 1800

TGATGAGTTC TTTTCTTTTA GACATATTGA TTGAAATTTC ATTTATAAGT TTGTCAGTAA 1860

ATGCCTGTGT TTCGAAATAT AATTTTTTCT TATTTAGTAT CATAAAAAAT TACTCACTCT 1920

TTTTTTTTCT TTCTTTTGCT CGTTTTATAA GAAATTCTTG AAAATCAATC AATTCTTTTT 1980

CTAATTCTTC TATGTCATCA GAACTTAAAT CATTAGTGTT TAATCTGAAA TGTGAAATAA 2040

TATTTTGTCC AGCTTCTTTT TGATTATTTG AAGGTGTTGT AGGGTTATCA GTGCGTCCTG 2100

TTAAATAATC AATTGAAACA TTAAAAAAAT CAGCAATAGC AGTAAGTTCT TCTTCTCTTA 2160

ATGGACGTAT TCCAAGTTCT ATTTTATTAT ATACACTAAC ATTAATGTCT AGTGCGTTTG 2220

AAATTTCTTT TTGAGTGATT TCTCTATTTn CTCTtAATTG TgTaTCTTTT CGTAAGTnTT 2280

CATACTATAT CTGTCCTTAT CTAATGTTTA T 2311

(2) INFORMATION FOR SEQ ID NO: 269:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7989 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

CAACACCAAT TTTATATAAT AATAGCGCAG CACCAGTAGG CCCCCACCTA ATACTTAAGC 60

TTGTAATTGG TCGCGACTCA TCAGTAGGTG AATCCTTCGT AATCATTCCT CCGGTATTTG 120

TACTACCATC AGGATGAATG ATCAAACTCT CTCCTTTACC ATTTTTCCAT ATTCCAACTA 180

AAGTAGTAAA ATCGCCAGCA TTAATCGCTG TAGTATCTAA ATCAGCTACA GTTAGCTGAC 240

TATTGTTTCC ATCAGCTGTC GGCTCTTCAT TTTTCACTGA TTCGTTAGAC TCGACACTGC 300

TGCTCGCTTG TTCGTTTGTC TTATCAATTG TTTCACTCGT TTTTtCATTC GTCTCTTCAA 360

CTTTAGAGGA TGCCTCCGTT GAAGCGGTTA CTTTCTTTnC TTTTGGTTCC CGACAaGCAG 420

TTAaTGATAA AACCGCTAGC ACGACAAAAA aCGCTAACAT TCCCTTTTTC ACGCCTATTC 480

ACCCCGTTCA ATGATCTATC TTACTAGACT ATTGTACTAT TAATTCAACT GTTTAGACAA 540

ATATTTATGG AACAGACAAA AGTTGAACCT GTCATTGAAT GCTTTGTGTC ACATGGACCT 600

ATCAATCAAC TTTCTTTTTT CTATGTATTA TATGAAAAAA AGTGAGCCAT CAGTCAAAAT 660

GACCGATGGC TCACTTTCTA AATCTTCAAA TAAACGAACA AATGGATTAT TCAGGTAAAT 720

CTACATTATG ATGGACTTTC TGAACATCTT CTAAATCTTC TAAAACGTCT AACATTTTTT 780

CAAAGTTCCC TAAATCGTCG CCTTCTAATG TTACTTCGTT TTGTGGAACC ATTTCTAACT 840

CAGTAACCGT GAATTCTTCA ATGCCTTTTT CTTTTAACGC TTCTTGTGCA TGGTGGAAAT 900

CTTCTGGTTC TGTGTAAACG ATAATTTGAC CGTCTTCTTC AACAACATCG CGAACATCAA 960

TGTCTTTTTC CATTAAGTAT TCTAAAATTT CATCCGCATC GTCACCAGCA AAGCCAATGA 1020

TACCTGTATT ATCAAACATG TAGGAAACAG CACCGCTGAC ACCCATGTTT CCGCCGTTTT 1080

TACCAAATGC GGCCCGCACA TCTGCTGCTG TTCGGTTCAC GTTGTTTGTT AAAGCATCAA 1140

CAATGACCAT TGAACCATTT GGCCCAAATC CTTCATAGCG TAATTCAGAA TATTGTTCAT 1200

CTGCTGAACC TTTTGCTTTT TCAATAgsGC GATCAATAAT ATGTTTTGGT ACATTGTATG 1260

TTTTGGCGCG TTCGATTACA AACCGTAATT TTTGGTTTGC ATGAGGATCA GGGTCTCCTG 1320

ACTTCGCTGC TACGTAAATT TCAATCCCGA ATTTTGCATA AACACGGCTA TTGTTTGCGT 1380

CTTTGGCAGC TTTTTTCTCT TTAATATTTG CCCATTTACG GCCCATATTC TCACTTCGCT 1440

TTCTCTGATT TGTTGTTAGT ATCTTATCCC TTGTTTATTA TACTGGGGAA aCTGATTTTT 1500

GTTAAGTATT TtTTAGGTAT CTGGCGAAAA TCGTAAGGGA AACCGACCAA AATTTTGTTC 1560

TTTTAAAGAA GAATCTCCCA TTTGGTATAG TTGATgGCGC GAATATCTAA CGAAAGCATT 1620

GCGCGGTTTT CyTTAGATAC TTTGGaAATT ACCGGCAATT GAATCTGTGC yTTGGkTTCy 1680

TTTAGATAAG CTTGTCCCTT TGGCGTAAAT CCTAGCACAT GGAGATAACT GkTTTGcCmA 1740

ACAGTTTCAA CTTCTTCTTT GGTCATATTT AACAATATAT AGGTTGCTAA TCGTTGCAAT 1800

CTTGTCCAAG TATAACGCTT CGTTTTCATA CGTTCCATTA ATTCGTGAAA AGAACCTGCT 1860

GCTTGGATTT GATCTTTTAG TCGATATTCA AGTCCTTCCG TCATTTGGTA AATGGCTTGT 1920

AACTCTTCTA ACGAGGATTG AACAATTTTA TACTTTAGGA AGGGCCAATA ATTTTCCCAT 1980

GAAATCATTG TCTGTGTTTG TAAATGTTGT GCGGTTATAG AAGGCACCGT TGGCAGCACT 2040

TGCGTTATTT CTTGTTGAAA AACACTTTGC CTAATCGCCG TGGCGCTAGC AACTTTCCCA 2100

GAAATCGTTG CATCATGAAA GCCCGCTTGT TTTCTAGCAA TTGGGTAGAG CGTCATCGGG 2160

GTTGGATACG TGGCATTTTC CTTTGCATAA CTCATCCCCA AAATATGATT AGGCGAAGAA 2220

AAATCAAACC TGCTTTCAGG ATATACTTGA CGAAAAACAG CAGTCATTTT TTGTGGATAA 2280

CTCAATTGTT TGTCTGTCAA TGCGTGGAAC GTTTGATCAA TTAAGGATTG ATTTTCTTGG 2340

ACAAACTGGC CGAAAGCCGC ATAATCAATT GCGCTTGTAC TATCTGTTCC AAAACATAAG 2400

CTCTCACATT GCAACGCCTG CAATAACTTT ATACCGCCTT TGGCAAAGTA ATCCGCAGAC 2460

TGTACCGACC AAGCTGTCGG CAATTCAATG ACTAAATCCA CACCATTTTG CAAAGCTTCT 2520

TCTGCTCTGG CCCACTTATC TAGTAAGGCT GGTTCTCCTC TTTGTAAAAA ATTTCCACTC 2580

ATTATAGCAA TCACTACTTC AGCTCCGCTT TGTTGGCGAG CTTGTTGGGC ATGATAGCGA 2640

TGTCCATTAT GAAAGGGATT ATATTCCACG ATAATGCCAC AGGCTTTCAC TCTAGCCACC 2700

TCCGTTATAA AAAAATGGAC TTATCAGCAA AGAGACAAAC GTTTTTAAGC GCTGTCCCAA 2760

GCTCATAAGC CAAATCTTAT TAATCTCTTT GACAAACAAA GAACCAGCGT TCACTTGTTT 2820

CAGTTGGTGC TTCATCTGTA AAATCTGCAT AAACTTTGAC GTTCATGAAA CCAGCACTTT 2880

CTAACATTCG TAAATAGTTA TCTAACGAAT AGGTTCGTTC TTGATGCAAT TCATCATACC 2940

GAACAAAGTG CTCATCTGTT TCCTGTTCTT TCACAAAAAA TGTTAAGAAA TGTTCGATGC 3000

TGTGTTCTTT TTCACCAGCG TAACTATCCC ATAGAAAAGC AAATTCTTCT GTTTGATAAT 3060

GATAGCTATA TTCTGGAAAA ACTGAATCAA TTTTGTAAGT AGAATGAACA TCAAAAATGA 3120

ACGTTCCTTC CTCTTCTAAC GCTTGATAAA CTTCATCGAA CACTTGTTGT ACTTCTTGGC 3180

GaTTAGCCAT ATAGCACAGT GAATCAGAAA AACAAGTGAT GGCTTGGTAT TGTCCCACTT 3240

CAGATAAATC TAGCATATTT CCCTGAACAA ATTGAACTGG CACTTCTTCT TCAAAGGCCC 3300

GTTCACTAGC AATCATCAAC ATTTCTTCTG ATAAGTCTAA GGCCGTCACA TCGTAACCAC 3360

TTTTCGCAAA ATCAACTGCT AAAGCGCCTG TACCACAAGC CATTTCAAGA ATTTGTTGGG 3420

TTCCTTGTGG GAGATGACGA TTGGAAAAAT CCAACCATTT TTGGTATAGG CTTTCGTCCA 3480

TCACTTCGTC ATAGACGAAA GCAAACGTTT CATAGGCCAT TTTAGTCGAC CCAAGCGCTT 3540

AGATCAACCA TTGGGGCATC TGACCATAAT TTTTCTAAAT TGTAAAAGGC ACGTTCTGAA 3600

GATTGGAATA CATGGACAAT AACATCGCCT AAATCGATTA AAATCCATTT TCCGCCTTCT 3660

TTACCTTCCA CACGTTTGAC TTCAACTTGA TTTTTGGCTT CTTGTTCGTT AATTTCATCA 3720

ACAATCGCAT TGATTTGACG TTCACTATTG GCTTGGCAAA TTAAAAAATA ATCTGCTAAT 3780

AATGAAATCT CTCTTACGTC TAATGCGACA ATTTCTTCTG CTCGTTTTGA ATCTGCTGCT 3840

TTGACAGCAA TTTCTAAAAT TTTCTGGCTA TCTATGATAG TTCCCTCCTA GTAATTCTTT 3900

CTTCTTTTAT TTTTGATTAA CGACCCATTG ATTGTAGGTC TCTAACGTTT TTGGATAAAT 3960

TTTATGTTCT TGTTCAATCA AATGTAACAA GGTATGTTTC GTTTCATAAG CAACCGCTGC 4020

ATCTAAGTCA ACAAGGGCAA TTTCACGGGC TTCTTTCACA CCTGGGAAAA TCACGTCCTG 4080

GTTCAATGGT AATCTGCGAC ATAAATAATT TTATCTAGCA AGCTCATCTT AGCAGCCCCA 4140

GTTGtATGCA CACGaATTGc CTGCAAGATT TCTTCGTCAT CAATGGCTAA TTCCCGTTGA 4200

ACAATATCCG CCCCAACTAA GCCATGCCAA ATTTCATTGC CATAGTTTAA TAAAGCCAAA 4260

TCAAAACCAT CTCGGCGAAT AATCAACTCG AATTCATCGT TTGGACGTTC TTTGGCGTAA 4320

TCATGCGTCA AAGCAGCAAT ACTGGCTTTC GCTTCATCGG CACCATATTT CGCTGCCAAG 4380

GCCACAGCTG TTTCCTCGAC TCCTAACACA TGGCGAAAAC GGCGTTCACT CATTTGCATT 4440

TGAACTTTTT GCATTAATAA TTCACGTTGT GCAGCATCAT AATTTGTTAT CCAGCTCATC 4500

TTGATACAGT CCCTTCTCTT GAATATAGTT TATCACATTC TCTGGAAGAA GATAGCGTGT 4560

TGAGCAACCA CTTTTGACTT TTTGTCTAAT CAAAGTGGAA CTAATCGCCA TCTGCGGAAC 4620

ATCGACCCAA ATAATTGGAT AGGTAGATTC CGTTGGATAA TTAGGACGCC GAATGCCGAC 4680

AAATTGAACT AAATGGAGTA AGTCATCAAT CCGATGCCAT TTGGGCAAGT ATTCCACCAT 4740

ATCTCCACCA ATTATAAAAT AATAATCAGT ATCTGGATTG GCTTCTTTTA AAGCTTTCAT 4800

CGTATCATAC GTGTAACTTT TGCCTTTGCG AATTAACTCA ATTGGCTCAA TATCTAAGCA 4860

GGGATTATCA GCCACTGCTA GTTCTAACAT CGCTAACCGA TGTTCACTGC TAATTGTTTT 4920

CTTTTCGTCT ACATGTGGTG GTAAGTACGT AGGCATTAAG TAAACTTTGT CCAAACCTAG 4980

TTGATTTTGA ACTTGATCGG CCATGACAAG ATGTGCTAAA TGAACGGGAT TAAAATTACC 5040

ACCTAACAAG CCGACTTGCT TTCTTGTTTG AAAACGTAAC GGTTCCTCTT GCAAGAGAAC 5100

GTCCGCTCCT TGACGCGCTT GTCTTTTTTC TCCCATTGTT CATCCCTCCA ATATTAAATC 5160

TCTTTGACTG CTGCTGAAAT TTTTTGATAT TTTTCTTTAC TTGATGGCTT GAATAAGACC 5220

AACACGCGAC CAATAATTTG TACCACATCG CAATGAATGG CTGCTGTTAG TTCTTCTGCA 5280

ACCTCTTCTG CTACTTCATC TGTATTTTGT AATAACGTAA TTTTAATCAG TTCACGTTTT 5340

TCCAACGCTT CATCAATTTG TGTAATCATT GCGTCATTCA AACCACCTTT GCCTACTTGA 5400

AAAATAGGTT GTAAATGGTG TGCCTGACTA CGTAAATAAC GTTTTTGTTT TCCTCTAAGT 5460

TTCAAATTGT TTGCCTTCTC TCTAAAATTT AAATTAATGC TTTTCTACGT AAAACATCGA 5520

CACCTTTAGG CGCCCAACCA GCGACTACAC CAGGTTCTGT GACAGTAATC CAACCTAGAC 5580

CAGCAAACAC AATATCCGTT TTTTCTTTCA CGGAAAATTC AAAGCGCACT AATTCAGGAA 5640

ACTCAGCGAC TTCATCGGCA CGTGGAGGCT GCAATAAGCC ACCAACATGT TTTTCATAAA 5700

AAGCATCGGC GGTTGCTGTT TTTGTTCGGT GTAATTTTAA GTCATTGGAA ACGTAGGCAA 5760

TGAAGGAACT TTTAGCACCT TGGACATAAT CAAAGCGAGC CAAGCCACCT AAAAACAAGG 5820

TTTGTTCAGC ATTTAATTGG TAAACTTTCG GTTTGATTTC TTTTTGCGGT GCAATAATTT 5880

TCAAGTCTTT TTTaCCTAGG TAATGTGCCA TTTGATGACG GTGAATAATT CCGGGCGTAT 5940

CGATTAAGAA GTGGCCATCG TCTAATGGAA TTTCAATTTT ATCTAGCGTT GTTCCAGGAA 6000

ATTGTGATGT CGTAATCACA TCTTGAACAC CCACTGTTTG TTGAATGATT TGATTGATTA 6060

ATGTTGATTT TCCAACATTG GTCACACCTA CGACATAAAC GTCTCGGCCT TCACGATATT 6120

TTTCGATCGT ATCTAATAAC TCTTGCATTT CTTGTGGTTT CTTGCCACTT GTTAGTAAAA 6180

CATCAACAGG GCGTAATCCT GCTTCATGGG CCCGTTCTCT CATCCATTGA ACCATTTTAG 6240

GTTTTTTTAA TGATTTAGGC AAAATATCCA CTTTATTGCC AACTAATAAC ACAGGATTAT 6300

CTCCGATAAA ACGATGTAGT CCTGGAATAA GAGAGCCATT AAAGTCAAAA ATATyCmcTA 6360

CGTTCACAAT CAGTGCATCT TCTGTTCCTA AACCATTCAA TAAACGTAAA AAGTCATCAT 6420

CTGTTAATTG GACATCTTGA ATGTCATTAT AATGTCTTAA GCGAAAACAA CGTTGGCAAT 6480

AAACTTCCCC TGTTTCCATT CCTTTTTCGA AAGCTGTTTT TGGTGTATAG CCTAATTCAT 6540

GAGGATTTTC CGTTTGGATA ATCGCACCAC AGCCAATACA ATGAATTGCC TCTGTCATTC 6600

TAATCCGCCT CTCCATGTCA TTTCTGGATG TTTTTTTGAT AAGTATTTCA TAATTTTCCG 6660

TTCAAAGAAC CGATTGATTC TAGTATTCCA GCCATCTGTA TCGACAATCG GTTGAACTAA 6720

AACATTGCGT ATACCCGCTG CGTTAgCTCC CCGAATATCT GTCATAATCT GATCGCCGAC 6780

CATTAACATT TCAGATGGCT TCAAGCCTAA TTTTTTTTCT GCTAATTTGA AGCCACGTGC 6840

GGTTGGTTTC AAGGCACGTG CTACATAATC TAAGTCAAAC TTTTCAACAA CGCGTTTAAT 6900

CCGGCTGTCT TTGTTGTTTG AAACAACGAG GACAGTAATT CCTGCATTTT TCATTTCTAA 6960

TAACCAAGTT TTTAATTCTT CTGTTCCATC GGGATTATTC CAAGCAATTA ACGTATTATC 7020

TAAGTCAGTT AGTACTGCCT TAATCCCTAG TTTTTTTAAT TGTGCGGGGG TAATTTTATA 7080

AATCGCATCA ATCATCCATG TCGGTTTATA TTTTGAAAAC ATCGTGACTC CTCTCACTGA 7140

CACTATGTAG TTTCTTCGAT TTTAAGAGAC TGAAACATAA CTAAACAATT ATGCTTCAGT 7200

CTCTACCTAT TGTATTATAC TGCATTTTAC TGAAAATTTC TATAAAGGAC GGTCGTTCCT 7260

TGTAAAAGAA CAAGTAAAAT AAAAGAAAAA ATAAAAAAAA GCCACCCATC AACCCCTGAA 7320

AAGTTCGTTG ATGAGTAGTA ACTTCTAATA TTTGCGAAAA CGTTGGTATC TTTTTTCTAG 7380

TAACGTTTCC GTATCTAATT TAGCTAATTC AGTTAATTTT GAAATGAACG CTTTTTGTAA 7440

CATGCGATTA ATTTTTGCTT GTTCCAGAGG CTGACCGTTC ATAACTTCTG GAATTACTTT 7500

ATCTACAATA GCTAATTCCT TCAGCTCTGT CGCTGTGATT TTCATCAATT CTGCTGCTTC 7560

TTTCGCACGA CTGCCATCTT TCCACAAAAT AGAGGCGAAA CCTTCTGGCG AAAGAACCGC 7620

ATAAATCGTG TGTTCCAACA TCCAAACTTC ATCTGCGACT GCTAAAGCTA AAGCACCGCC 7680

ACTACCGCCC TCACCAATGA TGACTGAAAT GATTGGCACT TTCAAATCAG ACATCTCTAA 7740

CAGGTTTTGC GCAATGGCTT CCCCTTCTCC CCGCTCTTCG GCACCAATCC CACAATAGGC 7800

ACCAGCAGTG TTCACAAAGG TAATGACTGG GCGATTAAAT TTTTCCGCTT GTTTCATTAA 7860

ACGTAACGCT TTGCGATACC CTTCAGGATT AGGTGAACCA AAATTACGTT CAATATTTTC 7920

TGGTAGATTT CGACCTTtTT GGaTACCTAC CACTGTAACT GGGTTTCCCT TGACAAAGTA 7980

GCAACACCA 7989

(2) INFORMATION FOR SEQ ID NO: 270:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5136 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

AAACACAAGG GAAATTTGAT TATTCCTCAA AATATAGCGG TGTGCTTTGT GAAAAACATT 60

GGCATTTAGA TGAACAGCGC TACCATGCAG ATCCTAGAGC AATTCATTTT ATTCGCTTAT 120

TTTCGCAAGT CTCCTATGAA AAAGTCCAAA ACATTCAAGT AAAAGAAGAA ACTAAGAAAA 180

GTATTCGAGA AACGATTGAT ATGTTATACG ATGAATATGT TGGTTTGCAC TTAAAAAGTA 240

AAAAATTTAT TGATCAAATG AAAACGTGGG AAAATACGTT AAAAATTCCG CCACGAAAGA 300

AGGAAGAAAA ATAAACATwA ACAAGCAGAC GTTTTtCTAA AAAAAATTTG ACAACTTCCC 360

TCAGACACCG TATTATGATA GAGAAGTTAA ATACGTACGA AGAAAAAGAG AAGTAAAAAG 420

AACCCTCTGT TAAGCGAATC TGGGAGAGTG GGAGCCAGAA ACACGGAACT TTTGAAAGGC 480

ACTTTGGAGT ACGACAAACG AAGCTGCCGA TGAACACATC GGAAGTAGGG TGGAACCGCG 540

ATAATTATTC GTCCCTATGT CTATTAATAG GCATGGGGAC TTTTTTCTGT GAATCACTAC 600

GATTGGCGCA AATGCGACAG AGTAGATGAT GAACAGTACA AGGTGACCAA AGGGAACGTT 660

GTCTCCTCTA TGTTAAAACA AACTGCTTAG CCAAAAACAT GGAGTAGATG ATGAACAGCA 720

CAAGGTGACC AAAGGGAACG TTGTAACCGA ATAAAAAAAG CAAACAGTTA ACACCAAAAG 780

AAAAAATAGT AACGTTTATT CAAAAAAGCT CAATTGGCGC AAAGGCGACA GAGTAGATGA 840

TGAACAGCAT AAAAATTATC CCATTCTACA AAATTGTTTC TCTGTAGCCC TTTTTAAATT 900

TACATTcCAG AAgCAAAGCC AATCCCTATG TCTATAAAAr ATATTCCAAG GaGGaAAACA 960

ATGaAAAACa AGTTAACCGT ACaAGAAATG aTTTTAACCt TGCaAAAATT TTGGTCaTCT 1020

AATGGCTGTA TGTTAATGCA AGCCTATGAT ACAGAAAAGG GAGCCGGTAC CATGAGTCCT 1080

TACACGTTTT TACGTGCCAT TGGACCTGAA CCATGGAATG CTGCCTACGT AGAGCCTTCT 1140

CGTCGTCCAG CTGATGGACG CTATGGTGAA AATCCGAACC GTCTATACCA ACATCATCAA 1200

TTCCAAGTTG TAATGAAGCC TTCTCCTGAA AATATCCAAG AACTTTATCT AGAAAGTTTG 1260

AAATTATTAG GAATTGATCC ATTAGAACAT GACATTCGTT TTGTAGAAGA TAACTGGGAA 1320

AATCCCTCAA TGGGTTGTGC TGGTTTAGGT TGGGAAGTTT GGTTAGATGG CATGGAAATT 1380

ACCCAATTTA CGTATTTCCA ACAAGTCGGG GGCTTACAAT GTCATCCTGT CACTTCTGAA 1440

ATTACTTACG GGTTAGAACG CTTGGCATCA TATATCCAAG AAGTCGAAAG TGTCTATGAT 1500

TTAGAATGGA CACAAGGCGT TAAATATGGT GAAATTTTTG TTCAACCTGA ATATGAACAT 1560

TCAAAATATT CCTTTGAAAT TAGTAACCAA GAAATGCTTT TAGAAAACTT TGACAAATTT 1620

GAAAAAGAAG CTAAACGCTG TATTGAAGAA AGTTTAGTTC ATCCAGCCTA TGATTATATT 1680

TTAAAATGTA GCCATACATT CAATTTATTA GATGCTCGCG GCGCTGTCTC TGTAACGGAA 1740

CGTGCAGGAT ATTTAGCCCG CATTCGGAAT ATGGCTCGTT CCGTAGCGAA AATTTTTGTG 1800

GCTGAACGTG AAAAATTAGG GTTCCCATTA TTAAATAAAG ACCAACACGT ATCCAAGGAG 1860

GCAGAATAAC ATGGCAAAAG ATTTATTATT AGAAATTGGC TTAGAAGAAA TGCCCGCACA 1920

TGTTGTCACA CCAAGCCGTA TCCAACTTGA AGAAAAAGTA ATCAAATTTT TAGATGAACA 1980

TCATTTAGAC TATGAAACAG TGCAATCCTT TGCCACACCT CGTCGTTTAG CCGTGAAAGT 2040

AACTGCGATT CCTGAAAAAC AAGCGGATGT TGAAGAAGAA GTAAAAGGTC CAGCGAAAAA 2100

AATTGCTTTA GATGCTGAAG GAAATTGGTC AAAAGCTGCA CAAGGTTTTG TACGCGGCCA 2160

AGGAGTTACA ACTGAGGACA TTGTTTTCAA AGAATTAAAT GGTGTCGAAT ATGTCTATGT 2220

GACGAAGTTT ACCAAAGGCC AAAGTGCCAA AGAGGTATTA ACAAAATTAA ATGACGTGAT 2280

TACAAGTCTA ACTTTCCCGG TAACAATGCA CTGGGCAAAC TATGATTTCG AATATATTCG 2340

TCCAATTCAC TGGATTGTGG CACTACTTGA TGATGAAGTA ATTCCTTTCA AAGTGTTAGA 2400

CGTGACAACA GGGCAAACTT CTCGTGGCCA CCGTTTCTTG GGAGACGATG TGACGTTCCA 2460

ACATGCTAAT GAATATGAAG CAAAATTAAA AGAACAATTT GTAGTGGTTC AACCGAATGA 2520

ACGGAAACAA ATGATTGTTG ATCAAGCGAA CGCTCTCGCT GCAGAAAAAA ATTGGCAGTT 2580

AGCTTTAGAT GAAGAATTAT TGGAAGAAGT TACGAATTTA GTGGAATATC CAACTGCATT 2640

TGTAGGTAGT TTTGATGAAA AATATTTATC TGTTCCAGAT GAAGTATTAG TGACTTCTAT 2700

GAAAGAACAC CAACGTTATT TTGACGTTCG TAACGACCAA GGCTTATTAA TGCCGCATTT 2760

TATCGCTGTT CGCAATGGGG ATAATGTTCA CCTTGAAAAC GTGATTAAAG GAAATGAAAA 2820

AGTGTTAATT GCTCGTTTAG AAGATGCAGA ATTCTTTTAC AATGAAGATA AAAAATTAAC 2880

GATTGAAGCA TGTGTTGAAA AACTAAAAAA TGTGACTTTC CATGAAAAGA TTGGCTCAAT 2940

TTATGAAAAA ATGCAACGGG TTGCTTTAAT TGCACAAATC ATTGGTCGAA AAGTTGGCTT 3000

ATCTGAGGAT GAATTGGAAG ATTTAAAACG TGCCTCTGAA ATTTATAAAT TTGACTTAGT 3060

AACAAATATG GTTGGTGAAT TCCCAGAATT ACAAGGAATT ATGGGAGAAA AATATGCCTT 3120

ATTACAAGGC GAAAAACCAG CCGTAGCAAC GGCGATTCGA GAACATTATC TACCAACTTC 3180

CAGTGAAGGT GAATTGCCAG AAACAGCAAT CGGGGCTGTT TTAGCCTTAG CCGATAAGCT 3240

TGACAGCGTC TTTTCATTCT TCTCAGTTGG TATGATTCCA ACAGGTTCTA ACGATCCTTA 3300

TGCAcTTCGT CGTCAAACCT ATGGCGTTAT CCGAATTATT GAAGATAAAG GCTGGACCTT 3360

CCCATTAGTT CAATTACAAA CAGAAGTTGA TGAAGCAGTG AATCAAGATG TTGAGAAATA 3420

TGGCGTTTTA CTAAACGAAG GACAAGCAGA AGTCGTTGAA TTTGTGAAAG CTCGCTTACG 3480

TCAATTATTA ATGACGAAAA ATGTTCGCCA TGACATTATC GATGCAGTGG TGTCTGCTGA 3540

GCAAGCCGAT TTATCGAAAT TATTTGCTTC AGCGAATATT TTAAAAAGCC GTTTTGAAGA 3600

TCAAGATTTT AAACCATCAA TGGAAGCCTT AACGCGTGTT ATTAATTTAG CGAAAAAAGG 3660

TCAAGAACTT TTAGGTGATA CTGAAGAAGG GATTGATCCA AGTCTGTTTG AAAATAAGGC 3720

AGAAAAAGAA CTGTATCAAG CAGTTAACGA TTTATCTGAA GCATTTGCGA CACGCACGAT 3780

TGCAGAAAAC TATGAAGCAC TAGTAAATCT ACGTCCGTTA ATTGATGCGT ACTTCAATGA 3840

AACgATGGTT ATGGTTGAGG ATGaAAAAgT CaAACAAAAC CgATTaAAAC AATTAATGCA 3900

AATTGCTAAA ATGGCGTTAT CaATTGCTAG CTTAGATTTA TTAATTGTGA AATAAAAAGC 3960

GCTGTTTTTC GACCGCTAAG GAGGAACCGA CAGTATGCTA GAAACAACAT TTGAACAAAT 4020

TTTAACACAA TTATCAAAGC CAGCTGTTCG TGCGCTAACG AATGAAAAAA TCGACAGTGT 4080

GGATGAACTC TACGCACGAG GCCGAAAAGC GTTACTTTCG TTGCATGGCT TTGGTCCAAA 4140

GTCTATTCGA ACTATTGAAG AGATGACTGG GAAAGAATTG AAGTAAAAAA AGGATCGCCG 4200

TTTAGAGGCG ATCCTTTTTT TATTTACGTT TTTGTTTTCC AGCATTACGA CCTTTATTGC 4260

CGTTTGCTTT TGGTTGGTTC GTTTTTTGCG TTACTTTGTT TGTTTCTTTT GGTGTCACAT 4320

CCTTGATAGG AGTAACCACT TGTTTTGGTG GGTTATTTTT CAATTCTTCT TGTACTTGTT 4380

TTTTAATTCT TGGACGTAAG AGGATATTTG TGATGGCAGA TTGAATACAT GTGAAAATAC 4440

CACCAACTAC CCAGTAAAGC GTACGCCTGC AGGTGAGCTA AAGGACATGA AAACAATCAT 4500

TAATGGTGAA ACAATCAGCA TTGATTTCAT TGTTTTCTTT TGTTCTTCAG GAATACCAAT 4560

CATTGAAATA TAGCCTTGAG CTAAATAGGC TAAACCAGCT AATGCCACTA AAATCATATT 4620

TGGTTGCCCT AAATTCATTC CTAAGAAGCT TGCCGTGGCA ATACCTTTTG TATAACGAGC 4680

AGCAAAGAAC AGAGAAGAGA AAATAGGCAT TTGAATTAAT AAAGGCAAAC AACCAATTCC 4740

GCCAACCATA CTCACATTGT TTTCTTTATA AATCTTTTGT AGTTCTGCTT GTGCGGCCAT 4800

TTGTTCTTCA CGCGTACTTG CTTGTTTCAT TTTTGCTTGT GCAACATCCA CTTGGGGTTT 4860

GATTGCTTGC ATTTTTTCCG TTTGGATAAA GCTTTTTTTC GATTGATGTA AGCCTAACGG 4920

TAAAATAATG ATCCGTACAA TAATGGTGAT AAAAATAATC GCCCAACCAT AGTTCGAATT 4980

AAAGTTATCG ACAAGATAAG TAATTGCATT ACTCATCGGT AAAACAAGGA AGTTATAGAC 5040

AAAACCTTCA CCAGTTGGTT GTCCATCTGC TCCTGTTTTC ACACAGCCTG ATAAAAACAT 5100

TACTAAAAAT AGTAAGCCAG AGCCGTAAAG CCATAC 5136

(2) INFORMATION FOR SEQ ID NO: 271:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13808 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

GGCGAAATTG TTGAAAAAAA ACTAGTTACA GCACCTCTGT CAACTGAACA AATTGATCAA 60

GCAATGGCCG AAATGACAGG TGAAaTCACA CAAATCCCAC CGATGTTTTC TGCGGTGAAA 120

GTCAATGGTA AACGTTTATA TGAATATGCG CGGAACGGGG AAGAAGTTGA ACGGCCCCAA 180

CGAAAAGCTA TGATTTATTC ATTTGAACGC ACTAGTGAAC CAATTTTTAA CGAATCCGCT 240

CATACACAAA GTTGGCGGTT TAAAGTGGTC TGTGGCAAAG GAACCTATGT GCGGACACTA 300

TCAGTGGATA CTGGCAAAAA ATTAGGCTAT CCTGCCCATA TGTCGGATTT AACGAGAACG 360

GCAAGTGGCG GTTTGCAAGc tGCGCAATGC TTAACGTTAG AAGAAGTAGC CAAACAAATG 420

GCTGCTGAAA CAATCGATCA ATGCCTGTTG CCCATTGAGA CAGCAGTAGA ACAATTCCCA 480

CGAATTGATT TATCAGATGA ATTGTATCmA AAAGTAAAAA ATGGCATGCG TTTGCACmAA 540

AAAGAATTAG GrATAAAGGC AaTGCCAGAA TCCTTAGTCG CCTTATTTTA CCAAAATCAA 600

GTGGTTAGTT TATACATGCC ACACCCAACG CATGATAAAT TATTGAAACC AAGTAAAGTT 660

CTACGGAACA ATTAAATAAA AAAGAAGGAA GTTTTTGCAC ATGCAAGTTA TTCAACTACA 720

TCATCCCTAT GAACCCAATC AAATTCCTAA TGAAGAAGTC GTGATGGTCC TAGGCTTTTT 780

TGACGGTGTT CACAAAGGCC ACCAAAAAGT AATTGAAACA GGTAAAAAGA TTGCGGAGGA 840

AAAAGGCTTG AAGTTAGCTG TGATGACCTT CAACCAAcAC CCAtCAATCG TTTTTCAAAA 900

AGTtTTACCA GAAAATATGA AATATTTAAC TAGCTTgGAA CAAAAAGAAC GCCTAATGGC 960

CAACTTGGGC GTTGATATTT TATATATTGT CGAATTTACT TCAGCGTTTG CTCAATTAAA 1020

GCCGCAAGAT TTTGTCGATC AGTATATCGT TAATTTAAAT AGTGCAGTGG CAGTCTCTGG 1080

GTTTGATTAT ACGTATGGTC CTAAAGAAAT TGCTGGAGTC AAACAGTTAC CAACTTATGC 1140

GCAAGGCCGT TTTGAAGTGG TGACAGTACC AAAAGAAGAA ATGACTGGTG CCAAAATTAG 1200

CTCAACAAGA ATTCGCGAAA AAATGGAAGC CGGCGAAATG GAGGAGGTCA CGGAGTTGTT 1260

AGGATATATT TATGAAACAG AAGGAACAGT TGTCCATGGC GATGCTCGTG GCCGTTTATT 1320

GGGTTTCCCG ACCGCTAATG TCAAAGTGAA AAGCACCGTG CGTTTACCAA GAATTGGCGT 1380

CTATGCTGTG CAAATCGAAG TTGGTGGCAA ATGGTATGTC GGCATGGGCT CAATTGGTCA 1440

TAATGACACC TTTGGAGAAG GACGCGATTT AACCGTGGAA GTCTATATTT TGGACTTCCA 1500

TCAAGACATC TATGGTGAGC AAGTGGTGGT TCGTTGGAAT CATTATTTAC GTGACCAAGT 1560

GAAATTCAAT GGGGCTGAAG CCTTGATCGA CCAATTAAAA CAAGATGAAC GAGATACGGC 1620

TGACTATTTC CAACAATCTG AGGACAAATA ATGAAGATTC GTCAAGAAAA ACCTGCAGAG 1680

TATCAAGCTG TAGAACGTTT AACGTATCAA GCGTTTAAGG AATTAAACTT GACCGAAAAC 1740

TGGCGTCCCA CAGAGCATTT TATCGTGCAT TTATTAAGAG AAAGTGCCGA CTTTATTCCG 1800

GAGTTGTCCT TAGTTTCAGA AaCTGATCAA GGAAAATTAA CGGGTCATAT TATGTCTTCA 1860

AAAGCCAAAC TACGATTACC TGATCATACA GAAAAGGCTG TTTTGACCGT TGGGCCGCTA 1920

AGTGTCCATC CAGAAGCACA GAACACGGGC GTAGGCTCAG CGTTAATTAA GCATTCTGTT 1980

CAGAAGGCCA AAGAATTAGG CATCGGCGGA TTGATTATCC TGGGCCATCC AACTTACTAT 2040

ACGAAGTTCG GTTTTGTTCC AGCGACTACT TTTCAGATTA CGTTACCGGA AAAAGAAACA 2100

TCAGAAGCAT TATTGGCTTT GGAATTACTG CCAGGTTATT TTGGCACAAG TGGTGGGGAG 2160

TGGCATTTTT CAACATGCTT TGCTTATCCA GAAACACACC CTGCAGAATT AGAAGCCTTT 2220

GAAGCAGAAC TCGGAATAAA TGAATAAAGA ATAATAAGAc GCGACAACTA CTAGTTAAGT 2280

ACTGGTGGTT GTCGCGTCTT TACTACTTTT TATTACAAAG TAGTTGACAG AACCGGTACT 2340

CAGTATTATT ATGTAGTGGT ACGTAGTACT AAAGAGTAGT TTAGGGAGGA AAAGCAGTGA 2400

AGCAGACAGA GTTGCTAAAA GGCATTTTAG AAGGTTTAGT TTTAGCCATC ATCCAAAGGA 2460

AAGAAACATA TGGTTATGAA ATTACAAAAA TTTTAAATGA CCAAGGGTTT ACTGAAATTG 2520

TTGAAGGAAC CGTTTATACG ATTTTACTGC GTTTAGAAAA AAATCAATGG GTTATAGCTG 2580

AAAAAAAGCC kTCAGAAAAA GGGCCAATGC GGAAGTTCTA TCGCTTAACC TCATCAGGAG 2640

AAGCAGAACT CGCTGATTTT TGGCAGCGAT GGACGTTGCT TTCAAAACAA GTAAACAAGA 2700

TGAAAAAAAA TGGAGGGATA GAGCATGTCA AATTTTAAAG CGTTGATCAA AAAAGTTGTT 2760

GGGGATAAAA AAGAGTACAA AGAATATAAA CGACGGATTG CAGCATTACC AGTAGAGTAT 2820

CGACAAGTAT TTCAAGAAAT TGAAAAATAT GCTTGGCATT TTTCAGATCA TAGTGGAGCC 2880

AATATGTTCA ATGCGTTGAC TGATTTACTT GATTTGTTTG AAGAAGGAGC TGCTAACGGA 2940

ACACCTATCA AAAATATTAC GGGAGAAAAA GTAGGTGATT TTGCTGAAAC AATTGTCAAT 3000

GAAGTCGCTG GCAAATGGAC TGATAAACAG AAAGAAAAAC TGAATCATCA ATTTTCTAAA 3060

CAATAGTCAA AACAAAAAAA GGAGAATAAA CATGGTTGAA AATTTTTTTA AAGAAGGAAA 3120

GTTGCTGGTT ATTCCCAAGA AAAATGCGAA AAAAGTCCAA GTGTTTCAAC GTGTGGCCAA 3180

CCAATTTGAA TTTGGGAAAG AATATACGGA AAAAGAAGTC AATCAGCGAC TTAGAGAAAT 3240

CTATGAAGAT TATGCACTTT TGAGACGGTA TTTAGTAGAT TATCAGTATT TAGAAAGAGA 3300

TAAGTTTGGA AAAATTTATC AGAAATGTGA GCAACACACA AAAATTATAT AAATGGCTTT 3360

TTTTAAAAGT AGCGAAACTA AGTGAAATGA ACCAAGTCTA GCTACTTTTT GTTATGTAAG 3420

CTTAGTAATA GTTCTATAAA AAACTCTCTA TTTGTTTTCT AACTCAGACT GGTAAGATTG 3480

ATTTAATTAG AAAAAGATAG AGGTGAGAGA ATGAATCTAC GTACTTTAGG TCTAACAAAA 3540

ATTCTTTTGA CCTTTATTAT ATTAAGTATT ATTGGCGCGT TAATGATTTA TAGCGCCAGT 3600

AGCTATGACT TGTTGATGCA AGGCGTGAAG CCCACTGCTG TGTTTATTAA ACAAGGCATT 3660

ATTATGTGTC TCAGCTGGGT GCTAATGTTC GTGATTTACA AAGTTCGACT AGAAGTGTTA 3720

TTTAATAAGA AAATTGCGAT TGGACTGCTT TTAGTTTCGG TACTATTACT GTTAATGGTT 3780

CGACTACCTT TTTTTGGTGT AGCAGTGAAT GGCGCACAAC GTTGGATTTC TCTCTTTGGA 3840

ATCCAGTTTC AGCCATCAGA ACTTTGTAAT TTTGCGATTA TTTATTACTT GAGCTGTTAT 3900

TTAGGTGAAA AAGAAAATGG CTTGACAACG AAGCAGTTAA GGAAACAATG GCTGTTTGTT 3960

TTGGTCGTTG CATTTTTAGT TTTGATTCAA CCCAAAGTTG GCGGGGCCAT TCTTATTTTA 4020

GTTATTGGCA GTGTCTTGAT TTTCTCAGCA AGTATTCATG CAAAATTTAG TGTAATTGCT 4080

GCGGGGATTG TCGTTGCTAG CGCGGCGCTT CTTTCTAAAA TAATTATTTT TCTAGGCGAT 4140

CACCGCTACT TACCACATTT TTTTGCGCAT GTGTATGACC GCTTAGTTGT TTTGAAAAAT 4200

CCGTTTTTAT CCTTTCATGA TCGAGGATTT CAGCCTTCAA TGGCCTATTT AGCAATGTAT 4260

AACGGTGGGT TTTGGGGCAC TGGACTAGCC AATGGTATGG TAAAAAAAGG TGGCTTGCCA 4320

GAAGGTCAAA CCGATTTTAT CTTTGCGGTG ATTGTGGAAG AACTTGGTTT GATTGGCGGC 4380

TTACTGTTGC TATTTCTTTT GTTATTTTTA GCGGCCTCGA TTTTGCGGAG TAGTTGTGTA 4440

ATAAAGAATC ATTGTTACGG CTTGTTTTTG TTAGGGGTCG GGACCTTGAT TTTAGCTCAA 4500

ACGGCGATTA ACATTGGTGG CGTTTTGGGT TTGATTCCAA TGACGGGAAT TCCTTTACCG 4560

TTTGTTAGTT ACGGTGGGAC AAGTTATTTA ATCTTTTCAG TGGCATTAGG AATCGTCATT 4620

AAAATTATTG CAAACGAAAG GCGGCAGTTA AATGGTCAAT ACAAAAAAAT TCAACTTACA 4680

TCTTAATTAT GATTTATTAT TGCCCATTTT TTTACTCACA CTTTTATCTT CTGGTGTGCA 4740

ATATTGGATT GCCGTTAATG AAGGAAAAGA TGGAACTGTG CCaGCATTAA AACAACTTTT 4800

CTTTATTTTT GTTGGCTATG CAGGCATGTT TTTAGCTAGT CGGCTGTCAC AGAAATTTAT 4860

TTGGAAAGTA GCTCCTTTTT TCTATGGGTT TTCATTAATC TTAATGAGTG CTTTGTATTT 4920

TTCCTATGAC AAAGGAATGT ATTTATTGAC GGGCACAAAG CGCTGGTTAG ATTTAGGGTT 4980

TATTAAATTT CAGCCATCAG AAATTGCGAA AATTGCGTTT ATTTTAATGT TGGCAAAAAT 5040

TATTGTCCAA CATGAGCAAC AAGACTGGTC TGATAAATGG CGTTCAGACA AGCAACTTCT 5100

AAAAAAAATC GTTGCGGTTA GTGTCCCTGT TTTTTTCTTG ATGGCGGTAC AAAAAGATTT 5160

TGGGACATCC CTTGTTTTCG TAACGATTAT TCTATCTCTT TTAGTTATTT CAGGAATTGA 5220

TCGCAAAATT TTAATCATTA TTTTCAGTGC TTTAGCAACG TTAGGCGTGG TCTTGATTCT 5280

GTTAGTCTTT ACAGAGTGGG GGCACAAAGT ACTCTTCTTT TTACATTTTA AACAATATCA 5340

```
ATTGGATCGA ATTTTAGCAT GGATTCATCC CTATGATTAT GTCGATAAAA TCTCCTATCA 5400
ACAAGTACAA GGCTTATTGG CAATTGGTTC GGGTGGTTTA TTTGGCAAGG GTGTGCATGG 5460
GATTGAAGTG TATGTTCCTG TTCGTGAGTC TGATATGGTT TTTACTTTTA TCGGGGAAGC 5520
CTGGGGCTTT GTGGGCAGTG CCACCGTTGT TTTTCTCTAT TTTTACTTGT TTTATCAAGT 5580
TTTAGTAGCT GGCTTGCGGA gCAATTCGCG TTTTTGTATG TACATCTGTG TCGCCCTCAT 5640
TTTTTCGCTG GTCTTTCAAA CGGTGGAGAA TATCGGTGCG GTGATTGGAC TGTTGCCGTT 5700
AAAAGGTATT CCGCTTCCGT TTTTAAGTCA AGGAGGCACT TCGTTAGTGA TGGCTATCAC 5760
TTCTTTAGGA TTTGTCaAAG GAAGAGAAGC CGCTACCTGA AAACACAAAT AAAAACCAAG 5820
ACTGGAACTC GCTTGCCGAC AAGCTCTAGT CTTGGTTTTT ACTATTGAAA TTGTTTTTGG 5880
CATTCAGCCA AAAGTTGTTG AAGTTGTCGA CTCATGTAAC GTTCTTTTAA ATAATAAACT 5940
GCCATGTTTC GTTTAGGTAG TTCTTGGTAA GTTAGTGTAT GGATCGAATT GTCTAAAAGT 6000
TGAATATTTG TGTATTCGGG GACAATAGCG AAACTATCGG AATGAGACAC CAAGTGAGTA 6060
ATAAGTGAAA GAGTATCCAC TTGAATGACT TTCTCTGGAG AAAAATTAGA AAACGTAtAT 6120
CTAGCCAATT GGTTAATTTG TTTTGCTCGG AAAATTCGGT ATTTAAACGG ACGAAGGGGT 6180
ACATTGTTAG TTCTTTTAAA ATTTCTTCCG TCGTGGCTTT TGCCAAAGCT GGATGATCGT 6240
TTCTTAAAAT GAATTGATAA TTTGCTGTAA ATAAATCAAT TTTACCAATA TGTTCAGGTA 6300
ATGGTAAGTC AGCAAAGGAA ATACCAAAAT CTGCATCCAT TGTTTCCAAG GATTGAAGAA 6360
TATTTTCTGT AATAATAGTG TGGACTTGTA AGTTGGATGA ACGATCCGTT AATGTTGTCA 6420
TTAGTAAGTC ATTCAATTTA TATTCTAAAT CACAGAGATA AGTGGCCACA TAAATATGGT 6480
TGGAAGAATG ATTAATTTCT TTTAGCTCTT CTTCAATTGA GACCAACTCA TTGGTGATGC 6540
GCCAACAwTA TTTTyCTAAA aTTAAwCCaG CTGGCGTTAA GGTGGTAGTT TGGCCTTGGC 6600
GATTAAAAAG TTCCGCGCCT AACGTTGCTT CAAGAATTTT TATTTGCTTA CTTAATGCAG 6660
CTTGAGAAAT TGCTAAATTT TCGGCTGCTT TAGTAAATTG TTTCACTTTG CATAACTCTA 6720
AAAAATAATG TAAACGAAAT AGTTCCATGG GAATTCCTCC TTTTTCTATT CCTAGAATAA 6780
ATAACGTAAT TGTTCGAATA AATAACGGTG TGCTTCACAA TGTCCGTGAT GATTATTCTG 6840
AACTAAATAG ATGGcTCGTT CTGGaCTATC TTCCAGAGGa ATACCAACTA AGTTACGTCT 6900
GTTTAGTTCa TCAGTGaGTG AATAGGGAAG GATGCTACAG GCTTTTTCTC GTTGGACCAT 6960
GTCTAAGACT AAATGAAGAG AAGGAAGTTC GATGACGATT TGCCGATTTT CCCAATTGAT 7020
TTTACGTTCC CAACGTTTGA ACACTTTTTT TTCATATACT TGATCGGAAT ATGCAGCAAA 7080
AGGTAAAGAA CGAATTTCGT CAATCGTCAC GCTATCGGCA TGGGCTAAAG GATGATTTTT 7140
ATTGACACAA AGGAGATAGC TTTCTGTATA AATTAATTCT TTCTCTAATG GAATCGAGGT 7200
ATCCAAAAAT ATCCCAATCC CAATGTCAAT TAAATCTGTT TCCAAGGCTG TTTCAAGATC 7260
GTAATAATAC TTAAAATCAA TACTTACATA CGGATAGGAT TGATTAAATA ACAACAAGAC 7320
```

TTCTCCAAAT GCTTTATCTA TATCAATTAA GTTTGTTCCA ATCCGAATGG TCGTTGTCGA 7380

TCCTTGGGTC ACTAATTCAT TAATCTCTTC TTGCGCTTCC TTAATGAGTG CAAAAATCAT 7440

TTCTGTTCGC TTGGCGACGA CTTCGCCTGC TGGAGAAAGC CGACAATGTT TCTTGTCTCT 7500

TAAAATTAAT GAAGCACCTA ATTCACGTTC CAACTCCTTG AGCTGGGTAC TTAAATAAGG 7560

TTGACTCGTA TGGAGCTTTT GAGAAGCTAA AGTAAAACTC GCATACTTTG AAACTGCCAT 7620

AAAATATTCT AGTTGTTGTA ATTTCATTTT CCTCACCGCC GTTTCTTTAA CTATACAAAA 7680

AATAGCAGAA GAAAAGCTTC TCGACAAGGC TATCTGATGG TCTTTAGATT AATTTAACCA 7740

GTTATGATTC TATGAATAAA TACGAATCAT ACCAGAAAAC CATAGTAAAT AAGCGGATTA 7800

TTTTTCTTTT TGTTACTGAA TGGACAAAAA GTTATCTACT AAAAACTTTT TCTATATCGT 7860

TATTTTTTCA TGGAAGAATT ATACTTGCTG TTGTAATTAA ATAAGATTTG TTAGGTGAGG 7920

CTCCTATACA AACATAGGCT ACTGCCCAAA AATGTCGAGA GACGCCAATG GGTAGAACAG 7980

GAACTGTCGA ATGAAGGCTT TTCTTAAGGT AGCTAAAAGC AATCGCTTTT TACGTTGTAT 8040

AGTGCTAAAG CTCAACGATA AGAACGATTA TTCTTTGACA CGTtGAGTTT TCAGCGTGTC 8100

TTTTTTTGTG GAAAGGGTGA GGAAAATGGA AGATGAACCG ATCGTCAATT CAAGTCAACC 8160

ACAGATAAAT GAGGTGAAAG AAAATGGAAA AGAAAATGCG TCACTTGGGC CATGACTATA 8220

GTTACAAAAA GTTTGCCCAT TCAGAAGTAG AAAAAGTTTT ACAACGTTTT GGCTCAACTT 8280

GGGATGGCAT TAAACACCAA GATATTGAAG AATTACAAGA ACGTTATGGC CGCAATAAAA 8340

TTATGAATGA AAAAAAAGnT TCGCGGcTGC GACTTTTTAT AAAATCGTAT ATCACGCCGT 8400

TTACTTTGGT GTTACTAGCA CTAGCCACGA TTTCATTTTT TACAGAATAT GTCTATGCTG 8460

CGCCAGAGGA AAAAGATGTT ACTGGCGTGT TAATTATGTT AGCGATGGTC aTCTTAAGCG 8520

GCACGATGaG TTTTGTCCaA TCAGTCaAAT CaAGCAATGC aGTTGAAAAA TTACAAAATA 8580

TGATTAAAGT TACAGCGACA GTCATTCGTG ACAAGAAGCA AATGGAAATT CCTATTGAGG 8640

AAGTTGTGTG TGGCGACTTA GTCCAACTAT CCGCTGGCGA TATGATTCCa GCCGATTTAC 8700

GTTTAATCCA ATCTAAGGAC TTATTTGTTT CCCAATCATC CTTGACTGGA GAGAGTTTTC 8760

CCATTGAAAA ACACGCGACT CATCAAAAAG ATAGTGATCA AATTGATACA GAATACGACA 8820

ATCTTTTGTT TTTAGGAACA AATGTAATTA GTGGCACTGG TTTGGGCATT GTCATTAAAG 8880

TTGGTAATCA AACTTTATTC GGAAGGATGG CTTCTGATAA CGGAGAAGAG CAGCAGCAGA 8940

CGAGTTTTGA AACAGGAATT AATAAAACAA CGTGGGTGCT GATTCGCTTT ATGTTGGTCA 9000

TTACGCCAAC TGTTTTTCTA ATTAATGGGT TAACAAAAGG TGATTGGGCG GAAGCGTTAA 9060

TGTTTGCGAT TGCCACAGCT GTCGGCTTGA CACCAGAAAT GTTACCTATG ATTGTAACAA 9120

CCAATTTAGT CAAAGGTTCT CGAGAAATGG CTAAAGAAGG CACGATTATG AAAAATGTCA 9180

ATGCCATTCA AAATTTTGGC GGGATGGATG TACTTTGTAC GGATAAAACA GGTACCTTAA 9240

CACAAGATAA AGTAATTTTA GAATACCACT ATAATATTGG GTGCCAAGAA GATCAAAAAG 9300

TCTTGGATTT AGCATTTCTA AACAGTTATT TTCAAACAGG TTTGCGTAAT TTAATGGATA 9360

ACGCCGTTAT TCAAGCGGCG mCACArGAAA GTGATATACA ATCAGATGAT TTTTATAAAG 9420

TAGATGAAAT TCCTTTCGAT TTCAATCGAC GTCGAATGAG TGTGATTATC AAAGAGTTTA 9480

AAACGCGAGA AACACGCTTA ATTACCAAAG GGGCGGTTGA AGAAATGCTC TTGGTCTGCA 9540

CACAGGTACT TTTAAATGGC GAAATTGTGC CCTTAACAGA GACGCTGCGT CAAAAAATTA 9600

CCAAAGACGT TGAAGCTTTA AATCGAGATG GTTTGCGCGT TCTAGCAATT GCCGATAAAA 9660

AAGTCGAAGC GGCTGAATGG GAGTATACAA CTAAAGATGA ATCAGAGTTG ATTTTACAAG 9720

GCTATCTTGC TTTCTTAGAT CCACCAAAAG AAACAGCGGC GGCAGCTATT CACGCCTTGC 9780

ATCAACATAA TGTAGCTGTG AAAGTATTAA CGGGGGACAA TCAGTATGTG ACGCATTCTG 9840

TTTGTAAAGA AGTTGGTCTG GCTGGCGAAA AAATTATTAC AGGAAATGAA CTTCAACAAT 9900

TAAACCAAGA AGAATTAAGA AAAACCGTTC AAGCGTATAA TATTTTCGCC AAAATTACGC 9960

CTGACCaAAA AGTTCGTATT GTAAATGCGT TAACTGAAAA TGGTCAAACT GTTGGCTTTT 10020

TAGGTGATGG AATAAATGAT GCTGGAGCAA TGCGTGCTGC TGATGTAGGA ATCTCTGTTG 10080

ATACAGCTGT CGATGTGGCG AAAGAATCGG CGGATGTAAT TTTATTGAGA AAAGATTTAA 10140

TGGTTTTGGA AAAAGGGATT TTATCTGGAC GGCGCGTGTT TACGAACACG ATGAAGTACG 10200

TCAAATTAAC AGCAAGTTCT AATTTTGGCA ATGTTTTTTC CGTGATCCCA GCAAGTATCT 10260

TTTTACCATT TTTACCGATT GCTCCGATTC AATCGCTATT ATTAAATCTA ATTTACGATA 10320

CTTCGTGTAT GTCAGTTCCT TGGGACAAAG TAGATGAAGA ATACGTAGAA AAACCAAAAA 10380

AATGGGAACC AAAATCAATT GGTAATTTTA TGCGTTGGTT TGGCCCAACG AGTTCTATTT 10440

TTGATATTGT TACCTATTTA TTTATGTATT TCATTGTGTG TCCGGCTATC CTAGGTGGCT 10500

CATTCTTTGA ATTAAATGGA GCAGATCAAT TGCTCTTTAT CGGTATTTTC CATGCAGGGT 10560

GGTTCATTGA GTCCTTATGG TCCCAAATGC TCGTCTTGCA TTTCCTACGA ACAGAAAAAA 10620

TGCCTTTCCT TCAAAGTAGT GCTTCTGGAA tTATGACGTT AGcACtaCAG CAGGGaTTGT 10680

TCTCGGgAAC TGtTTTACCC TTTACCGCTT TCGGTGcAGA ACTAGGATTT GTCGGTTTAT 10740

CACCAaGCTA TTTCTnGTAT TTAATCCCGA CAATTGTTGC TTATTTAGCA CTGGTTGCgT 10800

TCATTAAAGT TTTATACGTC aAACGTTACG GTCaATTACT GTAACTGATT TGTACTATTA 10860

TATATAGAAA AATTTTTTTA CTTTTCTTTC aAGGTTGAGG AGGTGTCGGT TTTTATGACA 10920

GCTCCTCTTT GTTGTTTTAA GTGACTGTGC GAATTTTTTT ATGGATTTAT TTGTCATTTA 10980

CAGTTAGTAT GGCTAGTAGA GACTTTCCAA AAAGAACGAG CTTTTTGGTA ATAAGGAGAT 11040

AAAACGAGCT AATGACCAAA ACAAACCAAC TTTCAACTTC AACGCCGATG TTCGATCGGT 11100

CGGCCTATAT TCATATTCCA TTTTGTGAGC ATATTTGTTA TTATTGCGAT TTTAACAAGG 11160

TCTTTTTAGA AGGACAACCA GTTGATGAAT ATATTCAGAG TTTATTAAAA GAAATTCAAC 11220

TTACGCAAGc ATTGTATCCA GAACAAGAAA TGAAAACcAT CTATATTGGC GGAGGCACGC 11280

CAACTTCATT ATCAGCGAAA CAGTTAGATG TGCTACTAAA AGGGGTTCGC GAACAGCTAA 11340

CGTTTGATGA CCGAAATGAA TTTACGGTGG AAGCAAACCC TGGTGATTTA ACCCAAGAAA 11400

AATTACAAGT CATGAAAAAC TATGGCGTCA ATCGTCTTTC GATGGGCGTT CAGACATTTG 11460

ATGATCGCTT GCTCAAAAAA ATCGGTCGCA AGCACACCGC CGCAGATGTT TATGAAACAA 11520

TGAAATTTTT AGAGAAAGAA AATTTTACCA ATGTCAGTAT TGACTTAATC TATGCTTTAC 11580

CTGGACAAAC ATTGGAAAGT TTTCGCGATA CTTTGACAAG AGCCTTAGCA TTGGACTTGC 11640

CGCACTATTC GCTTTATTCA TTAATTCTTG AAAATAAAAC AATGTTCATG AACTGGGTTC 11700

GTCAAGGACG GTTACAGCTG CCTGAAGAAG AAATCGAAGC ACAAATGTTT GACGAAACGA 11760

TTGAAGCAAT GGAGAAAAAA GGACGTCACC AATATGAAGT CAGTAATTTT GCTTTAACAG 11820

GAAAAGAAAG CCAGCATAAT CTGGCTTATT GGAATAATGA CCACTACTAT GGATTTGGCG 11880

CAGGTGCCAG CGGCTATCTT GGGCAAACTC GTTACAAAAA TCACGGTCCA ATCCAACATT 11940

ATTTGAAACC CTTGCGAGAA AATCAATTGC CAATTGTGGA GACTGAAGAA TTAACTCGTC 12000

TAAATCAAAT AGAAGAAGAG CTATTTTTAG GCTTACGTAA AAAAGTAGGC ATTTCTAAAC 12060

AAAAATTCCA ACAAAAATTT CAGGAACCAA TTGAAGCAAT TTATGGAGAG GTCATTCAAC 12120

GCTTAATCAA AGAAGAATTA TTAATTGAAG AAGCAGATAT CTTACGTTTA ACGAAAAAAG 12180

GGCTATTTGT GGGGAATAAT GTTTTTGAAG CATTTTTGTT AAGTGAAAAA GAGTAAAAAA 12240

ATAAAAAAAT CATCGAAGTG TGACAAAGCA TTTCGATGAT TTTTTTATTA CAAAAGTTGG 12300

TATTGATGGT TTTTGAAGGT GGTTTGCAGC GATTAGCACT CTTGCGTTAA GAGTGCTAAA 12360

AAGTTGCGTA TTTCCCTTGA CAAAAAGGTA AGGACTTGTT ATATTAGTAT ATGTAGTTAG 12420

CACTTAGATG TTAAGAGTGC TAATGAAGGG TGGTAATAmC GATGATTACA GAAAGACAAC 12480

AAAATATCTT GAGACTTATC ATCCAAAATT ACACGAACAC GGGCTTaCCT GTCGGTTCTa 12540

AGAAATTGAT GGaGGATGGc ATTGCCTCTA GTTCTGCCAC AATTCGTAAT GATATGAAAG 12600

CATTAGAGGA ATACGGCTTA TTAGCTAAGA CCCATTCTTC TTCAGGGCGA ATCCCCTCAA 12660

TGGCTGGTTA TCGTTACTAC GTGGATCATc TTTTACmACC mACACAAGTA GAAGAGAACG 12720

AACTTAGAgA ATCCGTCAAT CTTTTGGAAA AGAATtTCAT GAGATTAATG ATATTATTCG 12780

ACAATCAGCT GAAATTCTAT CAGAATTGAC AAGTTATACT GCTTTTTCAT TAGGACCTGA 12840

AGTGAAAGAA CGGAAATTAA CCGGCTTTCG AATGGTTCCT TTGAATGATC GGCAAGTCCT 12900

TGCTATTATA GTGACAGATA AAGGCAATGT TGAAAACCAA GTATTTGCCA TTCCTGCCGC 12960

TGTTTCTAGC CAAGATCTTG AAAAAATGAC ACAAATCATC AATGATAAGT TGGTTGGACA 13020

ACCTTTGTTG ACCGTTTATC ATCGTTTGCG GACGGAGATT CCGATGATTT TACATCGCTA 13080

TTTCCAAACA CCAGAAGGAA TGATGAACTT ATTCGATGAG ATGTTAGGAC ATGCCTTTGA 13140

AGAAAAAGTC TTTGTCGGTG GTCGCATGAA TTTGTTAGAT TTTGGCATCA AACAAGATAT 13200

TGAACAATTG AAATCTGTCT ATTCCTTTAT GCAAAATTCT GATGAGCTGA CGCATTTATT 13260

AAATGGTTCT GCAACAACAG AAAATCCAAT TGTTTTTCGT ATCGGCTCAG AAATTGGCAA 13320

CAATCTCTTA GAAGATATGA GCATGATCAC TGCTACGTAT GAAGTATCAG GACATGGCAA 13380

AGGAACAATT GCCCTGCTAG GACCAACCAG TATGCCGTAT TCAAAAATAT TCGGTTTGGT 13440

AGACACGTTT CGACATGAGC TAGCTTCACA ACTAGGTGAC TATTACCGTT TTTTAGGTAA 13500

TTAAGCAAAC CTTGTTAGGA TAAGAAGTAG ATGCATAAAG AGAAAGGAAG TTGCTGAAGT 13560

GAGTAAAAAA GAAGAGAAGC AAGAAGAATT ACAAGAAGAA ATGGAAGCTG TTGATGCTGC 13620

TGGCGTTTCT GAAGTAGAAG TTGAAGCAAC AGAGATTGAA AATCTTAAAG CAGAGCTTtC 13680

CGAAATGGAA GACAAGTTCT TGCGGGCAAG AGCCGAAaTT GCTAATATGA GCmATCGAAa 13740

TAAAAcGrAC GAGAATTATt AGTcGTTcCG CTCACAAGAT TwGGTAAAgA TTTACCACTA 13800

TTGTAATT 13808

(2) INFORMATION FOR SEQ ID NO: 272:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3379 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

AATAAAGCAT CGTTACGCTC TTTAATTTTA GCAGCAGAAT TCGCTGTGAT TATTGCTGTA 60

CTTTCACAAT TTACAATTCC ATTGGGGATT GTCCCCTTAA CTGGACAAAC ATTTGCCATC 120

GGCTTAACCG CTACCGTTTT GGGCAAGCGA ACGGGGACCT ACGCTATTTT AATTTACTTC 180

TTATTAGGAT TAATTGGTTT ACCCGTTTAC GCTGGGATGT CTAGTGGGTT CGGTGTACTC 240

TTTGGACCAA CTGGCGGCTA TTTAATCGGA TTTATCTTTA ACGGCTTATT AACTGGTTAT 300

ATTTTAGAAA AAACAACCTT TAACTATACT TGGGCCATTA TCGCAAATAT TGCCGGCGCC 360

TTAGTTACCT TAGTTTTCGG TATGATTTGG TTGAAATATT CCGCAAATCT GCCGTGGcCA 420

AACGCTTTCG CTGGCGGGTT TGCTCCGTTT ATTATTCCTG GTATCATCAA GGCTGTCGCA 480

GCTGCTTATG TTGGTATTTT AATTCGCCAA CGCTTTATGA AACGTTTTTT AGCACACCTT 540

TCTTAAGGGG AAAACTTTTT TCTTTTCCTT GTCTGTAGTA CACTAAAAGG ACAAAGGAGT 600

GAAGTCCATG TCCTCATTAT TAAAACGATT GGTTCAGTTG GTTTTGTTAG TCGTCGCTGT 660

CTTGCTGATT CGACACTATG TTTTCTCCCC TGCTGCGGTG AACGGCTCTT CAATGGAACC 720

AACACTTCAT AACAACGACC GTTTATGGGT GACCTCGATT AAAAAACCAC AGCGCTTTGA 780

TATTATCGCT TTCCCTAGTC CTCGCAACGG CCAACGAGTA GCCAAACGTT TAATTGGTTT 840

ACCTGGCGAA ACAGTCGAGT ATCGCGATGA TACCCTTTAT ATTAATGGTG TATCACTCAG 900

TGAAGATTAC TTAGCAAGTG CTAAACGAAA TGTCTCTAAA AATGAAAATT ATACCCAAGA 960

TTTTACGCTA GAGACCTTAG AAGCCACCCA ATCCCTGACC GTTCCAGAAG GCATGTATTT 1020

TGTCTTGGGG GATAATCGCC CGCGCTCAGA CGACAGTCGT TATTTTGGCT TTGTTAAACA 1080

AGCGAGTGTG GAAGGTGTTT TGACTTTTCG TTATTATCCA TTAGATAAAA TTGGCTTTCC 1140

ATAAAATAGT AAACCGCCCT CAAAAAGTCT AACTTTTTGG GGGCGGTTTA CTATTTTTTT 1200

AAGGTCCAGT GACTAAGTTC CACGTCACCG TTCCTTGATA TTTCTCATTT ATTTTCCCTT 1260

GATTGGCAGG GACTTCTAAT TTCACATTGG CAAAATCAAA ATCTAATTGA TAGACATCAT 1320

TACCTGTAAA TTGGTTATTA AGAATTACTG GTGTAGATGT ACCGTCACTC GTCAAAGTTA 1380

AGGCTTTCGT AACCCCAACT TTATTTGTTA ACTCTGTGAC TTGATTATAG TTTTCGATTG 1440

CTATAACATC CGCACTTCCT AATAATAAAC GCGTACTAGT TGGCAAACTA tCAGcTGACG 1500

CTGTTAACTG TGACAATTGT GCTGTTAACT GCCAATTAGG TTGGCTTGTC TTTATTCGTA 1560

GATATGGATT ACGTGTGCCG CTGCCATAGT AATCAGCTGC TTTTTTCAGT CCGTACTGCT 1620

TAGTGCTACT GTTAATCGTA ATTTTCCCAA AATCAAACAT TGGTGTACTA ATGAAACCTT 1680

CTTCTACTGG TGTTGGCGTA ATCGTTTGAT CTTCTCCTTG AATACGTACA AAGTTACTAG 1740

CTGTCACTGC AGAAAAATTT CCACTAGCTG TCATTTCTGC GGTTAAAAGT TGTCTTGGTT 1800

CTCCCGTAAT TTTCGCTGTG TTCACTGTCA CTaAAGTTCG CCGTTCACAG GAATCTCTAT 1860

ATCTAAGTTA ACTCCTGCTT GCCATTGTTC TGCCGTCACT AGAATAATTT TATTTGGTGT 1920

ATTACCAATG CGAACAGCTA ATTGTGTTGG TGGTGAAATA CCTGCCGCCC ATTTGGCAGT 1980

TGGGGCTAGT TTAACGTTGG TCAATGGAAC TTCACTGGTG TTTTTTAGTC TAACTGTCCA 2040

AACCATAGAA GCTTCATTTG GGATCACTTC TAGTAACCGA CTAACCGTCA TTTCTGCTGT 2100

TGGTTGGACT TCCTCATAGA CCGCTGTCAT GTCATCTTTA TCGTCGTAAG TCACGGCGTA 2160

ACTTGGTGTT TTAGTCGTTG TCAAGGTGCC TGGTTTGGTT TTACCTTTGT ACCAGCCTCT 2220

GAACTTATAG AATTTGTCTC CTGTTGTATA AGTGTCTGGT AAGGCTTTCG TACTAGCGTA 2280

AGTGAAATTA TCACTATCAA TCACTGTTTG TTTTCCTTGC GTGAAGCCTG TTGGCGGTGT 2340

AATTTTAGCG CCGCTTGGAT CCACAAAGTT TTCAGTGACT CGGCGATTTT CTAAATAATA 2400

GTAGATTGTT CCCTCTGGAG AAGCTAGGTT ATAGTTTTTT CCGACTGTCC CTGAATAAAT 2460

TCCCATAGCA TACAACGTGT TATTCGGGTC CCATGTCCAA AAAGCTCGTC TGGTTCTAAA 2520

AAAAGTTCCA GCGTCAGTTA ATTTGAAGGC TTCTGCTCCG TCAGCTTGAG AAACAACCGT 2580

ATATGGCGCT TCTATATTTT TCCGGATCTC TTTGCTCATA GGATACTTTG TTTTGTCTCC 2640

CGTATAAGTA GCTAGCGGTG TAATCGACAT TGTCTTATAT TGTTTAGGAA TTGTATGATT 2700

AACAGACCTT GCACCGTATG TCTTAAGCTT ATCTATCGGG AAATCATATG ACTTGCCAGG 2760

AATCGATAAT TTTTTTAATT TACTCATCAC TTCATCAGTT CCTTGGATTT TCCCTAATAC 2820

TGTTGATTCA GTTTCGTCAC TTTCACCTAA TTCGGCTTCT ACAGAAAAGC CTGCTGTATT 2880

CAATAGGTTA TCTGCCTCAT CAACATAACC AAAGTACACA TCTTTGGCTG GTAAATTATA 2940

CGTTTTCATC ACGACTTCTT CATAGACAAC CGTCAAATCA TCATCATCAT TATAAGTGAC 3000

TGGATAACTG GGAGCTTTAG TCGTCGTCAA GGTGTCTGGT TTGTTTTTAC CTTTGTACCA 3060

GCCTTTGAAC TGATAGGTTT TCGTGCCGAC TTTATAGGTA TCTGGCAAAT CTCCGCTTTG 3120

CTTAAAGGTG TACGAATCAC TAGTAATTAC TGTTTTTTTC CCTTGGGCGA AACcTGACGG 3180

TGCGGTGATT TTAGTACCAT TGGTATCTAC GAAGTTTTCC GTAACaCGAC GATTTTCTAG 3240

GTAAAGATAC ATACGTGCAT TTTCATCTAG TGTATAGaCA CTCGGTTCTA CTCCTTTACC 3300

AGTAATAGTC GCACTTAATA GTGTTGCAAG TCTAACATTT ATAGGCATTG ATGACCAGTA 3360

TATACCGCTC ATCGTAAAC 3379

(2) INFORMATION FOR SEQ ID NO: 273:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1314 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

TAGATTGTTT AAAAAAACTA TCGAATAATA GAGATTAAGC TTGCATTTTA GCGGTGTTTC 60

CTGnAAAGTA AGCTTGTGTG GAAACACACC ATCTATCACT TGGTTTAGCG AGTCACCAAC 120

GCTTCACTCA GTGATCAGAC GAATAATAGA GAAAGAGGTG GAACAAATGG CAATTTCACA 180

AGAAAGAAAA AACGAAATCA TTAAAGAATA CGCACGTCAT GAAGGAGATA CTGGTTCTCC 240

AGAAGTACAA ATTGCTGTCT TAACTGAAGA CATTAACCAA TTGAACGAAC ATGCGCGTAC 300

ACACAAAAAA GATCACCATT CTTACCGTGG TTTAATGAAA AAAATCGGTC ACCGTCGTAA 360

CTTGTTAGCT TACTTACGTA AAACTGACAT CCAACGTTAC CGCGAATTGA TCCAACGTTT 420

AGGATTACGT CGTTAATCAT TACGTATAAA GCGAGGTTCA ATCAGAATCT CGCTTTTTAT 480

CTTATTGTCA CTATTGAAGG TCCTCTTTGG AAAGTTTTTA GGAACTACTA ACCAAATGAA 540

AGAGCAGCAA CGACTTGTTA CTTTTTCATT TGTTTAGTAG CCGTCTGACT TTTCCAAAGA 600

GCAAGAGTAA AGGAGAAGAA TTATGACAGA AAAACAAGTC TTTAAAACAA CTTGGGGCGG 660

CCGCCCACTA GAAGTAGAAA TCGGCCAATT AGCCAAACAA GCCAATGGAG CTGTCTTAGT 720

ACGTTACGGC GATACGGTCG TTTTAAGTGC TGCTGTTGCC TCTAAAGAAG CAAAAGATGT 780

AGACTTTTTC CCATTAACTG TAAACTATGA AGAAAAAATG TATGCGGTAG GAAAAATTCC 840

TGGCGGCTTC ATTAAACGTG AAGGTCGTCC AAGTGAACGT GCGACATTAA CCGCTCGTTT 900

AATCGACCGT CCTATTCGTC CAATGTTCTC AGAAGGCTTC CGTAACGAAG TCCAAATCAC 960

GAACATCGTT ATGAGTGTGG AGCAAGACTG TACACCAGAA ATGGCTGCAA TGTTTGGGtC 1020

TTCCTTAGCG TTACAATTTC AGATATTCCG TTTGATGGAC CAATTGCTGG TGTTGACGTT 1080

GGTCGTATTA ATGGCGAATA TGTTTTAAAT CCAACGGTAG AACAAGCGGA ACAAACAGAT 1140

ATTGAATTAA CTGTTGCTGG AACAAAAGAA GCCATCAATA TGGTTGAAAG TGGTGCCAAA 1200

GAAGTATCAG AAGAAGATAT GTTAGGTGCC TTACTTTTCG GTTTTGATGC CATTAAAGAA 1260

CTAGTGGCGT tCCAAGAAGA AATCGTTGCC GCTGTAGTAA ACCGGGGATC TTAA 1314

(2) INFORMATION FOR SEQ ID NO: 274:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1234 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

GCCGTTTTGC ATACAGAAGG ACCACTGTTA GTCATGGCTG GTGCTGGTAG TGGAAAAACG 60

CGGGTTTTAA CACACCGAAT TGCCTATCTA ATTGAGGAAA AAGAAGTAAA TCCATGGAAT 120

ATTTTGGCGA TTACGTTTAC CAATAAAGCT GCTAAAGAAA TGAAAGAGCG TGTAAATAAA 180

TTACTGGAAA CGGGTGGCGA AGATGTATGG GTCTCCACTT TTCACTCCAT GTGTGTAAGA 240

ATTTTACGTC GTGATGTGGA TCAGATTGGT TACAACCGTA ATTTTACAAT TATTGACCCA 300

TCGGAACAAA AAACATTAAT GAAACGCATC TTAAACGACT TAAATATTGA TTCAAAAAAA 360

TATGATCCTC GCTCAATTTT AGGTACGATT AGCAATGCCA AAAATGAATT ACAAACGCCT 420

GAAAAAGTTG CTGAAATGCA AGGTTCCTTA TATGAAGAAA TTGTCGCAAA ATGCTATGAA 480

GCCTATCAAA AAGAATTGCG CAAAAACCAA TGTATGGATT TTGATGATTT GATTATGAAT 540

ACGATTCGTT TATTTAATGA ACAACCAGAT ACGTTGGCTT TCTACCAAAA TAAATTTCAC 600

TATATTCATG TGGATGAATA CCAAGATACC AATCATGCAC AATATACCTT AGTCAATTTA 660

TTGGCTGCAC GTTTCAAAAA TCTGTGTGTC GTAGGTGATG CCGACCAAAG TATCTATGGT 720

TGGCGCGGGG CGGATATGCA AAATATCTTG GATTTTGAAA AAGATTATCC AGATGCCAGT 780

GTAATTTTGC TAGAACAAAA TTATCGATCC ACCAAAAACA TTTTGGCGGC TGCGAATGAT 840

GTCATCAAAA ATAACCGTAA CCGCCGCGAG AAAGAACTGT GGACAGAAAA TATAGACGGT 900

GAAAAAATTG TCTATTATCG CGGGGACACA GAACGGGATG AAACACAATT TATTGTGAGC 960

CAAATCCAAA AAGAAATGCG GGAAAATGAC CGAATTTATG GTGATTTTGC GGTATTATAT 1020

CGGACCAATG CTCAATCCCG TGTAGTGGaA GAAATGCTCt TGAAATCTAA TATTCCGTAT 1080

ACCATGGTGG GCGGACmTAA ATTCTACGAC CGTAAAgAAA TTAAAGATAT TTTGGCTwAT 1140

TaAGTTTGAT TGCCAATCCA GATGATcGAT TAGTTTTGAA CGAATTGTCA ATGAACCAAA 1200

ACGAGGGATC nGTAAAAGTC AATTGGAAAA TnAC 1234

(2) INFORMATION FOR SEQ ID NO: 275:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6415 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

```
AAAGGACCGC CTAAACCAAC AnCGACCTGT CGGAGCGnAC GATCATTTTA GAyCCAtGCC      60
CGTTsCCAAA TCACCCGTCG aGAATCAATC TTTAATTCAT TTCCTTGCTG TAAATGATTA     120
TCTAGTAACG CAGCTAAGGC ATTATTGGCT GTCGTAATCG CATGCATATC ACCAGTAAAA     180
TGCAAATTGA TTTCTTCCAT TGGTAATACT TGTGCGTAAC CGCCACCTGT TGCGCCACCT     240
TTAATCCCCA TCACAGGACC TAATGATGGT TCTCTTAGGG CAATAACTGT TTTTTTATTG     300
ATTTGATTTA ATGCATCTCC TAAACCAATG GTCACGGTTG ACTTTCCTkC TCCTGCTGGT     360
GTTGGATTAA TCGATGTAAC GAGGATTAAT TTACCTTCTG GTTGCGCTTC TAAGGATTGA     420
AGCGTTGGAA AATCTATTTT CGCTTTGTAT TTTCCATAAA GTTCGAGCGA ATCTTCTGAA     480
AGACCGATTG TTTCGGCAAT CTCTTGAATA GGTTTTAATG TTGCTGCTTG TGCAATGTTA     540
ATATCTGTCT TCATAACGGA CCCCTTTTCA ATTCTTTATC GTCTTTGTGA CACATTGTTT     600
CTTTTAAATT ATAGCTTGTA TTGAATAGAA TGGCTAGAAT AACCTTTTTT AAACAACCCA     660
ACCATTGACG TTTGGTAGCC TTTTCTCGTA TAATGAAAGT GTCTTTACAA ATCTAAAAGG     720
AGCGTGAATT GGATGCAAAA AGGAATTGTC AAATGGTTCG ATAACAGAAA AGGTTATGGT     780
TTTATCGTTT ACAACGAAGA AGATGAGATT TTTGTCCACT TCACTGCTAT AGAAGGTGAT     840
GGATTCAAAT CACTTGATGA GAATCAATCT GTAGAGTTCG AGATAATGGA AGGCAATCGT     900
GGGCTACAAG CAGCACATGT CAAAAAAATA GATAACTAAA AAAGCTGAAC CAACAGCGCT     960
TAGCTCCAAA AAGTAAGTAA ACGTAGCGTT TGTTTCAGCT GATTATTTTA AGATTTTGAG    1020
AGTGGGTCAC CAGTCGTTTC GACTTATGAC TCACTCTTGT TTTTTTCTTC TTTTTTATGA    1080
AGGCCCAAAG AATGATCTTC TGCTACCTTC AATTCAGGCT TTTTTACCTT AAGTGCCTCA    1140
AGAAATAACG GAATCGTTTT CTTTTTCATT AAAAAGACTT TTGAGCCTTC CTTAAATTCT    1200
GAGGAAAAAA TTTCGATGCT TTTTTCATTG TAGGTTACTT TTTCTATCTG AGAAAGCGGA    1260
AGTGCTTTTT TGCGAAACCA CAAAACATCT CTGACAATTA AAGATGTTTC TGTCATATTA    1320
AAAGACCGCG CAATTCCTAG ATAAGCAAAC ACAAAAAATG GAACCATCAC CAGATTACTA    1380
ATCAAGTAAG GACCATTATT TTCCAATGCT AAAATTAAAC TAATAAATAA TATACAAAAT    1440
GTGCAAGACC AGTAAATAAT TGTTGgATGC TAATTCTGGg CTGgCCAATA GgAATTTCTG    1500
TTTTTTCATA AAGgTGCCCT CACTACTTCT gGATCAACGC CGCTAAATTT AAATATGTTA    1560
TAATAGTAAC ATACTTTTTT TAAGAACmAA AAGAAATTAG GAGTAATTTT AwGtTAAGaA    1620
TTTATGtAGA TGCTGCAACA AAAGGCAATC CAGGTGAATC TGGCGGCGGT ATCGTTTACC    1680
TCACCGACCA ATCGCGGCAA CAATTGCATG TGCCTTTAnG CATTGTTAGT AATCACGAAG    1740
CCGAATTTAA AgTGCTGATT GAAGCTTTAA AAAAAGCGAT tGCCAATGaA GACmATCAAC    1800
aAACCGTTCT TCTCCACTCA GATAGTAAAA TTGTTGTCCm AACmATTGAA AAAAACTATG    1860
CTAAAAATGA AAAGTACCAG CCTTATTTAG CAGAATATCA ACAACTAGAn AAGAATTTTC    1920
CTTTGCTCTT AATCGAATGG CTACCTGAAA GTCAAAACAA AGCGGCCGAT ATGCTTGCAC    1980
```

GGCAAGCATT ACAAAAATTT TATCCCAATA AAAAGTAGCA CTGTTTACTT AATGCTTTTC 2040

CTTTATTAAT TTGATAATTA AACACGTGGA GCAAAAATTC CaAGTGATTT TTGCTCCaCg 2100

TTTAAAAACA GATAAAcGGT TCTGTCTCGA CTTCTTCTTA TAGCCaCTTA TTCTTTTGTC 2160

GTTATTyCCG CAAATkGCCC ATtGGTTAGC GAAaGGATTG CTTCAGGCGC TAATTCAATT 2220

TGCATGCCAC GTTTGCCTGC AGAAACAATA ATTGCAGAAT ATTGTTGAGC TTCTTCAGCC 2280

AAATAGGTGG GAAATTGTTT TTTCATCCCA ATTGGTGAAC AGCCCCCACG AATATAGCCT 2340

GTGGTTGCTT CTAAATCTTT TAAGTGAAGC ATTTCTACCT TTTTGTTGCC ACTAGCTTTC 2400

GCTAATTTTT TCAAATCTAA TTCTTGATTA CCAGGAATGA CAGCCACGAC TGGACCTGTT 2460

TTATTACCTA CTGTAACCAA GGTTTTAAAA ATTCGTCCTT TTTCAATACC TAAACTTTCA 2520

GCGACACTTT CTGCACTTAA ATGATCTTCA CTCCAAGCGA ATTCATATTC TTTATATGGT 2580

ACTTTATGTT GTTCCACCAT GCGCATCGCA TTCGTTTTTT GTTGTGTTTT TTTCTTTGCC 2640

ACTTTTTTCA CCTTCTGAAT CACTTGTTTT GTTCGAATGT CTTTAAATAA TTTCTTAATA 2700

CTTGTACGGA TAGTTGGCCT GGTGCAGACG CTTGCTGGGC GGAGCCAAAA GTTAGTGCCG 2760

ATCCGAACAG TTGACCTGTT ACGCGACTAA TCATCCCTAA CTGTCCCATT GACATAGTAA 2820

CGATTGGAAC GGATGCATAA TGGGTGTACA TTTCATTGGT TGCTGAAAGT AACGTAAGTA 2880

CATCTGTCGC ATCTTGTGGC ATCACAGCTA TTTTACAAAT ATCTGCCTGA CGCATTTGcA 2940

TTTGGCGTAa CCGCGCGACA ATCTCCTCTT GCGACGGTGT TTTCTGAAAG TCGTGATTAC 3000

AGAGAACAAT TTTAATCCCC GCCTTTTTAG CTTCGTGAAT TAACGTATCT GCCGCCAGCG 3060

GATTCGCAAA TAATTCAATA TCCAATAAAT CCAACGCACC CTTTTTTACT AATTCATGAT 3120

ACAATGCAAA ATAATTTTCC TCGGAAAAAG CCATTTCGCC GCCTTCTTTT TGTGTACGAA 3180

AAGTTAACAA TAACGGTTTT TGCCCCAACC GTTCCATTAC TTGTTGTGAA AGATTACATA 3240

CATCGGAAAA ATCAGCGACA TTTTCATAAT AGTCTAAACG CCATTCCACC AAATCACAAT 3300

CTAGTGTTTG GCTCGCCGTT GCTTCGGCCA AAATATCCTC GGCTGTTGGA GCAACAATGG 3360

GCACCACAAT TTTGGGATTC CCTTCACCTA TTCGCACATT TTTGACTATA ACTGGCTTCA 3420

CAGTTCCACC TAACTTCCTA GTTTGTTTTT CTCTCTTCAT ATCATACCAT TTCTATTCTT 3480

TTTTGCAATT TAAAGAGCTA CAAAGACACC TAGAATTAGG AATTTCCTAA CACTAAATGC 3540

CTTTGTTTGG GTTATTTAAC TCTTTCTATT GTTACGTTAA ACTTCCGTAC TTTGTAACTG 3600

ATACATATCC GCATAAAnGG CACCTTCTGC CAGTAACGTT CGTGCGTTCC TCGTTCTACA 3660

ATCTGCCCTT TGTCTAAAAC TAAAATCAAG TTTGCGTCAC GAATCGTCGA AAGGCGATGC 3720

GCAATCGCAA TTGTCGTTCT yCCcTTGACG CATTTTGGCT AAtCCTTCTT GAATCaATCC 3780

tTCTGTTTCT GTATCAATAT TAGCGGTTGC TTCATCTAAA ACTAAAATTT TGGGATCAGT 3840

CACAATCGTT CGTGCAAACG AAATTAATTG ACGTTGCCCg CTAGAATAAC TCGCTCCTCG 3900

TTCAATGACT TTCGCATGAT AGGTATTGGG TAACGTGTGA ATAAATTTAT CTGCCTGAAC 3960

AAATTCCGCA GCCTGTTTAA TTTGCTCATC TGTAATATTG GGATTCAGTA AACGGATATT 4020

TCCAGCAATA TCTCCATAAA ACATGAAAGC GTCCTGTAAT ACTAAGCCCA TCTTTTCCCG 4080

TAATTCTGTC ATGGGAAAAT CACGAATGTC ACGATCATCA ATTAAAATCT GCCCTTCATA 4140

AAATTCATAG AAACGCATGA GTACATTAAT AATTGAACTT TTACCACTTC CTGTATGACC 4200

GACTAAAGCG ACGGTTTCCC CGGGATTTGC GACAAAGCTA ATATTTTTCA ATACTTCGTT 4260

TTTGCCGTCA TAAGAAAAAG TAACATTGCG AAATTCAATT TTCCCGCGAA TAATTTCCCC 4320

ATTCGCACCT ACTGATTGTT GCGGCGTTAA TTCTTCCGTA TCCATAATTT TTAAAATACG 4380

GCTACCTGCG ACAATCCCAT CTGTAAAGAT ACTTAGGAAG TCCATCATTT GCGTCATCGG 4440

ATTAAAAAAA GCTTGGACAT ACGTTACAAA AGCATAAATC ATCCCCGCTT CAACAGGTGA 4500

ATGTAAAGCA TCGATGCCAA ACATAGTTAA GGCCAATGCA ATCGCCAACG TATACAACAA 4560

ATTAATGaTG GGACTTAATA ATAACGaATT CATACGAATC ATTGcTaCTC GTGTTTTTAA 4620

ATAATCATCA TTGGTTTCCT CAAATTCTTT TTCCAAACGC TTTTCTTGGC GAAATTGTTG 4680

AATGATTTGC ATACCAGAGA TGTATTCATT TAGTTTGGTG TTTAATTGAC TTAATTTTTC 4740

CCgCATACTG CGATATAATT TTGAACTAAA TTTTTGGTAA TACCAAATCA CAACTAATAA 4800

AATCGGCAAG AAAATTAGAC AATAAAATGA GATTTCTGGA CTGAGTTGAA ACATTGCAAA 4860

AAATGACGAG ATTACGGCAA AAATTCCTGT AATCACCATT AGGAATACAT ACCAGAATTC 4920

AAACAATGTT TCTGTGTCAT TTGTAACACG AGACACCGTC GAACCCGCTG GTGTTTGATC 4980

AAAATAACGC ATACCTAAAG TGTGTAATTT TTCAAATAAT TTCACTCGTA TATATTGATA 5040

CGTTTTTAAA GAAGCCaTCG AATACAAATA CCATTGGAAA AACCAAACAA TACTTTTGAC 5100

AATGACACCA AAGAAATAAA TAGCAGCAAA AGTTAGAATC ACTTGCATTG TAGCAGTTTT 5160

AGGCGTTAAA TAATTATCCA TAAAAACTTG aATGACTCTA GGCAAAAGAA TATTGATAAT 5220

TGATAAGCCA AATGCAAAGA GAATAGCTGT GTAAAACATG CGTCGAAACG GTTTGGCAAA 5280

ACGAAACAAA CGCTGAACAA TCGTTAATTG TTCTTTCATC GGTAATTCTC TGGACCATTC 5340

GGATTGATAC TTTTCTTCCA TTAACTCTCA CTCCCTTCAA TTTTCGCTTC TAATTGCTGT 5400

TTTTCCCACA TCCGCTGATA CCAACCCTTC TGAGCCAGTA ACTCTGGATG CGTCCCCCGT 5460

TCAATAATTT TCCCTTCATC CAACACTAAG ATTTCTTTGG CATGCATGAC ACTACTTAAA 5520

CGATGCGCTG TAATAATAGT GGTTTTTTCT TGGCGCGTTT CTTTTAGATT TGATAAAATC 5580

GCCTCTTCAG TCTTCGCATC GACTGCAGAT AAGGCATCGT CTAAAATTAA CAATTCTGGT 5640

TCAACAATCA AGGCCCGTGC AATAGAAATC CGtTGtTTTk GTCCTCCAGA AAGCGAAACG 5700

CCACGTTCCC CAACCATTGt GTCATAGCCT TCAGGGAATG CTTTAATTTC TTGGTTAATA 5760

AATGCTAATG CTGCCGCTTG TTCAACTGCT GCTTGTTCCA AGTGTGGTTT AGCAAAACGA 5820

ATGTTATCTC GTACCGTCAT GGAAAAGAGA AAATGGTCCT GTGGCACGTA CCCCATTGCA 5880

CCCATCAATG CATCTAATGT GTAATTTTTA ATGTTATGTT TTCCAAAGGA AATCGTTCCT 5940

TGGTATTGGT CATATTCACG CATCAACAAT TTCAAAATAG TGGTTTTACC AGCGCCTGTT 6000

TTGCCAACAA TCCCCAAGGT TTCTCCTTCT TGCAACGAAA ATTGGATATT TTCCAACGCC 6060

ACCGTCTCTT CTTTTGGATA AGAAAAAGAA TCAATCTTCA TTGAAATTGT CCCTTGCGCC 6120

ATTGTTTTGA TGGCATCTTT TCGTTCGATA ATATGCGTTT TTTCATGTAA TAATTCGTTC 6180

ACACGATCGT AACTAGCATT CCCACGTTCT AACACATTAA ATAAGCGCCC AATCGCAAAC 6240

ATTGGCCAAA CTAGCATCCC AATATAGGCA ATGAAAGAGA CCAACTGACC TAAACTAATC 6300

GTTCCTTCCA TGATGAAACG ACCACCAaTA ATAATTGkCA AAACATAAGA AaCACCmATA 6360

ATCAGCGkAA TAAaCGGaTC AAATrGgCAT CTAAAAAGTc GTcGTGAATT TTGAC 6415

(2) INFORMATION FOR SEQ ID NO: 276:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20633 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

CCTAGGACTG GCCTAAGATG GGGCCCCTCT TTCTAAACCG AAAGGGTCTT ATnGGAAAGG 60

GTAACTCTAT GGTAAATGGA GGTnCnTCTA AATTTCAACC GATTTGGCAA AAAATCCAAT 120

AGCTCAATGT TTGACCAGAA TAAATCAAGT TAGGATTTGC TAATCCGTTT AATGCAGCTA 180

AGGTTTGATA AGTCGTACCA AGCTTAGCTG CAATACTTGA TAAATTATCA CCGTATTGAA 240

CTGTGTAAAT GTTGCTTACT ACTGATCCAT TTACTTTCAA AATTTGGCCA GGGTAAATAA 300

GATTTGGATT TGTCAATCCA TTTAATGCAG CCAATCTTTG ATAGTTTGTT CCGTATTGAT 360

AAGCAATGCT TGATAATGTT TCGCCGTATT GTACCACATG AGTCGCTTCT GGTTGTTTAT 420

CAGGAACAGT TGCCGCATCT GGCAATAATT CAATATTGCC TTTGCTGATC CATGACAAGA 480

TACCTTCaAG CAATACTCTG CTcTCAGTTA CTTCTTGTAC TTTATAGCTG TTTCCTTTTA 540

CCCATTGCGG AATAGtTCGC CAGTTGCCCA AGCATCAACA CTAAATTTCA CTTTAACAGT 600

GTCGCCCACT TTAACATCGG AACTTGGCGT ATTTTCTACT TCTTTACCTT CCTCAATAGC 660

TGGTGTGTTT GTTGCTGGyT GGTTATTCTT TGTATAACCA TTATCAGTAA TACCTGTTAA 720

ATCAACGTTA CCATCTAGCC CTCCTGCAAC GTAGGTTGAC GTGAACTGAA AAATACCGAT 780

ATTTTCAAAT GaAGGAAAAT AATTGTAATT TGGATAAGGT GTCACTTCAT AATCTGGATA 840

TTCTGCCATC CATAATTGAT ACTTCTTTGC AATTCGTGAT AAATCATAAG ATGAAGTAAG 900

ATACCCTTTG TATCCGTAAA GCATCGGTGT ATATCCAGCC TCTTTGATAT AGTCTAACGC 960

CCATAGCGTT ACATCCGTCG ATTGAACGCC GTCTTCCGCA TCTAAGGCGA CAATTGATCC 1020

CTTTGGTGTT TGAACTTTAG GTAAAAAGTA ATCTAATACT TGCTTTGCAT TTTCGTAGGT 1080

AAGGACGTTT TGCCACCATA CATACGTATG CGCTCGTTTA CCTTGAGCAA TCGTACTAGC 1140

TACTTGACTA GAATAAGTAA CTTGATCATA AATCCCATAG TTGTTTTGCC CACCAATTTG 1200

TGAAATTGAG AATTTATCAT GTGCATAACC AAATGTTGCT TGATACCCAT TCCAAATAGA 1260

CAAATCAACG CCTTGGtCTC CTTTTGCAGC AAATGTGCTA ATTGGTGATA CAAAAAATAA 1320

GACTACCAGT AACGTTGCCA ATAGTTTCTT TTTCATTTAT TTATCTCCTT TCCTATCTGA 1380

TAATCCAGGT GTTGTGGGAT CTGTCACAAT ACCTAGAATA GTTAGCACAA CAAATAATGC 1440

ATTAACAACA TCTAACAGTT GCTTATTAAT CATTTCAATT TGAAATTTAT ATCCAAAAGG 1500

AACTGCTACT ACTTGAATTA ATAATAGAAC TGCTGGAATA ATCGAAAGCC AAAATTGTTT 1560

GTTCTTTATT CTTGATTTCC AATTAATCAT TTTTATTTCC TCCAATCCCT CTAAAGAGGG 1620

TTTTATTTTG CTCTTCCAAT CGGCTAATGC GTATTTCATG GTTATTTAAA CGTTCAACAG 1680

CTTCCTTTAG TTCTTTCATG CTATCTTCTA ATTGAGAAAA GACATGATAA AATTTCATTA 1740

ATGCGAAAAT AATtCCCCCT AAGAATGTAA TCACtGCTAA CCATtGTtCT ACTGTTAAGT 1800

cATCCTGCAC CTACTTctAC TTACAAGAAA aACCGCTTAG CTTTCGCTAA ACGGTTCTCC 1860

ACAAATTTTT GTATATTCTT CTTTTGTTAA GCAATTCATA TTCACGTAAT CTACTAAATC 1920

TTGTTTTGTA TAACAATTCC AATCATACAA CTGTTTAATA TTATCGAAAC CTGGAAAAGC 1980

ATTCGTTTTC ATCTTATTCC GCTCCTTTCG TAAGTTCAGC GACTTGTTTC ATCAATTCGC 2040

CAGTCATTCT TTGTGTTTGT TGAATAACTT GATTCTGCTG AGAAACTTGC TTCATTAATT 2100

CCGCATTTTG CTTTTGTAAA AGTTCCAATT CCGTAGGTGG CGTTGGCTCT GGTTCTGGCA 2160

CATTATCAGG ATCGTATATT AAGCTTGTTC CGTCCCAGCG ATAATTAAAA AAATCTGTTG 2220

GCTCTTTTTC TACTTCAAGT TCAATTTTAT TCGGTTGTTC CATTGTGGAA TATCCGTCTA 2280

AATAACCTTG AATATCATCA ATCCAAATTT TCATGTTCTC TTTCCTCCTA GTATTCATAA 2340

ATCGCACTTA ACGCAAACAT TTTACTGCCT GAACCAGCCG TATCACTAGA AGAGGCATTT 2400

AAGTCGTTTC CAGTAATTTT AGTATCCTTA ATATATAAAT ATTTCCGTAC TAAATTAGTA 2460

TAGTATCCAT GTAAAAGGaA AATAACTCCC CCAGAACCTG GATTCTGCGC AtGCTGTTTA 2520

GGCACTAAGA AAAAGTGATA ACACGCACCG TTCAAAGCTC CTTCTTTGGT ATATTCTTGC 2580

CATTGTAAAA TCCAACCATT CTCACACTGT GAAAGAGACT TAGAAGGTAC TTGCCCATTT 2640

CCTGCAGcTC CGCCATACCA TGCACCTGTC CACAACGGTT CTTTTTTGGT AATTTTTTGA 2700

TATGCTTTGT CCGCATCTAT TTTTTTTATA TAAGATTCAA GCCCATCAAT TGCGGCAGTA 2760

TGTGTTTTAA GATAGACTGG TTTTCCTTTT TCTTTTAATT GAACAATATC TGTTGTCATT 2820

ACACTTCCCC TACCTTTTCA AACGTAAtTG CTGGCAATCC ATCTAGCTTt GTTTTATCTt 2880

CTTTAGACAT CAAGCCATTT TTTATTGAGG TTGCAACGTC TGTCGTTGTT GCATTTTGCC 2940

CTGCTGGACC TTGCGGACCA ACGTCTCCTT TATCTCCTTT TGGACCTTGT GGACCTGGGT 3000

CTCCCTTTTC ACCTTTCAAT ACTTCTGGTT TCCCTTCCAC AGCATTCCAA TGTGTTTGAG 3060

GAAATACCTG TGTTCCTCCT TGTTTTACTT TAACAATATC TGTCATTCAA CTTCCCCTAC 3120

TCTCTCAAAC GTAATATCAG GTATTCTGTC AATGGCTTCT TGAACTTTTT GGTCAACATA 3180

TTGTTGATTC ACTCCGCCGC CATCGCCACC ACCAGTTGCT GAAATAACAC CATCTTCTGA 3240

AATAGAAATA TTCGCTCCAG CAGTATAACC TTTCAACTCT TCCAGTTTCG ATTTTAGTTC 3300

AGTGGTGAAA TTTTGATCTG TTTGCTTTAC CGCAGACAAC GTTCCGTCTT CTGTAATTTC 3360

TAACAGTTGG CCAACCTTTA TTCCGCCCAG TTCATCTGCG GTAGCGATTG GAAGAATGTA 3420

CACGCCTCCC TCGCCATTTG ACAACCGTTG AAACATTTCA GCAGTGATAA TACCGTCTGT 3480

TTCTTCTGTC GCATAAGGAA GTTCTGTCAG TGCATTTTCT AAGCCTAGAT CTGCTTTAGT 3540

GATAATTACT GCCCCAGTAT ATCCATTAAC AGATAATACT TTTGATTGAC CCGCAATAAT 3600

TTTTTCTAAT CCTCGAACAG CGGATGCATG TGTAATAGGA TAAAACTGAC GTTCCACGCC 3660

ATTTTCATCG GTTTCCATCA TTCGTTTTGC TTTAACCACT TATTTCACCC ACTTTTTCAA 3720

ACACATAAGC GTTCTGTTTT GTATCATCAA CTGTTGCGAT AACCAATGCC CCATCGATCG 3780

CAGGATAATC AACTGTTCCA ACAATTTCTG TTTCATGATT CAGTGAAAAA GCATCATCTT 3840

GTAAAATAAT CAAGTCACTT ATTTCGCCAT ATTCTAACGT ATATAAGCGT TTCTCTAATT 3900

TCTGATACAA ATATTCCATA TCTGCCAATA AACGTTCAGA AATTGAATTA TGGCGCACTC 3960

CTTGAATGTC TACACGTGCA TCCATTAGCT CGGCTAACAT CGTACCGCCA GGATCAATCG 4020

TTTTTAAAAT ATCTTTGATT GATTCGAACC ATGaAGTGAA ATCTGTTTTT TGCGCATCTC 4080

GCCACGCTTC GAACTCTTCT TTTCTAGCAT TCATCCAATC AGTAAAATCA CCCTTATTTT 4140

CGTTGATAAA AGCGTTCATG TCCGCGATTA AATCTTCAAT GGACTGCCAA TAAGAACCCA 4200

TTTCACCTTC TGTTTTCGAA GCAGCATTCA CTACAAAATA AGAAAAGTTT TGCGTTGCAC 4260

CAATCAGGTT ATCACCTTTA TGAATACTGA AATATGCTTC TTGTCTGTGT AACGACTGCA 4320

TAGAATATTC ATCAAAGGTA TACTGAATAA TCCCTTTTTT GGCATTCACA ATTTTTGCTG 4380

AACGTTGAAT CGGATATTTA TTATCAATAA CCGATTCaAA AAAAAACTTC GCAACCTGTT 4440

AAATCAAGTG GCAAAGCATT TTCAACTAGT ATGGCTTCTA AGACTTCTGT GTTTCGATTT 4500

CCTTGTCGTA CATTCTGAAT CCCAATGTAA TTGTAAGGTT CAGTTGTACT TAGCGTTGCT 4560

TGCCATTTAA CCATTGAAAA ATCCTCCTyT CGTTATTTTG GTGGAATAAC AATCGATTGA 4620

ATAGAATTAG CAAAATATAA TCGGTCATAT TTTGCGACAA TTTGyCCTTG CTCGGCrTTC 4680

TGTTCTATsG TTTGGATACG TCCGTTATTT AAGCCGTAAA TCACGCCCGT GTGACCATAT 4740

GTTGGGTCTA CTGTCCAACC TGTTCCCCAT TGGCCACCTC GTCTAATATT GACGATTGCT 4800

CCTACyACTA AATCTTGATA CGTTGGATTT TGGATTACTC GCCAACCTAC CGCATTCCAA 4860

TCATATGCTT CACCAATATC TGCAGCAGAT GATGTATCAC CAATTACATG TGAAAAGCCA 4920

TAAATTGTTC CTGCACCTAA ACCACAGCCG CCCATAAAAC CAGAATATTC GGCTGGAACG 4980

GCATAACATT GCCCATTACC AAGCCATTTG CCCATTAAGG TCTCCAAATG TTCTATGCCA 5040

GCTTTTCCTG TTGCAGTAGA AGCTTTCAAA TCTTTGAATT TGTCATACCA TrCTTGTGCA 5100

TAGGTTTGTC TTTCTGGATG TGCTGCAGCT GGACGTTCAA AGTTTAATTC AAACGCATAA 5160

GCAGCTGTTT TAGGCGAGCT GACAACTTTA AATTCATCAA CTGTTAATGG ACTTACTTGT 5220

CCTAACCATT GCCCATTGAA CATACACCAA TTAATTAATT GAGCTTGGGC TAATGACGTC 5280

CTATAGTCTT GTTTGATACC TGCAGCTGCG ATTAAGCGTT GTACATATTC TCGGCCATTC 5340

CAAGTTGGTG CGCCTACCAA TGGATATGCT GAACCGTCCC ATTGAACCCA TCCGTAAGCT 5400

GGACCGCCTA TTTGTTCGGT ATCTGGGTTC ATACTTGGAC CAACTTCTCC TTGTACATTT 5460

CCGAGGATAC CTGCAGCAGC TGCTTTGCTG TATCCGTTAG CTAATAGGTA ACTCCATAAG 5520

TCCCAAGCAA ATTTATCTGC ATCGCTTGTA ACTTCTGATG GATAACCACC TGTACCAGCT 5580

CCAGAACCAC CACCACCATT TTGACCAGGG ATAACTTCTT TGCCGCCCAC AATCAATCGA 5640

TCAACTGTCA GAATTGCTTT ACTTCCATTT GGACCAAAAA AGTTAAAATT ATTTCCAACA 5700

AAAAACTGTG TAGGACCAGT AATTAAATGT CCTGTGCCTG ATTGGTTAGA CAAACCAATA 5760

ATTTTTTGGG GATTATCTGC GACTAATAGT AATGAATTCC CATCAGAAAC TATTGGATTT 5820

CCATTTTTAT CTGCTAACCC TGGAAAAGGA TTTCCCTTTG TTCCCATCGT GCCAACGTGA 5880

CTATTACCAT TCCAAAACTC CATCCCTTTT TTAGTTAATT CCATGATTTT TTTCTTATTA 5940

TTCCAAGCTT GTAAAATACC GTTTACCATA CGTAAAATAT CTCCACAGCT ATTAAATGAT 6000

GTTTCAAAAA TATTAGATTT TATTTTGCCA GCTCCTATAA AGTCCGCATT CAATACTCCA 6060

TTAATTCCCC ATGCATTTTT AAATGGaCCT TTCCAACCAG TTCTTGAGAA TCCAATTCCT 6120

TTGTGATTAA TTGCAATCAC ATCTTTTGCA GTATCCGTGG ATTCCGTATC TAAGTAATAA 6180

TGGGTGTTAG GTCGATTTTT AGGATACTGA CGTATACTTC CGCCTTCAAC ACCGTTAATC 6240

AAATCAGTAA TATAATCTAC AAAATCACTC ATATACTCTT TTTTTGTCAA TGTTTTTATA 6300

GCTTCTTGAA AGTCTTGACT TTGCTGTTTA TAAAAAGCAA CTTGGATATC TCCCGCAGTG 6360

ATTTTTATTG TTTtTTCTGC TAAAGCATCA TAGACAATTC CTGTAACTTT CGTTTGAATG 6420

TCAATATCAT AAAGCTTGTG GTACACAGTG AATGTATCGA ATAAATTATA GTTACGCATC 6480

tTAGCAAATT CTTTTGCTTC TTCTGAATCT GTAAGTTTCT CaATTTCTAA TTCAATAGAA 6540

ACTTTAGGCT TATCACTTCC TGGATATAAT GTAGTGAAGT ATTTACTTGC CACTTTATTT 6600

AAGCTAGCTA TATCTTTTAC TCCTTGATCT TCAGTAAACT GAATGTATTG AGCGTAAACA 6660

TCAGGATACT TACTGATATA TTCGCTTTTA ACTGCATTTC CATAAATCCG TTGAGAAGTT 6720

CCGTCTGCTC CACTTTGAAG CTCTGCAAAT GGCAAAACTT TAGTAACAAT TGATTGCCAA 6780

TCAAATTTAA TGGTTAATCC TTTTAAATCT TTACCATAAC GAACAGTTCC AACGTTATCT 6840

CGTCCTCTAC GCCTTAGCAA AGATAATTTA AAAGGCTCTC GTTTTATTTC TCCGCCCCAG 6900

TATTGAAGTA GAGAACCTTG TTCCCCTGCA ATACAATTCA ATACATTTCT AGCTTCAAAT 6960

GTAGTGCTAG AAGCTGTATT TATATCAGAA TATAGTTTGA TATCACAAGG TTCGTCCATG 7020

TTCTGTTCGA TTAATTTCAT TGCTTCTGCA CCATTACGAT TATCAACTGT TACTAGCCTC 7080

ACTTGTCTGT TTCCTAGCTT ATAAGTACGA GATTGGGCAT AAATAACAAT GCTATTAGTA 7140

AAAGTATCTT TAAACGTTTG TTTGATCTCA AAAATGTGGT ATTCTTCTAA GTCATTTGGC 7200

TTTGCTTTAA TTTGATAGCC ATTTTCGAAA TAATCACTAA ATCTGCTAAT CGCTGGATAG 7260

TCCATTTCTA GTTCATATTT TCCGTTTGCT TCTTCAGTGA TTTCGCAACG TGTCGCATCA 7320

ACAAGACGTC CTAATCCATT TGTTGaAAAA TCTTTTTCTC CAGGTTTAAA AATAACTGGA 7380

ATCAAACCTT TCGCCTCCAA TTCGGCTGAA CCTTAAACTC TGTTACTTTA CCAGTCCAGC 7440

GAAAATTATT CTCTCCACAT TTTAAAATCG GATAATCTTT AAAAAGTGTT TTATGATCCA 7500

AGATTTCAAA TGCTCCACCT GATTTTCTAT AAGCTTCTTG TTTTTCTGAA TCTATAATGA 7560

TGTCACCGTT AATTGCTTTT AATGAATATG ATTGATTATT GATAAAAAAA GAAATATCCC 7620

CAGACCCCAA AATCTGAATA ATGGGTTCTG AAGGATATTT TTCTGTATTG ATTAACTGAT 7680

TAGGATTACT TATCCAATAT tGGCCAATAC GATTTTTCTT AAAAGGTCGG ATACTTACAG 7740

TAAATTCAAA AGGAATTAAA aCCCCGCTTT TTCTTGTTCC TGTAAATTTT GGTGGACTCG 7800

TTaCAATCGC CTGATaAATa TAATGCTCAT CAAAATAGAC AATAAAATCA GAATAGTTTC 7860

CCATATCGAG CCAAAAGGAA ATTTCATCTT CTAAAAAAGA AaCTyCTtGk AAAGtATTtG 7920

CTTTCGCATA GCATGTAATG GTACGTTCTA CATTTTTATA ATATGCAAAA TCAACGGCTA 7980

TTGAATCATT ACCCATTCGC TCCCTAAGCT CTACCACACG TCCTGCAGAA AGTCGTTCAG 8040

GTCTTTCTCT CATAAATACA TTGAATTCAG AACTATGTTT TCCATTAAGA AAAAACTGTC 8100

CTCTTTTAAA TTCCACCAAA AGCACCCCCA GTkGCATCAC TATCTCTATT TTTAACAATT 8160

TGgAATATAC TTAACGAGGT CTTTAGCCAT ATCCaTTAAT tGTTTTTCAT TTAATTTTCC 8220

CATAGCTTGT ATATTGATAT TGAAAGTATC ACCGCCAATA TTAGTCGTAG CATTACCTTT 8280

ATTTTTAGCT AAGCTTTCTG TTTGAGCTCC TTGTTGGTTG ATATATCTGC CAGTAGTAGA 8340

AAAATTTGGT AACTCTGTTG GTAAATCGGT CATTTTTTTC ACTGATTTGT CAAGCGTTCC 8400

TTTTTCTTGG TCAATACCAG CTACAACACC TAATACAATA TTTTTACCAA TCATATCCCG 8460

CATCCATCTT GAAGGTGAAT GAATGCCTAA AGCACCTTTG ATTTTTTCTT TAATATTACC 8520

AGCAACTTCT TTAATTTTTT TATTCACAGC ACCAATCATT GAACCAATAC CGTTAACTAA 8580

TCCTTGGATA ATATTTTTAC CAATTTCAAA TAAATCGATA TGGCGCATAT CATTAAAGGT 8640

TTGCTTCACA TTTTCAACTG TATCACTAAC GCTTCTTTTA AGATTATTCC ACGCATTTTT 8700

AGCACCTTGT ACCAAATTGT TGAAAATATT AACTGTTCCC TGTTTTAAGT TTTCCCAACC 8760

GTTAATGATG CCGTCTTTTA TACCTGTCAC AAGATCAACA ATCCACTGTT TAAAATTATT 8820

CCAGGTATCT TTTGACCATT GAACAGTTGC GTTAAATGTA TCAACTGTGC CTTGTTTTAA 8880

GTTATTCCAA CCATCAATTA CACCATTTTT AATGTTTTCT ACTGTTTCAA AGAACCAAGT 8940

TTTCAAACTT TCCCATATTC TAATTGCTTC AAATTTAATA TTTATCCACG TTTCGATGAT 9000

AGAATATTTA ATTTCAATCC AAACGTTGAT CGCTCCATAT TTAATGTCAA TCCAGAGTAA 9060

GGTGAAAAAT AACTTCACAT CAATCCAAAT CTTTTTAATT GTCAACATCA ATCCATTAAA 9120

AATAGAAGTG ACTGAATAGG AAATAGCTGT AACAGTGTTA TAAAAGATAT TTTTAATCCC 9180

GAACCAAATA GTCTGAGCAG CTTCAGCAAT ATTATCCCAA ACGGCAATCA TGTTTTCTTT 9240

TGCCTCTTCC CATCCACCTG TGATCATTGA TGTAATGAAA AGAATTGGAG CTAATAGAAC 9300

ATTTTTTAGA ATGGTGACGA CATTTTCAGC GATCATTTTG ACATTTTCAA TGTTCGCTTT 9360

CATAGCGTTA ACAACCATTT TAAACGCATT TTTGATTCCT GTTACATATG GACCAATATA 9420

TTTCCAGACA AAATCAaATG CTGTTGTGAA AACATCTGAT ATTGATTTTC CAACACCCTT 9480

AAACCAaTCT TTCACATTAT CAAAGCCATT TTTAAAACTT TCTCCAACAC TTTTAGCACT 9540

GTCAGCAGCA CTTTGTTTAA TATTTTCCCA TGTATTTTTT GAGCCTcTTT TGkTGaATyC 9600

CAAAaGTCCA CTGAAAAATT CCTTGGTACC GTTCCACTTA TTTTTAACCC AGTCGGCTGC 9660

ATTTCCAGGT ACTTCTTTCA TCCATGTGCC AGCATTCGAA AAAGCCTCTT TTGTGCCATC 9720

CCACATGTTA CTGAAAAATT TCATTGTGGA ATCCCaAGCT TTCACAACTG TTTCTGCAGC 9780

ACTAGAAATA AATTCTTGTA TATTTTTCCA AATATTTTTA ACAGCATCCC TAAACCCTcG 9840

TTAGTCTTCC AAAGATAAAT AAATCCTGTA ACTAAACCTA CAACTGCAGC TAAAATAGCG 9900

ACAAACGGAT TCGCCAACAT AGTTGAATTA AGTATCGtTG CGCAATTGAT AATCCTTCTG 9960

TTGCTTTTTG CCAAGCCGTG AATGCTGCAC TTACTTTTTT AGCAAGCATC AACGTTCCAA 10020

TACTGCCAGC TAAACCTGCA AGCAATGGAG CATAAGGTTT TAACGTATCA TACAATGTTT 10080

TGGCTGTTTT AATCATTGGC GGAATCATCT CGGCAAATTT AGATAAAGCT GCTTCCATTT 10140

TTGCCCCTTT GTCAGCAATG ATTTCACTAA TACTTCCAAA ACCTGCACTT TTTAAGCCTT 10200

CGTCAATTTT AGTTACAACG TTAGCCACAC CACGAACGAT TGCAGTCTTC ATGTTAGCTA 10260

AACCTGTTTT AATACCAGCG GTAGAATCTT TAGCAATCTG TTCTAATGAT TTAAGACCGC 10320

CACCGCCTTC TTTATTTAAT TTGATTAAAG CATCTTGAAA TTCTTCAACT GAAATTGAGC 10380

CATCAGAAAG CCCTGCTTTC ATCTGTCCAG CTGTTAATCC CATCTGTTTT GCTAAAGCGT 10440

TTAATGCTGG TCCTAAACCA CTATTAATCA TTGAATTCCA AGTTTCAGCA TCTACTTTAC 10500

CATTTGAGAA CGATTGGGAC AGCTGGATAA TAGCATTTTC TACCATTTCA GCAGAACCAC 10560

CAAAACCGAG GATACCGTTA TTTAAAGCTG CGAAAATCTG TTCTGATTTA CCTAAATCAT 10620

TTGTTGATGA AGCGATTAAT TGAACACCTT TAATAGCGCT ATCTAACGGC GTAGGCAGCC 10680

CTTGGATACT TTTCTTTAAG CTATCCATTG TTTTAGATGT TTCACCAGCT GAAAAGCCCA 10740

TATTTTCAAA CACACGATTT GCGTTATTTA ACGTATCTAC ACGATTAATG GCCCCGTCGA 10800

TGTTGCTGGT AATCAATCCA ATACCTTTAG AGATAATTTT AGTTGcTCCG CTAGCTAAAA 10860

AGTTACCAAC AAAAGACGTC CATATGCTCC CAAGAGATCG GCCGCCTTTT TGTCCTGTTC 10920

TATCAACTTC AACATCAAAG CCTTGTAACT TTTTTACTGC TGAATTTAAT CCTTGTGAAA 10980

AGCCAGATTC ATCCAGTATC ATTTTTAAGA CTAAGTCTTC ATTGtTCAAA AAGTACCCCC 11040

TCCCTCTTAG AACATAGTAT TTTCATCAAG ATATTTGATA TTTTCAAATT CGtCTACAGC 11100

ATCTTTAAAT GCATAAATTT TCmAAAGCTC GtTTAAATCT GTGTTTTCGA TCTCATTTAA 11160

AGTCCACCCA TTCTCAAGAA GCGAACTTTT TAGTTCTGCT TCTCGATATT GTGGCGTGTA 11220

CTTAAAATGA GGATGATATA AAAGTTCCGT TACTTTTTTT TCTGTTCAGA ATAAATTGCA 11280

TCATAACCAG AAGTAACAGA ACCTAACAAT TGAGCTGTAA TTTTCAATAA TTCACGAGCA 11340

TCCATACCGT CAATATATTC TTGTCCAGTA AACTGTTTTT CAAAAATAAC gTCAGCAATA 11400

AAGTCAtAGC ATTCTCTTAA AATAGGACGA ATTGCTTCCA TATCATTTGT TTTTGkTGcT 11460

TCyTCTAACC TGATTTGTAA ATCCGTACCT GTATCCatGA CTGAACCTGG TAAAAAtTCT 11520

GCCGACGTGA ATTGTTTTGT AGTATATTTG CTTCCATCTT TAATCATTAA TTTAATTTTT 11580

TGTTGAAATT TACTTGCCAT TTTAATTCCc TCCrTATAAA ATaGGACGAC AAGGTCGkCs 11640

TAAACTGTTA TTTTTAATCT GCTGTTGTTA CATTTAATGT ACATTTTGCG GTAAAGTTAC 11700

CATCTTCTGT TGTGACTACG ATTTCCGTTG TCCCTTCTCC TACAGCAGTA ACCTTCCCTT 11760

GAATTGGTGT TACAGTTCCA ATAGCCTcAC TTTCTGaACG GAACTGATAT TTTTTATTTG 11820

AGGCGTTATC TGGTGTAATT GTCGGTGTTA AGGTTGCTGT TTGGCCAACT TTTAAATTTA 11880

ACTCTGTTTG GTCTAAAGTT ACACCAGTAA CAGCAGTAGT ATTTTCTTTA CGTGGaTCCA 11940

TTACCTCAGT AAACCAGTTT TTAATCATCT CTAAGTCAAC ACCTTCATCG TCTTCATCCA 12000

CGGAATACAT ATAACCCAAC CCTGGAACAT CAACGAAAGA CCCCGTCCAT TCTGGATGGG 12060

TATAAGATAC TGAACTTCCT TCTAATGTAG ATGTTTCATC AGATGTTAAA GCAAATTTTC 12120

CTTTATAGAA AATCGTATAG CGATATTTAC CGTTCGATTT TCGGCGACGA TAAGCAAATG 12180

CGCCATCTGA TGCAATATCA TCCGCAGACC GCAATACGCC ACCCTTTAAT TTTTTTCCCC 12240

CTGTAATTTC AGCTAAAACT TCATTTtGGT AGCCGTTTGT TTCTAAAGTA ACTTCTGCAC 12300

CACCAAATGC AACATACTGA TCTTGAACTA CACTATCGCC ATAGTCAGGC GTTGTTTCTG 12360

TTGTAACATC TGGTTTGATA CTTACAGCAG TACCGATTGT AATTGGCGCT CCGTAAACTG 12420

GAAAAGCGCC CGTTTCGTCT GTTAGTGGGA ACCACGTTGG CTTCTCTACA GAAATGACAC 12480

TTACATTTTT CTTTTTTGCC ATCTATTTTT CACTCCATTC AATTAATTGA GGGAACGCAA 12540

CATTAAAATT GATATGTTGA ATTCCGTCTG TTTTAAACGT TTGATAATCT TCTGGAAACA 12600

ATTCGTTTCC GTCCAAATTC AACACATTAA AAAAAGCCCC ACAGCTTTCT GTTAGGCTTT 12660

TTACTAATTG TTTATCTTTC TTACTATCAA CCAGTGCAAT ATCAACATTG TATGCTTTAT 12720

TTTGAACGTT TTGACCTACA TTATCAGTCA TGCTTTCTTC TACGCTCAAA ACGAAATAAA 12780

ACGGTTCTGA CGATTGCATC ACATTATCAA GATAGATAGG TGCATCCGAG AACTGTTTTA 12840

TTGTGTCAGT AAGCATTTTT AAAATTTTAT CGTACATATC TATCCTTTCT TAACAATGAT 12900

AATCGCCATT TGTTTAAATC GTTTCGGAAT ATATGTTGCA TTAGCTAATT TATTAGATTT 12960

TTGCAACATA AATCTGCCTT TAACAAATCC GCCATTTTTT GTTCGGTGGC CATCGTTTAC 13020

ATATCTAAAA TATTTTTCAT TATTAATCAA TGCTCCCACG ATACGACCaC TAGACAGCTT 13080

tCTAGCTTTA ATGaTTCGrT AACCTCGTCT TAAATCaCCC GATTTAATtG GtGTCATAGG 13140

TACAATTAAC TGATAAATTT TAGCTAACGA ATCATTCACC ATTGCAGCAC CTTCTTTTTC 13200

AGCAATGGGT GTCATTTTCT TTAAATTTGC AATAACTTTA TCAGCATTTG ATTTCATTCk 13260

AAGATCGCTT TTACTCATCA ATTGAACTTC CTGATAGCAC TACTTCTATA TGGCTTGGAT 13320

AATAAAAAGG TTTTTTTGAA AATAACACAT GTTTTTGACC TGTACCTTGA GTAATAGTTA 13380

TTCTATCTCC TTTTTTAACT TTTATATTAG GTTCTAAAAA AAGTTTTTGT TCTTCATAAG 13440

AGATATTAAA CTTAAACGCT TCTTTGTTTT CTATTACAGG TAAGTTTCCC ATACTTCCTT 13500

GAGAGAAAGC ACAGGGTAAC TTGCCATCAT GAATTGGAAA ATAAACTTGT yCAGTAATTC 13560

CGCTTyCCAC ATTTTCAATA TCACTTATtC GCTCAATAmC ACAGGTATCA AAATAAGTAG 13620

CTGCTAAAAC TTCTGCTTCA TTCAATAGAA AAACACCCCG CTATCACAGC CTAAAATACG 13680

TTTAATGGCA CTGCTATAGT TCTTCATGAG TGATTGTATG TCTTTTGATT CAACTACATA 13740

ACTAATTGAT GTATCACCAC GTTTTACACT AGCTACAGAC TTATCTATTT CGTTTTTTAA 13800

AGCTTTATAG ATAACCTCAA TTATAAGTGG TTCAAACTCG TCCCAAGCAA TATCAATTTT 13860

ACAAGTATTG TAAGAATTGA TTTCAAAGAT AACAAGGTTT AAAACAGACA AAATCCTATC 13920

TTCAGAAGCG TTAGGTAGCA TCAATTGAAT TTTCTCAACG ATTTCTTCTT TTTTTTCATC 13980

AACCATAAAG CATCATTACC TAAACTTTAA AAGTATCTGC AGCACGTTCT AATATTTCAA 14040

TAGCTTCTTT ATCGTCTTCT GAAACTATAA ATTCATTATT yTCGTTTGCT GTGaTAAATT 14100

TTTTTGTTTT AGGATGCATA AAGCCCACAA AGTTTTtCTT GTCAAGTACA CGATAAGTTA 14160

CTTCTTTTTT TGCTGTTGCC aTTTTTAATT yCCTCCTTkG TTATTATATT TTAGGCTTTC 14220

AAATTCAAGA TTGCTCCAGA ATTAGAAGCT TTGTATTCAA TTGAATACTC ACCAACTAAT 14280

CCAATCCGTC TTGAATCTGT TGTTTTTGCT AATTCTTCCG CACGCCATTC ACGTAATGGA 14340

CGTAATTTTA CATAATTAGT ATCAATAGCT GCGATTGTTC CATTAGGTAA ATTAGGTTCA 14400

ATTAACGCAA TTCCTGAACC GTAATTTGAG ACAATATTTC CAAGTTGCAA TCCAAAAGTA 14460

AGTTTATCGC CAAATTGCAC AATTTTTGTT GATTTTTCAT CCAACTGATC AATCATTAAT 14520

TCTTGCATAT CAGGTGCTAC TAAACATAAT TTTTCGCCCA TGTATCCTTT TTGGAACATT 14580

GTTTTAAATA AGGCATCAaT ATCTTTTCTT GTTACTGCCC CCGCAGCTGC TGTTTCGACT 14640

TTATTCGTTG AGCTAATCAA ATTTAAAATT CCGTTCATCT GACGACCTTT AGAACCAGAT 14700

TCATCAGCCT TTACACCAAC AATCAATTTA CGATTTAAGT CAATTTTCAT TTCTGTAGCA 14760

CGAAGAGCTA CTTGGcTATT CAATTCATTT CCTACACCAT CTACATTAAT AGCATCTAAT 14820

GTGCCAGATA CAGAAGTTGa TTTTCTGAAA ATTTCAGTAT AGTTGtTAAA CCATGkACGA 14880

TCAGATTCCG CATCTGCGTA TTCTCCGCCT TCTAATTGAG CAGATGAATC ATCATTATTC 14940

ATGCCATATT CTCGCCATTT AATCTCAGTT GACTTGGCAG CTTCAAcTTk GCCTGCGCCT 15000

AACAAATAGC TTAAAAATGG TGTGTTTGGG AACTTGTAAt GCGTTaACTT CCkGTGAAAt 15060

ATCtAAATAC TCAaGATTAT TaAgGAACTT TTTTTCaTTT GTAGTTTCCT CCTAATCGAT 15120

AAAAGCTTGT AATTTTTGTC CTAATGCTAC CTCTGGATTT TCAAACGATT TTGTTTGATT 15180

ACCAGTTCCC ATGTTGGTTT GTTGCGATTT ATTACCAAAA GCTTTAGTCA TTTCTACATT 15240

TTTAATAGCT TCTGCATGCT TATCATTTAT TGCTTCTAAA AGCTCTGTAA AGCCTTCTAC 15300

AGCCTTCTTA GTAAATTCAG TATCTGAACT AACAAGATTA TTTAACATAA ATTGAGAAAT 15360

AGAATCTTTC AAATCTCCGT CCCAATCTAA GCCAGCAATT TTTTCTGCAA CAAAAGCCTT 15420

ATTATCACTA GTCACACGTA ATGCCTTTTC AGCTTCAAAT TCAGCCTGTA ATTCTTCTAA 15480

TTTAATTTGT TCAGGAGTTT TATTTTTCTT AGATTCTTCA TACTCCTTGA TTGTTGTTTC 15540

CTTGATTTTT TCAAGATTAT TTTGTTTCCA AGCTTCTAAT tGTTTATCTG CAGCTGATTG 15600

TGACTGTGAT TGTACAAATT TTTGTGCTTC CTCATTTGAT TCTACAAATG CCTTAAAATC 15660

ATCGAAAGTG AAGTTTGTAC CACCGTCTTC TTCAGCAAAC ATTTGTAAAT TCATTGGTAA 15720

TAGGTTTGGT TTCATTTTGT TTCTCCTTTC GCCCCACGAT TCGACTAATC GCCCCGCATT 15780

GCTTTAGATT TATTTATTGC GCCCCACCAT TCAATTAAGC CCAGCATTGC GCTAGTTTAA 15840

CGTCATTTCG GACAAAATAA AAAGCCTAAC TTTCGTCAGA CTTTAATTGC TTTTCTTCTC 15900

TTAGTAAATG CTCTTCATAA TCTGCATCTA AGTAATCATA GGGgATCCAT CTAATCACTT 15960

CCAATTCTTA TGGACCAGTT CAGCACCTAA CATTTGATAA TCAGTGACAG CATCTTTTAC 16020

GTTTTGCAGA GTCCTAGACA CAATCGAAAT AGTTAATTTA CTTTTTTTAC CTGGTAACGA 16080

ATACAAAATA TCAACGTGAC AATAATTACC ACCCCAAACT GATTTAAGCT CATCTTTGAC 16140

GATATTACCG TTGCCATCTC TCAAAGTGTG TTTGGTTAAG TATCGTTCGT TTTCTTGTTC 16200

AAAAGCTTTT TTATAGGCTT TTTCTGTACC ATTGGTAACT TCTAGATTTA ATACTGCTTC 16260

GAATAATCCT TTCATGATCT CACCTCCAAT TTAGGGTATA AAAATAGCAC TCAAAGTTAT 16320

CCTTTAAGTG CTTAGTAAAG TCTGCTATGG TAATCTGACG GGACGTGGAC CGATTTATTT 16380

TCTTTAATAT GCTCATCAAT GAATTTCGCT AATTTTTTTG CACCTAGAAT AGAAAAATCG 16440

AAGTCATCAC TCTTTGTTAC ATCGATATAT TCATATATTG GAAACTCATC ATGGAAATGT 16500

TGTTCATACT TTTTAACTGC ATCATCAATA ATAGACCATG CGCCATCTTC GCAACTAAAC 16560

ACTTATTTCA CCGCCTTTAA AATATCATCT ACTAATTTTT CCCATTGGCT TACTGCTGTC 16620

GGAAAAATCT CTTTCATTAA TTCTCTTGAC TTATCGTTAG CAACCACTTC TGTCATATGT 16680

GCAAAAAATT CCGTTTCTGT GCTACCTGTT GTTTGCCAAT AGCGCTTGCC GTGACCAAAA 16740

CCTAACGGAT AGTCTCCTAA AGAACCTATA GATTCAATCA TATCTGAAAT AGCGGATACT 16800

GAGTTAGGAT TTTCCCTAAC TATTTTTGAT AATCGkTCTC yAGTTTTTtG TAAGTTCTCT 16860

GTATnnnTTT tATAAATTTC CGCCCTTAAG TCATCaTAGG CCTTTTTCTT ATCCGCATTT 16920

CTAGGTCTAT TACCTAAATC ATTATAGGAT GGCAAATCTC CGTTTACATA TTTCCAAAAA 16980

TCGCGTTCTA GCGCTTCTTT AATGTTATAC TCAGAAAGTG ACGATGCGTG TGTTATATGC 17040

TCTTCTATAA AAGTGGTTCG TCTTCCTAGC TTTCTTTTTA TCAGTTTTCC TGTCGGCATC 17100

GTATTTTTTC CTGTTAATAC TTCAAGCCCA AGATGATCCA AAGCATGTCC TATCTCATGA 17160

TAAACTGTTT GATAAGGATT AACACCCCTA TCCCCATCAA ATGATTTTTG ACTAAGTTGA 17220

ATTGATTTGC CTTGTGCAAA GGCTCTTACT TCTTTGATGT CTTTAAACTC CAACTTATGC 17280

CCTATCACAT TTATTAAGTT TAGCACCCTT TGGTCAGTTA TCCCATCTAA ATGATCTAAA 17340

AAATTAGAAT AATTTTCTTT TCCTAAAGCT TCTTGCATGT TTGTTTTGTC TAATACGGAT 17400

TTTAAATCGA GTTTAGTTGC ACTAACTTTT GGATGATTTA ATTTAATACC CCACTGCTTA 17460

CTTGCGAATT CATCCAATAA ATCACCAAAC ATCTCTTCAT ACAATTTATC TATGTCATCG 17520

CTAATTtCAG GAATAATTGG TATTTCAGTA CmACGGCAAC GTCCATGATA CGGTGGATGC 17580

CAATCATCTT TAATCTCTTT TCCATGACGT CCACCACAAA TAGAACAAAC ACGCTCATCT 17640

TCTGCCGACC AGCTTTGTGT TTGCTTAACA CCTATATCCT TTAGCGATTT TCTTACACCT 17700

TCTACCGCAA AATGTGAATA TTCCGTTCTA ACAAGATTTT CAATCGAACG ATTAAACTTT 17760

CCTTGTTCCA ACTTAAACAT ACCGCTAATA ACACCATCGT TTTTCATCGT TCTAAGAGCT 17820

TCCACAACTC CTTCACCACT TGCCAATGAA TTAATAATGG AATTGCTCAA ACGTTGCTCT 17880

AGGGTTGATA TATTGCCCCA TAAACGAGAT GAAAATGTTT TTCCGCTCCA CGGaTAGTTC 17940

ATGATGTTTT CTAGTTCATT CTTAGTTAAA CCAGGTGCTG AACCGCCTAA TAATTGTATC 18000

AACGCATTAG AATTAGAATT GTAGATTCGT TTTGTGATAT TCTCTAAGTC GTTATTAAAT 18060

TTACCGTTAA CATCACTAGC TATTGCTTCA CCTGCAATGG TAGAAAAAAT GTCTGCTCGT 18120

AATTGTAACA AGCGATTAAC TTTCGCATAG TCAAAGGATG GAAAATATTC ATCAATGAAT 18180

TGTTTATAAA CTTCATCTGA TTCCATCAAC TTTTCATAGT TCTTCTCAAT ATACTTACGG 18240

TACTTCTCTT GGTCTCGTTT GCTAAAATCT TCTAGCATCT CACTTTGAGT AATACCGTGT 18300

AAATCAGCTT GTGCCAATAA CTGTCTTTGA ATTTTAACTA AAGCACGTTC GAAAACAGAT 18360

TCTAGTTCAC TAAGAGTTTT CTTTTCCAGT TTCAAGCGTG CTTTGTCTTC TAATTCACGA 18420

CGTTTTTCCC AGTAGCGTTC ACTAACTGTT GCTTTCTTCT TCGTCATTAT CCGCACCACC 18480

TAGTTTGTAG TCACCACTAG GATAtCTTGA CCTTGTTCTA AATTCATCAA GTCCATTTCA 18540

TAATCTGGGT CTTTAGCAAA TGGAATCTGA TTAATAATTG TTCGTTTAGA TAGAAACGGC 18600

GAAAGTTTTG GcAATGCATC AGCAAGATAG CCAATATCTG TTGGcAAACT GCGACTGAAT 18660

GTAAACACAA TTTTAGACAC ATCTAAATCA AGCTTATCGT TAAATTTAAT GAAAGCTGCC 18720

ATCGTTTCAG CGACTTCTTT TAAACCTTCC CTGAAATACT GCTCTTTTGT ATTTGTTTTT 18780

GCTTCTAAGG CAATAATTTG CCATTTACGA GCTTCACCAG AAGCGTTTGa TTTAAATACC 18840

TCATCATTAA AATCAATTGA TTTTGTGATT GTATAGAACA TTTTTTTTAG CTGGTTTAGA 18900

TGATACTCGT TGAAATCTTT ATTAATATCT TTCGTTACAT ACCCAACCTT AGCTTGTGGA 18960

TCTGGTAAGT TAAGGATACC TAATTGTTCC ATCATTCTTT GTGCTTTTTC TTCATCTAAT 19020

ATTGAGCCGC TAATGGCCAT GTAAGCAAGT TTAAACTGTT CAAyTcGTTT TGTTGGTCTG 19080

ATAAGCTTCT ATCAAATGCA TCAGAAAGTT CTTCCGCCAC TTCAAAATCG CAATAACGAT 19140

TCGTATTGTT TTTAAATTCT GATAGGTAGA ACGTTTCTAG CGGGTTTTCT ACTTCCTcAA 19200

TTAATTTAAA TGTTCCAGAT ACACTGACTA AATTAAATTC AATGTATCTG CTATATATGC 19260

GTATTCTTTC TTTAGTAATG aCTTkCATTT CTTCAAAAAA TTTTTTTTGA TGTGTGtCGT 19320

AtTTTTCACG AATAAAGATA tCTGCATTTT CGTATTTTTC AGCTTTCCAT GGTTCGATAT 19380

TGCTCGCCCA TAATTGCCAA CCTTCCTTGG TTTTAACAGG TTCTAACAAA CGAAAAGCAA 19440

CACCACAAGC TCCTTGAAAC CGAGCTGTGT CAGAATCAAG CATGGCAAAC CGCATATCAT 19500

TCACTAACTC TGTCAGCCGT TCGAACTCTT TTGGTGTTTT GGTATTTTTT AATGTATTAT 19560

TCAAAAATAA ATCTTTAGCA CGTTGCATAA TTGATCGTCT CTGCTCTGTA ATATCGTAAT 19620

CCCACTTAAT TGGAATACCT GTGAAATGGT CCGCTGCTTG ATCGACAATA GTATTGTATA 19680

AACCAGCATG AAGTTTATTA TTCACTTTTA TAATCTTTGT GTTTGGTTTA GGTCTTCTAT 19740

CGATCTCATT TTGTTCGCTT GTATAAGCTT TGTATTTGCG CTCTCTATCA TCAAAAAATG 19800

GCTTCATTTC AGTAATAAAA TCATTAGGAT CGAAAACTTC TTCATTAATT TGTGTAGAAT 19860

ATTTTGTTCG TAATCTTTTA TACCGCTTCA AACTTAAATT GTTTTGAAAC ATTTTCACAC 19920

CTCCTAATAT TGGATAAATC TGACTTGGTT ATCATTTGCT AATGTTTCTA CAATGCCTGT 19980

TACAGCATCT GGTGCATCAT CGTGTCTGTT TTTACCTTCC CTTTGGTAGG TCGTCATAGC 20040

TTCATAAAAT TCTGGCCATC TTATTTTCCA ATCTGATGGG AAATACACAT TATTTTCTAC 20100

AAGTGCCGAA TTAGAAAGAA TACGTGATTG TTTATTTGCC GATTGATGAA ACGGCTCGTA 20160

ATAAGCACCA CGATATCCTC GTTCTTTAAC AATTCTTTCT GAATTACGAG AAAAACCACG 20220

CCCACCAGAG TTAGACTCAA TGCGGACATG GTTTACTTGG TTATTTTTAA TTTGTTCAGC 20280

GTGTGCTGTT TCCGTTTTTT CCATTGGTTC TTTTGTAAAT AGAACATCCA ATACTTCTGC 20340

TTTATGATCT TCTGTCTCAC CAAAAACAAT TGAACAAAGA TTATCAGCAC CAGTGTCTGC 20400

TGTATCGGTA TAATTCCATA TTTTAATGTA ATTAGAGCGT GTTTCATATG TTGAAAATTT 20460

TTGGTACAAA CGACCTTTTA AATCAATCGG TTCTTGTTGA TAGTTAGCAG AAGCAATATC 20520

TGCACCCATC GTTTTCTTTT TACGGAAATA TTCTTCTTTA GAAAGGACAG ACTCACAAAG 20580

CATGGTGTCT GTTTCTcGTT ATATGCTTTC ATGCTAATAT GTTTTACTTT ATA 20633

(2) INFORMATION FOR SEQ ID NO: 277:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1663 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

ACATTGnTGA AAGAATTGTT GACGGAAGAG GATGATCAAT TTAACTTTTT CTCATTTGAT 60

ATGGAGGAAA CACCTTTAAG TGCAGCGATT GCTGAAGCGG AAACAATTCC GTTTTTTGGT 120

GATTATCGCC TTGTCTTTGT GGAAAATCCT TACTTTTTAA CTGCCGAAAG AAAAACAAAC 180

GGTGTGGAAC ATGATGTCAA TCTTTTATTG AATTATTTGG AAAGTCCATC ACCAACAACT 240

GTTCTTGTTT TTATGGCCTC ATATGAAAAA TTAGATGAAC GCAAAAAAAT TACCAAAGCA 300

TTGAAAAAAA ATGCCACAGT AGTTGACGTC AATCCAATGG ATGAAAAGGC TATTCGTCAG 360

TACATTTCTC aAACGATTGA GAGCGAaGGG TACACTATTC GCTCCGATGC TTTTGATTTA 420

TTACTTCAAT TGACAGACTT GAACTTATCA AAAATTATGG GaGAATTACA GAAGCTGTAT 480

TTATATGCCC TTGAAACGAA GACAATTACA CGAGGTGCGG TTCAAGAATT AGTCCCTAAG 540

TCTCTGGAAC ATAATGTTTT TGACTTAACC AACGAAGTAC TTTCTGGAAA CAGTGAAAAA 600

GCTGTTCAAT TGTATGAAGA TTTACTTTTA CAAGGAGAaG AAACCATTAA GTTAAATGCA 660

ATCTTATTGA ATCAGATTCG CCTTTTTCTG CAAACAAAAA TTTTGGCTAA GATGGGCTAT 720

CAACAAGCGA ATATTGCTGA TACATTAAAG GTTCATCCTT ATCGAGTCAA GTTAGCGTTA 780

CAACAAGTCC GACGATTTGA ATTGTCTCGA TTAGAAAGAC TTTATGATGA ATTAGTTGAA 840

AATGATTACC GCATGAAAAC GGGCCAAATG GATAAAGAAT TACTATTCCA ATTATTTATT 900

TTAAAATTGA CTGCACAAGC AAAATAAGAA AATGCCAGAG GCAATcGTCA TACAACGATT 960

GTCTCTGGCA CTTTCTTATT ATTTAGCTCC TAATAGTTCT TCTTTTCCGT GAACAAGCTG 1020

TGTCAACATG GCACAGAAAG GTGTTGCTAC GTTGTATTTT TGTCCTTTGC GCCAAACGGC 1080

ACCATTAATG TAATCAATTT CTGTGAGACG ATGATTTTTA ATTAAATCTT GATACATAGA 1140

CGGATAGTGC AAGCCAATGC CATTTGGATC ATAAGTTTGA ACAATATGTG TATAGACTTC 1200

TGCTTGATCT AAATAAATGG CTTCTTTTTC TGCGACTGCT GCAAATTCAC TGATTAATGT 1260

TTTAACAAGG GATTCACTAA CAGGTAGTGC GCCAAATTCG GCAATATTAC AATCAAGGAT 1320

TGTACACAAA CCGTTTAGCG TGCCATTGAC ACAGGCTTTC CGCCAAATAG AGTAGCGAAC 1380

ATTGCTACTA TAGCTTGGaT TTAAACCAGC TTTTTGGAAG ACATCAACGA CTTCTAAAGC 1440

AAATTTTTTA CCAGAAGGAT CAATATTTTC TAATTCAATT TCGcCATCAC CTAAAAGTTT 1500

AACACGTCCC GGTCCTTCTA AACCAGCTGT CCACAkGGTA ATGCCAACCA AGATGTTTtC 1560

TTTTGGtACA wATTTtCCtA GTACATCTTC ATGACCTAAG CCATTTAATA AACATAGTAC 1620

ATACGTTTTT TCAGTAATCA TTGGTTGAAT CGCTTTAAAC ATT 1663

(2) INFORMATION FOR SEQ ID NO: 278:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1587 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 278:

AAAAAACAnG GAATCCCTTC CCAAAATCTT CTTTCATTTT CAAGAAGCCT TTAAATAACT 60

GTAAGTCATC TTGTGTGGTw AACTTAATGT TTTTATCTGA GCCTAAAGCA AAATGCAGTG 120

TTTCGCCTAA ATCCaCCATC ATAGTATTCG TGTATGACGA TTCAGAAATA CCAATATTCT 180

CACGGAATGC TTTTTCGTAA GCCCAAGTTA ATTTTTCAAA TTTATAGGCT TGTGGGGTTG 240

ACACGCGACG TAATGTTTCA CGATTAATAT ATTGTCGTGT TGTTTCCTCA GTTTCTTTTA 300

CAAAAATTTG TTCATTATAT GGCAACGAAG TCACTGCATT TCCGTATTCT TGACACTTAA 360

CAATTACATC AGATAAAACT AATTCATCAA CTAATGGACG AATCCCGTCA TGGATAACAA 420

TAGTATCTTC TGGGTTTGAA TGTTCTTTTA AGAAATTAAC CCCATTGTTA ATTGATTCTT 480

GACCAGAATT TCCTCCTGGa ATAATCCATT GTAATTTGTC AATATGGtAC TCTTTTGCAT 540

AGGCATCTAG TGTTTGTTCC CAACCTTTTT TACACACAAC TAAGATTCGA TCAATGAGTG 600

GATGCTTTTG AAAAGATTCC AATGTGTAGA TAATTATCGG TTTTTCCTCT ACCATAATAA 660

ATTGTTTAGG AATCTCCTGC CCCATTCGTT TACCGACCCC ACCGGCAATA ATTAAAGCTG 720

TAATCATTTG ATACAAAATC CTTTCTCTAA CTAGTCAACC ACTAGTTTTA TTTCAATTTT 780

GTTTwACTTA TATATTCATC AACACAGCCA AAACTGCATT CAATCTTTTA CATTTTCTCA 840

TAACTATCAG TAAGTTTCAm TAATTTATCG TAGACCTTAA TCGAACTGTA AATGAGAGTG 900

TAAAATATTT TGTGTAAATG AAAAAATCCA TACAAAAAAG GAAGTCGCTT CTGTAGAATA 960

AAGTTAACGA CAACCAATTC ACAGAAAAGA GGACTTCCCT ATGAATGATT TTACTACAGA 1020

AATTGTGCAA ACTCTAGTCA CTAAAGGCGA TTTAAATGAA TTATTCCGTT CGCACTTAGA 1080

AAAAGCGATA AACACACTCC TACGGACTGA ATTAACGGCT TTTTTAGATT ACGAAAAATA 1140

TGATCGCACT GGTTTTAATT CAGGTAATTC GAGAAACGGT TCTTACTTTC GATCAATCAA 1200

AACCGAATAT GGTGAATTAA CATTGGAAAT ACCTAGAGAT CGTAATGGTG AGTTTAAACA 1260

ACAAACTTTA CCAGCCTACA AAAGAACAAA CGATACATTG GAAACCACTA TTATCCATTT 1320

ATTCGAAAAA GGTGTTACGA TGTCTGAAAT TGCTGATTTG ATCGAAAAAA TGTACGGTCA 1380

TCACTATACT CCACAAACCA TGTCCAACAT GACTAAAGTT CTGACTGAAG AAGTAAATGC 1440

CTTTAAATCC AGAGCCTTAA ATGATAAGTA TGTCGCTATT TTTATGGACG CTACTTACAT 1500

TCCACTAAAA cGTCAAACcG TATCCAnAGA AGCGATTTAn TnTTGCCATT GGTATACGAG 1560

TAGACGGCAC TAAAGGAGTA CTGAGTT 1587

(2) INFORMATION FOR SEQ ID NO: 279:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15363 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

```
CATTAAAAAG ACGTTCATCA AAAACGAGCC CAAGACTTAC TTGACCTAAC GCTTGTGCGC     60
CTCCTAAAAG AAAAAATAAC ACCGTCTGCA CAAAGACATA CAAGGAATCC ATCCCTAAAA    120
TGTTTGGCGT TAGGAAATGA TTTTGCGTGA CTGTTTGAAA ACTAATAGTT GAAAAAGCCG    180
CAGCGATTCC AACAAAAATA AAAGCCAATA ACTTTTTGCC TCGTAACGCC AAAGCAAAGG    240
ACCAATTGCC ATACGTGTTA TAAGTTAAwT AmCAAAGGCT CACAATTACT ACGGCAGCTA    300
AAAGCAAAAG TAATGAAAGC CATTTTTTTA AAGGCCGTTT CATTGTTTGC GCCCCCTTAG    360
TAATAATAAA ATGAACAAAA TGCTCCCTAA TGTTCCGACA ACCACGCTGA CAGGAATTTC    420
ATAGGGGGCA ATCAATACTC GTGCTAAAAT ATCACAGATT AACAAGAACA AACTCCCGCA    480
AATCCCAGTT AACCACAAGG TATTTTTCAT ATGATCGCCA TAACGTAAAG ACACGAGATT    540
GGGAATCACA ACCCCTAAAA ACGGAATATT CCCCACCATC AATAAAACGA TGGCACTTGC    600
TAAAGCAACC AGCGTCAAAC CTAAAAACTG AATTCGCTGG TAATGGACGC CAAGACTATT    660
TGCAAATGCT TCACCCATAC CAACGACAGT GAAATGATAA GCAAATAAAT AAATAACAAA    720
CCACAAAGGA ATCATTAAGA ATAACCATTC ATAATTCCCT TTCATAACTG TGGAAAAATT    780
TcCTTGTAAC CAAGCAGACA TATTTTGAAC CAATTGGTAT TGATAGGCAA AAAATGTGGC    840
AATTGAGCCT AGAATGTTAC CAAACATCAC GCCCACCAAA GGGACCATCA CTTGATTCTT    900
ACTTGGTAAT AATCGAATGA GCGCTAAGAA TAATAAGGTC CCGCTTAGCG CAAAAATAAA    960
AGCAATTGCC GTTCGATAGA GAAACGGAGC TGTTGGGAAA AATAACATTG CGACTAAAAT   1020
ACCCAACCTT GCACTATCGG TTGTTCCAGC GACACTAGGA GCGACAAATT TATTTTGTGT   1080
GAGTTGCTGC ATAATCAATC CAGAAATACT CATCGTACTA CCAGCAATGA TTAAACTCAA   1140
CGTGCGGGGA ACACGGGTGG TAAATAAAAC AAATTGTTGC TGCGCATCTA ATTGAAACAA   1200
CTGAGAAAGT GCAATGTCCT TCACCCCTAT CACTAATGAC AATAAACTGA AAGTAATAG    1260
TAGAACAAAC AAACCAATCT TTTTCATGAG AAAACTCCGT TTTTATTTAA TTGAGAATGA   1320
TTCTCATTCT TAATTACAGT GTATGAGAAA ACCATCACAA ATGCAACCAC TTTTTTTCCT   1380
TTATTAAAAA CTTCTTCTCA AGCCGACTGA CCAACGACTT ATCTAAGATA TGCTATAATG   1440
GAGGGAAACG TTAGGAAAGG AGTGTTACTG AATGGCGAAT GATGACTTAC AACAACGTCT   1500
GGATAAAGGA ATGTACGGTA CGCCCCTCGT TAATCCTGAA gAACAACACA AATACATGGG   1560
CACATTTCGT GAACGCTGCC GTCTCTCCAT GACCGTCGCT GAAATGAAAG ATGCACAGAA   1620
CCAAAAACAT TTATTAGAAG AATTAGCGAA CATCCtGAAG TTACAGTACT TTTAAATGGC   1680
GAAATCTCTT CTGACCTGCA AAGTACCTAT ATTAAATTAT TAAATCAGCA TGGCGCAAAC   1740
TTTAAGATTG TTAACAATTT TGTGGAGAAC AACCCTGACT CTCTCGGCTT ACTGTTAGCT   1800
GAAAAACATG CAGTCGATGA ACCTGTGATT GACGTTACAG AAAAATATCC TCAAGCAACA   1860
GAAACGCCAA AAGAAGAGCC AACAGCTAAA AAAAGCTTTT GGCAAAAACT ATTTCACTCA   1920
TAAAAAGCTC AAGGAAAGTT ATGAGCGCTT CCATTTTAAA TAGAATTATG CTATTATGAA   1980
```

```
AATGAAAAAG AGAGGTATGC GGGTACATAC CTCTCCTTTT ATACCAACAC CAGTTATAAG      2040

AACTGGTGGC TTACTGAGTT AAGTTTTGAA TCTTAACCGT TCGGCCTACG AAGCTTGTGA      2100

ACGGTTATTT TTTATTGTTT TTCGTTGCTT CAAGGATACC GAAAATCAGT AGTGCAACAA      2160

ACGCTGCAAA ACTAATCATC AATTCCAATG CCGCTTCGAT TGACAAAAAA ATGCTCCTTT      2220

CGTTAAAGTG CATCGGATAC ACATAGGCAT CACCTGCCTC TCTAACTCAG AGCCACCGTC      2280

ATACAACTTT CATTGGTAAT AGCTTTAGTA TAACCTATTG GGTGTACCAA GAAAACGTTT      2340

TTATAGGAAT TTTCTATAAA ATGATATAAG TTATGAGCGC TTCCATTTTA AATAGAATTA      2400

TGCTATTATG AAAATGAAAA GAGAGATATG CTACTACATA CCTCTCCTTT ATACCAACAC      2460

CAGTTATAAG AACTGGTGGC TTACTGAGTT AAGTTTTTGT TGTTCTTTAA CCGTTCAGCC      2520

GTCTAAGCTT ATGAACGGTT ATTTTTTATC GTTTTTCGTT GCTTCAAGGA TACCGAAAAT      2580

CAGTAACGCA ACAAGGGTTG CAAAACTAAT CATCAATTCC AATGTCGCTT CGATTGACAA      2640

AAAAAATGCT CCTTTCGTTA AAGTGCATCG GATACACATA GGCATCACCT GCCTCTCTAA      2700

CTCAGAGCCA CCGTCATACA ACTTTCATTG GTAATAGTTT TAGTATAACC TATTGGGTGC      2760

ACCAAGAAAA CGTTTTTATA AAGAATACCG CCAAAAAATC TATTATGAAT TTTTTGGCGG      2820

TATTCTTTTT TTATTGAACC ACTTCTGGGT ATAAAAGTTT TAAAACGTCC ATTGTCAAGC      2880

GTCTACCCTG TCCAGCGTAT GGATACACAT GCATGTGCAT GGCTGGATTA AAGTTCGCTT      2940

CGATAATCCC GTACGTTAAG CTATCACGTG TGCCTTTTAC GTCTTTGTCA GGAATGATTA      3000

AATCAATGCC ACAAATTTTG GCTCCTAAAG CAGCTACGGC CTCAATGGCG ATTTGTTTAT      3060

AACTATCATC AATGACATCG GTCATATCAA TCGAATCCCC GCCCGTGCTA ACATTAGAAT      3120

TTTCTCGCAA GTACACGATT TGCTCTTTTT CTGGCACAGA ATAGATAGTT AAACCTTGTT      3180

CTTTCAACAT TAGTTTTTCT AAATCATTTA ACTGGATTAA CTCTAGTGGT GCACGGTGAT      3240

TGGTCCCCCG CAATGGATCA CTATTTTTAG CGGCCACCAA TTCTTCTACA GTGTGTTTGC      3300

CATCTCCGGT AACATTGGCT GGCACGCGCA ACATGATGGC TTTTACATCA TTATCTAACA      3360

CAAAGAACCG ATATTCTGTT CCAGGTAAAA ATTCTTCAAT TAAAACCGCT GTGTCCTCTT      3420

TAAAAGCAAT CCGTAACGCT TCCGTAAAGT CTTCCAACGA AGCTCCTTCT TTAAAAATCG      3480

TAATTCCTAA ACCGTAATTC GTTGATTTTG GTTTCACGAC AAACGCTTTA TTGGCGTAGC      3540

GTAAATGTGC TTCTTGTGCC TCAATAAAAG ATGAAAATTC TTCACCGCCA GGCACATGGA      3600

AACCTGCTGC AGACAAAATT TTCTTTGTCA CGGTTTTGTT TTCCATAATC AATGGTACCA      3660

CGTAGTTATC TTTGCTAGTC ATGTTGGCAT TTTTGACGTA TTCAATGTGC TCGCCATGTT      3720

GCAGTTTCAA AAATTGTTCT TGTTCATCTA AAATTTCGAT TTCCAAACCT TTTTGAATCG      3780

CATCAAATAA GAAAATTTGT GTGGATAATT CCATCTCACG GAAACCAGCT AATTGATAAG      3840

GGCGTTCAAA GGCCATACTT TGATATTGGT TTCCAAAAAT ATTTCCCAGT TCACTGTTGG      3900

AGTTTTCTTG AATAATCGTC CACATTTTGC CAGAAACGGT GTCTTGTGGA TTCCGCAGTT      3960
```

```
GTTGATAATA CTTACCGACA ACTTCTTTGC CTTTGCGAAT GCCTAGAGCA TCTAACATAT     4020
CAATCATTTC TGAAAAAATC CGATCGCCTT CTGCAATTAA CTTAATCGTT TCATGAGGAT     4080
GACCAAGAGC CACTTGTTCA TTAAAAATAT CCCCAGTTTT CACCCATTCA TCCGCTTCTT     4140
CTTTTTCATC TGTCCACAAT AAATACAACA TGAAATAATG TAAGAAATCC ACAGTATCTT     4200
CCACAATGCC TAAACGTGAA AAAGGATTTA AGTCTAAATT ACGCAACTCG ATATAGCGAA     4260
TACCTGTTTT AGGCAGATCA GACATTTGTT TCCCGCCACG TAAGCGCACA GGCGCATAAA     4320
ATTCTTTTTC TTCAGAAAGT AAACCATTTT CCACCATGCG ATGAATATCT TCTAAATAGC     4380
GTTCCAATGA GGCATACGAT ACTTTCACAT TGTCATGATT TCTGTAGCCA TACGTACTAT     4440
TCCGAATACT GCGAACAGGT TCTTGCGGTT GGTCGTCGTA GACCCTAAAG TAGCGGTCTT     4500
CACTAACTGG TGAAGCCCCA AAAAGATACG TAATTAACCA GCGATAACGT AAAAAGTTAC     4560
GGGCAACTTT CATGTACACT TTCGTTTTAA ATTGTTTACA ATCTGTCACT TCGGATTGTT     4620
CATCATATAA TTGCTGAATC AGGGCTTGGT CATATTCAAA ATTAAAATGA ATTCCGCTGA     4680
CCATTTGTTT TCGTTTGCCA TACTCTTTTG CCAAATAACG ACGATATAAC ACCGCATCAT     4740
ATTGATCTAA TTTAGCAATT TTAATCTCTT CATCTTTTGT TGGTAATTGT GGCGGCATAC     4800
TTAATGGCCA CAGCATTTCA TCTTCTGGAA TCGAACGACG AGCCACATCG TGAATGGCAT     4860
CTAAAAAACG AAGCATTTCT GTGCCGCTAT TTGCTACAGG CGTGATTAGT TCTAATTGTG     4920
TTTCACTAAA ATCTGTTTGA ATATATGGAT GATAAGAACG GTTACCAAAA ACCGTGGGAT     4980
GCTCAGTTGT CGCTAAAAGC CCTTCTCGTG TACTACGTTG GTTTTCTTTT TCTAAaCCAA     5040
AACGAGCCAT CAATACGTAA GGACGAACAT tTTTCTTTTG CATTAATTCT CTATAATTCA     5100
TCTTGTCGCC TTTCTTTCTA CAAACTATCT GTTTCTCTTA CTATTATATA GGAAAACGAG     5160
TAAATATGCT GTTTTTTTGA ACGTTCGCTT GGAAATTTTC GAGACACTTT TAGAGAGCGC     5220
CTTTTTCCGT TATACTGAAA ACAACTAGTT TATAGAAAGT GGGGATTGCT CATGAAAAAT     5280
CGTGCATTAC TTTTAATTGA CTTTCAAAAA GGCATCGAAA GTCCGACGCA GCAATTGTAC     5340
CGTCTACCAG CAGTGTTGGA TAAAGTCAAT CAGCGTATTG CTGTCTACCG CCAACACCAT     5400
GCGCCCATTA TTTTTGTTCA ACATGAaGAA ACCGAATTAC CTTTTGGCTC aGACTCTTGG     5460
CAACTTTTTG AAAAGCTGGA TACACAACCC ACAGATTTTT TCATCCGAAA AACACATGCG     5520
AATGCCTTTT ACCAAACAAA TTTAAATGAT TTaTTGACGG AACAAGCCGt TCAAACGTTA     5580
GAAATCGCTG GTGTCCAAAC CGAGTTTTGT GTGGACACTA CAATTCGGAT GGCTCATGGT     5640
TTAGGCTATA CGTGTTTGAT GACACCGAAA ACCACTTCCA CCTTGGATAA TGGACATTTA     5700
ACCGCGGCAC AAATTATCCA ACATCATGAA GCTATTTGGG CTGGCCGCTT TTTGACTTTT     5760
TTGAGCCTGT GACAAAAATC ACTTTGGATT TTTGTTCCGG GCTCAAAACA TGATAAACGG     5820
CGGGAACAGA ACCAACTCTA CGCGTTGCTC CGATCTCAAC AACTTCTAAG AGCTAAAGCT     5880
CTAAGAGTTG AAGGTCcTTC GGAAATAAGC CGAAATTGTT GAGAATCACG AAGAACGTAG     5940
```

TGATTCGATG ATGAACAATA CCTACTTGGT CTCCAAAAAT TAAAGAACAC CGGGTAGGTA 6000

TCATGCAACA TCAACTCGTA CCTCGTTGTG TTTTATGACA ATTTTCGGAA ATTTCTTCTT 6060

ATTGCTGAGA AACACTATTT GCAAAGCAAA TAGATGTTAA TCAGTACCTA CTTGGTTCTC 6120

GGAGCTAAAC GGTTCTGTCC CGACCTCTTT TCCAAAAAAT TGTTTGTTCA CTATGTTTTC 6180

TTGTCGTATT TCATGTTAGT ATAAGGAAAA AGTGTTCAGT TGATTAGTTG AAGATTTAGA 6240

GTGTGGGACA TAAGTCGATT AGGCTTGTCT CTCATGTTCT GTCTTTTTTT ATGTCATCAT 6300

TCACTAAACT AATAAATAAT TGGAGGTATT TTATGAACAA ATTTTTTACA GCCCAACGTT 6360

TAAAAGACAC CGCCTATGTA ACAGTCGGTG CATTCATTTT GGCCATTTCA ATCAATGCGG 6420

TCTTATTGCC AAATAAACTT GTCGCAGGCG GAgCAATGGC ATTAGTATTG TTATTAATTA 6480

TGTTTTCGGG ATTAGTCCCG CCATTGTTCT TTACGCCATC AATATTCCTC TTTTAGTGCT 6540

TTGTTTCTTA CTTCTTGGGA AAGAGGTCGG TGTCAAAACC ATTTATGGCA GTCTTATTTA 6600

TCCATTTTTT GTAGGAATTA CTTCAGGGAT GCCTGTTTTA ACACATAATA TTTTTCTAGC 6660

AGCATTATTT GGCGGTATTA TTACAGGTGC TGGCCTAGGC TTGGTTTTCC GTGGCAATGC 6720

CTCAACTGGT GGGACAGCAA TCATTTCACA AATTGTCAAT AAATACTTCA AAGTTTCCTT 6780

AGGTATTGCT ATTTTATTCG TCGATGGTTT AGTCATTTTA TCAGCAATGT ACGCTTTTAA 6840

TGCAGATATT GTGTTATTCT CTTTAATTTG TTTATTTACA ATTGGTCGTG TCGTTGACAT 6900

GATTCAAGTT GGTCTGGTTC GCTCGAAGAA TGTGATGATT ATTTCACCAA AATATGTCGC 6960

AATTCAAGAA CGCTTATTAC GTGAACTAGA TAAAGGTGTG ACCTTGGTGC CAATCGAAGG 7020

CGGTTATCGA AACGCGAAAG GCATGCTCTT AATGACCGTT ATTCGTGAAA AAGATTTTCC 7080

CCGCTTAAAA GAAGCCATTC TTGAGATTGA TGAAGAAGCC TTTTTGATTT CAATGAGTGC 7140

GAGTGAAGTC TACGGAAAAG GATTTAGCTT GAAAAAAGTA GCTGATTCTT ATGGCGTTGA 7200

AGCAACTAAT GCCAATAATT TACAATAGTT TTTTGTAGGA AAAGrGCCGC AAAATCTGTA 7260

GAGATTTTGC GGCTCTTTTT ACATCTTCAT CCAGACAGGT TCTAAGCTAT TGCCATCTAG 7320

GTCTTGCACT TCTAAGCCAT ACATTTGCTC TTCTGGTATT CCCATATCAA CTCGGTAGCA 7380

GTTGCCTCCG TTAGCTTTGG CGGTTTCGCC AAACTGTTTT ACGGCTTCGG CACTCTCTAA 7440

ACTAAAGGAA ATTAATGCCC CACTGGTTGT TTGCGCATCA ATAATTTGCT TTtCTTGGAT 7500

AAATTTTCCA TAAAAGTCAT GGTTCAACAA CATAATCCAA AATTGATCAT CCCACATCAT 7560

CGAACTAGCT TCCTCTGTTG AAAATTCCTC GTTCTTTTTG AATCCAAGTT TTTCATAAAA 7620

TGCTGTAGAA CGTTGGACAT CCGCTACTGG GAAGTTAACA AAAACCATTG TGCCCATGTG 7680

ATTCATCCCT TTCTTTTAAT CATCTACGAC AGCTAACTGT GCGTTGCGAC AAAATCAACA 7740

AGCTTGTTTG ATTTATACTA GCATATTCAT AAAAGAAATG TTACTAAAAC GTTTTCTTAA 7800

GGTAATTCTC GATAAGCCAA GTGCTGAATG ATTTCACTAT GTTGTGGATA ATCGTTTAAC 7860

AATTCCGCTT GAGTAAATAA CTCGAAGTCT TCAAACTCAA AGCCTGGTGT AACCACACAA 7920

CTGACTAACG CAAATCCTGT CGCTACTGTA GAGCCAAAGA TTGTGCCGGC AGGAACTGAG 7980

AACGAAAGCT GTTGCCCTTG AACTAAGTCT TTCCCTAACT GAACAGCTTG ATACGTTCCA 8040

TCAGGAAAAA TACAGTGAAC GGTTAACGGC TGACCATCGT GAAAAAACCA TAATTCATCC 8100

GCTTTTAAAC GATGAAAATG TGACGGACTT TCTGTATTTA ATAGAAATAA AATCGCTGTA 8160

TGCAACGCCC GTTCGCCTCG TTTGGTCGGA ATCAATTGTT CACTTTTATT GGTTTGTTTA 8220

AAATAACCGC CCTCAGGATG CGCTTCTAAT CCTAACTTCT CAATCCACAT TTGTTCTGTC 8280

ATCTTCTCCC TCACTCCTTC ATTTTTTCAG AAACGATTTA GGAAGAACTT AGAACATCCG 8340

TCCCAATCCT AAAATAAACG GCGGGACAGG AGTAGCTTCT TCATTCTTGA AGCTAAATAC 8400

TCCTGTCCCA CTTTCTTCCC AAACCTGTTT ATTTTCTACA ACTAAACATC TTTTAATAAG 8460

CCTTTAAATA AGCCCATTGC GAAATCGTTG GCATTGAATG GTTCTAAATC GTTAATGCCT 8520

TCACCTAGAC CAACTAATTT CACTGGCAAG TGTAATTCAT TCCGAATAGC AATGACAATA 8580

CCACCTTTGG CTGTTCCGTC TAGTTTGGTT AAAACTAAGC CAGTGACATC GGTTGTTTCT 8640

TTGAACTGTT TGGCTTGTGT CATCGCATTT TGCCCAGTAG TTGCGTCCAC GACTAAGAGA 8700

ACTTCGTGAG GCGCATCTGG AATTTCTCGT TGAATGACTC GTTTAATTTT TTCTAATTCC 8760

TTCATTAAGT TGACTTTATT TTGTAAACGA CCTGCTGTAT CCACTAGCAA AACATCGGCC 8820

TGTTCTGCTT TGGCACGTTC CACTGCATCG AAAACGACCG CTGCTGGATC GCCGCCAGCA 8880

TTCCCACGTA CAACTTCAAC GCCAGCCCGT TCGCCCCAAA CAACTAATTG ATCGATGGCG 8940

CCCGCTCGGA ACGTATCGGC AGCTGCCATT AAAACTTTCT TTCCTTCCAA CTTGTATTGA 9000

TGGGCTAATT TACCAATGCT GGTTGTTTTA CCGACTCCAT TAACACCGAC AAATAAAATC 9060

ACGGTTAAGC CATTTGGCTG TAAGTTAATC GCATTATTTT CATTAATTCC TGCTTCTTCA 9120

TACAAATCAA CCATTTTTTC AATGATGGTA TTTTGAACTT GAGCTGGTTT TTTCACATTT 9180

CGCAGTTTCA CTTCTTGTCG CAAGGCTTCT GTAATTTTTA AGGATGTATC AAAGCCTACA 9240

TCTGCACCAA TTAACGTTTC TTCTAATTCT TCAAAGAAAT CTTCATCGAC GCTTCGGAAA 9300

TTAGCAAACA ATTCGTTCAA CCGTTGACCA AATGTTTTAC GCGTTTTTTC CAAGCCTTTT 9360

TCATATTTTT CTTGGACCTG CTCTTCTTTT GACTCTTCAA TTTCTTCTGA AACCTGCGTG 9420

TCTGTTGGTT CTGCCAAAGG AGTCACGACG CTACTTTCTT CCTCAGGCGC TTCCACAACA 9480

ACTTCTTGTG TTTCCTCTGT CAGGATTTCT TCTGGCTCTT CTTCTAACGC TTCTTCTACA 9540

TCAACTAATT TTTCCATTTC AGCGTTATCT GTTAAATCGG CTGTTTCTGT GACTTCTTCC 9600

GCTGTCGTTT CTGGCGTTAT TTGCGCCTCT TCTGTTGCGG TTTCTTTTGA TTCAGCAGGC 9660

GCTGATTCAT CACTTGGTGT TTCGTCTACT GTTTCTTCAA CAATCTCTTG TTTTTCTTCT 9720

TCCTTTTTTT CAGCGGAAAA AGCTTTTTTA ATTTTATCAA AAAATCCCAT TCTTATTGGC 9780

CTCCTTCCAA CGGCGTTAAA ATATATTTTT CGACCGCTTT GGCAACGCCA TCTTGGTCAT 9840

TCGTGTCTGT GACGACATCT GCGGCTTCTT TAACTAATGG AATCGCATTT GCCATGGCGA 9900

```
CACCTAAGCC AGCGTATTCG ATCATCGGTA AATCGTTTTC TTCGTCCCCT AGGGTCATGA    9960
TTTCTTCTGC ACGGATTCCT AAATCTTTGG CTAACAAGGA AATACCATAG GCTTTCGTAA   10020
TCCCTTTTGG CATAAATTCT AATAAGTTAT TTCGCGTTTT AATAATTTCA AAGCGCTCAT   10080
AAAAAGAGGC TGGAATTTCC TTGATGCGTT CATTTAAATA CGCTTCATCA ATCGCAACAA   10140
CAACTTTATT ATAAATTCGA TTCGGGGTTA ACTCTTCTAA TTTGTAGGAC TCAAAGGTTA   10200
ATAATTTGTT TAGAGAACTA TAAATTGAAG GATAATTAGG TGCAGAAGGC AACTGCATCA   10260
CGACCTCATC GGATAAGACA TCAAACGGCA TGTTGAGTGT GGTTGCTAAT TGGTACAGCT   10320
CATGAATCGC TTCCAAGGGC ATCAATGTTT TCTCAATAAT GGCCCCTGTA TCATTTTTTT   10380
GAACCAACCC GCCGTTAAAA GTAATGCTAT AATCGCCTTC ATCTTGTAAA CCTAATTCTT   10440
CTAAGTAAGG ACCAATGGCT GCCAAAGGTC GACCTGTGCA GATAACAACT TTCACGCCTG   10500
CTTGTTTCGC AGCCATCAGT GCTTGTTTAT TCCGTAAAGA AATCTCTTTT TGGCTATTTA   10560
ATAATGTTCC ATCTAAATCA ATAGCAACTA ATTTTATCAA ACGTTCGCTC CTCCTTGCTT   10620
CTGGCTAATT GCTTTTTTCA ATTGCGCCAC CTTCTTTGAC CTCTTCCAAG CGAACTGAAA   10680
CGATTTTTGA AACCCCAGAT TCTTGCATAG TCACGCCATA TAAAACATCT GCGGCTTCCA   10740
TTGTCCCTTT ACGATGCGTC ACCACAATAA ACTGTGTACC ATCTTCAAAT TCGCTTAAAT   10800
AATGGCCAAA ACGAGCAACG TTGGCTTCAT CTAAAGCTGC TTCTACTTCA TCTAAAATAC   10860
AGAACGGAAC TGGTCGAACA CGAATAATTG AAAAGAGTAA CGCAATCGCC GTTAATGCCC   10920
GTTCGCCTCC AGACAACAAG CTTAAATGTT GTAATTTTTT CCCTGGTGGT TGCGCCTCAA   10980
TCTCAATCCC TGTATTCAAG AGATCCTCTG GGTTGGTTAG CACCAGTTCT GCCCGTCCGC   11040
CACCAAACAT ATTAGGAAAT ACCACTTTAA ACTGGCCACG AATTGCTTCA AAGACTTCTT   11100
TAAACCGTTC TTTCACTTCT TGGTCCATTT CATCCATGGT TTCAAATAAT TGTTCTTTCG   11160
CATTTAATAA ATCATCTCGT TGACTGACTA AAAATTGATG GCGCTCATCG ACTTGTTCAA   11220
ATTGCTCGAT GGCACTTAAG TTAACAGGTC CTAACCGCTC AATTTCTTGT TTTAGGCGTT   11280
TCACTGTTTG CTGCGCTTGC GCTAaATCAT CGATTGGGAA ATACGCTTCA TAGGCAGCTT   11340
CAAAGGTTAA GCTGTATTCT TCCTGTAAGT AACTTAACGA GCTATCTAAT TGCATTTCTG   11400
CACGATTTTT TAGGACTTCG ATTTGTGTTT TATCTGCTAA ATACTGTTGT TGGCGTTTGT   11460
TTTCTTCCGC TAATTTTGTA TCTAACTCGT CTACTTCTTC TTGCAACGCT TGTCGTTGAC   11520
TACGTGCCGT TTGTAAAGAG GTTTGCAACG CTGTTTGTTT CTCCGCTAAT TGAGCGACTT   11580
GTGCCGCTAA GCCTTCTTCG GTTAACTGGT GATCGCTAGA ATGGCTACTT AATTGTTGTA   11640
GCTGTTGACG GATGGCCGTT TCTCGAATCA ACAATTCGTC CAATTGTTCT TGTTTGTCTT   11700
GCTTTTGACG AGCAAAATGG GCACATTGTT CCGCCGCTAC TGCTTGTTCT GCTTGAACTG   11760
TCGTCAAGCG TTCTTGCGCT TGGGCTTTAA AGGTTTCCAT TTGGCTAGCT TCTTGTTCCA   11820
CTTGTTTCAT TTCAGCATCT AAACGTTCTT TCGTTGCCGT TAAGTTTGCT TGTTGCTCTG   11880
```

```
TCAATGTGGC TTTTTTCGTT TGATATTCGG TTAAGAATTG ATGCAATTCC CGTGATTCAT  11940
ATTCAAACAA GCGTTTTTCT TTTGTTAAAC GAGTAATCGT TTCTGTTTGA TTGGCTAATT  12000
TATTGTCAAT TTCTTGTTGT TTTAAGCGAT TTTGTTCACC AGCAGAGCGC AACATTTCTG  12060
CACGTTCCGT GGCTGTTTTC ACTTCTTGCG AGAGCGCTTG GACTTCTTGT TCTACACTTC  12120
TCAGTTGTGT TTCTAATTGA GTCATTTGTT CAGTAATTGT TTGAAGTTCT TGCGCTTGTG  12180
AAAATAGGCT TCCTTGGTTG CCACGTTTAT TGGCTCCCCC AGTCATCGAA CCGCCAGGAT  12240
TCATCACATC GCCTTCCAAT GAGACAACGC GATATTGGTA ATTGACTAAT TTTGCTAACT  12300
GGTTGGCACT CGTTAAATCT GCCGCTAAAA TCGTGACGCC TAAGAGATTT TGAATAACGG  12360
TTTGTACTTG TTCTGGATAA CGAACTAGTT CACTGGCAAT CCCCACGAAG CCCGGTGCAC  12420
CAGCCAAACG ATTCTGAACC ATCGCCGATA CAGAGCGTGG CTTAATGGTA GTCAATGGTA  12480
AAAAGGTTGC TCGGCCACTG TGTTGTTGTT TTAAGAAGGT GATGCCTGCT CGACCATCTT  12540
TCTCATTTTC CACAACAATA TGTTGCGCCG CACCACCTAG GGCCGTTTCA ATCGCCAACG  12600
TATATTCTTT AGGCACTTCA ATTAACTCAG CCACTGCGCC AACAATCCCA GTTAATTGGT  12660
TTTTGTGACG CAATACAGCT TTCACGCCTT GATAGAAGCC AGCGTAATTT TCTTGAATTT  12720
CTTGCAAACT TTTTTGGCGT GCCTtCGCTT GTTGCACTTG ATTCATCgCT TGGTACATAT  12780
CGTTTTGACG CTGTGCTAAt CGTTCACGCT TAGCTTCAAG CGTTGCTTTC AACGCAGTGT  12840
ATTCTTCCaA CTGTTCTTGT AAGCCTTGTT TTGCAACTTT TTGCTCTTTT TCTAGCGTCT  12900
CTTTCATTGC GAGAGCTTCA ACCATCTGTT CTTCTAGCGC TTCGTGTTTT GCTAACGATT  12960
GTTGATTTTT AGCTGTTTCT TGTTGATATT GTCGTTCTAA ATATTTTAAG TCATTGGCTG  13020
TATTGGCTTG TTCTTGCATG ACTTCCACAT ACTGACTGCG CAATTCTTCC ATTAATTCTT  13080
TAGAAGACTT ACTATATTTT TCCACATCTT TCGTTGCCAA AGCCAATGCT TCTTTTAACG  13140
TCTGACGTTG TGCCGTTTTT TCTGCGATTG CTGTTTCCAG CGTCTGCAAT TCTTCTCGAT  13200
AACGAACAAT TTTTTCGGCC GTTTCTGCCA ATGTCTCTTC GTATTCGCTT GCCGTTTGGG  13260
AGGTGTGTTT GGAACGTTCA ATTAAAACAT TTTTTTGTCC TTCTGCTTGT TTCAACGCTT  13320
CTGTCACTTG TAATAATTGT TGTTGTTCCG TTTCAATTTG TTCATCTAAA CGATTGCGCT  13380
TGCTGCGTAA ACGTACCAAT TTGCCTTCTA AGTCATGGAC TTGCTTACTT GCGCCAGCTA  13440
ACTTTTCTTC AATGGCTGTT AATTCTTGCG TTTTCGTTTC CCAAATAGCT TTTGCTTCTT  13500
GAATTTCGGT CACAGTTAAA TTGACATCAA TTTCGGTAAG TTCTTCTTTC AAAGCTAAAT  13560
ACTTTTTAGC CGCATCTGCT TGGGCGGCTA GAGGAACTAA TTGGTCTTCC AACTCGTAAA  13620
TGATATCTTG GACACGGCTT AAGTTGTCCT CGGTTTCAAA AAGCTTTTGT TCAGCCTTTT  13680
TCTTTCGTTG TTTGTATTTT AAAACGCCCG CTGCTTCTTC AAAAATCCCT CGACGGTCTT  13740
CAGGTTTACT ATTAAAAATC GCTTCAACTT TCCCTTGTGA AATAATTGAA AAAGATTCTT  13800
TGCCTAACCC TGAATCCATA AATAAATCTT GAATGTCTTT TAAACGGCAC GCTTGTTTAT  13860
```

TAATAAAAAA GTCACTTTCG CCCGTTCGTT TTAAACGACG GGTGACACTA ATCTCACTAT 13920

AATCCAATGC TAAATAATGG TCGCTGTTGT CTAACGTCAC GGTAACTTCC GCAATATTTA 13980

GCGGCTTACG ACCTTCCGAA CCGGCAAAGA TAATGTCGTT CATTTTGCCA CCGCGAAGGT 14040

TTTTCGCAGA TTGTTCACCT AATACCCAGC GAACGGCTTC CGTGATATTA CTTTTCCCAC 14100

TTCCATTCGG ACCAACCACT GCTGTCACAT CATCTTCAAA TTCAATAATG GTTTTATCTG 14160

CAAATGATTT AAATCCTGTA ATTTCAATTC GTTTTAAATA CACGAATATT GCTCCTTCCA 14220

GCCTTACTGA GGAATACTTT TCAGTGCCCG CTCAGCGGCA TCCTGTTCAG CTAACTTCTT 14280

CGATTTTCCT TGGCCTAACC CAATGAGTTC ACCATTCATG TAAACTTCAG TGAAAAAGGT 14340

GCGGTCATGA GCAGGGCCTT CTTCTTTAAT TAAGCGATAT TCAATTGAAA CATCGCCTTT 14400

GCGTTGTAAA ACTTCTTGTA ATTGTGTTTT GTGATCCATC TCATGTGAAA AAGCACCGGC 14460

ATCAATTTTC GGAAAAATAA CGTCTTCAAT AAATTTcTTG cTGCGCCAAC TTTTTGGTCT 14520

AAGTAGAGGG CACCTAAAAA GGCTTCAAAT AAGTCACATA ATAATGATGC ACGTGTTCGT 14580

CCGCCCGATG CTTCTTCTCC TTTACCTAAT AAAATGTAGT TGTCGAAGTG ACATTCTTTC 14640

GCAAATTTGG CTAAACTATC TTCCCGAACG ATGgTGCGCG CATCTTCGTT AATTTTCCTT 14700

CTGGAAGTTC TGGGAATTTT AAATACAAAT ATTGTGAAAC AATTAATTCT AAAACTGCAT 14760

CTCCTAAAAA TTCAAGACGT TCATTATCGG ATAATTTTAA ATAGCGATGC TCATTCACAT 14820

AGGATGAATG AGTAAAAGCT TGCTCTAATA GATTGACATC ATGGAAAACA ATGCCGTAAC 14880

GTTCTTTTAA CTCTGTTGTT AACTGATTGT CCATTCTCCA CATTCCTTTA CATTTATTTA 14940

CACCTTGTTT CATTATACTT TTTTTTTCAC TTTTTTTAAA TGGTTCTATT CAGAAAAAGC 15000

CTGTTATTCA CTGTGTTCTC ATTATTTTTT AGAAAAAATA AATAAATCAC CTCTCCAGCT 15060

ATCAAATCAG CCACTGCTAC CCCATAAAAA AAGAGAAAAC GTAGCAACGT TTTCTCGGAA 15120

TTTTTTGAAT AAACGTACCA GCAAGAGACA ATAAAAAAAG CATTCAGTTA TTCACTTTCA 15180

CTAAATATTA ACATCTCATT TAAAGATTGG GATAATTCTT CTTTTTGTCG TCTTTTTAAT 15240

TTTGTTTTAG TCGCATCATG ATTCTTTACT GTGCCTATAT GAACCTGCAC CTCTTTCTTA 15300

TTATTTATAG CAACAATAAA AGGTACAGAA AGCGACTCAT CCCCAGTATA TTTTTGATAA 15360

AAT 15363

(2) INFORMATION FOR SEQ ID NO: 280:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3214 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 280:

CACTATTTTA TTTATATTAA TGCAACGTtT CTATGTAGAA GGCGTGAcTT CTGGCTCTGT 60

CAAAGGCTAG AAAAActCGT TTGGTTTATA GCCAAACGAT GTTTTTTATT TTTTGGCGTT 120

GTTTTCCTTA CACGTTTTTC TTCGTAGCGC TTGATAATTG GACAAAATTT TCATGACCGC 180

CAAAAAGTGT TGATTGACCT CTTGATAATC CTCTAAGGAA AGAACACTAA GCGTTTCAGA 240

TAGTTGCTGT TTTACTTGAT TTTCTTTTTG CTTAAAGAGC TTCTCGCCTT CTTTTGTCGT 300

ATAAACATTA ACCACGCGGC GATTTTCTTG ATTATGGCGA CGTTGAACAA TCGCTCGTTC 360

TTCTAACCCC GCAATAGAGC GACTCACTTG aACTTTCGAC GTTCCTACTT GGTTGGCTAA 420

CTCCGTCATA CTAATTCCTT CATTTGTTGC CACAATTAAT AAAATTAACA TGTCTAACTT 480

GGTTAATTTT AGATCCTCCA AGTCTAGGTC TTTAAAAAGC AATTGGCTCA TATATGAAAA 540

AATTTCAGTA AAGACACTAA TATCTAAATT TGATGAATGC TTCTCTTTAT CCATATGTTC 600

TCCCTCGTTT CCCACCTAGT GTATACTAAA TTCACTTGAT TATCAAATGG TAATGATTGA 660

TTTTTGTTAA CTATTATGTT ACGTTATACA CTAACTTAAG GTTGTATTGT TGATAAATTG 720

GAAGAAGGAT GCGAAAAATA TGCTAGAACT CAAAGAGATr AAAAAATATT ATACTGTCGG 780

CGGAACGACC ACTAAAGCTT TAGATGGCGT CTCTGTCGCT TTTCGAAAAC AAGAATTTGT 840

AGCCATTCTG GGTCCCAGCG GCTCTGGTAA AACGACGATG TTGAATGTGA TTGGCGGATT 900

GGACAATTAC GATTCTGGTG ACTTAATCAT TAACGGCAAA TCAACAAAAA ATTTTAAAGA 960

AGCCGACTGG GATGCTTATC GGAATAACTC AATTGGTTTT ATTTTTCAAA GTTACAACTT 1020

AATTAGTCAT CAAGGGATTT TGGATAATGT CGAGCTTGGT ATGACACTGA GTGGTATTTC 1080

AAAAGCAGAA AGACGCCAAA AAGCCGAAGA TGCACTCATC CGTGTCGGTT TGAAGGACCA 1140

TATCCATAAA AAACCTTCCC AATTATCTGG CGGTCAAATG CAACGGGTTG CCaTTGCTCG 1200

CGCTTTAcTn AATGaTCCCG ATATTTTACT TTGTGaTGAA CCaACTGGTG CATTGGaTTC 1260

AGAAACCAGT CTGCaAATTA TGGCACTTAT CCAAGAACTT TCGAAAGAAA AATTAGTTAT 1320

TATGGTGACC CACAATCCTG AATTAGCTCA CGAATACGCT GATCGAATTA TTGAATTTTC 1380

AGATGGCCAA ATTCGTcATG ACTCGCGTCC TCATATTGAA CATGAAAAAA AAGAACCTTT 1440

CGAATTGAAA CGCACCAAAA TGAAATTCAC GACTGCGCTG CGTTTATCTT TTAATAATAT 1500

CCGCACTAAA AAAGGGCGAA CTTTTTTAAC CGCCTTTGCT TCAAGTATCG GAATTATCAG 1560

TATTGCGATT GTCCTTTCTC TTTCAACAGG ATTCCAGAAA CAAATTGATA AAACTCAGTC 1620

GGAAACGTTA GCGCAATTTC CGATTACCAT TTCTAAAATA GCGACCGATC AAGATCCAGA 1680

AAGTTTCAAT AATCAGGACA AAAAAGGCAC CTTCCCGAAA GAAAAAGAAG TGACTGCTAA 1740

AATTAGCGAT GCGGACCGTG CGCAACACAC GAATTTAATT GATCAAAAGT TTATTGACTA 1800

TGTCAATCAC ATTGATCCTG AATTAAGTAA TAATATCGGC TATACTCGTC TAGTCAATAT 1860

GrACTTATTA AGAGAAATCA ACGGGAAAGC ACAAGCTGTC AAATTTTCAA aCkmyrCACC 1920

CGACGGcCAA TCAAACGCGA TGGCTTCTAT GATGGCTGCC CAAACTGGTG TCGGCGTTTC 1980

TGCTTTTCCT AAACAACTAG AAAATGGAAA AAACAATTTC TTAAAAGACA ATTACTCGTT 2040

GCTTGAAGGA AATTATCCAG AAAAAGAAAC CGATGTCGTT TTAATCGTCG ATAGTAACAA 2100

TACCACCAAT ATTAATGCTT TAAAAAATCT TGGCTTTGAC GTGAAAGATA ATCAAAAGAT 2160

TGCCTTTTCA GACATTGTTG GCACAAAAAT GAAATTAGCC AATAATAATG CATTTTACAC 2220

CAAATTACCG ACAGGCAACT TCATTCCTAA CCAAGATTTG CAAGCTGTTT ACGATAACCC 2280

AGAAAATACT GAATTAACAA TTTCAGGGAT TTTGCGAATT AAAGATTCTT CAACAATGAA 2340

TCTATTAGCA CCTGGAATTG CTTATAGCGA TGCACTTTCA ACTAATATGA TTGCCAAAAA 2400

CAAAACCTCT GACATTGTTA AAGCGCAAGA AGCCAGCAAC ACTAACGTTA TGACTAACGA 2460

AACATTAGAT GCTTCCGCGA AAGAAAACCT GTTGTCTTAT TTAGGAGCGA ACGAAATCCC 2520

TTCAAGTATT ATGATTTATC CAAACGATTT CAAATCAAAA GATAAAATTC TTGATTATTT 2580

AGATGCTTAT AACAAAGGAA AAGATAAAAA AGACCAAATC ATCTATTCTG ATTTAGCTGG 2640

AACAATGACC CGCTTAACAG GTGGTTTAAT GGATGCCATC ACCTATGTGC TGATTGcATT 2700

TGCTGGGATT TCCCtTGTTA CAAGTATGAT TATGATTGGG ATTATCACCT ATACTTCTGT 2760

CATTGAGCGG ACCAAAGAAA TCGGTATCCT GAAAGCGTTA GGAGCTCGTA AAAAAGATAT 2820

TACCCGGGTG TTCGATGCCG AAACTGCTAT TTTAGGCGTT GCCTCTGGTA TTCTTGGTGT 2880

TGTGATTGCC TTTCTCGCAA CATTTCCGAT CAATGCAGTG CTCTATAACC TGACTGATTT 2940

AGAAAATGTT GCCACCTTAA ATCCAATTCA CGGGATTATT TTGATTGTAA TTAGCACAAT 3000

CTTAACCATG ATTGGTGGCC ATATCCCTGC CCGAATGGCA GCGAAAAAAG ATGCGGCAGT 3060

TGCTTTACGT GCAGAATAAA TAACAAAAAA TAATCCTTTA ATTTCTCTTA AAGGATTATT 3120

TTTTGTTTGT AATTTTTATA AAACCACCGT ATACTAAATG ATTGTGAATG CTTAATAAAA 3180

AGAAACGAGG TAATTTCATG TTAAAAAAAC CATT 3214

(2) INFORMATION FOR SEQ ID NO: 281:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5830 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

CGCGGCTTGA TTTAGAGGAA GTCCAACAAG AAATTAACgA TGCGAACGAA tTGCAAATGA 60

AGCGAAACAA TCGTGGCAAG CAAAAAAAGA ACAACACAAA GAAGCAACAT CGAAAGAAGA 120

ATAAGGAGCA AAGACTGTTT TCTGAATAAA GGAAAACAGT CTTTTAAAAT ACTAAGAAAT 180

CACGGTTTTT TATGGTAGTT TGCTAATTTT TTGTTATTTT ATCGTAAAAT ACGGTAAACT 240

AAGAAGGAAA CTTTGAGAAA ATAGATAAAT TATTATTTGC TAAAGGAGCA AACTTCTTCT 300

AAAATAAAAA GGaAGTTTAA AAGAAAGTAT GGTAGCAGAT GGAAATACAA TTTTTAGGAA 360

CAGGCGCTGG TGTACCagCA AAACATCGCA ATGTCACAGG aATTGCGTTG aAATTATTAG 420

aTGAACGAAA TGCAGTTTGG CTCTTCGATT GTGGGGAAGG GACTCAATTG CAAATTTTAA 480

AAAGTAGTAT TCGCCCAAGA AAAATTGAAA AGATTTTTAT TACACATTTA CATGGGGATC 540
ATATTTTTGG TTTACCTGGT TTGTTAAGTA GTCGTTCTTT TCAGGGCGGG ACAGAACCTT 600
TAGAAATCTA TGGACCAGTT GGGATTGCTG ACTTTGTCAA GACTTCTTTA CGGGTCTCGC 660
AATCGCGGTT GTCCTATCCA CTGAAATTTA TTGAATaACG AAAGAAAATG ACgTTATTTT 720
TAAAGATAwA CAATTCACTG TTCGCTGTAA TATCTTGGAC CATGGCATCA CGAGCTTTGG 780
CTATCGAATT GAGGAAGCAG CGCATGAAGG AgAwTTmCAA GTAGAGAAAC TTCAAGCTTT 840
AGGGATTCCT TCTGGGCCGT TGTATGGCAA ACTCAAACGT GGTGAAACGA TTGTTTTTGA 900
TGGACAGGAA ATTAATGGAC AAGCGTTTGT GGGAGAGCGT AAACCTGGAC GAATCGTGAC 960
GATTTTAGGA GACACTCGCA AAACCAAAAA TAGTGTAACT TTAGCACGAC GAGCGGATGT 1020
TTTAGTCCAT GAAAGTACGT TTAATAAACA CGAAGCAAAA ATGGCAAAAG CCTACTTCCA 1080
CTCAACCAGT CAACAAGCAG CTGAAGTTGC AAAAGAAGCG CAAGTAAAAC AACTTATTTT 1140
AACGCATATT AGTGCCCGTT ATTTAACCAA AGAGGCCTAT CAGCTTCAAG AAGAAGCCCA 1200
AGAAATTTTC CCCAATACAA AAATCGTAAA AGATATGGAT ATCATTGAGA TTCCATTTGC 1260
CAATGAAGGA GGAGCTTAGA TGGATTTAAC AAATAAAGTC GTTGTCGTAA CTGGAGGTTC 1320
AGCTGGTTTA GGTGAACAAA TTTGTTATGA AGCAGCTAAA CAAGGGGCGG TCGTCGTTGT 1380
CTGCGCCCGC AGAATTAATT TGATTGGTAA AGTTCGCGAG CAGTGTGCTG TTTTAAGCGG 1440
ACGAGAAGCA TTTTCTTATC AATTAGATAT TGCCGATCCA GAAAGTGTTG AACGTGTGGT 1500
TGAAGCTATT TCAGCAGAAG TTGGTCCTAT AGATGTTTTA GTCAACAATG CTGGTTTTGG 1560
CTTGTTTGAA AATTTTGTTG AAATTGACTT GGCTGTTGCA CGCCAAATGT TTGATGTCAA 1620
CGTTCTAGGA ATGATGACGT TTACTCAAAA AGTAGCTATT AAAATGATTG AAGCTGGGCA 1680
AGGACATATT ATCAACGTTG CTTCAATGGC TGGGAAGATG GCTACAGCTA AATCAACCGT 1740
TTATTCTGCA ACGAAATTTG CTGTGTTAGG TTTTTCAAAT GCTTTACGTT TAGAATTGAA 1800
ACCGTTGGGT GTAGCAGTAA CGACTGTCAA TCCAGGACCA ATCCAAACAG AATTCTTTGA 1860
TAAAGCGGAC CCGACAGGCA CATATCTAGC TGCGGTGGAT AAAATTGTTT TGGACCCTAC 1920
GAAATTAGCG AAAGAAGTGG TTGGGAGTAT GGGGACTTCT CGACGTGAGA TTAATCGTCC 1980
ATTTGTCATG GAAGCGGCAG CCCGCTTTTA TACATTGTTT CCACATTTAG GAGATTTCAT 2040
AGCGGGAAAT ATTTTAAATA AAAAATAAAG GAGTCGATCA AGTCATGAAA AATCAATGGC 2100
GTGTTATTTT GGGCCTTGTG TTGGTCTTAA TTGTTGTTAT TTTTGCTGTT TTAAATAATC 2160
AAGCTGTTCC AGTAAACTTT GGTTTTACTA AAATTAGTGG ACCGTTAATC TTAATTATTT 2220
TAGGTTCTGC AATTATTGGC GCACTAGTTG GATTACTAAC CTCCACAACC ACCATTTGGA 2280
ACCAACGTAA GGAATTAAAA GCTGTACAAA AAGAGTTGGA TATTTATAAA AATGATATGG 2340
ATAAATTAGT CAAAGAAGAA ACGGAAAAAG TTCAACGTTC TTTTGATAAT CAATTGGCTG 2400
ATTTGCAAGC CAAACAAGCG GCAGCACCAC AAGTCTCAGA ACCAGTGGTC AACGAACAAA 2460

CGTCAAACAC AGAAGTAGAC GTAACGCCTG TGTCAGGTAG TCGTATTGAT CGTTATGTGA 2520

AACCTCGTGT GAATGAAGAG GAACAAAAAT AAGGAAGAGA CTTTATCACT AAGTTTCTTT 2580

TAAAATAAAA CTGTGATATC TGCTATTTAT CTAACAGATA TCACAGTTTT TTTCTCAGAT 2640

TTAAAAAAAT GCTTTCTTTA GTTTTTGGAA AAATTTTGCT ATAATTACTT GGTTAGGAGT 2700

GACGTCGTGA AAAAATCAAA CTATCAATGG CAGTTACAGA CAAAGACAGA ATTACCTGTA 2760

GAATTTATAG AACAATTAAA AAAAGAAGAA ATTAATCCAT TAATTGGCCA ACTTTTATGG 2820

CACCGCAATA TCCGGACAGA AGAAGCATTA CGTAAGTTTC TACATCCGAC TATCGAAGAT 2880

ATTTATGATC CGTTCTTAAT GCATGATATG GAGAAAGCGG TAGCCCGCAT CCAGCAAGCC 2940

GTCGAAGCGG GTGAACAAAT CCTTGTCTAT GGGGATTATG ATGCAGATGG CATCACTAGT 3000

ACCACGGTTA TGAAAGAAGC AATCGAATTA GTCGGTGGCA TGGTTCAGTA CTTTCTACCG 3060

AATCGTTTTG TTCATGGCTA tGGGCCTAAT AAAGaCGTAT TTGCTGAACA AATTGAACAA 3120

GGAGtCCAAT TAATTGTAAC CGTTGATAAT GGTGTAGCsG GACATGAAGC AATTAACTAC 3180

GCTATGGCAC AAGGAGTAGA TGTAATCGTT ACTGACCATC ATGAATTGCC TGAACAATTA 3240

CCAGAAGCCT ATGCGATTGT CCACCCTAGA CATCCACAAG GTGACTATCC TTTTGGAGAT 3300

TTAGCTGGTG TTGGCGTGGC TTTCAAAGTC GCTACTGCGT TACTTGGTGA ATTGCCAATT 3360

GAGTTATTAG ATTTGGTAGC AATTGGGACG ATTGCTGACT TGGTTTCATT AACCGATGAA 3420

AATAGAACGT TTGTCAAAAT GGGCCTTCAA ATGATTCAAA CTGGTGATCG AATTGGTTTA 3480

GACGTTTTAT TACAAGAAGC GGGTGTGAAA AAGGAAGCTG TTTCAGAAGA ATCAATTGGC 3540

TTTACCATTG GACCTCGTTT AAACGCATTA GGTCGATTAG GGGAAGCCGC ACCAGGTGTT 3600

GAATTGATGA CTACATTTGA TGAAGAACAA GCGCTGGAAA TCGCTAAATA TATTGATCAA 3660

CAAAACAATG AACGAAAAGA CATTGTAACG ACGATTGCCA AAGAAGCCCT GGATTTAATC 3720

GACCCGAATG CGCCGGTCCA TATTTTAGCG AAACAAGGCT GGCATGAAGG CGTCTTAGGG 3780

ATTGTCGCGG GCCGCATTAT GCAAGAAACG GGGAAACCAA CCATTATTTT AGCAATAGAT 3840

GAATCTGGTA CGACTGCTAA AGGCTCTGGA CGCAGCATTA GCGCATTGAA TCTCTATGAG 3900

GCGTTGAACG AAgTGCGTGA GCAATTCACT CATTTTGGTG GGCATCATAT GGCAGCGGGA 3960

ATGACTTTAC CTGTTGAAAA TATTCCTTTT GTTCaAGAAC ATTTGGCACA CTTTATTGAA 4020

AAAAATCAAA TTGATATGGC TAATGGACAA GAGTTGCTCA TTAGTGAGAG CTTAGCGGTT 4080

TCGCAGGCCA CGACTACTTT TATTGATCAA TTGCGTATTT TGGCGCCTTT TGGGACCGAT 4140

AATACTGTTC CAACTTTTGT ATTTAAGGAA ATCACGCCTA CGCAAATTCG GCAAATTGGT 4200

GCAGATAATG CACATCTGAA ATTTCAAATG AATCAAGAAG GGGCACAATT GGATGCAATT 4260

GCTTTTCAAA TGGGGCCTCA AGCGGATGAA CTAGCACAAG GAACCGCCGA TGTGGCAGGC 4320

CAACTTTCCA TCAATGAATG GAATGGTCGT AAGAAACCAC AATTAATGGT GACTGATTTT 4380

GCTGTATTGG GAAGACAATT ATTTGATTTT CGTGGCAAAA ATAATCAAAC GAAACCAATT 4440

CCTTCAGAAG CAACAGCTTA TCTTTTGTTT GATGAAAAAA ATCAAAAATT CATTTCAGAT 4500

CCAACCGCTA ATATTATTGT TTGGTCAAAC CAAGAAGAAT TGGTGGAAGC TGTTTCTCAA 4560

AATCAAATCG AACAACTGGT TTTTGTCGAC TGTCCTGTGG AAGCAATCAC GGTTAAAGAA 4620

ATTGTTGAAG CGACTGAGAT ACAACGTATA TACATGATGT TTATTTCACC AGAAGAAGCT 4680

TATTTAAACG GTATGGCTTC ACGTGAACAA TTTGCAACAT TGTATAAATT TATTTTGCAG 4740

CAAAAAGAAG TAAATCTACG TTCCCAGCTA TCGAAGGTGG CAAATTACTT AAATATCCAA 4800

GAAAAATTAT TAATTTTCAT GATACAGGTG TTTTTGACT TAGGATTTGT TACAATAGAA 4860

AGCGGTGTTT TAAACAGTAT TGAGAAACCC GACAATCGAC CATTGACTGA AAGTCAAGTG 4920

TACCAACAAC GATTGAAAAA AATAAAAACA GAAGAATTTC TACTTTACAG TGATTGTCAA 4980

ACAATACAAC AATGGCTGTG GAATGAGGAG GACAAATAAA AATGGATTTA AGAGATTACA 5040

TTGCAAGCAT TCCAGATTAT CCTGAGAAAG GGATCGTGTT CCGTGACATy TCTCCATTAA 5100

TGGCTAATGG GGATGCCTAT CGTGAAGCAA CGAAACAAAT CGTGGATTAC GCGAAAGAAA 5160

AAAGAATTGA TATGGTTGTA GGACCGGAAG CTCGTGGCTT TATTGTCGGC TGTCCAGTGG 5220

CTTATGAACT AGGGGTTGGC TTTGCTCCGG TTCGTAAAAA AGGAAAACTT CCTCGTGAAA 5280

CCATCGAAGT GACTTATGAC TTAGAATATG GTTCAGATAC ATTAACGTTG CACAAAGATG 5340

CCATCACACC AGGCCAACGT GTTTTGATTT GTGATGACTT GTTAGCAACT GGTGGTACAA 5400

TTAAAGCAAC CATCGAATTA GTCGAACAAT TAGGTGGTAT TGTGGTTGGT TGTGCGTTCC 5460

TAATTGAATT AATGGATTTA CATGGCCGCG ATAAAATTGA TGGCTACGAT ATTGTTACCT 5520

TAATGGAATA CTAATCTAAT TTAATTAGAA AGAGCGAAGG ATTTAATCCT tCGCTCTTTC 5580

TTTGACATAT TTATAAAAAC GGAGCGTAGC TTGGGGGGAA CACACCGTTT TATTATTTAT 5640

TAAGGGAACT TACGATAAAC GTATCTTGCT TGCTGTTGTT TACGTGTACG TGTATAACGC 5700

TTCTCATCCC ATGACATAAA CCAAAGAATA AATAACGGCG TCACGATCAT AAAAGCAAGA 5760

CGTAATCCTA ATGTTCCnAA GAAAATACCT ATAACCAATA CAAAAAATAA AAATGCCAAC 5820

ACGACGAATT 5830

(2) INFORMATION FOR SEQ ID NO: 282:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1959 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

TATTTGACTG CTTAAACGTA TAATGAAACG CCGAGTCATT TGTTGGTATC CCTAAACCAT 60

TTTGTAAAAA TTCAATCGCT TCCTGTTTAG TATAAGATTC TGCTTCTAAT TGCGTTCCTA 120

AATGCGGATT TTTTTGATCA ACAAATTGTC TTGCTTTGTC ATTATTAGCA GACTTTTCTA 180

```
TAACGACTTC CTGTTGCGAA TTCCTACTTC CTGAACAGCC AGTTAATACA AATAAAATCA      240
CCAACAACCC GAAACACTTA CTCAAAGTTT TCTCCTCCTC TCCCTTTTCA TTAGTGATAC      300
CTGATGATTT CTTTTATATT TTAGCATGCT TGATATAACG AATGGTGCTC AAAAAAGACA      360
GAATCAACAA GTAAAAtTGT CgACTCACCT TTCTCAACTA AtATTAAAAA CCATCTGTCT      420
CAATATTTtC GAATTtCCTG kAAATTACTG ATAAATAAAA aCAGCTCCTT TTGCGtATAC      480
yTATATTATC AGCACCTAAA gGAGGAAmTT TcTTGAAGcA AAAAAAtCTA CTCACTTATC      540
AATCACTTGT AGCCTTGTTG CTaGTATTCA GTCTCTTTAG TTTTGACCCA TCGGTTTCTT      600
TTGCCACCCG TTCTGGAAAA ACGCCTGTTT CAGTTGAACT AGAAATTGGA GGGCTGCCTG      660
GTGATGAATC GGTTGACGAT GCCATTGATC CTGATTTAGA AAATCCTAAC ACTAGTTTTG      720
ACTTGCTCTT TATTCCGAAA GAATTTACTT TTGAAACCCA AGCGATTTCT GGTGATTTAA      780
CGGCAGTTCC TATACGACCT TCTGAAGGCA ATACTCATAC AATGAGACAC TTCGGGATGG      840
GGGATGTTCG CGGAGCcTTA ACTGGTTGGC ATGTCACTGC TGAAATCCCA CACATGCGAA      900
ACGAAGCTCA TTCTTTACTA GGCATCATTA CGTTTCAATT AACAGGCAGC TATGCTCGCT      960
ATGACAAAAC GTTAAGGCGT TTCTTTATGA CAAATACAAT GTATGGACTG GATTTTTCAG     1020
AAGATCCTGC TGCACCCGAC TTCCCGACCA ACCCAATTAT TATCGGTAAC GGTGCGACTT     1080
TAATGAGTAA CGCTGGTGAT AGAAAAGGAC AAGGAATGTG GAGCGGCCGC ATGACCGATA     1140
TTTCTCTTGC AATCCAAACC CCTGTTTCAC AATTATTTCC TGGCGCATAT ACAGGTTCGA     1200
TTATCTGGAA TTTAATCTCT GGACCTGTAT AAAATAAAAA AACGCTTAGA AAACTAGCGG     1260
AATCACTACT CTTCTAAGCG TTTTTTTATA TCATAAGGAC TATTTGCTCC TTCTATTTGT     1320
CTAAGCATGC TTTACCACTT TTCGTTTAAC GAATCACTTG GCCCCAACGG TAATCAAACT     1380
GATACTCTTG ATTACCATTA TTCAGGTTAA TTTTGTATAA ATAACGCGGG ACTTTAATGA     1440
ATAACGCATC CTCTACTTGA AAAGCTTCAA TAGGTACaTT TTCTTTAGTT AAAAGCAACG     1500
GTGTTTGCCA CTCGTATGTT TTTTTCTGCC AATTGTATAA GCCAACTATC GCTkGATTAC     1560
TTkGTCTTTT CTTGTATGAC AGAACATAAT AATTTGCTTC ATCTTTCTTG CCTAATAAAA     1620
CTGGTGCTAT TAGTTGCTCT TCTAGTTTTC CTTGCACTGT TGTACGCTCT TCCGATGACA     1680
GCTCTTGACC AATTTGACTC TCATTTGAAT CTACCGTCCA TTTAGGCGCT TCTTGTGTTT     1740
GAAAGACTTC TTTGAAAGTT GGGTCTTGTT TTAACTGCCA ATTTTTCAAA TCTAATTGAT     1800
AATATCGATT GTtTAATCCA TAAATATAGA GATAGCGATT ATCCATAAAA CGATAATCAA     1860
CGAAATCAGA GCTTAAATAG TCTTTTAATT CAGGGAATTt CyTCACTAAA TCTGCTtCAG     1920
AAAACTGkTT TTTCCtGTTT GnACnAGnAA GAAATGGCC                            1959
```

(2) INFORMATION FOR SEQ ID NO: 283:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3485 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 283:

GATAACCTGA CGAGCCATAA GATTGATTAT AAAAACCCTC ATAAAGTTAC TTCTGCACAA 60

GTTGGTTCAT ATAGTAAAAC AGAAACAGAT AACCTATTCA TCAATAAATC AGATGCTGAA 120

AACGGCTTAT TGGTAAGAAA AAGTATTGAA ATAACAGACT TAAATAATGC TAAAGAACCT 180

GGCATATACT CGATTCCAGC TACAGGAGTT GAAAACAAGC CGCTACCAAA CTCTGGAAGT 240

TTGTTCGTTA GTAAAGACCC AGGAGGAGTC AGACAACTAT TTCAAACGGA AAGGACTATC 300

GTTATTCGTC AATTTGGTGG TATTCCTTCA AAGTGGACGG ATTGGAAAGA AGTAGCATTT 360

AAACCGAATG TTGTGAATTT AACTGAACCT CAACGTATAG GsGGAACTAA aGAATTTGCG 420

GATATTCCAT TAGTAAATGG TACAGAAtAG CTTTGAAAGA AGATATATTT TTTTATCAAA 480

AGACTGGGCT TGATGAAGTC CaAGCAGAcT ACAAAGCTTC TTTTAGGGAA GAAACAAATA 540

TGTTTCTTAC ACGAGAAGGA AACAGAGTTG ATGCTTATTT ACGAGTTAAT GTGGTGGATG 600

TCACCAAACT AAAAACTGCT TTTGTTCCGA TTTTTCAAAT TCCTGATGGA TTTAAAATTG 660

ATCTTAGCAT GAGGGAGAGT TTCTGGAATG TTCCATTGAC AGTAACTCAA TACACCTACC 720

CACAAGGTAA CTATGGTGCA TTATATGAGA TGAACGTGAA AGGAATCCGA TTCGGAAGTG 780

ACCGTCTTGG TAATCATTAT CTATACGGAA GTTGGCATAC TAATGATCCA AAACCAGATG 840

CTAAATTTAA GTATATGCTT TATGTGAACT TATATAATTT GTCTGAGGGA AGAGATTTTG 900

TTTCAGGTTC ATACAAAGGA GAAATTTATT ATGTAGCTGT TGAAATTGAT GGACAGTTAG 960

GrGAAAAATT AAAAGTATCT GATGGTTTTT ATAAGTATAC GGTAGGAACT AAAATAAATA 1020

AGGcTTCTCA AAATGCATGG GTGATTGGTT ATGATCAGTC AGGAACAGAA GTAACTAGAA 1080

GCAAAATCAA AATATTATAG AGAGTAAGGA GAAAACAAAA AATGAAAAAA ATTTGGAAAT 1140

ATGGACGTAC TGGAGGGCAA GAGCTACAGG TATCTGATGA TTTCCCAATG CAAGTTCCTT 1200

TTACAGATGT AGTTCCTTTA CCAACCGTCA ATTTAGAAGA CCAATTTTTT ATTCCATCTG 1260

AGAATAGATG GAAAGAAATT TCTAACCAAT TAGATAAGGA AAATTTGGAT AATTTAAGTA 1320

TATTATATAA AAACCTTGAA AAGGATAATG AATTATTAAA AGCTAAAGCA GATAATCTTG 1380

CTCTTCTAAA TTCTAAGCTA ATGCTCAATG ACCTTAATAT CCAAAAAGAA AATACCCTCT 1440

TGAAAGCTAA AGCAAATGAC CTAGCTGAGA TTGGTGCGAA ATCAATGTTG TCCATTGTAC 1500

AAATTACTGG GGAAATAGGG AAAATTAATG AGCAACTTAA AGGAGGTGCT AAATAATGTT 1560

TACTTTTGAT GATGTTAAAT TGATGTATGA TTGGGGTCTT TATACTGATG ATGAAGTAAA 1620

GCTATTTGTA CCTACATGCA TTACAGAAGA GGAGTTTAAC GAGATTGTAG GGAAAGAAGG 1680

TTAGTCAGTT GGAGTTAGAG CAAAAAGTAA AAGAACATGA AAAACGTCTT GGTGATCACG 1740

ATAGAGAAAT AGGTCGTTTA GATAGACGAA CGATGACTTT ACAAGAGCAA CTTAATGCAA 1800

ATTTAGTTAG ATTAGATGAG TCAAATAAAT TCTTACGTGA ACAAAATATG AAGCAAATGG 1860

AGCAAAATAG TGAAATTCTA AATGCTATTT TGAATAGAAA CAGTGAAGCg GATGAAAGAA 1920

AAGACGAACT AAAAAAACTT AACACTGAAA ATATATGGAA AGTAATACTA GCTATATyTG 1980

TTTCTAGCGG AGCAATAACT ATTTTATTTA ACTGGTTAAG CACATTTTTA GGAGGCGCTA 2040

AATGAAAATT AATTGGAAAC ATAAATTCAC AAGCAGAAAG TTTTGGGCTG CAGTGACrGG 2100

rGTAtCaTtG CTTTGkTaGC aGTTTTCAAT GTAGATGATT tAmCaTCTGa AAAAGTGGTC 2160

ACTTTAGTAG CAGCTATTGG CTTATTAGCT GCATATATTG TTGGCGAAGG ATTTGTTGAT 2220

TCAAATAGAG ATAATTAAGA AGTCATTCTA AAATGGCTTC TTTTTTTATA TAAAAAATTA 2280

AGAAAGAGGT TTTAAAATGA AAAAGTTTAG TAAGTTTTTA TTATCATTAG TTGTAGTTAC 2340

AGGATTATTG TTACCAACTG CCGCAGATGC TTATCAAGTG GAACAAGATC CTATCGATTT 2400

TGGCGGATAT TTTCCAGGTT ATGCGACTAA CGAATTAATT GTCTTGCACG AGTCAGGAAA 2460

TGGGAACAAC GTTGGCCTAA ACAGTCTAGA CAATGAAACG GCATATATGA AACGCAACTG 2520

GACGAGCGCT TATGTTTCAT ATTTTGTCGG TTCTGGTGGT CGCGTGAAGC AGTTAGCGCC 2580

AGTTGGCCAG ATTCAATGGG GAGCGGGAGC GACAGCCAAT GCAAAAGCAT ATGCACAGAT 2640

CGAACTTGCT CGAACGAATA ATAAAGAAAC ATTCAAGAAA GACTATGCTG CCTATGTCaA 2700

TTTGATTCGT GATTTAGCAA CACAAATTGG TGCaACATTT GACTTGGATG ATGGAACAGG 2760

ATATGGAATT GTGACGCATG ATTGGATCAC AAAAAATTGG TGGGGGGaTC aTACaGATCC 2820

ATACGGATAT TTAGCACAGT GGGGaATTaG CaAAGCmCaA CTAGCACAAG ATTTACAAAC 2880

TGGACTTCCA GAAGATGGTA GCGAAGTTAT TGTAAATCCT GGCAAGCCTA ATAAACCAAA 2940

ATATAAAGTC GGTCAGCACG TTCGCTTCAC AACAATCTAC AAAAATCCAG ATGCGCCAAT 3000

TTCTCAGcaT ATCAATGCAA ACACATTGTG GACTCAAGTT GGaACCATTA CACAAAAAAT 3060

AAATGGcCGT AAAAATCTAT ATCGCATCGA AAACAGCGGC AAACTTTTAG GTTATGCAAA 3120

CGATGGTGAT ATTGCGGAGT TTGGGAAAAC AGCAAACCAA CACCAGCTAA AACATTCACT 3180

ATTGGTGTAA ATGAAGGAAT TGTGTTGCGT ACTGGATCAC CTAGTTTGTA TGCGCCATTT 3240

ACGGCGTGTG GCCAAAAGGT GCACAATTTA GATATGATTC GGTTCGTGTG GCCGATGGCT 3300

ATGTTTGGTT AGGGGGTTCT GATTCAAACG GAACTCGGAT TTGTATCCCA GTTGGCCCAA 3360

ATGATGGGCA ACCCAGACAA TACTGGGGTA CTGGATATTA AGAACAGCTT GAnTTTTCTG 3420

TTAACCCnTA nGTTTAGGAA TAAACTTACA CTTATTTAAA TTTCTCCTGA GTCGCCTTnC 3480

CCCAA 3485

(2) INFORMATION FOR SEQ ID NO: 284:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 914 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

AGAAGGAAAA AAGAACTGAA AAAGCTTAAA AAAAGCTGGG AAAAAAGTGG AGGAAGAGGT 60

AAACCACCAA CTTCTCTATC TACGTTACGA TGTGGTCGAA TACTCCAAGA ATATATTGAG 120

TTTTGCTTAT TTGATATGGA AGAAAATACT CGTTTAGCTA TGTATCAGCC AAAAGAAGGC 180

ATTTATACAC AAAATGAAAC ACAAATAAAA CGAGTGATTG GGTGGCTTGA ACCGAAACAC 240

AATGCTAGAG CGGTAAATGA TGTTATTTTT CACATTTGGA AAGAAGCTAA AATAAAGCCT 300

AAAACGGTCT CAAGATATCT TATACCAGTG AAAAACGGGG TATTTAACCT TAAAACGAAA 360

CAACTAGAAC CGTTTACGCC CAATTATGTA TTCACATCAA AAATTGCAAC CGCATATGTA 420

GAAAATCCGT CGTTACCTAA ATTTAATGAT TGGGATGTAG AGAGCTGGTT GAACGAAATT 480

GCTTGCGGAG ATGCACAAAT AACTACTCTG TTATGGCAGG TAATAAGTGA TGCGATTAAC 540

GGGAACTATT CAAGGAAAAA ATCAATCTGG CTTATGGGTG ATGGATCAAA CGGAAAAGGA 600

ACGTACCAAC AGTTATTATA TAACCTAATT GGTCCACAGA ATATCGCAAC GTTAAAAATA 660

AATCAGTTTA GTGAGCGTTT TAAACTTGCT TTATTAGTAG AAAAAGTGGC AGTTATTGGG 720

GATGATGTAG GAGCrGGAAt TTATATTGAT GATAgTTCGG AATTTAATAG TGtAGTGACT 780

AAkGrGACTA TACTAATTGA AgTTAAAAAT AGAATGCCCT ATAGCGCTCG AATGTATGTG 840

ACTGTTATTC AGTCTACCAA TGAGATGCCT AAAATTCGTA ATAAATCTAA TGGAACGTAT 900

CGCCGTTTGC TTAT 914

(2) INFORMATION FOR SEQ ID NO: 285:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1053 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

GGCATAGCCT TGAAGAAGCC TATCAAGCGA TTTTGTCGGC GCCAAGTTTT GATGCTTGGA 60

TTAAAACCAC ACGCCCTTTG AAATTAGACA ACAATCAATT GTGGCTAGAA GTCCCTTCTG 120

CAGTTCATCG TGATTACTGG GAAAAAAATC TTTCGGCAAA AATTGTCGAA ACTGGATTTA 180

AATTGACCGG TGCAGAAGTG ATGCCCCATT TTGTTGTTGC CGATGAAAAA GACGCAGCAC 240

TGGCACAAGA ACTCGAAGAA CCTGCCGAAG AAGAAGTTGT TTTTAGTGAA CAAAGTAAAA 300

AAGCGATGCT TAATCCTAAG TACACTTTCG ACACCTTCGT TATTGGTAAA GGAAACCAAA 360

TGGCTCATGC TGCAGCGCTT GTTGTTGCCG AGGATCCAGG TTCGATTTAC AATCCCCTTT 420

TCTTCTATGG TGGTGTAGGG TTAGGGAAAA CCCATTTGAT GCATGCTATT GGTCATCAAA 480

TGTTAGTGAA TCAACCAGAT GCCAAAGTCA AATATGTTAG TAGCGAAACG TTTACGAATG 540

AATTTATCAA CTCAATTCAA ACaAAGACAT CCGAACAATT TCGGAAAGAA TATCGCAATG 600

TTGACTTATT ATTAGTCGAT GATATTCAAT TTTTGGCGGA AAAAGAAGCA ACATTGGAAG 660

AATTTTTCCA TACCTTCAAC GATCTTTACA ATGAAAATAA ACAAATTGTT TtAACAAGTG 720

ATCGCCCGCC GAATGATATT CCmAAATTAC CTGnAACGAT TAGTTTCTCG CTTTGCTTGG 780

GGTTTGTCTG TCGATATCAC CCCGCCTGAT TTAGAAACAC GGATTGCAAT TTTGCGCAAA 840

AAAGCAGATG cCGAGCGTTT AGAAATTCCG GaTGATACAC TAAGTTATAT CgCTGGtCaA 900

ATTGaTTCCA ACATCCGTGA ATTAGAGGGT GCTTTGGTCC GTGGTTCAAG CCCTTTGCCA 960

CCTATTATGG AGAAGGATAT TACCACTAGT TAACCGGCGG AGCCTTGGAA ATCCCTAAAA 1020

nCAGTTGGAG TAAAAAAnCA ATnGCCnATT TAC 1053

(2) INFORMATION FOR SEQ ID NO: 286:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2883 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

CGATTAGTGC CTAAAGACAA GCTCACGCAA TTAACGAAAC AATTCCTTTA CTTTTCGGAG 60

TCCTTGGGAT TTTATTTATG GTCGGCCAAT AAGAAGCAAG AACGAATAGA ATTACTTTGT 120

CATATTGAAG AGAGTCAATG TCAACAATTT TATAGAAGAA AACAATCATG GGATTTTTAC 180

GAwAAACrAT TACTAGArAT CTTCCGrCTA CCGACCGCTA ATTTAAATTT ACTGCCAGAT 240

AGACGGCACA CTGGCCAATT GATGCATGAT TATTATCAAA AATTGACGCA ACAACTTGGT 300

TACAGAAATG CTCAGTTGcT TAGAACACAA AGCTTTTTAT ACACGAAAGG TTGTCACCTT 360

CTCCAGTTaC CTCCTTGGTT TTACTATCCT GGATTGAAGC TAATACCTTC TTCTGAAGAA 420

GATATTCATT TAAAAATGCT TGTTTGGGAA GCACTAAAAA CCGAAGAAGA GCGTTTAGTG 480

ACCAAACAGG AACTATGGCG TATCTTGGAA GCTATTTTTT TGGAAGAAGG GAACTTTGTT 540

TGGCAGCCAC TGCCGAATAT TGGCTTAAAA AAGCTTTTTA AAATTGTGGG AAATAACTTA 600

TTGAGATGGT TACAAGAATG TTATGTGTTA ATTAAAGTAA ACAATAAGTA CCTAATAACT 660

AGTATTTTCC TAAGAGAAGA TCCTGAAAAG CTCATTCATT GGCTAAAAAA AGTGAAAAAA 720

CAACATTTTT TTAGTCACAC GTGTTAAATT ATGATAGTAT AATGAGTAAA ACGTTTCTTA 780

AAGGAGAGTA AACTATGAGC GAAATTAAAC AATTACCAAC ACGTGATGAA GkTCCAACTC 840

CATTGACTTG GGATTTAACC AAAATTTTTA AAGATGATGC TGCTTTTGAT GTTGCGTATA 900

ATCAGTTGAT AGAAGAGCTA AATCAAGCAG AATCATTTAA AGGTACCTTA GGGGATGGTG 960

CAGAAGCTTT TCTAGCTGCG TTAGAATACG TTTTAGATGT CTATCGAAAA GTGGAAACAT 1020

TATATGTCTA TTCTCATTTA AAAAATGATC AAGACACTAC AAATACAGCT TACCAAGCTT 1080

TATATGCCAG AGCTAGTTCT TTATATGCAC AAGTTAGTGA AGCTGTTTCT TGGTTTGACC 1140

CAGAGGTTTT AACATTAAGT GATGAACAAA TCTGGGGATA TTTTGAAGAG CAGCCGAAAT 1200

TAGCTGTTTA TCGTCACTAT ATTCAAAATA TTTTAGATGA ACGTCCCCAT GTTTTATCGA 1260

TGGAACAAGA AGCTTTACTG GCGGGGGCTA GTGAaATTTT TGGTGCCTCA AGTAATACAT 1320

TTTCAATTTT AAATAATGCA GACTTAGAAT TTCCAACTGT GCAAAATGCT GAAGGCGAAA 1380

CGATTCAACT TTCTCATGGC GTTTATGGTC AGTTAATGGA AAGTGTCGAT CCATCTGTTC 1440

GTGAAGCAGC ATTTAAAGGT TTGTACAAAG TATACAAACA ATTTAGAAAT ACATTAGCCT 1500

CAACTTTAGG TGCACATGTT AAAACACATA ATTATAAAGC AAAAATTAGA AATTATGATT 1560

CTGCTCGAGC AGCCTCTTTA GCAAGTAATC ATATCCCTGA AAGTGTTCAC GAAACATTAG 1620

TAGCTGTAGT AAATAAACAT TTACCTTTGC TACATCGTTA TGTAAAATTA AGAAAAAAAT 1680

TATTAAACGT AGAAGAATTA CACATGTACG ACTTGTATGC GCCTTTGCTA GGTGAAGCAC 1740

CAATTCGTTA CAGCTATGAA GAAGCAAAAG AAAAAGCAAT TGAGGCTTTA AAACCACTAG 1800

GTGAAGACTA TTTATCTATT GTTAAAGAAG CTTTTTCAAG TCGCTGGATT GATGTGATTG 1860

AAAATCAAGG AAAACGAAGC GGCGCATATT CTTCAGGAGC TTATGATACA GCCCCATACA 1920

TTTTAATGAA TTGGCATGAT AGTTTGGATC AACTATTTAC ATTAGTCCAT GAGATGGGCC 1980

ATAGTGTTCA TAGTTACTAT ACAAGAAATA ATCAGCCGTA TGTTTATGGC GACTATTCAA 2040

TTTTCTTAGC TGAGATTGCT TCAACCACAA ATGAAAATAT TTTAACAGAA TATTTATTAC 2100

AAACAGAAAC AGATCCTAAA GTACGTGCGT ATGTCTTAAA TCACTATTTA GACGGCTTTA 2160

AGGGAACCAT TTTCCGTCAA ACGCAATTTG CGGAATTTGA ACATTTTATT CACACGGAAG 2220

ATGCTAAAGG CACGCCATTA ACAAGTGAAT ATTTGAGTGA GTATTATGGC GAGCTCAATG 2280

CTAAATATTA TGGACCAGAA GTAGTTAGAG ACGAAGAAAT CAGTTACGAA TGGGCGAGAA 2340

TTCCACATTT TTATTACAAT TACTATGTTT ATCAATATGC AACTGGCTTC TCAGCCGCTT 2400

CGGCATTGTC TAAACATATT TTAGCTGGAG AAGAGGGAGC TTTAGAGAAT TATCTTAACT 2460

ACTTGAAAGC AGGAAGTAGT GACTTCCCAA TTGAAGTGAT GAAAAAAGCG GGCGTGGATA 2520

TGACACAAGC CGCGTATATC GAAGATGCAA TGAAAGTTTT TGAAGAACGT TTAACGGAAT 2580

TAGAAGCTTT GGTTGAAAAA TTATAGACAA AAAAGTGGGA TATAAGTCCA GATGACTTAT 2640

ATCCCACTTT TTAACTCTTA AGATAATTGG TTAACAGCTT TCGTCGCAGC ATCTTTTTTA 2700

TAATAAATGT AAATGTCCAT TTGGAAAATG ATAATTGGAG CGACGTTTAA ATTCAAAGGT 2760

TGCGTCACCT kGGTTCAGCA ACGGTTCGTT ATGTGGATGA ACACGACAAA GTTGTTTAAT 2820

TTCATCCATG TTTTCAAAAT TTTCGACTAA AATGCCGAAA TCGTTCTCGC AAGTATAGTT 2880

GTA 2883

(2) INFORMATION FOR SEQ ID NO: 287:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3765 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

CGTTAACCnA GGGTCTAGCT GTCTGnTAAA GCTAGAACCG AAGGACTTAG AGnCGGATTA 60

CCCCACAAAG TCTTCCCACA AwTTTkGATA CTTTACgAGT kGAAACgCCT GATACATACc 120

ATTyCCAAaC ATTGAAGCCc ATGAGGGCTT TTTCGTTTCc GTTGATAACG TTCcAAACAC 180

TGTGGGTGAA AAATGGCCaT CACGTGTTCT GGGTACTTTT AATTCTAGCG TGCCTACACG 240

TGTCGTAAAr TGCGCTCATA ATAGCCATTT CGTTGACTTT GTCGGTTTTC TGTTCGTTCA 300

TATTCTTTTG CTTGAATATA TTCTGTTCGT TGATTTTCCA TTAGTTGATT AAATACCGTT 360

GTTAAAATAT TTTTAGAAAC GTCATCCTTT ACAGAATATT CAATAATGCT TTGAATCTCT 420

TCGCTTTTCA GTGTAAAATG TACTTGGGTC ATGTAAAAGT CCTCCTGGGT ATGTTTTtGT 480

CGTTAAAAAC ATtGTACCGT AAAAGGACTG TTATATGGCC TTTTTACTTT TACACAATTA 540

TACGGACTTT ATCtATTTcA GCCTGTATCA TAGCTAAACA AATCGAGTTG TGTGTCCGTT 600

TTTGGGCGTT CTGCTAGCTT GTTTAAAGTC TCTTGAATGA ATGTATGTTC TAAGTCAAAA 660

AACTCTGACA GCGCCTTTAT ATAGCTTTCT TTTTCTTCTT TTTTTACTTT AATGATCGAT 720

AGCAACAATG ATTTAACACT AGCAAGTTGA ATGCCACCAT TTCTTCCTGG TTTAATCTTA 780

AAGAAAATTT CCTGATTCGC CTTCAGTACC TTCAGCAATT TATCTAATGT CCGCTCAGGA 840

ATGCCTAGCG CTTCTCTAAT CTCTTTTTTA GTCGTCACCA AATAAGGCTT GTATACATCA 900

CTTTTTTCGC TAATATAAGC CATTAAATCT TCTTTCCATT CTGACAAATG AACACGTTGA 960

CGTTCACTTC TTTTTTTCTT GAATTTAAAC CACCCTTGAC GGACAAATAA ATCTTTACTG 1020

GTTAAATCAC TTGATACCCA AGCTTTGCAA AGAATGGTAA TGTATTCCCT ATTAGCCCCT 1080

TGATAGTTTT CTGAATAGGC ACTTCTAACA AGTTTAATCA CTTCTTTTTC TTCTAAGGGT 1140

TGATCTAATC GATTATTAAA CTCAAACATA TTATATTCGC ACGTTTCGAT TGAATAGCCT 1200

GAACTAAAGT AGGCTAAAGA GAGGGTAAAC ATGACGTTAT TACGCCCTAT TAAACCCTTT 1260

TCTCCTGAAA ATTTCGTTTC GTGCAATAAG AGATTAAACC AGGGTTCATC TACTTGTTTT 1320

TTGCCTTCTG TACCGCTTAA AACCGTTAGA CTTGAACGAG TAAAGCCCTT ATTATCTGTT 1380

TGTTTGAAAG ACCAATCTTG CCATTCTTTG AAAGAATAAC GGTAATTGGG ATCAAAAAAT 1440

TCTACATTGT CCGTTCTTGG TATaCGAGCA wTCCCAAAAT GATTgCACGT TAGATCAACT 1500

GGcAAAGACT TyCCAAAATA TTCTCGGATA TTTTGCgAGA TTATTTgGCT GcTTTGaCAG 1560

ATTtAAATTC TGATTTTGAA GTCACATAAC TGGCGTTTCT AAAACAAAAT ATGcTTGaTA 1620

ACCtTTATCA GATTTGATAA TTAACGTAGG CATAAAACCt AAATCAATAG CTGTTGTTAA 1680

AATAtCGCTT GCTGAAATAG TTTCTTTTGC CGTGTGAATA TCAAAATCAA TAAAGAAGGT 1740

ATTGATTTGT CTTAAATTGT TTTCAGAATG TCCTTTCGTG TATGAACGGT TTTCGTCTGC 1800

ATACGTACCA TAACGATAAA CGTTTGGTGT CCAATGCGTA AATGTATCTT GATTTTCGTG 1860

AATCGCTTCT TCGGAAGTCA GAACAACGCC ACGTCCGCCA ATCATGCTTT TTTTTGAGCG 1920

ATACGCAAAA ATAGCCCCTT TGCTTTTACC TGGCTTGGTA GTGATTGAGC GAATTTTACT 1980

ATTTTTAAAT TTGTACTTTA ACAAGCCGTC ATGAAGCACA GTTTCTACAA CAAAAGGGAT 2040

ATTCATTCAG CTGTTCTCCT TTCTtACGAA AATTAATTAG TTAGAAGCTA CGATCAAAGT 2100

TGAATCACAA CAAAAAAGGC AATCAACTAA GTTTTTCTTA ATTGATTGCC TGGTATCTTC 2160

TTAAAGACTT GAAATTCCCT CAAAAACCCG ATATAATGGG TTTACAGATA TTTAAGTATC 2220

TGATTAATAA AGTAATTAAA TACTTTACCA AATTTCGGGT CTCGACTTCT TTAATTGATT 2280

GGTGGTAATC AATTAAGGCT CGCAACTTAT TTTCTtGGCG GAAAATAAGT TGGTCGTGGC 2340

TCTTTTTTTG TATTCTTTAT TCAGTTCGTT GTTTCGTTAT ATCTAGTATA CCGCTTTTAA 2400

AAAAAATAAG CAACAATTTC GTTAAATATA TTAACGACAA TCATTGCTTC TCTTCTTCTA 2460

TTTGCAACTC TACTTGTTCT AATTCATTAG GAATTAAGTC AATAACTAAG GTTGAGTTTG 2520

TCAATTCTGC AATTCGTTCC AATGTTTTTA ATGTCGGATC TGATTTTCCT GATTCAATCA 2580

ATGAATAATT TGGTTCTGTT GCAAAGTTTT AAATCTACTA TCAAATAAGG TAGAATAATA 2640

GAAAAAGATA GCAGGAGGAA TGACGATGAA TCATTTTAAA GGAAAGCAAT TTCAGCAGGA 2700

TGTGATTATT GTAGCCGTGG GCTACTATCT TCGTTATAAC CTTAGCTATC GTGAAGTTCA 2760

AGAAATCTTA TATGATCGTG GCATTAACGT TTCTCATACG ACGATTTATC GTTGGGTGCA 2820

AGAATATGGC AAACTACTCT ATCAAATTTG GAAAAAGAAA AATAAAAAAT CCTTTTATTC 2880

ATGGAAAATG GATGAAACGT ACATCAAAAT TAAAGGAAAA TGGCATTATT TGTATCGAGC 2940

CATCGATGCA GATGGTTTAA CCTTGGATAT TTGGTTACGT AAAAAACGGG ACACACAAGC 3000

AGCCTATGCT TTTCTTAAGC GGTTAGTGAA GCAGTTTGAT GAACCGAAGT TGTtAGTCAC 3060

AGATAAAGCC CCCTCTATTA CAAGTGCCTT TAAGAAACTA AAAGAATACG GCTTTTATCA 3120

AGGGACAGAA CATCGTACCA TTAAATACCT GAATAATTTG ATTGAACAAG ACCATCGTCC 3180

AGTAAAGAGA CGCAATAAAT TCTATCGAAG TTTACGCACT GCCTCTACCA CGATTAAAGG 3240

CATGGAAGCC ATTCGAGGAT TATATAAGAr AACCCGAAAA GAAGGCACTC TCTTCGGGTT 3300

TTCGGTCTGT ACTGAAATCA AGGTATTATT GGGAATCCCA GCTTAAATCA TAGATACCGT 3360

AAGGGATTTT ATTCTTTATT TAAAACTTTG CAACAGAACC ATTGTGTAAA AGTAAAAAGG 3420

CCATATAACA GTCCTTTtAC GGTACAATGT TTTTAACGAC AAAAACATAC CCAGGAGGAC 3480

TTTTACATGA CCCAAGTACA TTTTACACTG AAAAGCGAAG AGATTCAAAG CATTATTGAA 3540

TATTCTGTAA AGGATGACGT TTCTAAAAAT ATTTTAACAA CGGTATTTAA TCAACTAATG 3600

GAAAATCAAC GAACAGAATA TATTCAAGCA AAAGAATATG AACGAACAGA AAACCGACAA 3660

AGTCAACGAA ATGGCTATTA TGAGCGCAGC TTTACGACAC GTGTAGGCAC GCTAGAATTA 3720

AAAGTACCCA GAACACGTGA TGGCCATTTT TCACCCACAG TGTTT 3765

(2) INFORMATION FOR SEQ ID NO: 288:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7043 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

nnATAAATGT TGAATAAAGT TTGATAnTGT TTTATTATGT TTTCGAGAAG GGGATTTGAG 60

TTTAAAGAAA GGGACAGAAT TTTAATAGTA GATACACGAG GTAAGGAGGA AAACAAAATG 120

CGTGTCATAC TTGGTTCAGA TTTAGATGGA ATAAAATTAA AAGCAGAAAT GAAACAATAC 180

TTATTGCAAG AAAAAGTTGA GGTAATCGAT AAAAGTGAAA GTGCTTCGGA AGACTTTATA 240

GAAGCAACGC TTGCTGTTGC TCATGAAGTT TTAAAGGATA CTGAAAGTTT AGGAATCGTT 300

TTTGATGGTT ATGGGGCAGG TAGTTTTATG ACTGCTGCTA AAATTAAAGG AATGATTGTC 360

GCTGAGTTGT CAGATGAGCG TTCTGCATAT ATGGCGCGAG AACACAACAA TGCTCGGATG 420

ATCACGGTGG GCGCAAAAAT TGTTGGAACG GAACTAGCCA AAAACATAAT AAAAGAATTC 480

TTAACAGGTC ATTATGCAGG CGGGCGTCAT CAAATACGTG TAGATATGTT AAATAAAaTG 540

GCTTAaTAAA TAATAGAAAG GAAGCGTTTT TATGAAAATT GCCATTGGTT GTGATCATAT 600

TGTGACGGAT ACCAAAATAG CAGTTTCTGa TTTTTTGAAA GCAAAGGGAT ATGAAATTTT 660

AGATGTAGGC ACGTATGATT TTACTCGAAC TCACTATCCT ATTTTCGGAA AAAAAGTGGG 720

AGAAGCCGTA ATGAGCAATC AGGCTGATTT GGGTATATGT ATTTGCGGAA CAGGCGTTGG 780

aATTACCAAT GCGGtTAACA AAGTGCCGGG TATCCGTtCT GCTTTAGTTC GCGATATGAC 840

TACAGCGTTG TATGCGAAAG AGGAATTGrA CGCAAATGTT ATTGGTTTCG GTGGGAAAAT 900

AACAGGAGAA TTCTTGATCT GTGATATCGC TGCTGCCTTT ATTGAAGCAC ACTATCATCC 960

AACAGAAGAC AACGAACAAC TAATTGAAAA AATTAATCAA GTTGAAAAAC AACATCCAGA 1020

ACAACAAGAT GCGCACTTTT TCGATGAATT TTTAGATAAA TGGCAACGAG GAGTCTATCA 1080

CGATTAGTAT TTAATCAATG AAGGTTACTT AAGAAAGTGG TCTTAAATAA TAATCATTGG 1140

TATTTGAGTT AAATTCCCCT TCATTAATTT TTATTGCTGT TTTAATGTGG CACTGTAAAG 1200

AGCATTATTA AGAAAAAGAA GTAACAAAGG AACAAATTGA AGGCTATACA GGTAAsCATA 1260

AAAAGTACAC GAGTTCTTTT CTCGTGTACT TTTTATGGTT TCGCTGCTTG TTTTATTTTA 1320

GTAATTATTT TTCTATAAAA ATAGTCCCAG AATTATTAAA ATAATCCCGA TAACTAAACT 1380

GCCGATGCCA GAAGCCGTAC CATATTTTTT ACCTAAATTT ATATCAGTAG TATTTTTAGG 1440

AATCAAGTTT GAGAAAAAAT AAAATTTTTT TACCAAGATT AGATATGCAA CAAGTAAAAG 1500

TAATAATCCT AAGAGAaGAT AACTAAGaTT TTCAGCTAAT AAATCtrGAT GAGCTAATAA 1560

CaATATCCCC AAAAAGAGGC CTACCAATCC ATATTTCATC GTTTTTTCTT TTGCTGACAT 1620

ATAAACATCC CTACTTTACT TTTTAATTTA TTAAGAGCAT AAAACCATAG ATTCGATATA 1680

AGTTCATTTT GTTGTATCCA AAAGAGCATG ACCTGGAATG ACCTATATAA TTAGAAAATT 1740

GGAGTGATAT ATAGTTATTT TACCATAGCT TAAATGAGTT GATTAACTCC GAATAATGTA 1800

CAACTATGTA CACCTATTGT GGGTGGAATG TGTGTTGGGA CTTTGCTAAC CTAATAGTTG 1860

AAAAAATGAA TCAATAAATT TAGAAAAGGG ACTATTCAAA AGAAACTTTT TTGGGAATTG 1920

CTATATTATT TATTTTTTTA AGCGTAATCA ACTAAATGTT TTTATGGATA AAGAAAGTCA 1980

TTGTTCCTTG TCATCAAGCT ACAACAGCCA AATTCATTAG AATTTTTTTA GTGAATGACT 2040

TTAATTTTAG AAAGGTACTG TTGTTAAACA TTCATGCTAT CAAAGGAGTG TGAGGCGAAC 2100

GAAGCAAGTA TTTTCaTTTT ATAAAAGTTG TATTATGGAC TCTTAAAAAG AGGTGATTTA 2160

AAAAATTTTG AAGTGGAAAT GTTAAGAAAG GAGAGAGGAA ATGACTTATT TAGTTCTGGT 2220

CAGTCATGGC GATTTTGCTA GAGGGTTAAA AAATTCATTA GGGATGTTTG CTGGCGATGA 2280

GATAGAGAAA GTCAGAGTGT TTGGCTTACA ACCTGGTGAA TCTACCGATA ATTTTGGTAA 2340

GCGGTTTCTT TGTGAAGTTG AAAAATTACC GCAgGAAACG AGGCTAATCG TTTTGGCTGA 2400

TATTGTAGGT GGGAGTCCAC TTGCTACTAT TTGTAAAGTA TTAGATAAGC AAAAAATGCT 2460

AGAAAATTCT CTGATTGTTG GAGGAGTGAA TTTTCCAATG GCTTTAAATG CGCTTCTATT 2520

GAAAGAAACT AGTGAATTAT CTGATTTGAA AAAAATGATA GTAGCAGATG CTTGTCATGG 2580

CATTAGAGAG TTAACTATAA CTGAAGGCGA TGTAACTGAA GAGGATATCT AAAGGAGAAG 2640

ATGAACATGT CAATTTCTTT TGTTCGAATC GATGATCGAA TGATTCATGG CTTAATTACT 2700

TTACGTTGGA CTAAAGAAAT GCCATGTGAT GGCATTATAG CTGTGAATGA CAAAGCAGCA 2760

AGTAACCCCA TTTTAAAGGA AGCATATAAA GCAGCTGCAC AAGATAAAAA AACCTTTATT 2820

TGGACAATGG CGCATTTTTT TGATGTAAAG GATAAAGTAT TAGCGTCTAA GTCAAAATAC 2880

TTTTTGATTA CAAAGAGCCC TTTGGATATG AAAAAAATTT TGGTAGATTG GCATTTTGTT 2940

CCAAGTGAAG TAAAAATGAT TAATGTTGGA CCGGGGAATG ATCGCGAAGG GACAATTAAA 3000

TTAGGGGATA ACCAATCCTT TACAGCAGAA GAAGCAGCGG CTTTTGAAGA AATTGAAAAG 3060

GCTGGTTACA AAGTGGATTT TGCCTTGTTA CCTGATCAAC GAATTGGTTC TTGGGCGCAT 3120

TTTAAAGCTA AATTTGGTTA CTGATTCTAG AAGTTTTGAA AAGAGGAGAG TAGTATTATG 3180

ACAATCAACT GGATACAAGC GGCTATTTTA GGCCTATTTG CTTGTCTGTG TTCAAATTCT 3240

TGTATGGCTG GACAAGCGGT GGGAAATTAT ACGATTGGGC GTCCTTTAGT CGGCGGATTA 3300

GTTTGTGGCA TTATTTTAGG AGATTTAAAA TTAGGAATCG CTTGTGGCGT ACAATGCAAT 3360

TAGTTTACAT TGCTTTGGTC ACACCTGGAG GCACAGTGTC TGCTGATGTG CGGGCAATTT 3420

CATATATCGG GATTCCTTTG GCGATGGTGG CGATTTCGAC CAAAGGGTTA GATCCACTTG 3480

GCGGCTCTGC AGCCGACCTG GCTAAATCGG TCGGCACATT AGTTGGAACG ATCGGAACGG 3540

TATTATTTTA TTCTGTTGCC GCGTTAAATT TAGTTTGGCA ATCGTTTGGT TGGAAAGATA 3600

TAAAAAAAGG CAAGCTTGAT CATTTATATG CCATTAATTT TGGTTGGCCG TGGATTTCTC 3660

ATCTACTTTT TTCCTTTTTG CCAACCGTGA TTTTAACCTA TTTTGGGGCT ACGGCTGTTA 3720

```
CAGCAATGAG AGATGCATTA CCTATGGATG GTCTTGCAAT GAAAACGCTT TTTACTGTAG    3780

GTGGCATGCT TCCGTGTGTG GGAATTGCAA TTTTGTTACG TCAAATTGTT AATAAGAACA    3840

TTGATTTTGT ACCTTTTTTT GTAGGGTTCA CATTAGCTGC TTCATTAAAA CTCAATTTAG    3900

TCTCAATAAc GATTATTGCg CTTCTTTTTG CTGTAATTTT TTAwAAAATT GCTTTGATTA    3960

AAAACAATAC TAAACTGGTT GATTCAACAG GAAgTCTAAC GAATGATGAC GAAGAGGAGG    4020

ACATATGAGA TGGAAGAGTT GAAAGCAAAT AAATTAGACC AAAAAACGTT AAAGCATTCT    4080

TTTCATTTAT GGTTTTGGGG AGCATTAACT TGTTTTTCAC AACAGCATAT GCAAACCTTC    4140

GGCTATCTTT CTTCAATGCT TCCTATtATC AAAAAACTTT ATCCAAAACA TGAAGACCAA    4200

GTGAAAGCCA TTCAAGCATA TACAGCTTTT TTTAACACAA ATCCAATGCT TGGAACGGTC    4260

ATTGTTGGTG TGACTGCTAG TATGGAAGAG GCTAGAGCAA GTAAGAATGA AATTGATGGA    4320

GAAACCATTA ATGATATGAG GGCAGGATTA ATGGGACCGA TTGCAGGGAT TGGCGATTCA    4380

TTAGTTGATG GTACCTTGAT TCCTATCTTG TTAGGAATTT CCTTAGGAAT GTCCACGGGC    4440

GGTTCGCCAA TTGGTGCCAT ATTTTATATT ATTGTTTGGA CGTTAATTTC CTATTTTGGC    4500

CAACGCTTTT TGTATTTTAG AGGCTATCAT TTTGGTGATA AAGCAGTAAG TTTCTTAGTT    4560

GGGAAAGAAG GAGCCGCTGT TAGAGTTGCT ATCGGAGTTA TTGGTAGTAT GGTTGTCGGT    4620

GGGGTCCTTG CATCTTGGGT AAATGTAACT ACTTCTCTTA AATTAATAGG TGCTGACGGA    4680

AAAGTTTTCC TTAAACTGCA AGATAAAATT GACAGTATTT ATCCAGGATT ATTAACGATA    4740

CTTGTTACCT TGTTTTGTTG GTGGTTAATG TCCAAAAAAC ATGTTTCCGC TATTTGGACG    4800

ATGCTTATTT TAGTTGTTAT TTCTCTCGTT GGCGTTTTAT TAGGTGTTTT CAATCCAGGA    4860

TTAACTTACT GAACAATCAC CAAAAAGGAC TCTAGCTGTT AGAGCAAACT GACTTTTGCA    4920

TAAAACCGAT AACGAAAAAG CTAATAAATT AAATTCTTAC TCTCTAAGAC ACAGAAAAAA    4980

TCTTGTGATT AAAATAAAAA TATATAAAAT AGAAGAAATT TAGAGACTAG CTTCTATTTT    5040

ATATATTTTT TTGTTTTAGA AATTTAGTTG GTTTATTAGA ATTTTAGGAA AAATTTTACG    5100

AAAAGTGTCG CAAAAATGGC TCCTAAAAAC GGTGCAGTTC CGGGAACAAT CAACCCATAT    5160

TGCCAATCAT TATTGGTTTT ATTTTTGATT GGTAAAAGTT GATAGGCGAT ACGAGGACCT    5220

AAATCACGTG CTTGATTCAT TGCGAATCCT GTAGTACCAC CTAGCCCCAT GCCGATTGCC    5280

CAAACAAGCA AGCCAACAGC AATGGGTAGT TGTGTTTCGT AGGAATGTGC AATCGCTAAA    5340

ATAGATGTTA AAAAAATGAA TGTTGCAAAA GTTTCAACAA AATAGTTTCT TGGTAAATTT    5400

CGTTGGTTCG GATTTGTAGA AAAAATATTA CGAATGGCGA TCGGATCGAC ACTATCGGCG    5460

GATAGCTTGA AATGATCAGC ATACATAATA TAAACAATGA CGGCGCCACA AATGCCGCCT    5520

AACATTTCAG CGATGACATA GGGAATGAAG TAAGACCATG GAATCATTCC TAAAATAGCT    5580

TGAGCGAGGG CCATTGCTGG ATTAATACAG ACTCCGCCAA AGACGAACAG AACAACAGAG    5640

ATTCCAAAAG CCCAAGTAGT GATTGCAAAC ATGTGACCAG AACCAGCGTA TTTTGTTCGT    5700
```

TTTAGGACAT CGTCGCAGTG AACACCTACC CCAAATACGA TCATTAATGC AGTTCCCATA 5760

AATTCTGATA AGATATGATG AAACATAAGA TAAAGCTCCT TTAATCAATA GTTTTGTTAT 5820

TTAATGTTTG GACGAGGCGA TAGACTTCCT CGATTGTAAA ATTATATTGT TTAGCAATTC 5880

GTTCACAATC TTGGTATTCA ATAGACTCTT TCAATAGGCC ATGATAATGA TTTTTTTTAA 5940

TATGGACCTT ACCAAAAGGT GTTTCCAAGA TTTTGAAAGA ACGGTCCATC ACACTACGTT 6000

CCATTGTTTG AAAACGCATA CCAATTGTAG AAGTATGTTT AAAGAGTAGT TCTGTAAAAT 6060

ATTCTTTTTC CTCTGTTGTT GTTAAAACCG TTAATAGAAT GGCACTTCGG TTTTTTTTCA 6120

TATGAATGGG CGTGAAAAAA ACATCTAGAG CACCATGCTC TAATAATAGA TCCATTACGT 6180

AACCTAACTG TTCaGGTGTT TGaTTGTCTA tATTTGTTTC aATTTTTAAg ACTTGGTCaT 6240

GGTGTTCTGT CCTCTTTTTT TTAcTATGTG TTTCAGTAAA GATTGATCCT CTGAGGGCAT 6300

TGAACTTGCC TGtTTCCCGT TTGCCaAATC CGTAACCTAT TTTTTCAATT AATCGTTGTT 6360

CTGGTGCGAC AAATAGAGGA CTAAGTTCTT TAAAAATAGC TAAGCCAGTC GGTGTGACAA 6420

GTTCAGTTTT TATCTCAAAG TCTTGTTGAA TGATTAAATT CGTTTCTTTA CGTAATTCCA 6480

TCACAGCGGG AACTGGTATT GGCATGACTC CATGGGCAAT TGTGATAGTG CCACTTCCAT 6540

CTGTAATTGG CGTTGAATAG ACTTGTTCAA TCTCTAACTG TTCCCATAGG ATAAAAAAGC 6600

CAACGATGTC AACAATTGAA TCAATTGCTC CAACTTCATG AAAATGAATA GTGTCAATTG 6660

TTTGTTGATG AACAGCAGCT TCAGCTTTTG CAATGTCATA GAAAACATTT TTACTATGTT 6720

TTTTTACAAA ATTGGACAAG TCGCTTTTTT CAATTAAGGA ACAAACAAAA CGATCATTCA 6780

CGAAAAGAAA AATAATCTGA GTTTTGCACC AAACACAAGG TTTGATCTGC CTATTTTACT 6840

GAAAGATTCG TTTAAACCTG GAGACTATAC CTATACCATT AAGTTGAAAA ACCATGAGGG 6900

GAGCTGGACG TTCTCAAAAG ATTTCACGAT AAACAAAAAG CAAGCTGATC AGTATAATAA 6960

GCGTTCTGTT GATCATAAAG TCAAAAACCT TCACTGGCTT AAGTACCTCA TgGnCTTTTT 7020

AATTCTAGTT ATTTTAGGTA CCG 7043

(2) INFORMATION FOR SEQ ID NO: 289:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 817 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

TAATTATTAT CTCATTTGTA GATTAGTAAA AGGATACCGA AAnACATTTA GnAAAGCAAC 60

CACTTATTTT CAGAAAGTTA GGAAATCATG AAAATATTTG kTCyTCaTTT tCCtTAAAAT 120

TTTGtTAAAC TGtAAAGaTA TGAGTCsTAA CGAAgGAGTT AAATAGTATG ACAGAATCCA 180

AAACTGCCAT TATTTTTGAT ATGGACGGTG TCCTAGTAGA TAGCGAAgCT TATTATTATG 240

AGCGACGTAA AGCATTTTTa GCGGAATTTG ATTTGACCAT TGAGGGCCTA ACCTTGCCGG 300

AATTAGTCGG TGCGGATATG CGTTCATTAT GGCAAAAGAT TGAACAAGTC AACAAAAAAG 360

AGCTAGATAT TGCTTTTTTA AATGAACAAT ACATAGCCTA TAAAAAAGCG CATCCGATAG 420

ATTATTTAGC TGTGCTTGAT GAAAATGCGA AACGTGTTTT ACAATTTTTA AAAAGACAAG 480

GCTATAAAAT CGGATTAGCT TCTTCTTCCA CAAAAGACGC AATTGAAGAG GTTTTGACAG 540

TGGGGCAATT AAGTAGTTAT TTTGATGCGG TCGTGAGTGG GGAAGATTTC GAAGAAAGTA 600

AACCGGCGCC GGATATTTAT CTGCACACGT TACAAGAATT AGCTGTTGCG CCACAGGAAT 660

GTATAGCCAT TGAAGATTCA GAAAAAGGCA TTGCCTCTGC CAAAGAAGCT GGGcTAGAAG 720

TTtGGGcTAT GCGCGATGAA CACTTTGGGa TGGATCAAAG TCAAGCGGAT GCCTTCTTAA 780

cACmACTAAg TGATATkTGT AAAAAAaTTA GTGAAAA 817

(2) INFORMATION FOR SEQ ID NO: 290:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10397 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

ATGAAGAAAT ATATCTTCAC AGAAAGAAAC GGTATCTACA TCATTGACTT ACAAAAAACA 60

GTGAAATTAG TAGATGCAGC TTACGATTAC ATGAAAAACG TTGCAGAAGA AGGCGGCGTT 120

GCTTTATTCG TAGGAACTAA AAAACAAGCA CAAGAAGCAA TCAAAGATGA AGCAATCCGC 180

GCTGGTCAAT ACTATGTAAA CCATCGTTGG TTAGGTGGTA CGTTAACTAA CTGGGATACA 240

ATCCAAAAAC GTATTGCTCG CTTGAAAAAA ATCAATGCAA TGGAAGAAGA CGGCACTTTC 300

GAAGTTTTAC CTAAGAAAGA AGTTGCTGGC TTAAACAAAG AACGTGAACG TTTAGAAAAA 360

TTCTTAGGTG GTATCGCTGA TATGCCAAGA ATTCCAGATG TAATGTACAT CGTTGACCCG 420

CGTAAAGAAC GCATTGCTGT TCAAGAAGCA CACAAATTAA ACATTCCTAT CGTTGCAATG 480

GTTGATACAA ACTGTGATCC TGATGAGATC GATGTTGTAA TCCCTTCAAA TGATGATGCA 540

ATCCGCGCTG TGAAATTAAT CACTGCTAAA AtGGCTGACG CTTTCaTCGA AGGTAACCAA 600

GGGGAAGACC AAGCAACTGA AGAGTTATTC GTTGAAGAAA CACCAGAAGC AACTTCAATC 660

GAAGAAATCG TTGATGTTGT TGAAGGAAAC AACGAATCAG CTGAATAATA ATGATTACGG 720

AGCTGTCTCA AAGGCTAGGC GACAAAAAGC CTGATTCTCC TTTGAGATAG CTTTTTTAAA 780

AAGAATTGAT AATATCCGAG GAGGACACAA ATAATGGCAG ACGTTACAGC TAAAATGGTA 840

AAAGAATTAC GCGATATGAC TGGCGTTGGA ATGATGGACG CTAAAAAAGC ATTAGTAAAA 900

GTAGAAGGCG ACATGGAAAA AGCCGTAGAC TTTTTACGTG AAAATGGTAT GGCTAAAGCA 960

GCGAAGAAAA ATGACCGCAT TGCGGCTGAA GGTTTAGCAA ACGTAGCAAC TGTGGGTAAC 1020

GTTGCAGCAA TCGTTGAAGT AAACTCAGAA ACTGACTTCG TTTCTAAAAA CGAAATGTTC 1080

```
CAAGATTTAG TTAAAGACAT TGCGACTAAA GTTGCTGAAA ACAAACCAGC TACGATGGAA   1140
GAAGCAATGG CTATTAAAAC TGAAAAAGGC ACAATTGAAT CTGATTTAAT CGAAGCAACA   1200
ACTGTTATCG GTGAAAAAAT CAGCTTCCGT CGTTTTGAAG TTGTCGAAAA AGCTGACAAT   1260
GCTGCATTTG GCGCTTACTT ACACATGGGT GGCCGTATTG CTGTATTAAC AGTTATCGAT   1320
GGCACAACTG ACGAAGAAGT AGCGAAAGAT GTTGCTATGC ACATCGCGGC AATCAACCCT   1380
CGCTATGTCA ACGAATCTCA AATTCCTCAA GAAGAATTAG AACACGAAAA AGCTGTGTTA   1440
ACTGAACAAG CATTAAACGA AGGTAAACCA GCAAATATCG TTGAAAAAAT GGTTGTTGGT   1500
CGTCTACAAA AATTCAAAGC AGAAATCGCT TTGGTTGACC AACCATTTGT TAAAGATCCA   1560
GATATGACTG TCGAAAAATT TGTTGCTTCT AAAGGCGGCG AAGTGAAATC ATTCGTACGT   1620
TTTGAAGTTG GCGAAGGTAT CGAAAAACGT GAAGATAACT TTGCGGATGA AGTAATGAGC   1680
CAAATGAAAA ACTAAGAGCA TCAAACGCTC GGTTGAAAAG CCGTGAATCA TAAGTATCTA   1740
AACAGACAGC GATCCCCAAA GGAAGCTGCA TAAATTGATT ACTGCTTGAT TCAAGGCAGT   1800
CTTTCAGCCA TTTGGGAACG CATAACTTTG ATATGCGCTC CCTTTTTTTA GGCAAAACCA   1860
GTGGATCGCA CTGCTTGGTC TAAAAAATCA TTTTGGTGAA AAGTCATCTA AATGAGGAAT   1920
TGAATGATTT TTAGAGAATC TTGGAGGGAA TTATGATGGT TAAACCTAAG TATCAACGTG   1980
TCGTATTAAA GTTAAGTGGA GAAGCGCTAg CCGGAGAAGA CGGTTTTGGC ATTAAACCAC   2040
CAGTCATTAA AGAAATCGTC CAAGAAATTA AAGAGGTCCA TGAATTAGGT ATCGAAATGG   2100
CCATCGTGGT TGGCGGCGGA AACATTTGGC GTGGACAAAT TGGTGCTCAA ATGGGCATGG   2160
AACGTGCACA AGCTGATTAT ATGGGCATGT TAGCTACTGT GATGAATGCC TTAGCATTGC   2220
AAGATACATT AGAAAATCTT GGTGTGCCTA CACGTGTTCA AACATCAATT GAAATGCGTC   2280
AAATTGCAGA ACCATACATT CGTCGTCGTG CAGAACGTCA TTTAGAAAAA GGCCGCGTTG   2340
TTATTTTTGC CGGAGGAACA GGAAATCCTT ACTTTTCGAC AGATACAACG GCTGCGTTAC   2400
GAGCTGCAGA AGTAGATGCG GACGTCATCT TAATGGCGAA GAACAATGTT GATGGTGTTT   2460
ATTCTGCGGA TCCACGTGTG GATGAAACAG CGACAAAATT TGAAGAACTA ACACATTTAG   2520
ATGTTATTTC TAAAGGGCTG CAAGTAATGG ATTCAACAGC AAGCTCATTA AGTATGGATA   2580
ACGATATTCC TTTAGTTGTC TTTAACCTAA ATGAAGCAGG AAATATCCGT CGTGCCATCC   2640
TTGGTGAAAA TATCGGAACA ACTGTAAGGG GGAAATAAAG AATGTCTAAA GAAGTTTTAG   2700
CAACAGCCAA AGAAAAAATG ACAAAAGCGG AAGAAAGCTT ACGCCGTGAA TTAGGTCAAA   2760
TTCGTGCTGG TCGTGCGAAC GCAAGTCTAT TAGACCGTAT TCAAGTAGAA TATTATGGTG   2820
CACCAACACC AGTTAACCAA TTAGCAAGTA TCAATATCCC AGAAGCACGT GTATTAATGA   2880
TTACACCGTT TGACAAAAAT TCTATCGCTG ACATTGAAAA AGCTATTCAA ATGAGTGATA   2940
TCGGCATTAG CCCAACCAAT GATGGAAATG TCATCCGTTT AGTGATTCCA CAATTAACGG   3000
AAGAACGTCG GAAAGAATTA GCCAAAGACG TGAAAAAAGA AGCGGAGAAC TCAAAAGTTG   3060
```

CTGTCCGTAA CGTTCGTCGT GATGCAATGG ATGAGTTGAA AAAAGCTCAA AAAAATGGCG 3120

ACATCACTGA AGATGAATTA CGTTCATTTG AAAAAGATGT TCAAAAATTA ACAGACGATA 3180

GCATCAAAAA TATCGATGCG ATTACTGCTG AAAAAGAACA AGAACTTTTA GAAGTATAAA 3240

ACCAATCACT AATAAAAATT ATTGAATCCA GAGAAGCAAA TGCTCTGGAT TTTTTTGTAT 3300

AGATGAGAAA GGCGTATCTC AACGCTCTCT TTGCAAAAGT CATTCTAAAG TAATTATTTA 3360

AGAAAACGCT TTCTTTAATA AATTTTTTTA TTCTACCAGA ATTTTTTTCT TTGTGTAGGG 3420

AGGATCACTT AAAAATATAG AGAAACTTTG AAATATCAGC CTTTTAAAAT TTTTCTTCAC 3480

AAATTGAATC CTAAAATTTG ATGATAAGAG TTGTTATTTT TAGCAATTGT AATTTGCGAT 3540

TCATTCTCAA TTAGGCTAAA ATGAGTGAGA ACGGAAAAAC GATTAAAACA AGTTGGAGGG 3600

AATTGTACAT GTCTGTGTTA GAAATCAAAA ATTTACACGT ATCAATTGAA GATAAAAAGA 3660

TTCTAAAAGG AGTCGATTTA ACAATTAATA CTGGAGAAAT CCATGCAATT ATGGGACCTA 3720

ATGGAACTGG TAAATCTACT TTATCAGCAG CAATTATGGG AAATCCTAAC TACGAAGTCA 3780

CAGAAGGAGA AATTCTATTT GACGGCGTCA ACGTTTTAGA TTTAGAGGTA GACGAGCGTG 3840

CTCGCTTAGG TTTATTCTTA GCAATGCAAT ACCCAAGTGA AATTCCAGGA ATTACCAATG 3900

CTGAATTCAT GCGTGCAGCA ATCAATGCAA AACGTGACGA AGACAACAAA ATGTCAGTTA 3960

TGCAGTTCTT AAAAAAATTA GATAGCAAAA TGGAATTATT AAATATGCCT GAAGAAATGG 4020

CTGAGCGTTA CTTAAACGAA GGTTTTTCTG GCGGCGAGAA AAAACGAAAT GAAATTTTAC 4080

AATTATTGAT GTTAGAGCCA ACGTTTGCCA TTTTAGATGA AATTGATTCA GGCTTAGATA 4140

TCGATGCGTT AAAAGTAGTT TCTAAAGGGG TTAATGAAAT GCGCGGCGAG AACTTTGGTG 4200

CGTTAATTAT TACCCATTAT CAACGCTTAT TGAACTATAT CACGCCAGAT GTTGTTCATA 4260

TCATGATGGA AGGCCGCGTT GTGAAAACCG GTGGGGCTGA TTTGGCAAAA CGTTTAGAAA 4320

CAGAAGGCTA TGCTGGTATC AGCCAAGAAT TAGGTATTGA TTACAAAGAA GAAGCGTAAG 4380

GAGGACAAGA TAAATGAAAA AAGTAAATCT AATGGATTAT CTTGACGAAg TCACCGcTTT 4440

TTCACTAGAA CATGAAGAAC CAGCTTGGAT GACTGAATTG CGAACAACGG CTTTAAGAAA 4500

CGCTGACGAA AGTGAGTTAC CACATATTGA TCGTGTGAAA TTTCATCGTT GGCCTTTATT 4560

AAATGTTCAC ATGGAAAGTT ATGTGCCTTC TGAAGGAAAT GTCGCAAGTT TTGATCAAAT 4620

GAAAGACAAT CCTTTAATTG TTCAACAAGG ATCTTTTCAT GCTTTTGAGC AATTGCCAGC 4680

CTCATTAGCG GAACAAGGGG TAATTTTTAC AGACATTTTT ACTGCTTTAC AAGAGCACCC 4740

AGAATTAGTT AAAGAATACT ATATGACTAA AGCTGTATTA CCAGAAGAAG ATAAGCTGAC 4800

AGCAGCTCAT GCAGCTTTCA TGAATAGTGG CGTTTTCTTA TACGTTCCTA AAAACGTAGT 4860

GATTGAAGAG CCAATTGAAT CTTTGTTTAT TCAGGATTCA GATAGCAACC AACCTTTCTT 4920

TAAACATGTT TTAATCGTTG CAGATAATCA TAGTGAATTC AGTTATTTAG AACGTTTTCA 4980

ATCAACTGGT CATCACGCAG AGAAAGCTTC TGGAAATATC GTTGTGGAAG TGATTGCGAA 5040

```
AGATGGCGCA AAAGTTAAAT ACTCAGCCGT GGATCAATTA GGGCAAACTG TAACGACTTA     5100
TATGAATCGT CGTGGCTATA TTATGCGGGA TGCCTCTGTT GATTGGGCTC TAGGCGTCAT     5160
GAATGATGGC GATGTTGTGG CTGATTTTGA TTCTGATTTA GTTGGCGAAG GCGCTCATTC     5220
AGAAGTTAAA GTTGTTGCAA TTAGTGCAGG TAAACAAACA CAAGGCATCG ACACACGAGT     5280
GACAAACAAA GCGCCACATT CAATTGGtCA TATTTTACAA CATGGAGTTA TTCGTGAACG     5340
TGGTACTTTA ACGTTTAACG GAATTGGCCA CATTTTAAAA GGAGCAAAAG GAGCCGACGC     5400
ACAGCAAGAA AGTCGTGTTT TAATGCTGTC TGATAAAGCT CGTGGTGATG CCAATCCAAT     5460
TCTTTTAATT GATGAAAATG AAGTGACTGC CGGACACGCA GCCAGCGTTG GTCGAGTAGA     5520
TCCAGAAGAA ATGTATTATT TAATGAGCCG CGGCTTGCGT AAAGAAGAAG CGGAACGTTT     5580
GGTGATTCGT GGCTTCCTAG GTTCTGTCAT TACAGCAATT CCAGTTAAAG AAGTACAAAA     5640
TGAATTTGTA GAAGTAATTG AAGGGAAGTT AAACGCATGA TGGATGCAGC AACCATTCGT     5700
CAATCGTTTC CTATTTTATT TCAAGAAGTT AATGATGAAC CGCTAGTTTA TTTAGACAAT     5760
GCTGCGACAA CCCAAAAACC AACGGCAGTG TTAGATGTTT TACGACATTA TTATGAAACC     5820
GATAATGCCA ATGTTCATCG TGGGGTGCAC ACGTTAGCCG AACGAGCAAC GAAGGACTAT     5880
GAAGCCTCCC GCGAAAAGGT TCGCCAATTT ATCCATGCCA AAGAAACAGC AGAAGTTTTA     5940
TTTACACGAG GAACAACAAC CAGTTTGAAT TGGATTGCCA AAAGTTATGG TGACTTGGCA     6000
GTTACAGCCG GGGATGAAAT CGTCATATCT TACATGGAAC ATCATTCAAA CATTATTCCC     6060
TGGCAACAAT TAGCTCAACG TACAGGGGCC ATTTTGAAGT ATATAGACGT CACTGAAGAC     6120
GGCTTTTTAG ATATGGCAAG TGCACGTCAA CAAATTACAG AAAAAACGAA AATTGTTTCG     6180
ATTGCTCATG TTTCAAATGT GTTAGGGGTC ATTAATCCAA TCGAAGAATT AACACAATTA     6240
GCTCATCAAA ATGGCGCTGT GATGGTGGTA GACGGCGCGC AAGCAGTTCC TCATATGCCT     6300
GTGGACGTTC AAGCCATCGA TGCTGATTTT TATGCATTTA GTGGTCACAA AATGTGTGGA     6360
CCAACTGGAA TTGGTGTCCT TTATGGCAAA CGTCATTTAT TGGAACAAAT GGAACCTGTG     6420
GAATTTGGTG GCGAAATGAT TGACTTTGTT CATCTTCAAG AAAGTACTTG GAAAGAGCTT     6480
CCTTGGAAAT TTGAAGCTGG CACACCTAAT ATTGCCGGGG CAATTGCCTT AGGTGCCGCC     6540
ATTGATTATT TAACAGAAAT TGGTTTAGAG GCCATTCATC AACATGAAGC AGCCCTTGTT     6600
CACTATGTTT TGCCGAAGTT GCAAGCTATC GAGGGCTTGA CGATTTATGG TCCACAGGAT     6660
CCAAAAGACC ACACAGGTGT GATTGCTTTT AATATTGAAG GTTTACATCC ACATGATGTG     6720
GCTACGGCGT TAGATATGGA AGGGGTCGCT GTCAGAGCGG GTCACCATTG CGCGCAACCG     6780
TTATTAAACT ATTTGAGCGT GCCAGCAACC GCACGGGCAA GTTTTTATTT ATACAATACC     6840
AAAGAAGACG CAGATCGCTT AGTTGAAGCG ATTAAAGCGA CAAAGGAGTT TTTCCAGCAT     6900
GGCACTTTCT AAATTAGATA ACTTATATCG CCAAGTCATT TTGGACCATT CTAGCCACCC     6960
ACATCACCAC GGAACGCTAG ATGCGTCTAG CCAAACAATT GAACTGAATA ATCCTACTTG     7020
```

TGGAGATGTG ATTGAACTGG ATGTGGCCAT TGAAGACGGT GTGATTAAAG ATATCGCCTT 7080

TCAAGGAAGT GGCTGTTCCA TCAGTACAGC CAGTGCTAGT ATGATGACTG ATGCGGTCTT 7140

AGGAAAAACA ATCGCCGAAG CAACGGCCTT AGCAGAGGAT TTCTCTCAAC TTGTTCAAGG 7200

AAATGAAGTG GCCGAAGATG AAAAATTAGG CGATGCAGCA ATGTTAAGCG GCGTAGCTAA 7260

ATTTCCAGCT CGGATTAAAT GTGCAACCCT GGCTTGGAAA GCCTTaGAGC AGGCTGTTGC 7320

TAATAACGGC CAAGGCGAAg cCGGTcATTT ACATTGTGAA AAATAAAGAA AGGGTGACGA 7380

CGAATTGAGT ACAGTACCTG AATTAGAAGA ATATAAATTT GGCTTTCACG ACGATGTGGA 7440

ACCAGTCTTT AGTACCGGAG ACGGCCTAAC CGAAGACGTC GTTCGCGAAA TTTCTCGCGT 7500

GAAAGGCGAA CCAGAATGGA TGCTTGACTT CCGTTTAAAA TCGTTAGAAC AATTCAATAA 7560

AATGCCCATG CAAGAATGGG GTCCAGACCT GTCAGATATT GACTTTAGTA AAATTAAATA 7620

TTACCAAAAA CCGAGTGACA AACCCGCACG CGATTGGGAT GATGTACCTG ATAAAATCAA 7680

AGAAACCTTT GAAAAAATTG GGATTCCCGA AGCTGAGCGT GCTTATCTAG CTGGGGCTTC 7740

TGCTCAATAT GAATCAGAAG TTGTTTACCA CAACATGAAA GAAGAATTTG AAAAATTAGG 7800

CATCATTTTT ACAGATACTG ACTCTGCTTT AAAAGAATAT CCTGACTTAT TCAAAGAGTA 7860

TTTTTCAAAA CTTGTGCCAC CAACAGATAA TAAACTGGCC GCTTTAAACT CAGCTGTTTG 7920

GTCAGGTGGA ACCTTTATTT ATGTACCAAA AGGTGTTCGA GTTGATGTAC CATTACAAAC 7980

ATATTTCCGA ATCAACGCTG AAAATACCGG CCAATTTGAA CGGACCTTGA TTATTGTCGA 8040

TGAAGGGGCC AGCGTTCACT ATGTTGAAGG CTGTACGGCA CCCACGTATT CAAGTAACAG 8100

CTTGCATGCG GCCATCGTAG AAATTTTCAC TAGAAAAGAT GCCTATTGTC GTTATACCAC 8160

TATTCAAAAC TGGTCAGATA ACGTCTACAA CTTAGTAACC AAACGTGCTA AAGCTTACGA 8220

AGGTGCCACT GTCGAATGGA TTGATGGGAA CCTAGGAGCT AAAACAACCA TGAAATACCC 8280

AAGCGTTTAC TTAGATGGCA AAGGTGCCCG TGGCACGATG TTATCGATTG CCTTTGCTGG 8340

CGCTAACCAA ATCCAAGATA CTGGGGCCAA AATGATTCAC AATGCGCCGA ATACCTCTAG 8400

CTCAATTGTT TCAAAATCAA TTGCTAAAGA CGGCGGCGAA GTTAACTACC GTGGCCAAGT 8460

GACTTTTGGC AAAGACAGCG CAGGTTCGAT TTCCCACATC GAATGTGACA CGATTATCAT 8520

GGACGAAAAA TCAAAATCAG ATACGATTCC GTTTAACGAA ATTCATAACA GCCAAGTGTC 8580

CTTAGAACAC GAAGCCAAAG TTTCAAAAAT TTCTGAAGAA CAATTATATT ATCTAATGAG 8640

TCGCGGCTTG TCTGAAGCCG AAGCGACAGA AATGATTGTT ATGGGCTTTG TTGAACCATT 8700

TACGAAAGAA TTACCGATGG AATATGCGGT TGAGTTGAAT CGGTTGATTA GTTATGAGAT 8760

GGAAGGCAGT GTAGGTTAAA CCTGTTATAG GAGCTAGAAC CCTTGATATA TAAGGATTCT 8820

AGCTTTTTTA TTTTGCCTTG ATAGTAGAAA TGACCAGCTT TTTACTGATG AGCCATCATA 8880

TCAGCAAACT CTTGACCAGT TTTCTTAGGG GCGCTGTTTG CAATATCCGT ATAGATAGAT 8940

TACGTTACTT TTCTATTTAT GTTGATGAGT CTATTCCAAA TGATAAGCAA GCTATATCAG 9000

AAGCAATTGA ACGTCTGAAA GAATTAATCT ATCTATTCAG AAGAAGGTGG GAGAAGGGAT 9060

TACAGCAAGC TTAAAAGAGC CACTGGAAAC GGAGTACACT CTTACTCTAA TTGATGGACA 9120

mCCAACTCAA GCTAATCATA TCyTCGGTAG AaTGATACAa GTAATTGAAT GAATAGGATA 9180

GAGAACCCCA CCCCGTTGTT ATTTGACCGT ACTAAGGATa CGCTTTAAAT GAGCTACATG 9240

GCTCACATGG GGTATTCTTT TTTGACTTTT GCaAAAGGTT AATAAATTTT GATACaTGAA 9300

ACGAAaTGAA ACACTGCTTT ATAAACATTG ATATAATAGG GTTTCTGTGT TTCGTATGTA 9360

AATGAAACGA AAAATAGTTT TTAGTTGAAA ATACTTAAAA AGTAGTGCAT AGTTGTGGGA 9420

GTTATTGCAG ATGAAACAGA CCTACAAATA TTTCATGTGC AGCAAGTTAA AGAATACTTT 9480

TTTATGGTAT AATTGTAGAT GTCTAAATTC ATGTGGGGAG GAATTGTAAT GAGTAAAAAT 9540

GGTAAAGTGA TTTCAATTAT CAATATGAAG GGTGGAGTTG GTAAAACAGC GTTGAGCGTA 9600

GGAATATCAT CTTTTTTGTC AGAAAAAAAA GATGAAAAAG TACTTTTGAT TGATAGTGAC 9660

CCTCAATTCA ATGCAACTCA AGCGTTCATA GATCCAGAAG ATTATCTGAA AAGTGAAAAA 9720

ACCATTTTTA AGTTATTTAA GCCCCAGACA GAATTACATC AGTCTTTTGT AATGCCAAAA 9780

AGAGAAGAGC TGGTTACTAA AATAAACAAA AACTTAGATA TATTAATGGG TGATTTGAAT 9840

CTAGTATTAG TTAATAAAAG TAGCGATTCA GGTTTAATTA AACGTTTAAA ACGATTTATT 9900

ACAAAAmATA ACTTAAGGAA TTATTATGAT TATATTATAA TTGATTGTCC ACCAACTTTA 9960

ACGTTATACA CAGATAGTGC ACTAGTTAGT TCTGATTATT ATTTAATTCC TAACCGTATA 10020

GACAGATACT CAAATATAGG AATTAGCTCA CTGAATAGAG CTATTGGGGA TTTGATAGAT 10080

CAAGAAGACT TAGAATTAAA GTGTTTAGGA CTAATTTATA CGATGGTTCA AAATCATCTA 10140

TCAGATAAAC AAAAAGAAGT AAAAGCCGAA TTGGAATCTA ACAAAGCGAT GGAAAATATA 10200

TATGTGTTTA AATCTAGTAC TAGCTTAGTG AATGATATGC AAGTGGGAAA ACAGGGACCT 10260

ATTGCTACAA GGTATGCTAA ATCAAATAAA GATATAAGCG ATGTGGTAGA TGAACTGCAA 10320

GCATTACTTA GCAAAAATTC TTCAAAAGAA GGGGTGACTG TAGATGGATA ATAGATTTTT 10380

ATACAGAAGT AGTTTGA 10397

(2) INFORMATION FOR SEQ ID NO: 291:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4956 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

CAGTAAATAT TTTAGTTAAG AGATGAAACT TGCATGTGGA TTTGAAACAT GTAAGTTTTT 60

TATTCTGTTT TAATAGTTGA TTATTTTTTG TTTTATACGA TAATGAATCT CGTATGATTG 120

AAATTTTTTT AAAAAAGGTA TGTTTTATAT AGTGACAGAG AAGAGGATTA TAATGGGAAA 180

TCAAGAGAAC AAATTATACC ATTTTGTCGG AATTAAAGGT TCTGGCATGA GCTCATTAGC 240

GTTGGTCTTA CACCAAAAAG GCTACAACGT TCAAGGATCA GACGTAGAAG AGTATTTCTT 300

TACACAACGG GATCTTGAAA AATCAGGTGT CCCTATTTTA CCATTTAATG CAGATAATAT 360

TGATAAAGAC ATGATTGTTA TTGCTGGAAA TGCTTTTCCA GATACCCATG AGGAAATCGC 420

CCGCGCCATC GAATTAGGCG CAGAAGTAAT TCGTTACCAC GATTTCATTG CTCGTTTTAT 480

CGAACCGTAC ACAAGCATTG CTGTAACAGG GTCACATGGG AAAACAAGTA CGACTGGTTT 540

GCTAGCACAT GTATTAAGTG GTATCAATCC AACTAGTTAT TTAATAGGGG ATGGCACTGG 600

CCACGGCGAA CCAnATGCTG ATTTCTTTGC ATTTGAAGCG TGTGAGTATC GCCGTCATTT 660

CTTGGCTTAT TCACCAGATT ATGCGATTAT GACGAATATC GATTTTGATC ATCCAGATTA 720

CTACAAGAGC ATTGAGGACG TTTTTTCAGC GTTCCAAACa ATGGCTCATC nAGTCnAAAA 780

AGGAATTTTT GCTTATGGTG ATGATAAGTA TCTTCGCCAG TTAGAATCAG AAGTGCCaGT 840

TTATTATTAT GGCGTCAGCG AAGAGgAkGA TATCCArGCC CGATkGCTkG ATAATTCAkG 900

CCCTCTTCCA AAaTGAAATC AGTCATCATT TGGAGTGCTG GACCAACACC AAAAATAATC 960

CCTcCAGCTA AACTAAATAA aAcAAACAGT AAATTTAGTT TGaTTACTGT CCAGGAAAGG 1020

TAAAACAACC GTTGAATCCC TGTACTTTCC ATCTTCCTCT TCCTTTCTAT TATGTAAAAG 1080

GCTAAGCCAA AAGCGCTCAG CTAATGCCTC AGCCTCCTGT GATTATTTTT GAGATTGGAT 1140

ATATTCGTCT AATTGTGTTT GCATTTCTTT TTGAACTTTA TCCCAGCCAG CTGTTTTTAG 1200

GTCATCCATT AATTTTGGAA GTGTTTCTTC TGGATCAACA GTTCCTGTAT TTAAGCTTGC 1260

TGCGTAACGG TTCATAACTG TAGCAACGTT AGTGATTTCA GTTTTCACTT TATCATTTAC 1320

AAAAGTAAAA CCAAGAATTG GTgAATCTTT TGCTTCTTCG ATGCTCTTAT CACGTTCTTT 1380

AACCATTTCT TCAGTGACAG ATTCTTCTGG CCAAATGATT AAGTTGTTTC CTGTGTTCCA 1440

AGCACTCAAA TGAGTTGTTG GTGTGTAATC TTTCAACAAT TTCACACGAT CATCGCCAAC 1500

TTTTTCATAT TGTTTGCCTT CTTCACCATA AACAAGTCCG TTTAACAATT CTGGATTGCT 1560

GTTTAATAAA CCTAACAATT CAACAGATTT TTCTTTGTTT TTAGACGTGT TTGCAACAAC 1620

ATAGTTAGCC ATTTGCGCTT GAGCTGTTGT TTTTAATGGT TCTGTTAGTG GACGAGAAAC 1680

AAGTGGTTTG CCAGCAGCTT GTGTTAAGAT TGTATCACCA TAATCCATAG GTCCTTGTGT 1740

TTCTTGACGC ATAAACCAAG TATTGGTATT TAAGTCATAT GGTGTTGTAC TTGTAGCAGC 1800

ATCTGTTGGA ATCAAGCCAT CTTTATACCA TTGATGCAAG ACTTTTAAGT TATTAATCAT 1860

GTCTTTGTCG GCATATTGGT TAATAATTTT TGGTGAGCCA GTATCAGTTG TTTTTACTGC 1920

AAATGGATAT TGGTTACCAA TAGGGAAGTC ATAATTACCT GTTGCAAAGA ATGTTTGGCC 1980

GATAGCAAAA GCAGCAATAT TTGGkTCyTT TTTAwGGaAT TCTTTTAGAA CTTCCGTTGC 2040

ACTTTCATAA CTACCATCGA CTTTACTAAT ATCTAAATTG TATTTATCGA CATATTCTTT 2100

ATTAAAAGTT AAAACTTGTT GACCGTAAGA GTTACCTAAA ATTGGGAACG CATACAGTTT 2160

TCCATTAATC GTATTTCCTT TAATATAGTT ATCTGGCAAT TGATCATAGG CTTCTTTGGC 2220

ATATTTAGGT GCTAAATCAG TTAAATCAGC ATAGGCGCCT TTTTGTGCAT TCGTTGCATA 2280

ATTTTGTGCT AATGAAATAT CATAGCTTTC ACCAGAAGCA ACGATTGTTG ACATTTTTTG 2340

GTCCCAATCG CCCCAACCAA CAAATTCCAT TTTTAATTTT GCCCCAATTT TTTTCTCGAT 2400

AATTTTATTC GCATTATCGA TTAATTGGTC ATAATTATCT GGTTTGTCCC CAACACGATA 2460

CATTAATAAC GTTGGTGTAC TATCATCACC TTTTGACGCT GCATCTTTTG AAGATTTACC 2520

GCACGCGCTT AGTCCTACAG CTAAAGCTGT GCGCCAGCTA CGGCTAATCC TTTTTGCCAT 2580

TTTTTCATTC CTTATTTCCT CCTGTACAAT TTTATTCTTT CACGCCGCCG ATTGTTAAAC 2640

CACCCACGAA ATATTTTTGG AAGAATGGGT AAGTTAACGC AATTGGTAAC GTAGATAAGA 2700

CAACAATCGC CATTCGAGCC GATTCACTTG GTAGAACGCC TAAGCCTTCT TGTGCTGCTG 2760

CACCCATATC CGTACTTCTT GATAAGAACT CTAAGTTGTT TTGAATTTTC ATCAATAAAT 2820

ACTGTAATGG TACCAAGGTA TCTTTTTGAA TGTAAAGTAA CGCATTGAAC CAATCGTTCC 2880

AATAACCCAA AGCCGCAAAC AAACTAATTG TGGCAATCCC TGGAACGGCT AGTGGTAACA 2940

CAATGCTGAC AAAAATGCGC ATTTCACTTG CGCCATCAAT ACGTGCCGAT TCAATAATAC 3000

TGTCAGGCAC GGTTTTCTTG AAAAACGTTC GCATCACTAA AATATTAAAT GGTCCTAAAG 3060

CTAAAGGCAA AATTAaCGCC CAAATTGTAT CTTTCAATTG CAATAAGTTT GTCATTACTA 3120

AATAGTTCGC AACCATCCCT GGTGAAAATA ACATTGTAAT TAACGCAAAC AATGTAAAGA 3180

ATTTTTTAAA TGGAAAGTTG GAACGAGAAA TGACATAGGC ATATAATGAC GTCATTGTTG 3240

CATTGACAAT CGTCCCTGTC ACCGTAATGA AGACCGTTAC CCCTAAGGCA TGAATAATGC 3300

GTTGACTCAT TGCCCCTTGG AAAATGTAAC GATAGGCCTC CAATGAAAAT TCTTTTGGCC 3360

AAAACCGGTA GCCGTTTGCT GCTAATGATC CTTCACTTGT CAATGAAATC ATAATAACAA 3420

rGAAAAACGG AACCACACAG GAAATAGCAA AAATAGCAAT GACAATATTG AAAATCAAAT 3480

TCGCTGTCGG GTTGAAGGAT CGAACGTTAA CGCTAGTGAC TTTTTTCTTT TTCATAATCC 3540

TTCCTCCTAA AATAGTGCTG AATCTGGTTC AATTCGACGA ACAACTAAAT TCGCAATTAA 3600

TAACAAACAT GCCCCAACGA CAGATTGATA TAACGCTGCA GCTGCCGTCA TGCCTACATC 3660

ACCAGTAGCC GTTAAGCCAT TGTAAATATA GGTATCTAAC ACACTTGTCA CGTTGTAAAG 3720

CGCACCTGAG TTTCTTGGTA CGATGTAGAA CAAACCAAAG TCTGCTCTGA AGATACCGCC 3780

GATGTTTAAA ATTAATAAAA CAGACATCAT TGGTAAAATT TGTGGAATCG TAACATTTTT 3840

TATTTGTTGC CATTTGTTGG CACCATCCAC CATCGCTGCT TCATAGTATG TCGGATCAAT 3900

CCCCATCACA GAAGCATAGT AGATAATACT GTTGTAGCCT AGACTCTTCC AAACATTTAG 3960

GAAGACAAAA ATTACTGGCC ACCAACGTGG ATCTGCATAC CAGTTAATCA TTGTGCCGCC 4020

ATTACCAGTA ATCCATTGAT TCAAAATCCC TTTATCTGGA CTTAAAAAGG CATAAACAAA 4080

ATAGTTAATA ACCATCCAAG ATAAGAAATA TGGTAGCAAT GACATCGTGT GATAAACTTT 4140

AAcCAAGCGA CGATTTCTTA ATTCACTCAT AATAATAGCA AAAGCAATCG CGAAGAATAA 4200

GTTAAAGGCT AAAAAGACGA CATTGTACAA AATGGTGTTT CTAGTAATTA ACCAAGCATC 4260

TTTTGACGCA AATAAAAATT TAAAGTTTTC AAAACCAACC CAAGGACTTT CTCTTAGACT 4320

AGCTAAGAAA CCATCTGGTG AAATATGAAA ATCTTTAAAT GCGACCACAT TGGCTAAGAC 4380

TGGTATATAG AAGAAAAATA TAAACCAGAT AAATCCTGGT AACGCCATAA GGATTAAGGC 4440

CTTATAACGC CATACATTTT GAAAAAACCC GTTTTTCTTC TTCATTCGTT TCGCTCCTTT 4500

CATAAATATA ATAGCGTTTT CATAATATCT TTTTAAGGAA CAAATTATAG GATTTGAGAA 4560

ACAAATTATA GATTTTGTGT CATTTTAAGG TAAAAATTTA CCGAAATGAC TGTTCTTTGT 4620

GaAAACGTTT ATCTGATATT TTTTTTACAA CTGTTTAAGA ACCTTAAAAA CAATTGGCTA 4680

AACCTCACTC ATAAAGCGAT TTTGCTAGCG TGTATTTGTT TTATAAAATC TGAGAACTTG 4740

ATAGTTtCAC AAAGGATATC CTCTAATTAT TTTTGAGGAG GACTATTTTT GAATTTTATC 4800

CGTAAAAAAC ACTAGAGCAA TTCGTTTAAT TATAGTTTAT TTTTTTGCAG TTTTAATGCA 4860

AGCTGGTTCT TTATTATTAC TAGCTATTTC ACTACTATTT TTCTATTTGC CTTTATGCAC 4920

ACACAATATG TTGGATTTCT TGTTATAAAT CAAAAA 4956

(2) INFORMATION FOR SEQ ID NO: 292:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20561 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

TGAAGTAGAA GAAGAAAATT ATCCAGAAAA TAGCTCAGCA CTTGTTGCTC GCTTAGAATT 60

TCGAATTGTC TTTGACGAAT TTGTTTTATC AGGGGCGATT AGCCAAATTA ATCACATTAT 120

TGATCGTAAA ATTGAAAAAC AAGAAGATAT TTCGCAAGAA GAAGTCGATG AACTTGTGCG 180

TCCTTTATTT AGCATCGTCG AACGTTTATC TTATGAAGTA ACTGAAATTG CTCTGGATCG 240

TCCAGGTGTT CAATTGAATT TTCAACAATC AGAAGAAGCT TAAACTTTAA AAACAGCGTT 300

TATAACGGAT TCCGTTATAA ACGCTGTTTT TATTCTTGTG AAGATTTAGC AACCAAAGCT 360

TTTTTAAACT GTTCATAATT TTTTGTAGCG ATATCCACAT AAGCAATCGG CTGACCAACT 420

TTGGTGATTT TTTCGGGAAG ATAGGGAGCG AGCATTtCAT TGTACAAAAT GAACGCAGCT 480

TCTGGTACTT CCTCTGAAAG AACATCATAA CTGATAAAGT CmATTTCATA GGTTGGTTGA 540

AAAAAGGCTT TGATGTTTTG CTGCTCTTTA TTTGAAAAGA GCCaAGGaCC TCCATTTGcA 600

GGGCTCCCCa CAAAcAACGT CGCTCCaTCa CGGTTGACAG AAAGGATCCa ACAAATTTTT 660

TGCATACGCT TACACCTCCT AGAACTAAGT ATACCACGTC AGCTTAAAAG AAAGAGGACA 720

AAAAACAAAT TATTTTTACA GTGACTAGTA AAACAACGCT TGGTGATAGT TTTAAACTAT 780

CACAACGGTT ATTTTCTATC AAAAAGTAAG CAAATTATCC CACTATTTGC TATAAAATAA 840

GGTACAATAG AACTATCACT GGAAACAGTT TGCTGTTTTC AGGGTATTTA CTGGAGAGTG 900

AGTACCGGGA ACCTTTACTT CCTTTGATGT ACTTGTCATT AAAAACATAG GGAGGAACTA 960

AAATGACAAC ATTCATCACT GTATTAGTGA TCCTTGCATT TATTGCTGTG TTATTCGGTT 1020

TCTACCAAAT GCAAAAGAAA CATTTAAAAT TCTCAACACG TGTCTTTTCA GCATTGGGCG 1080

TGGGGATTGT TTTAGGTGCA ATTATTCAAT TTGCTTTTGG TACAGACAGT AAAATTACTA 1140

CGCAATCAAT GGAATGGATC GGGATTGTAG GTAATGGTTA TGTAGCGTTT CTACAAATGC 1200

TCGTAATTCC GTTAGTATTT GTATCTATTG TAGGTGCATT TACAAAAATG AAAGAATCCA 1260

ATAAATTAGG AAAAATTAGC TTTAATGTGT TAGCCACCTT ACTAGGAACA ACAGCAGTAG 1320

CTGCCTTGGT GGGGATTGGA ACAACCTTAG CGTTTGGTCT TCAAGGGGCA AAATTCACGC 1380

AAGGTGCTGC AGAAACGTCA AGAATTGCCG AGTTAGCGAC GCGTCAAGAT GCGATTCAAG 1440

ATTTGACGAT TCCCCAACAA ATCGTTTCAT TCATTCCTAA AAATGTCTTT GCTGACTTTG 1500

CGGGAACACG ACCAATGAGT ACCATTGGGG TGGTAATCTT TGCTGCGTTT GTTGGTGTCG 1560

CTTATTTAGG CGTTCGTCGT AAAGCGCCAA AAGAAGCTGA ATTTTTTGCA AATTTAATTG 1620

ATAGTTTATA TAAAATTACC ATGCGAATTG TTACCTTAGT TTTACGCTTG ACCCCTTATG 1680

GCGTGTTAGC GTTAATGATT AATGTTGTAG CAACAAGTGA TTTTGTTGCG ATTATTAACT 1740

TAGGTAAATT TGTGTTGGCA TCATATGTGG CCTTAATCAT TGTTTTTGCC ATTCACATGC 1800

TGATTTTAAT TACTTTAAAA GTCAATCCAG TTACGTATCT AAAAAAAGTT TTTCCAGTCT 1860

TGAGCTTTGC TTTTACTTCT CGTTCAAGTG CAGGAGCTTT ACCATTAAAT ATTGAAACAC 1920

AAACCAAAGC ATTAGGTGTC GATGATGCCA CAGCAAACTT TGCTGGTAGT TTCGGTTTAT 1980

CTATTGGACA AAATGGTTGT GCCGGCGTTT ATCCAGCTAT GTTAGCAACC ATCGTTGCAC 2040

CAACTGTTGG AATTGATGTC TTTAGTTTAC AATTTATTTT AATGTTAGTA GCAGTTGTCA 2100

CAATCAGTTC ATTTGGCGTA gcgGtGTCGG CGGCGGAgCA ACCTTTGCTT CTCTAATCGT 2160

TTTAGGCGCA ATGAACTTAC CAGTTGCAAT TGTTGGTTTA GTAATTTCAG TTGAACCACT 2220

TATTGACATG GCGCGGACAG CTGTTAACGT TAACGATAGT ATGGTTGCAG GCGTGTTAAC 2280

AAGTGCTCGT ATTCATGAAT TAGATCGTGA TGTCTTAAAT GATCGCGACG TTGTCTTAGA 2340

TGCAAATATC TAAATCAATA AAACGAAAAG TGTTTGAGCA ATGTCTTGCT CAAACACTTT 2400

TTTTGAATTT TAAGCGAAAG TGCAGTAATT ATTTAAAAGA ACGTATATAA TAAGTTGTTA 2460

ACTAAATTCA AAAGGACGTG AAGGAATGTT TTCAAACCAA GTAAAGATTA TGTTATATGT 2520

GAATAATGTT GAGGAATCTA GTCAATTTTG GCAAACGTTT GGCTTTGTTG AAAAGGAAAG 2580

AGAGGAAGTT GACGGCACAT TGGTGGTAGA AATTGCACCA AGTGAATCAG CAGAAGCGAT 2640

TATTGTTTTA TATGATTTGG CCTTTATTCA AAAACATTCA CCAGAAGTGG CGGGGAATAC 2700

CCCTTCTTTA ATGTTTGCTA GCGATGATAT CATTGGTCTC TATAAAAAAA TGCAAGAGGC 2760

TGGCGTTACT GTTGGTGAAA TGGTTCAATT GCCAACAGGC TTAGTTTTTA ATTTTGCAGA 2820

TAACGATGAT AATTACTTTG CTGTGATGGG GCAAGAAAGT AAATAAAAAA CAGCTTGGCA 2880

CCAGAATCAA TCTAGATTCC GGTGCCAAGC TGTTTTTTAT AAAAGTTGAA AGTTTTATAA 2940

ATTTTACTAA AATCTGTAAT TTTCATGAAA ATATAGCGTA TGATAATGCT GTTAATAGAA 3000

TAGGCGGAGG TAAGCGGTAT GCGTAAACGA CATGCAAAGA AAAGACATGG AGGAGTGAAT 3060

TGGCTTTTTA TAGTATGTTT GTTGGTGGTG ATTGGTGGTA GTGGTTATTT AATAAAAACG 3120

TTCTTTTTCA CTAGAGATTC ACAAGTTAGT CAAGAATCGA AAGTGGTCTT GGAAGAAGAT 3180

CGCCGAAGTG ATAATTATGC GAATTTAACG AAAGAAATAG TTGCACCAGA TAGTGGCGAA 3240

CTTGATCAAA AAATTCAAGA AACAAATTAT ATTGGTTCGG CTTTGATCAT TAAAGATGAT 3300

CAGGTTTTAG TAAATAAAGG ATATGGCTTT GCCAATTTTG AAAAGCAACA AGCCAACACG 3360

CCAAACACAA GGTTTCAGAT TGGCTCAATT CAAAAATCTT TTACCACAAC CTTGATCTTA 3420

AAAGCAATTG AAGAAGGTAA ACTTACATTA GATACAAAAC TCGCTACGTT TTATCCGCAA 3480

ATTCAAGGTG CTGAGGATAT TACGATTAGC GATATGTTGA ATATGACAAG TGGTTTAAAG 3540

TTATCAGCAA TGCCTAATAA TATCGTTACC GATGAAGAAA TTATTCAATT TGTTAAACAA 3600

AATACCATTC AAGTCAATAA AGGAAAATAC AATTATTCCC CAGTAAATTT TGTCCTTTTA 3660

GCAGGAATGT TAGAGAAAAT GTATCAACGT ACCTATCAAG AATTATTTAA TAATCTTTAT 3720

CACAAAACGG CTGGTTTAAA GAATTTTGGC TTCTATGAAA CCTTATTGGA ACAGCCCAAT 3780

AATTCAACAA GTTATAAATG GACAGAAGAT AATTCATATA ACCAAGTGCT CTCAATTCCT 3840

GCAGCTAGTT TTGCCCATGA ATTTGGGACT GGTAATGTGG ATATGACGAC AGGTGATTTG 3900

TATTGGTACT TACATCAATT AACGAGTGGA CATTTAGTTT CCACCGCACT TTTGCAAAAA 3960

TTATGGACGT CTTCTCAGCA AAGCTCTTAT CATGGCGGCA TCTATGTTCA TGATAATTAT 4020

TTACGTTTAC ACGGCGTTGA AGCGGGTCAA CAAGCCCTGG TTTTATTTTC AAAAGATATG 4080

AAGACAGGGG TCATATTGCT AACTAACTGT GTGAATCCAG CGAAATACAA AGAATTAATT 4140

GGTTCGTTGT TCCATGATGT AACCAATTTA ACTGTTAAAT TTTAAAGCAG AAGTAACAAC 4200

GATCACTCCC AAAGCCTTTT TAGGTTTCTT CTGAGAGACT TTATATTAAA ATAAAGGGAA 4260

GAACGATCAA AGGGAGTGAC GAAGATGAAA CAATCAAAAA ATAAACAAGT TGTTATTGGT 4320

GAACCAACGA AAAAAAACAT TTATCTTCAA GATATAGAAC AGGAAACGCA AGCTTTTATT 4380

CTCCAAAACA ATCCACAATT GACAGCTGAA TCGATGATTA CATTGGCGGA GTTACTTAAT 4440

TATCGTTTGG ATTACATGAA ACAATTGGTC GCAAATGATT CTGAAAAAAT CCAACAACTG 4500

AATGAATTGG TGGTTCAACA TATCAAAGAA GATAAGCTGA TTTCCAATAA TATGAATGAA 4560

AAGATGGAAA AGAAATTAAC CTTTGGCCAA CGAGCAGCGG ATACCATTGC TAAATTTGGC 4620

GGTAGCTGGG TGTTTATTGG CTTCTTTTGC TTCGTGTTGA TTGCTTGGAT CGTGATTAAC 4680

TCTACCACTT TATTGGGACA ACCATTTGAC AAATATCCTT ATATTTTGCT AAATCTCGCC 4740

TTATCTTGTT TAGCAGCTAT CCAAGCACCG ATTATTATGA TGAGTCAAAA CAGACAAGAA 4800

```
GCCCGTGACA GAGAACAAGC GAATAATGAT TATAAAATTA ACTTAAAAGC AGAAGTAGAA    4860
ATTAATTTAT TACATGAGAA AATGGATTAT ATTATCAATA ATCAACTAGA AAATTTGGTG    4920
AAAATCCAAA ATATTCAAAT TGAACTTTTA GGTGAATTGC AAGAACAACT CGCCGCGACT    4980
TCTGAAAAGG AGTAGTATAT AGAAGAGGTC GGGACAGAAC CGTTTAGCTC CGAGAACCAA    5040
GTAGGTACTG ATTAACATCT ATTTGCTTTG CAAATAGTGT TTCTCAGCAA TAAGAAGAAA    5100
TTTCCGAAAA TTGTCATAAA ACACAACGAG GAACGAGTTG ATGTTGCATG ATACCTACCC    5160
GGTGTTCTTT AATTTTTGGA GACCAAGTAG GTATTGTTCA TCATCGAATC ACTACGTTCT    5220
TCGTGATTCT CAACAATTTC GGCTTATTGT TTTCTCGCCA AAGGCTCGTC ACAGTTTTAT    5280
CCTTGTAAGT TCCGAAGGAC CTTCACCAAG TAGGTACGAT GATTTCATCT GTTCATGCTT    5340
CACAGTGAAA CATCGCAACT CTTAGAGCTG TAGCTCTTTT GAAGGCTGTA CTGTGACGAG    5400
ACGCATGTCG AGAAAACACC TTAGAAGTTG TTGAGATCGG AGCAACGCGT AGTGTTGGTT    5460
CTGTTCCCGC CGTTTATCAG GCTTTTTTTA TTGTGTGTAA ATACACAAGC AAAATAATCG    5520
CTTTCCCTTT TTTATTTACC GTATAATGGG ATAGAATGAC TTTATACAAA AGGGGAAAAC    5580
AAGATGAAAT ACCAAGTATT ACTTTATTAC AAATATACAA CAATTGAAGA TCCAGAAGCT    5640
TTTGCGAAAG AGCATCTAGC TTTTTGCAAA TCATTAAACT TAAAAGGCCG TATTTTAGTA    5700
GCGACAGAAG GGATTAACGG AACGTTATCT GGTACTGTCG AAGAAACAGA AAAGTATATG    5760
GAAGCAATGC AAGCAGATGA GCGCTTTAAG GATACATTCT TTAAAATTGA TCCAGCAGAA    5820
GAAATGGCCT TCCGCAAAAT GTTTGTTCGC CCACGTTCTG AATTAGTGGC GTTGAACTTA    5880
GAAGAAGACG TTGATCCATT AGAAACGACG GGGAAATATT TGGAACCTGC AGAATTTAAA    5940
GAAGCCTTAT TAGACGAAGA CACTGTTGTA ATCGATGCTC GTAACGATTA TGAATATGAT    6000
TTAGGTCATT TCCGTGGTGC CGTGCGCCCA GATATCCGTA GCTTCCGTGA ATTACCACAA    6060
TGGATTCGCG AGAACAAAGA AAAATTTATG GATAAAAAAA TTGTTACCTA TTGTACTGGC    6120
GGGATTCGCT GTGAAAAATT TTCTGGCTGG TTATTAAAAG AAGGATTTGA AGATGTTGCT    6180
CAATTGCATG GTGGTATCGC CAACTATGGA AAAAATCCAG AAACACGTGG CGAACTTTGG    6240
GACGGCAAAA TGTATGTCTT TGATGACCGA ATCAGTGTCG AAATTAATCA TGTTGATAAA    6300
AAAGTTATTG GGAAAGACTG GTTTGATGGG ACACCTTGCG AACGCTACAT TAACTGTGCA    6360
AACCCAGAAT GTAATCGTCA AATCTTAACT TCAGAAGAAA ATGAACATAA ACATTTAGGT    6420
GGCTGCTCAT TAGAATGTAG CCAGCATCCT GCCAACCGTT ATGTAAAAAA ACATAATTTA    6480
ACAGAAGCAG AAGTTGCTGA ACGTTTAGCT TTGTTAGAAG CGGTTGAAGT ATAAAAATAA    6540
TAAAAAAATA AACAAAAAAA CTTCGAAAAT TATTTTTCGA AGTTTTTTTG TTTATTTTTT    6600
TGTCTTTAAA AGTGTTGATT TTAAAGCTTT CTTAGCTAAA ACCATTTTAT TTTTTAAAAG    6660
CGTCTGAAAA AAATAAAAAA ATCTAATAAA TATTTCTTTT ATATAATGTT TATTGGCTGG    6720
GCCAGTCAAT GTTGAAAATG GGGAAGGAGG AATTCAGATG AAAATCATAA AAAGGTTTAG    6780
```

TTTGGTATGT TTAGGGCTAT TGATCATTGG GttGCmAACA AAAAgCGkTA TGGCTGAAGA 6840

AAATAATTAT GAATCAAATG GTCAAGCGAG CTTCTATGGT ACCTACGTTT ATGAGAATGA 6900

AAAAGAGTCA AATGACGTAG CGTATACCCA ACAATCAGAA GAACAGGGAA GAAACAATTT 6960

AGCTGCTTCT GGACAAGCAG TTTTACCTAA AACAGGCGAG TCTGAAAATC CGCTGTATTC 7020

CTTGATAGGA GTTAGTTTGT TGGGGATAGT CATTTATTTA ATTAATAAAA TGAAACGAGA 7080

GAAGGAGTTT ATTTAATGAA AAAAACTGTT TTACTTACAT CGGCAACCAT ACTTTTAGGC 7140

GTAGGGTTTC AAGCAATAAA CGCATCAGCA GCAGAAACTT TTCCGAAAGA ATACAATACA 7200

GAAGGCACAA TTACTTTTGA AGCAGGTGAT AATGAAATCA CACCGCCTGT TGATCCAGAA 7260

AATCCCGATC CTACTGATCC AGTAGACCCA ATTGATCCTC CAGGGCCAGG AACGGGTGGT 7320

GCCTTATCAA TTGATTATGG GTCAAAATTT AAATTTGGtA CACAAAAGAT TTCTACGGCA 7380

GATCAAACGT ATTATGcTGC GCCAGATGAA ATGAAAGATG GTTCAAAAAA ACCAACCTAT 7440

GTACAAGTAA CTGATAAAAG AGGaACGTTA GCAGGATGGA AACTAACGCT TTCCCAACCT 7500

GAACAATTTA AaACAGCTAC TGGAGAAGAA CTAGTCGGCG CTCAATTGAC GTTTACAAAG 7560

GCAGAAGCTG CCTCAATGGT GGATGAAAAA TATAAACCTA CTGAAGTCAG TAGTACTATT 7620

TCCTTAACAC CTGGTGTAAA TAATAATCTA GCGATGAACG CTAAAAAAGA AACAGGTGTT 7680

GGGACTTGGG TTTATCGCTT TGGTTCAAAT GAAGCAACAA ATAAAGAAGC TGTCCAACTA 7740

TTCGTTCCCG GAAAATCTGT AAAGCTTGCA CAACAATATT CAACGAAGTT AGTTTGGGCA 7800

TTAGAAGATA CTCCCGCTAA TAATTAAATA AAGGAGAAAA GCAATGAAAA AACAATTCAC 7860

AGGAATCGTT CTATGCAGTA CATTTGCAGT TGCTGGTTGG TCCTTACAAG CGAAGGCAGA 7920

GGAGCCAAAA GAGTATCGTT CGAATGGTTT AGTAGAATTT ATTCCGAACG TAGATCCGAC 7980

AGAGCCGGTT GATCCAGAAA ATCCCGATCC AGAAAAGCCA GTCAAACCAA TTGATCCGAC 8040

AGACCCAGAA GGTCCCAATC CTGGTACCCA AGGACCGCTG TCTATTGATT ATGCTTCAAG 8100

TTTTGACTTT GGGAAAAATC GTATTTCGAA TAAAGACCAA GTATATTTTG cAAGAGCACA 8160

GCAATATCAA GAAAATCAAA AAGAAACACC AAATTTTGTG CAAATTTCTG ATAATAGAGG 8220

AACCAATAGT GGTTGGTCTT TAACGGTTAC GCAAAAGGAG CAATTTAAAG CAACAAAGGC 8280

GACTTTAAAT AGTCAGTTGA CAGGTGCGCA AATTTCTTTA GCTAACCCGA CCGTTAATAG 8340

TAATGCGCAA AATGTTGTGA AACCAGAGGC AACAAATAAG ATTGCCCTTG TTCCAGGAAC 8400

CGCATCGCTA GTGGCAGCTG CGArCAAGGA ACCGGTGCAG GAACCTGGGC TACTTATTGG 8460

GGCAAAGTAG AAGTTGTTGC AGAACGCGAT GAAACCAATA CGGTACACAA TGTTAATGTT 8520

ACAAAAGATG TTGCGCTATC AGTACCAGGG TCAACCCCTA AAGATGCCGT GAAATATCAA 8580

ACAAAATTAT TGTGGACGTT GACGGATGTA CCAGGAATTT AAAAAGTAGA GGTGAAGAGA 8640

GATGCGAAAC AAACAAAGCC TAGTCTTTTT GGCAAGTAGT TTATTCTTAC TAATCCTTGC 8700

GCCTATTACT AGCGAaGCGA TTAGTAATCC TTCGACAAAT ATCAACGAAG TCAATCGGGA 8760

```
AATAGTGGAG GAATCGGCCA ATGATCAGGC AACTAGCGCT CCTTCTCACT CGGAAACATT    8820
AGAGCAAGAA CCTTCCCTAG AAACGACAAC AACAACCGAT AATCATCCAG CAGAAACAGA    8880
CCAATCAACG GCTACTTTAG AACCTAGCGA ACAACAAAAG GATGACGGTA CCAGTGCAAC    8940
GGAATCGGTG ATAGTTGAAC CACAGCCGCG CATGTTGgCT GCGCCATACG CTACTTCATT    9000
GCCGGATGAT CCTAACATCA TTCCAATCGA CAAGGTGTTT CAAGAGCCAA TTGGAAATGC    9060
AACAAGTATT TTAGAAGGTG GGAAGTTATT ACAACTAAAC CCAGCTGTTA AATCACAAAA    9120
AGGCGCTATT TGGTCAAAAA AACCAATCAG TTTATTGTCG GATTTTACCT TTAAAAGTTA    9180
TTTGTATTTA GGAAACGAGT ATGCAAATGC TGGTGATGGT ATGACCTTTA CCTTAACGAA    9240
TGATCCTAGA ATGTCCACAA CGCCTCAAGA AGTCATTGGT TCACCTGGGA TGGGAATCGG    9300
TGCATACTCC ACGAAAGCTG GTCAACCTTA TGTTAGAAAT GCCTTATCGA TTGAGTTTGA    9360
TACCTATAAA AATACTGGTA GTTCCAACCG AATGGACCGA GAAATTTCAC AAGATAAGGG    9420
AAATGGACAT TTAGCTTTTG TCACTCCGAA AGCTAATAAC AACAACTATA CAGGGGAGCA    9480
TTCAGGCGTC ACTGTTGCGC CAACCTATCT ATCTAATGGA ACCTGGCGCA TGTTGACTGT    9540
TCATTGGAAT GCCGCAACCA AAGCATTAAC ATATGATTTA GAAGGTGTGG GCACGAATAC    9600
GTATGTAGTT TCAGATTTAA ATGCACAATT TGGCGCAACA ACGGTTTATT GGGGCTTTAC    9660
ATCATCCACA GGTGGCAAAT ACCAAGAGAA TGCTTTAGCA ATGACACAAA TTCCTACCAA    9720
CGTTACGTCA CAAGCAGCCC TTAGCGTGAA TGGTCAGGAA TTTAGCTCAG CTGTTGAAGC    9780
AGTGAAAAAT GATCAGGTGA GACTACGGAA TACATTGAAT ATTGACAATG ACTTTATCGA    9840
AGATCGTCAA CCGCAAGTAA GCATTGACTT ACCTGATGAA TTAGCATATG AAGAAAATTC    9900
ATTAACGATT GATGGTAAAA AGGTAGCAGC CAAAGATCTG ACACAAACAG GGAATCACTT    9960
AACAATTGCT TTAAATGACT ACCTAGTTTT AAAGAAGGAC ATGATGATTG AATTAAAAAC   10020
AACCCTTCAA GATAATACGC CCGAAAAAGT TTTAACTATG AACTTTGACT ATTATGAAGA   10080
AGGAACCTTA CTGCAAAAGT CAAATAATGT AACGATTACT ATTCCTAAAC CAACAGAAAA   10140
AACAGTAACC GTTTTTTATA AAGACGCAGA AGGTAAAGAT ATTGCTCCAC CTAAAAGTGT   10200
AACAGGGAAA ATTGGTGCAC CCTATCAAGA AAAACCCCTA GATATTTCTG GCTATGTTTT   10260
TACAAAAGAC TCGGGCAATG CGGAAGGAAC AATGACAGAG GAAACAAAAG ATATTTATTT   10320
TTACTATCGT TTAGGGGAAC TATATTTTAA AGAGGCACCA AAACAGATTA CTTTTGGTAC   10380
GGAAAAGATA CGTAATCAAC CACTAATTAA ACTTGGGCAT CCCACCGAAG gACTCAAAGT   10440
CGTGGATGAA CGTAACGCAA ATAATTGGCG CCTGCAACTA AAGCAAACGC AACCGTTAAC   10500
AAATGATTCA ATTGTAATGC CAGATGTTTT TTCTTTCGTT TCAACTGCTG GAAGCTCTCA   10560
GATTACAAAT GAGGCAATTA CGCTTTTAGA AAGTAATCAG AAAGGTGAAA CAGATTTAAC   10620
AGCATTACTT GATGAATCAA AACAACAAGG AATTCAAATC AATGTTCCTG TAGCTTATCA   10680
AAGAGTTGGA ACATTTAAAG CTCGTTTATC TTGGGCGTTA GAAGATGTGC CAGGAAATTa   10740
```

AAGAAAAGAA TGGAGAAAAG aAAATGaAAA AAGTAACAAC AGTATGTTTA TCCACATTTA 10800

TTTTAGGGAG CTTAGGCATG ACTTCTCAAG CCTTTGCAGA AGATGGCGGT AGTTATACAT 10860

CTAATGGTAT CGTAGAATTT ACACCAAGTG AAGAGCCAAC TAATCCGGTT GACCCAACGG 10920

ATCCAACTAA TCCAGTTGAT CCTATTGATC CAACAGATCC TGAAGGTCCT AATCCAGGAA 10980

CAAACGGTCC GTTGTCAATT GATTATGCGT CAAGTCTAGA CTTCGGTGTT CAAAAAATCA 11040

CGTCGAAGGA TCAGACTTAT TTTGCTGCTA CCCAAAAATA CAAGACGGTA GGCGCAACCG 11100

ATGAAGTCAA AGAAGGGCCT AACTATGTCC AAGTAACAGA TAACCGTGGA ACCGAAGCTG 11160

GCTGGTCATT AAAAGTGAAA CAAGAAGGAC AATTTAAATC AACTTCTGGT AAAGAATTGA 11220

CTGGGGCAGC TATCACGTTT AAAAACGGAA ATGTAGTGAC TGCTTCAGAT TCAGGCAAAC 11280

CAACAGGCCC AGCCACAATC ACGCTAAATT CAGATGGTAG TCAATCAGAT GTAATGAGTG 11340

CAGCGAAAGG AAACGGTGCA GGAACTTATT TATTTGATTG GGGAACAGAT GCAACTACAG 11400

CCGCTAAAAG TATTGAGTTA ACTGTTCCAG GTTCTACAAC CAAATACGCT GAAAAATACG 11460

CCACAAAATT AACATGGACG TTGACAGACG CTCCAGGAAA TTAAACGAGT TTTAAAAAGA 11520

AGAATAGCGG GGAACATAGT TACCTGCTAT TCACTACATA TAGAAAAGGA GTAAATAATT 11580

TGAACAGAAA AAAACAACGA TTTCTTTTTT TTCCGATACT ACTAATTCTT TTAGCGTTAG 11640

TACCTAGCAA TGGTTTTGCC AGTGAATTTA ATTTTGCGGT TACGCCTATT CCATCAGAAA 11700

AACAAGTAGA TAAAGAAAAA ACTTACTTTG ATTTATTACT AGCACCAAAT GAAGAAACTG 11760

AGTTGAAAGT GAACTTGCGC AATGATACAG ATAAAGAAGT AAAGGTAGGT ATTTCCATCA 11820

ATAGTGCCAT TACCAATTCA AATGTAATCG TGGAGTATGG TGAAAATAAG GGGAAAAAAG 11880

ATCAAAGTTT AGCGTTTGAT ATAAAAGACT ACGTTCAATA CCCTAGCTCA GTTCGTTTAA 11940

AACCGAAAAG TGAACAAACG GTTTCTATTA ACGTCAAAAT GCCTAATACA CCCTTTGATG 12000

GAGTGTTAGC TAGTGGTATC ACTTTTAAAG AAGAAACGTC GGATGAAGGC AAACGCCAAG 12060

ACGATAAGAG TCAGGGGTTA TCCATTAAAA ATGAATATTC GTATGTCGTT GCATTATTGA 12120

TGCAGCAAAA TAAAAAAAAG GTTGAGCCGA ATTTACTTTT GAAAAAAGTC TCACCAGGAC 12180

AGATTAATGC TAGAAATGTG ATTTTAGTCA ATCTACAAAA TGATCAAAAA ACATATATTA 12240

ACCAAGTTGC ATTTTCGGCA GAGATTACCA AAAAGGGACA TGAAGAAGTT TTATATAAAG 12300

AGGAAAAAGC AAACATGCAA ATTGCACCAA ATACCAATTT TTCTGTTCCC ATTGCGCTAA 12360

AAGGGCAACC GCTTAAGCCA GGAGATTACC ATTTGTCTAT GACCGTTGTT GGGAATAAAG 12420

ATGCTGCAGG CTCATTTAAA AAATCAATAA ACAATGAATC GATATCTTTT AGAAATCAAT 12480

GGCAATTTGA AAAAGACTTT ACAATTAATG GTGAAGTTGC CAAAGAATTA AATGAAAAAG 12540

ATGTCACCTT AAAAGAAAAT CATTCGAATT TATATTTGAT GATTGGCTTA CTACTATTAT 12600

TGATAGTGAT TTTAATCATT GCCTGGTTGA TTTGGCGAAA AAAGAAACAG GAAAAAAACG 12660

AAAGAGAAAT ATAAAGAGTA GTTAATTTTT GATTTGATCA TCGGCAACGT TTTTGTTTGT 12720

```
TTTTGCTTAT CAAACATGCA GTTTATTCAA CAAACCTAAC TCAAAAAAAG AAATAATAAG    12780
AAGGAGTATT GTTAAATATC GTCTGATACC ACTTTACTAG TTATATTTTT TGTTTTGTTA    12840
TGTAGCTAAA GTATTTGGCA ATATATGTGA GAAAAATGGA TAAATACAAA CATTTACTAT    12900
TATGGAGTAC ATTGTTACTC TCGTTTTTTC AAGTAACCAA TGTGCAAGCA GCAGAAGAAC    12960
GGAATCCAAT TTATAATCCG ACCAATCCAA CAGAAGAAAT TTATCCGATT GAGAATACTG    13020
GCACACCAGA AAAAATGGAA GAAACAAATG AAAGTACTAT GACTAGCGAA ACGCAAGCGG    13080
AAGAAAGTAG CGAAGACGTT AGTAAAAAAG AAGTGACAAC CAATAATCCA AAAACAAATG    13140
AAGAAAGGCC TTCTTCCAAG AAAAGTCAGC ATCAAGTATT TTTCACGGAT GAATTATTTG    13200
ATGCTAGCCA TATCACCATT GTACCAAATA GCACAGGCGG TGCTGGAACG GAGGGGCAAA    13260
ACACTTTTTA CTTGAATTTC ATACGGGAAT TAGTAGGAAC GGCTATTTAT GGGAGAAAGA    13320
ACTGATAAAT GATGAATTAT GAAATTTTTT TAAATAAAGA AGATCAACGA AGTTATTCCT    13380
TGCTAAGGCA TTTAGAAGAA AGTCCCACGC TATCAGGTAC TTTTATTGCA CTTCGCGAAG    13440
AATTATCAAT GTCGAATTTT TTAGTCAAAA AAACATTAGA AAAATTAAAA GCAGATATTG    13500
ACAATCTAAA GTTAAGCGAC AGCTTAACTT TAGTCGTTTC TGATGTAGAT GTTACACTTG    13560
AAATTGATGG AAACTATTCA AGTAAACTTT TGCTCACGAA ATACTTAACA GAATCGCTCT    13620
CTTTTAAATT AGTGTTGTCT TACTTTAGAG GAAATTATAG TTTGACAAAG TTTGCTGAAT    13680
GTAATCATGT CAGCACCTCT GTGGCGTATA GTACCTACAG CGGCTAAAAA AAGCCCTTAA    13740
AGGAGTATCC AGATCTATTT TAGTAAaCGG GAAATTGTCG GCAATCAAAA AGCAATTTCA    13800
TTTTTTTTGT ATAAATTGTT CACGCTTTCC AATCAACCAA TTTCCGAATT ATATTCGGTA    13860
AAAGTCTATA ATGAAGCGAA ACGTGTATTA CAGTCTGTTG AGTTAAACTA TACTTTCACC    13920
ACCTATGAGC GACGCAATTT TTTCCATTAT TTAGCTATTA TGATTAACAA CGAAGGACGT    13980
GCGGTGGAAG GGATCGATAC GCGTGCACTG AATACATTTT CAGAGGAACT AATAAAAAAA    14040
TCGCAAGTGG CATGGGCATT GGCTTCAAAa TCATTAACGT ATACGATTGt TTTTTTCTTG    14100
TATTTGCATG GCAAACTAGA GAAAAAATAT GTCATTCATG AAGATCCTAC GATTGAAAGC    14160
CTAACTCGTG TTTTCATTGG AAGTTTTGAA AAAGCATTTA ATTGTTTAGA GGAGTCAACA    14220
AGGAATACTC TAGAAGAAGG GTTAGCAATT ATTCATTTCA ATGTTATTTA CTTTCCGATT    14280
AATATGTTTG ACGATTTTGA AATGGATTTA CAGTTTTTTA AGCAAACATA TCCGGAATTT    14340
TATTTTTATT TAATTGAATA TATTCGTTGG TTAACAGCAA AACATAAAAA AATTGCGAAA    14400
GCCAATCACT GTTTGTTTTT TAACTATCTA TTGTTGCTAA TTAATCATGT TCCGGTTCAT    14460
TTAATTGCAG AGCCAGCCAA GGTCCTAATT GATTTTTCTT ATGGAAAAGA GTATAACCAA    14520
TTCATCAAAA aAAATTTATC CGTCTATGTG AATTTAAATG TAGAAATTAT TGACCCGTTA    14580
TCAGATACCT TACCAGATGT GGTGATTACC AATTTAAATA ACTTGTATCA AGAAGAACAG    14640
AGTAAAGTTA TGGTTTGGCT GGATCCACCT CGTTCGATTG ATTGGGTCAA TTTAACGCAA    14700
```

TCATTATTGA CTATTCAAGA AGAAAAATAT CaACAACAAA AAGAATCGAC AAAGACAAGC 14760

GGTGATCCTA TCGAATAGnA TCATCGCTTG TCTTTGTGTT TAAGTCTTTA CAAAACCACC 14820

CAAGTGTTTT ATAATATGAA TATATATTCA TATGAATAGA AAAACATTAA TGAGGAGGCG 14880

CATAACAATG TCAAAACAAA CAGGTGCACA CACACATGAA GAAAAAGGTA AAAACGTTCC 14940

AGTTATTCTC TTTTTTACAG GATTGGCTCT TTTCTTTATC GGTTTATTTT TAGGGAACAT 15000

GCTCTTAGTG AAAAATATTC TTTTCTCATT AGCTGCTATT TTAGCTGGTT ATCATATTAT 15060

TGGGGAAGGC TTTGGAGATA CGTATCGGGA TACGAAAAAC AATCGAAAGT TTTCGCCTAA 15120

TATTCATTTG TTAATGACCT TAGCTGCAGT TGGTTCAGCA TTAATGGGAA GTTTTGAAGA 15180

ATCAGCATTG TTGATTTTGA TTTTTGCGGC GGCGCATTTC CTTGAAGATT ATGCCCAAGG 15240

CAAAAGTCAA CGTGAAATTA CAAAACTATT AAATTTAAAT CCAACAGAAG CTCGCCTGAT 15300

AACAGACGAC GGCTCAATTC AAACTGTGTC AGTTGAACAA TTGAAGATTG GTGACCGCGT 15360

GCAAGTCCTC AATGGTGCGC AAATTCCGAC AGATGGCGTC GTGATTGAAG GAAGTACCGC 15420

TGTGGATGAA TCATCTATTA ATGGGGAAAG TATTCCCAAA GAAAAAAATA GTGGGGACCC 15480

AGTTTTTGGT AGTACCATGA ACGGTTCAGG CACAATTGTT GTTGAAGTAA CGAAAGATAG 15540

TTCAGAAACA GTTTTTGCAA AAATCGTACA GCTCGTGAAC CAATCACAAG AAAATCAATC 15600

AGAAATTGCA AGCAAAATTA AACGATTTGA ACCGAAATAT GTCACACTAG TTTTAGCAGT 15660

TTTTCCACTA ATTGTTTTAG GTGGCGCATT GCTTTTCCAA TTGACTTGGG CTGAAAGTTT 15720

TTACAGAGGC TTGGTCTTCT TAATTGCTGC CTCCCCTTGT GCATTAGCAG CGAGTGCCGT 15780

TCCGGCAACT TTATCAGGAA TTTCCAATTT AGCAAAACAA GGCGTTTTGT TTAAAGGTGG 15840

CTCATTCTTA TCAAATTTAG CTGAAGTGAA AGCACTCGCT TTTGATAAAA CAGGCACGCT 15900

AACAAAAGGC AAACCTGAGG TCACAGATTA TCTTTTTGTT GATGGCTTAG AAGATCGCCA 15960

AGAAGAATTA GTAGCTGTTT TAACAAATAT GGAGAAAAAA TCCAATCATC CATTGGCGAC 16020

GGCGATTGTT AATCGCTTTG AAGCAGAAAC AACGGCTTTA AATTTAGAAG TAGAAAATAT 16080

TGTCGGGGTG GGTTTAGTAA CGACCATTGC AGAAACAACT TTTAGAATTG GTAAACCATC 16140

TAGTTTCGAA CAAGTGCCAA CAATCATTGA GAAACAAACA ACGAAATTAG CAAGTGAAGG 16200

CAAAACGGTA GTGTACTTTG CCGAAAATGA ACAAGTGATT GGTTTGGTGG CATTGATGGA 16260

TGTTCCTAAT GAAGAAGCAA TGAACGCCAT TCACTATTTC AAGTCACAAA ATATTGAAAC 16320

AACAATGATT ACAGGAGATG CAAAACTTAC GGGTGAAGCA GTTGGTCGCT TAGTAGGCGT 16380

GGACCAAGTT TTTGCTAATG TTTTACCTGA GGAAAAATCA GCGATTGTTG ATCAATTAAA 16440

ACGCGAAgTA GGAATGACAG GGATGGTTGG TGATGGCATT AATGATGCAC CGGCTTTAGT 16500

GAATGCTGAT ATCGGAGTGG CGATGGGGGA TGGTACAGAC ATTGCTATCG ATGTTGCAGA 16560

CGTGGTAGTG ATGAAAAATG ATCTTTCTAA ATTAGGCTAT GCTCACCGTG TCTCAAAACG 16620

CTTGAATAAA ATTGTGCAGC AAAATATTAT TTTTTCAATG TTAGTGGTAG CAACGTTAAT 16680

```
TATCTTGAAC TTTTTAGGTA TTGCGAATAT CGCATTTAGC GTACTTATTC ATGAAGGAAG    16740
TACCTTAGTG GTCATTTTTA ATGGCTTGCG TTTATTAGTT AATACAAAAT AAGTCAAATA    16800
GTCACTTGCA AGAAAGGAGA GAAAGCGATG AATTCTTTAA AAATGATTGA ACTTTCTCTC    16860
ACCGACGTTG TCATTCGGAT GGCTCTAGCC GTGGTTTTTG GAGCTTTGAT TGGGTTAGAA    16920
AGAAGTATTA AAAGAAAAGG CTTCGGTATC AGTAGTAATG CCATTTTATG TTTGGCCTCC    16980
TGTACGATTT CCATTTTACA GATTCAATCG GTGGATATTT TAGTCGATGT GGTGAAACAA    17040
AATAGCGCCT TGGCGTCCAT TATTTCTATG GATATTACAC GATATGGCGC CCAAGTGATT    17100
AGTGGTGTAG GTTTTCTCGG TGCGGGGATT ATCGTTTTTA GAGAAAGGAA AGTCTCTGGG    17160
TTAACGACAG CAGTAATGAT GTGGAATGTT ACTATTATTG GTTTAGTCAT TGGGATGGGC    17220
TTTTTAACCA TTGCGTTTAT CAATCTCGTG GCAACTTTGC TAGTATTAGC TCTGGCCCAT    17280
TTTAAACGAA AGACTCCTTT AAAACGATTG CAGCTTATCG TAAAAATGAA GGAACCTATG    17340
ATTAAAGAAA GCCGACTGCA TGAACAACTT CAAGCTTCTT TAGAAAAAGG GGAGCGCTTA    17400
GTAGAATTGA CTTTGGAATC CGAACAAGAG GCAGTGACAG CAAAGATTTC TTATGTAGGC    17460
AAAGAAGATT TCTGGCATTC CAACTTGCAC CATCTCTTAA AAGAAAAAGG GCACGTTCAA    17520
TCTGTCTCTG TCACGGAAAC AATTTAACCA TAAAATAAAG CACTTGGGcg CTAGCGCGCG    17580
CCAAGTGCTT TTGTTTAAAT ATTTAATACT TTATCTAAGA AATCTTTTGT TCGAGGGTTT    17640
TGAGGATTTT CAAAGACTTG CTCTGGCGAA CCATCTTCTA GGAAGTTACC GCCATCAATG    17700
AACATTACTC GATTAGCCAC TTCTTTTGCA AAGCCCATTT CGTGAGTAAC GATGACCATT    17760
GTCATTCCTT GTTTTGCTAA ATCTTTCATT ACACCAAGAA CATCGCCAAC CATTTCTGGA    17820
TCAAGGGCTG ATGTTGGCTC ATCGAAAAGC ATAATATCAG GATTCATCGc tAACGCACGT    17880
GCAATTGCGA CACGTTGTTT TTGACCACCA GATAAAGAAT CTGGATAGGC ATCTTTTTTA    17940
TCTGCCAACC CAACACGTTC TAAAAGTTCT AACGCTTTTT TCTCCGCTTC GTCTTTkGAT    18000
AAGCGACCTA AATCGGTTGG TGCTAAAACA ATATTTyCTA AAATgATAAA TGTGGAAACA    18060
AATTGAAATG TTGGAAAACC ATGCCGATAT GTTGACGGAC TTGGTTGATG TCCGTATTTT    18120
TATCTGTCAA GTTCGCACCA TCAATAATAA TGTCTCCGCT AGAAGCTTCT TCTAATTGAT    18180
TCAAGCAACG TAAAAAAGTA CTTTTACCAG AACCAGAAGG ACCGATGACG CAGaCTACGT    18240
CGCCTTCGTT AATAGAAACA TTAATGTCGT TTAAGACCGT GTTGTCACCA TATTTTTTTA    18300
CTAAATGTTC AACTAAGATT TTTTCAGTCA TATTAGTTCA CCTTCTTTTC AAGGTTTTTC    18360
GCTAATTTAG TTAAGCTAGT AATTAAAATT AAGTAGATTA ACGCAATAAC GAAtAAACCA    18420
TTGTACTTTG TAATGTCCGA GCAACAATAA TTTTTCCTGT TTGTAGTAGT TCGATTAAGC    18480
CAATTGCTGA AAGGATAGTT GTATCTTTCA ATGAAATAAC AAATTGGTTA ATGAATGAAG    18540
GAATCATAAT GCGAATCGCT TGTGGTAAAA TAATTTTTTG CATTGTCCGA TTATAGGAAA    18600
GACCTAAACT TCTTGAAGCT TCCATTTGGC CAACTGGAAC AGCTTTGATT CCGCCGcGCA    18660
```

CAATTTCAGA AATATAGGCA CTGGCATTTA AAGTTAATGT GATAATCCCA GCAATGAAGA 18720

CCGGGATATT AAAGCCAAGC ACACCAGGTA ATCCGAAATA GATGAAGAAC GCTAAGACCA 18780

TTAAAGGAAT TCCACGAATT AAATCAACAT AAATGGTTGA AAGTACACGT AATGCTTTGA 18840

TTGGTGAAAC GCTGAATAGC CCAAAAATAA CGCCGACAAT TAAAGCAAGT ACGAAGGAAA 18900

TCAACGTTAA TAGAATCGTC ATCCATAAAC CGTGTAACAA TTGTTTCCAG TTGTTTTGCA 18960

TGAAGCCAAC AAACGTAGAT TCGTCAGCTG TTTCTTCTTT GGCTTCGTTG CCATCTTTCA 19020

CGTAAGTGCC GATAATTTTA TCGTATTCAC CTGTACGTTT CATTTCTTTT AAGCCTTCGT 19080

TAAACATTTC AAGTAGTTCA GGATTTTGCC CTTTTTTGAC AGCAAAACCG TAAGAACCAC 19140

CTTTTTCACG AGGAATTGGT GTCGCTAGTG GTTGATTTTG CGCAACCCCG TAGCCAATGA 19200

CAGGATAATC ATCCATCATG GCATCAACTT CGCCAATTTC TAATGCACTG TATAATGCAT 19260

CAGTTGTGTC TAAATACTTA ATTGAGTAAT CATATTTCTT TTTATTTTTT TCTAAGAAAT 19320

CAGCACTTTC AGTACCAATT TTTACGCCGA CTTTTTTGCC TTTTAAATCA TCATACGATT 19380

TAATTTTGTC ATTGCCTTTT TTGACCGCAA TTTGAATACC GCTATCAAAG TAAGGAACAG 19440

AGAAGTCGAA GGCTTTTTTA CGATCATCCG TAATCGTCAT GCCTGCAACC ATCCCATCTG 19500

CTTGACCAGA TTCAACAGCT TGGACAGCAG AACTAAAACC AATAAATTTA AATTCAACAG 19560

TAAAACCTTG CAATTCAGCT GCACGTTTGA CTAAATCTAC ATCAATTCCG ACATAATCGC 19620

CTTGGGCATT TTGAAACTCA AACGGAGCAA ATGTAGAATC ACTAGCAATC ACATATTTCT 19680

CTTTTTTAGG AGTGATTTTT TTCATCTGTT CGCCAGCGTC TTGTGTGTTT GTTTCATCGC 19740

CTGTCGCTAA GTAGTTGTTT AAAATTTTAT CATAAGTACC ATTGTCTTTA AGATTTTTCA 19800

AACCAGCGTT GAACTTTTTG ATTAATTCAG GGTTTTGACC TTTTTTCACT GCGAAACCAT 19860

AAGAACTGCC TGTtTCTTTA TCGCCAACTA ATTGTAATTT TTGGCCATTT TTCACTGCGT 19920

AnCCAATACA GGATAATCAT CAACGATTGC ATCCGcTTCG CCGTTTTCTA AGGCTTTATA 19980

AAGACCCGTT GCATCGTCAA AATTTTTGAT GGTGTAGTCA TATTTCTCTT TATTTTTTTC 20040

AAGGAAATTG GCACTTTCAG TTCCAACTTT AGCCGCAACC GTTTTGCCTT TTAAATCATC 20100

ATATGATTTA ATTTTGTCGT TTCCTTTTTT AACCGCTAAT TGTAAACCGC TATCGAAGTA 20160

TGGATCTGAA AAATCAAAGG ATTTTTTTCG CTCATCCGTA ATACTCATTC CAGCAATCAT 20220

GCCATCAATT TgTTTAGATT GAAtCgyTTG GACTGCACTA TCGAAGCCTA ATGGTTTTAA 20280

ATCCACTTCA AAATCTTGAT CTTTGGCAAT CGCATCTAAT AAATCAACGT CAATACCAAT 20340

GTATTTTCCT TTTGAGTCTT GGAATTCAAA AGGGGCGAAT GTTAAATCCG TTCCAATTGT 20400

ATATTTCTTT CCTTCTGgCA TGTGgCTGgA GTGGCCACTA ATGGAATGTT AGCAGAGTTA 20460

CCATCATCAC AAnGAAAGAA AGTTAAAAGT GTTTCCTTTT CCATAAAATA CCTCCTAGTT 20520

TTGGCATATG GAACCATTTA TATnGTTAGG nnATAGGGAG C 20561

(2) INFORMATION FOR SEQ ID NO: 293:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1836 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

AACTCTGTAT CACCAATTTT TGTATAAAGC ATGGATTCAG CCCCTAATAA TTCAGAAACA 60

ACGACTTCAG ATTTAATCAC CGCATCTTTC ATTGCTGTTA AAGCCACTTG TTCGCTATGA 120

ATATCTTCTG GACGAATACC GAAAATTAAT TCTTTGCCTT CATAACCATG TTCAACAAGC 180

TCTTTGTTTC TTCCTTCAGG AATTGCTAGT TTTAAGCCAT GTTGGTTAGA AATAACCCCA 240

TTGTTTAAGG TAACTTTGAA GAAATTCATG GCAGGTGAGC CGATGAACCC AGCTACGAAC 300

ATATTTTTAG GTGTATTGTA AACTTCTTTT GGTGAACCAA TTTGTTGAAT AAAGCCATCT 360

TTCATAATTA CGATTCGGTC AGCCATGGTC ATCGCTTCGG TTkGGkCATG TGTCACATAA 420

ATTGTCGTAG TTTCTAAACG TTGGTGTAGT TTAGCGATTT CAGCACGCAT GGCTACACGT 480

AATTTTGCAT CCAAGTTTGA TAAAGGTTCA TCCATTAAAA AGACTTTGGC ATCACGGACA 540

ATTGCCCGGC CAAGTGCAcA CGTTGGCGTT GTCCACCAGA AAGCGCCGCA GtTTACGTTG 600

TAAGTACTCT GTTAAGCCCA GAATATCCGC AGCATTTTCC ACCCGTTTTT TAATTTCAGC 660

TTTATCATAT TTTCTTAGTT TTAAACCAAA GGCCATGTTA TCAAAAACAG TCATATGCGG 720

GTAAAGCGCA TAGTTTTGGA AAACCATGGC AATATCACGA TCTTTCGGTG CAACGTCATT 780

CATGACTTTG TCCCCAATTG ATAATTCACC TTCAGTAATA TCTTCTAAGC CAGCAATCAT 840

GCGCAATGTT GTTGATTTAC CGCAACCAGA AGGTCCGACA AAGACGATAA ATTCACGGTC 900

TGTAATTTCT AAGTTAAAAT CTGTAACAGA ATAATGTTCT GCATTGTCAT ATTTTTTATA 960

CACATTTCTA AGCGCCATTT CCACCATTTG ATACGTCACT CTTTTCTTTT TGATTTGTTA 1020

CTCATAGTTT AAATGAAAAC GCTATCTTTC GGCAAGGGAG AGTTGCACAA GAAAAAGAAG 1080

ACTTTTTCGT CAAGGTGACA AAGGGCAAAC ACTCTCAGCG GATTTATCGA TAAACTTTTG 1140

TTATAATATA ACAAAGACAG GAGTTGGATA ACATGAAAAT CGCTTTAATC GCTCaTGATC 1200

GTAAAAAAAC ATTAATGATC AAACTAGCAA CCGCTTATAA ACATATTTTA GAAAAACATG 1260

AATTGTATGC AACGGGAACG ACAGGCATGA AAGTGATGGA AGCAACTGGC TTACCCGTGC 1320

ATTGCTTTAA ATCGGGGCCA TTAGGTGGGG ATmAACAGAT TGGCGCAATG ATATCAGAAG 1380

ACAACATTGA TTTAGTTATT TTTTTAAGAG ATCCTCTTTC CGCGCAaCCG CATGAACCAG 1440

ACGTGACGGC TTTGATTCGT CTAAGTGATG TCTATGAAAT TCCTTTAGCG ACCAATATTG 1500

GCAGTGCTGA AATTCTTTTG AGAGGTGTCG AGGCGGGGTT TGCGGACTTC CGTGAGGTTA 1560

TCCATGAAGG GGATCGTCGA CCGTTAGCTT TTTAGAGAAC CTTAATGGTT ATTTTGACTA 1620

TTTTGTGGTA TATTTTATGA GGACTTTTGT CCAAATAAAA TAACAGAGTA GGAGACAAAG 1680

CGTTAAGTGC TAAAAGGATG GGATGTTGCC TTTTGGACGA AGAACATAAA CAGTGTGTTG 1740

CTAGAAAGCA AAGGTTCTTA CTTTGCTTAT CATCAGAACA ACACGnTCAG CTTTTTAATG 1800

TTCGCGGTTT TGTTTnTCGC ATTCGCTGCA TGGTTC 1836

(2) INFORMATION FOR SEQ ID NO: 294:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4607 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

TATTTTATTG ACAAAGAAGG GGACGAGTAC AAAGAAGTGT TTGATCGATT GCATGTTTTT 60

TCAGCTGAAG AATTTGAAAG TGAATTGAAA CGATTACTAG AAAAAGATAA ATTCATTAAT 120

TGGGAGTGAG TTTCATGAGA AGAATAGAGA AAGAATTCAA CAAGAAGCTA GCCAGTTACG 180

AAAGAGAGTT AAAAAAATTA GGATGTCTGG ATGATGAAAC AGGGCTTATT CCCATAAGCA 240

AAAGACGATG GCATGTCATT TGGTGGCAAC CTGTTACACT TGCAAAAACA ATCGTGAGAT 300

CTTTTCGATT AACGCTAGAT AATGAAAACT TATGTATTTT AGGAGATGTA GAAATCACTA 360

TTTATCATGA TGGAACATAT GGTATTTCTA AAGAAGCTGT CCCAATTTTT ATTAATGATT 420

TATTGTCTTT AAAAAAGTTA ATTACTATTT TCTATGGAAC ACCTTTTAAC TTAAATTTTG 480

AAAAAATTCG CTGCGTTAAT TTTAACAGGT ATTGTGTCTC TATTCCTGAA ATTTATGTCG 540

AGAAATTTGA AGTATTAATA AACTACTTGA TAATTTTAAG TAGCTGCTTA CATGAAGTGA 600

AAAAGCACGT AGAATATGAT TAACAAGCTT TTCTTCTTTC AGAAGAACAC CAATAGATTG 660

GTTGAGTTCG CATACGCTCA CCTAAGAGAG GAAAAGAACG AAGTCAAGCT TCGTTCTTTT 720

TTCTACGTCG GAGAACCGAC GTAGTGATTC AGACGCAAAA ACCATTCGTC TGAATCAGGC 780

CCGTTTTTAG GAAGGGCTAG TGTGTTTGAT TTATCAGTTC TCACAGCCGT GTAATTTCTA 840

AGATTCCCTA ATCGTCTATC AGTCACTCGG CGAGTATAAC CTAAAACATA ACCTGTCGGA 900

GGTCAAGGGC AAGTGGCATC GGCAAGCCGA CCCTTGTCCT CCGACAGAAC ATGTTTTCTA 960

GTCGGTTATG CCGAACGACC GATATACGAA AAGGGAATCT AACCAGAATT TCCCCTCTGA 1020

CCTTTGAAAA CTGAAAAGCG AAATCAAACA AACAGGCCTT CGTCTAAAAA CGAGCCTAAT 1080

CAAACCTTAG GGGTTGCACC CCTAACACCC GGCGCCTCAC TCTCTCAACC GAAAAGGTTG 1140

GTCGGAAAAG CGTTGACAAA AACATAGTTA TTTATTGGTT GTTTGAAAAT AAAAATTGAA 1200

AGAAGGAAAG CAAAATGCTA ACAACAAAAG AAAAAAACAG ATTAAAAAAA ATGGTAGAAG 1260

GAAACAAAAC GTTTCATTAT TCTTATGTCG ATAGACTTAG ACAAGATGTA AGATATTACG 1320

TTAATCAATG TGAATCAGCT GTTAAAGCTA GAGAAAGCAT GGAAATCTTA GAATTTATCT 1380

ACAGTCTATT TTCTGATAAA GAGATACCAG CATGGTACAC AAAAGCAGAC CTTGAAAATG 1440

ATAAAAAATC AATTGAAAAA TTGGAACGCT GGGCAGCTTA GCTGTTCAGC AGCAATCAGA 1500

AAGGGCGATT AGCATGAAAT TTATCAAAGA TGTAACGACA TTAGAAGAAT TGAAACGCGT 1560

ATATAAAAAG CTAGCTTTAA AGTACCACCC AGACATGGGC GGTACTGATA AAGAAATGGC 1620

GCAAATAAAT AACGAGTATG AAGCGTTGTT CGATAAGCTT AAAAATACAC ACAAAAATAA 1680

AGAAGGTGAG TACTACCAAA AAGAAACGAC AGAAACACCG CAAGAATGGC AAGAAATCAT 1740

ACACAAGTTG TTAACACTAA AAATGGAAAA TGTTTCAATT GAAGTGATAG GCGCATTCTT 1800

ATGGGTATCA GGGAATACAA AGCCTTATAA AGAAGAGTTG AAAGCTTTAA ATATGAAATG 1860

GAGCAACAAC AAAAAGTCCT GGTATTTAGC CCCCGATGGA TATAAAAAAC GCAGTAAGAA 1920

AAAATACAAT ATGAGCGATA TTCGCAATAT GTATGGTTCG CAGGTAGTTA AAGAAGAACA 1980

AGCCAAACAA AAGACATTAA CAGCAACAGC CTAAGGAGGA AGTACAATGG AACAACAACA 2040

CCCGACTATT CACACATTAA AAATTGAAAC CGAGTTTTTC AAGGCAGTAA AAGAACGGCG 2100

AAAAACATTT GAGATAAGAA AGAACGATCG TAATTTTCAA GTAGGGGATA TTTTAATTCT 2160

CGAAGAATAT ATGAATGGCA TGTATTTAGA TGATGAATGC GAAGCGGAaG TTATTTATAT 2220

AACGGATTAT GCACAACGTG AAGGTTATGT GGTGCTAGGA ATTGAATTAC ATTAGTAAAA 2280

AATTAGATTA TACGTCCAAA GAGGTTTTAC CTCTTTGGAC yaCCAAATAA AAAGAAAGAA 2340

GGAATCAAGA ATGATTAAAA CAAGTGCAGT AGGAAGATTA ACAAAAAGTT CAGAACTAAG 2400

AGTAACGTCA ACAGGGAAGG CCGTTGCCAA TTTTGCATTA GCTTGCAAAA GAGAAAGACT 2460

TGATAAAGAA GGGAATGCCC AAACGACTTT TATTCAATGT GTAATCTGGG GGAAACCTGC 2520

AGAGGTGTTA GCAAAATATA CGCAAAAAGG TTCATTAATT GCAGTAAATG GTGAATTGCA 2580

ATCACGTTCT TATGATGATC AACAAGGACA AACGCATTAT GTTACAGAAT TAGTCGTAGC 2640

AGGCTTTGAA TTTTTAGAAT CAAAAGAAAC TGTACAAAAT AGAGAACAAA AACAACAAAT 2700

GCCAGAAATG GCGAACGCCG AACCGTTTTA TGGTATGGAT GATATGGAAC CGCCTGCTTA 2760

TCATTAAAAT TTATCTGAAG GAGGGAGAAA ATGAAAAAGT TAATCGGTAA AAAGTGGCTG 2820

CTgCTTACAG CAGTAGCCAC TTTTTTATTA TCAGGATGCG CAAGTCTTGA ACAAAAAGCA 2880

CAGGATAGTG TAAAAGAAGT TACTGAAAAT GTTACTCAAA CTATTTCAAA CGATCAACGT 2940

ATACCAGCTG ATTTTGTTAG GCACGTGGAT GGCGATACCA CAGTATTAAA AATTGACGGA 3000

AAAGAACAAA AAGTTCGGTT TTTATTAATT GACACACCCG AGACTGTGAA ACCGAAAACA 3060

AAAGTTCAGC CGTTCGGATT GGAAGCTAGC AAACGCACAA AAGAGCTTTT GTCTACTGCT 3120

TCAGAAATTA CGTTTGAATA TGATAAGGGC GATAAAACAG ATCGTTACGG ACGAGCGTTG 3180

GGCTACATAT TCGTAGATGG AACATTACTA CAAAAAACGC TTGTAAGTGA AGGATTAGCT 3240

CGTGTTGCCT ATGTAAAAGA GCCTACAACT AAGTATTTGG CAGAACTAGA GCAAGCCCAA 3300

GAACAGGCTA AAAATGAGTC ACTCGGAATC TGGAGCATAC CAGGTTATGT GACACAACGG 3360

GGGTTTAGTA AATAAATAAG AAATCGCTAT GAATTATTGA AAGCTTTTGG CGAGAAATTA 3420

TAACACAATA TGTTGTCGAT AACAACATAT TGTGTTATAA TTATATACAT AAAGGAGGGG 3480

TAAAGGTGGC TACTACAAAA AAGAAACCTA TACATGTAAA TGTTGATGAA AATTTAAAAG 3540

AAGAAGCAGA ACAATTATTT GATGATTTAG GGTTAAATAT GACAAGTGCA ATTACGATTT 3600

TCTTAAAACA GTCTATTAAT GAGCAAGCAA TTCCTTTTAT GATTAATAAG GGAAACAAAG 3660

AGACTCTACA AGCATTAAAA GACATTAAAG AAGGAAATGT TCATGGTGGA TTTTCTTCCG 3720

TGGAGGATTT AATGGAGGAT TTAAATGCTT GAAATATTTT ACACGAACCA ATTCAAAAAA 3780

GACTTTAAGA AAGCAAAGAA ACAAGGAAAG AATTTAGAGA AGTTAAAAGA AGTATTAGTA 3840

CTTTTACAAG AACAGCAAAC ATTGCCGCCA AAGTATAAAG ACCACGCACT AACAGGAAAT 3900

TATATAGGTA CAAGAGAGTG CCATATTGAA CCAGATTGGT TACTGATTTA TAAAATTGAC 3960

GGTGATAAAT TGATCCTAAC TTTAGCTCGA ATAGGCTCAC ATAGTGAATT ATTTCGATGA 4020

ATCGGAAAGA TATTACAAAT CAATAAAAAA CACTACTAaA CGACCAGACG AATAGTAGTG 4080

TTTTTTTGAT TATTCGATTG TACTAGTAAT TTTAACACAA TaAAAAATAG TCATGCAAGA 4140

CCAGGGCGCA ACGATCGGTT CTGTTGCAAA GTTTTAAATA AAGaATAAAA TCCCTTACGG 4200

TATCTATGAT TTAAGCTGGG ATTCCCAATA ATACCTTGAT TTCAGTACAG ACCGAAAACC 4260

CGAAGAGAGT ACCTTCTTTT CGGGTTTTCT TATATAATCC TCGAATGGCT TCCATGCCTT 4320

TAATCGTGGT AGAGGCAGTG CGTAAACTTC GATAGAATTT ATTGCGTCTC TTTACTGGAC 4380

GATGGTCTTG TTCAATCAAA TTATTCAGGT ATTTAATGGT ACGATGTTCT GTCCCTTGAT 4440

AAAAGCCGTA TTCTTTTAGT TTCTTAAAGG GCACTTGTAA TAGAGGGGGC TTTATCTGTG 4500

GACTACAACC TTCGGGTCAT CAAACTGCTT CACTAACCGC TTAAGAAAAG CATAGGCTGC 4560

CTGTGTGTCC CGTTTTTTAC GTAACCAAAT ATCCAAGGTn AAACCAT 4607

(2) INFORMATION FOR SEQ ID NO: 295:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4738 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

TTCTCGCGGA TTTATTCGGA TTGAACACAA CAAACCCAAA AGGTCAAATG ATGGTTACGT 60

ATGATTTTAC TGCCTGTTTT TCAAAAGATT CAATTATTTT TACAAACGAG CAAAGAAAAA 120

CAAGTTGTCG CAAACCACGA AGATGAAATT AAACAGCTTT ACCAAGGCTT TGAAAAAGAA 180

TTCGGACGTC CACTTTCACC AATTGAATTA GAAATGATTG GTCAGTGGTT GAACACGGAC 240

CATTATTCCC CTGAaTTAaT CCGGTTAGCG TTGCGAGAaG CTGTTTTAAA TCAGGCGTAT 300

AGTTTGAAAT ATATTGAtCG TATTTTACTT GCGTGGGAGC GTAAAAATAT TACCACCAAA 360

GAACAAGTTG CTGCTGATCA GAAAAAaCGA AAAGATTCTA TGATTCAAAA TGAGATTGAA 420

CAACAAGGTC AGACACAAGA ATCGTTGCCA AAGGTTACAT TACACAATTG GTTAAATCCA 480

GAAGACAGTG AATAAAAAGG AGGTGGGCCA TGCTATCAAA AGAGAAAACA ATGGAAGCTA 540

```
TTGAAATTAT GTATGAGATG TTTCCAAATG CAGAATGTGA ATTAAAACAT AAGAATCCTT    600
TTGAATTATT AATTGCTGTG ATTTTAAGTG CGCAAGCAAC CGATGTATCG GTAAATAAAG    660
CGACTCCTGG TTTATTCGCA GCTTTTCCAA CACCAGAAGC TTTAGCAGct GCGCCGGTGG    720
AGGAGATTAT TGCTAAAATA AAAACCATTG GCTTATATCG CAATAAAGCA AAAAATATTA    780
AAGCGTGCGC ACAACAACTT TTAGAACGTT TCAATGGAGA AGTTCCGCAA ACCAGAGATG    840
AACTGGTCAG TTTACCTGGC GTGGGGAGAA AAACAGCCAA TGTAGTAATG GGAGATGCTT    900
TCGGCGAACC GGCTTTTGCA GTAGATACAC ATGTGGAACG AGTGTCCAAG CGCCTTAGAA    960
TTTGTAAGCT AAATGCGAAT GTTACTGAAG TTGAGCAAAC CTTAATGCGT AAAGTTCCGA   1020
AAGAACTCTG GGTGAAAACG CACCATACCA TGATTTTCTT TGGGCGTTAT CACTGCCTAG   1080
CAAGAGCGCC TAAATGTGAA GCCTGTCCGC TACTCTATAT GTGTCAAGAA GGCAAAGAAC   1140
GGATGAAAGG GAAATAACAC ATTTTTTCCT AAAAAAGcGA GCCACAAGTC CAACCGACTT   1200
GTGGCTCGCT TTCTTTTTTT GATTTTTAAG AAACGCAACA AATGATTGCG GTTcTTAAAA   1260
ATTGATTTAG TATAGACAGT GTGGtATATA CyAgAAAATt TATTTTTTTG AAGAAAGATA   1320
GGCGATGAAA TGACAGAAAa TGCACCCAAA TTTAAGCmAA TTtCyGCAGA AaTTGAAAAG   1380
AAGATTCGAG ACGGTCTGTA TGTAAGTGCA CAAAAATTAC CTTCAGAATA TGATTTAGCT   1440
AAAGAATATA ACTGCAGTCG CTTGACCATC CGTAAAGCGA TTGATGATTT GATCCGCAAA   1500
AATATTTTGG TAAAACGACA TGGTAAAGGT AGTTATGTGA TGTCGCAAGC GAAAATTCAA   1560
AGTGGTCGCG CTGGCTTACA AGGTTTTACT GAGGCAGCCA AAGCTTACGG GAAAAAAAGC   1620
CAGACAGAAG TCATTTCCTT TGAAGAAGTA GTACATCCCG CTGAGAAAAT TCGGGAGGCG   1680
CTCCAAGTAG GCAAAAATGA GGCAATTTAT GAACTGATTC GCCGCCGAAT GTTAGACGGC   1740
GAACCAATGA CAGTTGAAAA AATTTATTTG CCACAGGCAT ACGTACAAGG CCATACGAAG   1800
CAAGACTTCG AGGGCTCTCT TTTCTGCTTA ATCGAGAAGA ACGTCGAGAT TGCTTATTCG   1860
CATCAAGAAA TTGAAGCAAT CTTAGTTGAA GCGGAAATTT CAGAATTATT GAATGTTCCT   1920
GTGGGCCAAC CACTTTTACA AGTCCACTCT ATCACCTATG CGCTTGATGC AACTCCTATT   1980
TTATATGATG TCTCTTTATA TCGAGCAGAT CGGTACACGT TTAAAAACAC ACTGACCCGC   2040
TATAGCCCGT CTGAAAACAA CCAAGTGGAG CTAGGAGGTT CTTGGAACGA ATGAAGATCA   2100
AAGAAGAAAT AGCCGCTCAA AAAGATTTAT TTTATGAAGA CTTAAACAAA ATTATCGCGA   2160
TTCGAAGTGT GAAAGGGTCG CCTAAAAAAG AGGCACCTTT TGGCGAAGGA CCGAAAAGAG   2220
CCTTGGAAGA AACGCTGAAA CTTGCAGAGC GTTATGGTTT TCAAACTGGG ATTGTCAATG   2280
ACGCAkTggC TATGCGCAAT GGGGAACAGC GGAAGAATAT CTGGGAATTA TTGGTCATTT   2340
AGATGTAGTA CCAGAAGGTT CTGGTTGGTC AGTGCCGCCC TTTcAATTAA CGAAAAAAAA   2400
TCaACGTTTG TATGGTAGAG GAATTCTAGA TAATAAAGGT CCTATCTTGG CTTGCCTGTA   2460
TGGAATGAAA TTACTGAAAG AACTTGGTTA CCAACCAAAG AAAaCCATTC GCTTAATGTT   2520
```

TGGCACGGAT GAAGAAAGTG GGAGTGGAGA TATCCCCTTA TATTTAGAGA AGGAAAaCGC 2580

ACCCGTTTTT GGATTTACTC CAGATTGTAA ATATCCAGTA GTTTATGGGG AGCGAGGGAT 2640

TGTTAATTAT GAGATCACAA CGACCATCCC AGATGATTCA AGTGAACAAA TTGGTCAGAT 2700

TATAGGTGAT CAAGCAAAAG ACCACGTACC TGATCAATTA AGTGTGGTGA TTGCGGGAAA 2760

AACAACAGCA ATCACGGGAA AACGTGCTCC TTCCAATGCG CCAGAACTAG GCAAGAACGC 2820

GATTACTTTA TTGGCACAGA AAATTAGCGA GGAACAGTTA GTCAAAGGAA ATTTATTACA 2880

GTATTTCGAC TGGTTAACCG CTAGTTTTCA CGAAAAGCAC TATGGCGAAG GAGTAGCTCT 2940

GGACTTTAAG GATCAGGATA GTGGGCAATT GATTTTAACG CCyTATGyGT TGGAAAAAAG 3000

AGGACAGCAA TTGGTGTTAT CATTGGCsGT GCGTTATCCT GTTTCTATTA CAGAAAACGA 3060

AGTAACCACG CAGCTAACGA AGGCACTATT TCCAGAAAGT GAAGTGACCG TCATCCGCCG 3120

CCTCCCTAGT ACGCTGTTTC CAAAAGATGA GCGCAATGTT CAAAAATTAA CCAAGGTTTA 3180

TGAACAAATT ACTGGCTTAG ATGGGACGCC AGTCACAACT ACAGGTGCTA CGTATGCTCG 3240

CTTTATGCCG AATATCGTTG CTTTTGGTCC ATCATTTCCT GGTCAAAAAG GCATTGCGCA 3300

TAACCAAGAT GAATATATGG ATGAAAAAGA TTTACTGCTT AATCTGGAAA TCTATATGCA 3360

AGCGATGATT GCATTAACAG AAGCATAAAA CCAATAGAAG ATACACGTAT GAGAAGAAGA 3420

CAATGTGTTT CGTAGAGGTC GCATACGTGT ATCTTCTATT TTTCTGTATA AAATTTCATT 3480

TTCAGTATAT ACAAAACAGT ATATACTAGT TTATAATGGT GGAGAAATGT AAGCGTTAAC 3540

GAAAGGGCGG ATGGAAAATG ACTTGGGGTG CAATTGCGAC ATGGCGGATG GCACATGATG 3600

GGTTACTAAA AGCTACAGAA GAATTACAAC AAGGAGGTGC TGCAGGCACG GCCGTGGAAC 3660

AATTAATTAA AGAAGTAGAA GACTATCCTT TTTATAAGTC AGTGGGCTAC GGCGGTTTAC 3720

CTAATGAGGA AGGGATTTTA GAAATGGATG CTGCCTATAT GGATGGAGAC ACATTTGCAA 3780

TTGGTGCTGT GGCGGGAATT ACAGATGTTA AAAATCCGAT TTCAGTGGCT AAAGCATTAA 3840

GTAAAGAGAA GTTTAATAGT TTTCGTGTTG GCGCAGGTGC AACGAAATAT TCAATGTTGC 3900

ACGGTTTTGA AAGAAAGAAT ATGTTGACAG AACGCGCTAA TCAATGGTGG CAAAAGCGTT 3960

TAAAAGAAAT TCAGGAAAAT CAGTTGAACC CCTATGATGG GCATGATACT GTCGGCGCTA 4020

TTACGTTAGA CCAAACAGGA TCAATGGCAG CTGGCACTTC CAGTTCGGGA CTCTTcATGA 4080

AAAAAGCAGG ACGTGTTGGC GATTCACCAC TGTCAGGTTC TGGTTTTTAT GTTGATAGCG 4140

AAATTGGTGG TGCTGCGGCT ACAGGCTTAG GTGAAGATTT AATGAAAGGC TGTCTTTCTT 4200

ATGAAATCGT TCGTTTAATG GgGGaaGGGC GCTCGCCACA ACAAGcATGT GATCAAGCCG 4260

TCTATGCTTT TCATGAAAAA TTGACCCAAC GTTATGGGAA AGCAGGAGCT TTTTCACTAG 4320

TGGCCATGAA CAAACAGGGA GACTGGGGCG TACTACgAAT GTGGAATTTA CTTTTACGGT 4380

TGGTACTGAT ATGCAGCAAC CAGAAATCTA TATTGCAAAT CCAGGCAAGA ATCATACTAC 4440

AGAGATTCAG CCGATTTCAC AAGAATGGTT AGCGGCCTAT GAAAAAAGAA TCAAAGCACT 4500

AATTGAATAA ATGAGGAGAG TGAAGGAACA TGAGTAAAAA ACAAATTTAT CTTTTTTGTG 4560

ATGCAGGGAT GTCTACAAGT ATTATGGTGA ATAAAATGAT GGAAGTGGTG GAAAAACATC 4620

AAATGCCTTT AATGATTACT GCTTTTCCTA TTGCAAGAGC mCAAGAAGTG GTTGAAGCAG 4680

AGAAGCCTGT TGCtATTTTA TTAGGcCACA AGTCCGTTTT TTTATTAGGn AAAAACCA 4738

(2) INFORMATION FOR SEQ ID NO: 296:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3225 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

GGGTGCTGCC CCAATGGCGT ATGCACTTTG GACAAAACAT TTAAAGGTGA ACCCAACAAC 60

TTCTAGAAAC TGGGTTGATC GTGATCGTTT TGTACTTTCA GCAGGCCACG GCTCTGCAAT 120

GTTGTACAGC TTATTACATC TATCAGGATA CAACGTAACA ATTGACGATT TGAAAAATTT 180

CCGTCAATGG GATAGTAAAA CACCAGGGCA TCCAGAAGTG CACCACACAG ATGGTGTTGA 240

AGCAACAACA GGACCTCTAG GACAAGGGAT TGCTATGGCT GTTGGTATGG CAATGGCTGA 300

AGCGCATTTA GCTGCAACTT ATAATCGTGA TAGTTTCCCT ATTATGGATC ATTATACGTA 360

CGCAATTTGT GGTGATGGTG ATTTAATGGA AGGTGTTTCT CAAGAAGCAA GTTCAATGGC 420

TGGTCATATG AAATTAGGTA AATTAATTGT CTTGTACGAT TCAAATGACA TCTCATTAGA 480

TGGACCAACT TCTAAAGCCT TTACTGAAAA TGTTGGGGCT CGTTATGAAG CATACGGATG 540

GCAACATATT TTAGTTAAAG ATGGCAATGA TTTAGATGAA ATAGAGGCTG CGATTGAAGC 600

GGCTAAAGCT GAAACGGATA AACCAACATT GATTGAAGTT AAAACAGTCA TCGGTTACGG 660

AGCACCAAAA GAAGGAACAT CTTCCGTTCA CGGCGCGCCA ATTGGTGAAG AGGGCATTAC 720

AGCGGCTAAA GCTGTTTATG GTTGGGAATA CCCTGACTTT ACTGTGCCAG AGGAAGTCGC 780

TGCTCGTTTT AAAGAAACGA TGATTGGCGA AGGCCAAAAA GCTGAAGAGG CTTGGAATGA 840

GATGTTTAAG AACTATGAAC ATGCGCATCC AGAATTAGCC AAACAATTTA AAGAAGCTTT 900

TGCCAACCAA TTACCTGAAG GTTGGGAACA AGAATTACCT AAATATGAAC TAGGAACAAG 960

TGCCGCAAGT CGTGTAACAA GTAAAGAAAC GATTCAAGCT ATTTCAAAAG TTGTTCCAAG 1020

TTTCTGGGGC GGTTCAGCTG ATTTATCTGC CTCAAATAAT ACAATGGTTG CTGCAGAAAA 1080

AGATTTCGAA CCTGGTCAAT ACGAAGGCCG TAATATTTGG TTTGGTGTTC GCGAATTTGC 1140

AATGGCTGCA GCAATGAACG GGATTCAATT ACATGGTGGT AGCCATGTTT ATGGCGGGAC 1200

ATTCTTCGTC TTTACTGACT ACTTACGTCC AGCAATCCGT TTAGCCGCTT TACAAAAAGT 1260

TCCTGTAACT TACGTCTTAA CGCATGACTC TGTTGCGGTG GGTGAAGATG GGCCAACACA 1320

CGAACCAATT GAGCAATTAG CAAGTGTTCG TTGCATTCCA AATGTTCATG TGATTCGTCC 1380

AGCAGACGGC AATGAAACTG TTGCTGCTTG GAGAATTGCT ATGACTTCTA CGGAAACACC 1440

AACTATTTTA GTTCTAAGTC GTCAAAACTT ACCTGTTTTA GAAGGAACGT TAGAACACGC 1500

TTCTGATTCT GTTCAAAAAG GTGCTTATGT ATTGTCACCA CAAAAAGGTG AACAACCAGC 1560

AGGGATTTTA ATTGCGACTG GTTCTGAAGT AAATCTAGCG GTGGAAGCAC AAGCGAAATT 1620

AGCGGAAGAA GGCATCGATG TATCTGTCGT GTCAATGCCA AGTTTTGATT TATTTGAAAA 1680

ACAATCTGCT GAATATAAAG AAAGTGTGTT ACCTAAAGCT GTGACAAAAC GTGTGGCGAT 1740

TGAAGCTGCA GCAAGCTTTG GTTGGGAACG CTATGTAGGG ACGGAAGGCA AAACaATTAC 1800

AATTGATCAC TTCGGCGCTT CTGnCACCTG GCGGTCTAGT TCTTGAAAAA TTCGGCTTTA 1860

CTCCTGAAAA TGTGGTTAAT ACCTATAAaT CaCTATAAaT AAACGATAAA AAACCTGAAT 1920

GCTTGTGATA GCATTCAGGT TTTTTTGCTT CATTTCTTTT TGGGAGCAAA GAGCATAAAG 1980

TGAAAAAATA ATGCGAACAA TTGGAACCCA ATACCAGTTA AAGTAACACC AGACCAACCA 2040

TAGTTTTGCC ACATTAATGT ACCAATCAAT GAACCCAGTG AACCGCCGAT GAAATAGAAA 2100

AACATAAAGA CGGTATTGTT ACGATTGCTG GCTTCTTCTC CCAAATTTTG GACGCGTGTT 2160

TGATTGGCTA CTTGCCCAAa TTGAGTACCG ATATCTAAAA CAATAATTGC TAAGATTAGT 2220

AAGACCACAT GTGTTCCACC TAGAAAAAGT AAGATAAAAC TACCGAACTG CATCAACAGG 2280

CCAATCAAAA CAATCTTCCG CTCTGGATAA TAATCGGATA GACGACCAAT GATGGGTGCT 2340

ACGAATGCTC CCGATAAACC AAAAATTGCT AAAATACCGA CTTCTTTCGG TCCCCAATAA 2400

TAAGCTGGAC TACTAACAAA GAAAATTAAT GTAGACCAAA AGATTGAAAA GGTTGCAAAC 2460

ATAAAAAAGC CATTGACAGT TGCTTCTCTT AACAAAGGTT GCGATTTGAT TAGCTTAGGA 2520

AGACTTTTCA AAGAACCTAA ATAAGTTAAT CGATTTGAAG TTTGAGTATG TATCACTGTT 2580

TTGGGCAATT TTAATTGAAG TAGAACGACC AATGCTAACA CGAATATTAC GGCGATTAGG 2640

TAGACTGTTC GCCAAGAAGC TGCACTAGCA ATGAAACCCG AGATCGTTCG AGACAATAAA 2700

ATTCCTGTCA ATAGTCCGCC CAGCATTCGA CCTGTTACTT TACCACGGTT CGCTGGTCCC 2760

GCCATGACCG CAGCATAAGG AATAATAATT TGGGCGACAA TAGAGAGCAG CCCGATAAAA 2820

AAAGAAGAGA GTGCAAACAG GAGAAAACTT GGTGCAAAAA ATGCAGCACT TAGAGACAGC 2880

GAGGAGAGCG CGGCTACTCG AATGATTAAT TTGCGACGAT CAACGACATC GCCTAATGGA 2940

ACAAGAAAAA GTAAGCCTAG CGCATAGCCT AATTGGGTCA GCATAGTAAC AAAACCAATA 3000

GAACCAATAC TGACTTGAAA GGCCTGAGCA ATTTGGGTGC CAATTGGTTG AATATAATAC 3060

ATATTGGCGA CCACGACCCC GCAAGTTAAC GCTAATAAAA ACGTCACTTG ATTTGTTAAA 3120

CCCGATTGTT GTTTCTTCAT CTAACCACCT ACATTTCTAT GTAATTCATT TTGTAATACA 3180

ACAACTACTC GGCAATAGTT GTTnTGATTA CAAAAACAGA CTTCT 3225

(2) INFORMATION FOR SEQ ID NO: 297:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 682 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

AAGTCCTTGT TTATTGGAAT TTTACGACAG TCGCACCATT ACCACCTTGG TTTTGTGGCG 60

CAAATTCATA GCTTTTCACA CTGCGGTGAT TTTTTAAAAA TTCCGTTATT CCGGTTCTTA 120

ATGCACCTGT TCCTTTGCCG TGAACAATGG TTACTTGTGG ATAGCCCGCT AGAATCGCAG 180

CATCAATGTA TTGGTCAACT TCTGCCAATG CTTCTTCGTA GCGTTTGCCA CGTAAGTCTA 240

GTTGTGTGCC AACATGACTG CTTTCTGCAG AACGAACCGT TGTAACTCTT GGTTTCGCTT 300

CTTTTTGTGG AGCGACAGGC GTCATATCCT CTTCAGATAC GTTCATTTTT AAAATGCCCA 360

ATTGGACTTG CCATTGTCCT TTGCCGTTGT CTTTTAATAA GGTCCCTCGT TGGCCGTAAG 420

TATTGACAAT TACTTCATCG CCAGCTTTTA ATTTTTTCTG TTCTTtCGCC TTCTTCAGCA 480

CTTTGTTTTT AGCAAGCTTG GTTTCyTCAt GGTGTAAnTG AGAAAGTTGT GkTTTCGCAT 540

CAATTAACTG ATGTTCTTTG ACACCACCTT GTTGGCCACT TTCCAATTGC ATTTTGCGAA 600

TATCAGAAAT AATGGTTTCT GCGTTTTCTT CCGCTTCTGC AATAATTTTA TTTGCTTCTT 660

TCCGTGCTTn TTGTAATTCT GT 682

(2) INFORMATION FOR SEQ ID NO: 298:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4029 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

GCCAGTTGAG GAnTTGGTGT TGGCGCGGAA TACCnTTTnT CAAAAAATAC CAAGGAGAGA 60

ATTAAAAGGA GTTATCCCCT ATTATATTGT CCGATGCTTT CGCGGAGGAA GTTTTTAAAG 120

GCAATCCCGC CGCTGTTTAT GTTTTGGAAA AATGGTTGCC AATGCTGTAT GAAGAATTTT 180

TTTATACTAT TTTTTTGCCA TATTATAAGG GACGGGTTAA AGAAAGAACT TATAATAATA 240

GACTTAGTAT GATAAAAAAT ACTTTTGCCT ATTTTTTTAA AATGAAATTG AAAAGTATAG 300

CTCCTGTTCA TATACAGAAG TGGCAGAATG AGCTGTTGGA ACAGTATGAA AACACGTATG 360

TAAGAAATAT TTATGGACTG TTTCAGATGT CTTTAGATCG AGCTGTTGTG TTAGGAATGA 420

TTTCTTCAAA TCCAGCAAAG ATCGTTGGTA ATGTGAAAAA AAGTAAGAAA GAAATTGATT 480

TTTGGACAAA AGAAGAATTT GAAAAAGTTG TTAAGACTTT TTATGTAGAA GATTATTACC 540

AAAATTTTTC TTTTATTTGT ATCTGGTTAT TATTTATGAC AGGCATGAGA ATAGGTGAGG 600

CTACTGCTTT AACTTGGAAA GATGTGAATT TAGATAGAAG ATATTTAACA GTAAAAAAAT 660

CACTATATTA TAAAAATGCG CAAACTTATG AATTAGTGTC TCCAAAGACT AGAGCCAGTA 720

TTAGGACGAT TTATTTAGAT GAAGATACAG TTCACTATTT AAGAGATTGG AAAAAAAGGC 780
AAGATGATGT TGGAGGAATT GAATTTATTT TATCTTATAA TTCAGTTCCT ACTCAAAAAC 840
ATACAGTAAG ACATATTATT AAACGTCATG CAAAATTGGC AGAGGTACAT GATATTAGGA 900
TTCATGCTTT AAGACATTCT CATGCTTCTT TATTAATAAG TATGGGTACT AATGCTTTGT 960
TGATAAAAGA ACGATTGGGT CATGAAGATG TGCAAACAAC TTTAGGTACT TATGGACACT 1020
TATATCCAAG TAGCTCCACT GaAATAGCTA ATGAATTGAA GGGAATAGTT AATGTTGAGT 1080
TTACTAATCA AAATATGGCT TCGGAGGTTA CTAATCAATT TACTAAAGGA GTAAAAAAAT 1140
AAAGTGTGAT AAAAGTGTGA TAACGCAAAA GAAAACCCTC TATAAATCCT ATTATATAGG 1200
GGTTTATAGA GGGTTGTTTA TTATTCCCAC TCCACAGTTG CAGGTGGTTT TGAAGTAATA 1260
TCGTACACGA TACGGTTTAC ATGTGCAACT TCGTTCACGA TGCGCACAGA GATTTTTTGT 1320
AAAACATCCC AAGGGATACG GGCAAAGTCA GCGGTCATGC CGTCGATAGA AGTCACAGCA 1380
CGGATACCGA CTGTGTAATC GTAGGTACGG CCATCGCCCA TAACGCCTAC TGAGCGGATG 1440
CCAGGTAAGA CAGTAAAGTA CTGCCAAATG TCACGATCTA AGCCAGCTGC TGCGATTTCT 1500
TCACGTAGGA TAGCATCTGA ATCACGAACG ATTTGTAACT TATCTTCTGT AATTTCACCA 1560
AGCACACGAA TACCTAATCC TGGTCCTGGG AATGGTTGGC GCCAAACGAT TGCATCAGGC 1620
ATGCCAAGTT CAGTTCCTAA GGCACGAACT TCGTCTTTGA ATAATGTATT TAAAGGTTCA 1680
ATTAATTCAA ATTGCATGTC TTCAGGAAGA CCACCAACGT TGTGATGTGA TTTAATTGTT 1740
TGCGCAGTTT CTGTACCAGA TTCGATGACA TCTGTGTAAA GAGTTCCTTG TGCTAGGAAT 1800
GAAACGCCTT CTTCCCCAGC AAGTTTAGTT GCTTCGTCAT CAAAAACATA CACAAATTCG 1860
TTACCGATGA TTTTACCTTT TTGTTCAGGG TCAGAAACGC CTGCTAATTT GCTTAAGAAG 1920
CGTTCTTTCG CATCTACTTT AATAATGTTT AAGCCAAATT TACCGCCTAA ACTTTCCATT 1980
ACTTGTTCTG CTTCACCTTT ACGTAATAAT CCATGGTCAA CGAAGATACA AGTTAATTGG 2040
TCGCCAATGG CTTTTTGTAA AAGCACGCCG ACAACACTTG AGTCAACGCC ACCTGATAAA 2100
CCAAGTAAGA CTTTTTTATC GCCTACTTGT TCACGGATTT TAGCGACTTC CATATCGATG 2160
AAGTTTTCCA TGCTCCAGTC ACCAGTACAG CCACACACAT CAAAAGCAAA ATGACGTAAT 2220
AAGTCGTTAC CATATTCTGA ATGACGAACT TCTGGGTGAA ATTGTACGCC GTAGAATTTG 2280
CGGTTGGCAT CTTCGATAGA AGCGATTGGA CAATCGTTGC TTGTCCCAAC TTTTTCGAAT 2340
CCAGGAGCGA TTTCTGTTAC TAAATCACCG TGGCTCATCC AAACGGTTTG TTTCGCAGGT 2400
GTTCCTTGGA ATAATGTCGC ATCACCAAGG ATTTCTAATT CTGCTTTTCC GTATTCACGG 2460
TTTGCGGCAG GTTCTACTTT TCCGCCTAAG TTATGAGTCA TTAATTGCAT ACCGTAGCAG 2520
ATGCCAAGAA TGGGAATTCC TAACTCATAA ATTTCTGGAT CAATGCTAAA AGCATCTTTG 2580
TCATACACGC TGTTTGGTCC ACCTGAGAAG ATGATTCCTT TAGGCGCGAT TTTGCGAATT 2640
TCTTCTGCCG TAGTTCGGTG GCTTAATAAC TCAGAGAATA CGCCAAATTC ACGAATACGA 2700

CGAGTAATCA ATTGGTTAAA CTGACTACCA AAGTCTAAAA CAATAATTTT TTCGACAGAT 2760

GTCAAATCGG CAACGTTAGT CACGTCTATT ACCCCTTATC CATTTAAATT TATCTAAAGA 2820

AAAGGCGGTT AAAAAGTGGT CAGCACCAAG AATTAAGAAA GGATTTTCAA AAATTGTTTT 2880

ACGATTTTTA GAAAATTTTG GCTTAATTCC GCAGGGgCTA CTTTTAAACA GCCGAAAATC 2940

AGGCGGTTAA AAAGTGGTTC ACTTTTGTtC CGTACTCTTC TTTGTCTAGC TATACTTATT 3000

AAAACATTTT ACTCCTTGAA AGTCATGCAT TCTGCTGAAA ATTTTCAAGG ATATTTCTAC 3060

mAAAGATACA GGATATTTCT TAGCTCGTCA ACCCTTAATA TTTACGTAGA TAGATTTGGT 3120

CAATCAAATG ATTTTCCGTT TTATGCAAAA TAATGTCCGC ACGTCCGCGC GTTGGCAAGA 3180

TATATTCCTC CAAGTTCGGT AAGTTCACGG TTTTCCAAAC ATTCCGTGCC ATGGCGAAAG 3240

CATCTTCTCG TTTGCCAATT GCATATTGGT AATAATAATT ATTAGGATCT AAGAAAGCCG 3300

TATCTAATAA GGCCCCAAAA CGTTCTAGGT ACCATTTTTC AATTAAGGCT GGATCGGCAT 3360

CTACAAAAAT AGAAAAATCA AAAAAATCAC TCACATAAAT TTGTTGATTG GCGGGTAACT 3420

GTAATGTGTT AATGCCTTCA ACAATCAGAA TATCTGGTTG TTGTATCAAC TCATATTCGC 3480

CTTCGATAAC ATCGTAGACA CTGTGAGAAT AGACGGGTGC TTTAATTTCG TCTTTGCCGC 3540

TTTTGAtTCG TTTAAGAAAT TAATTAGCTT CTCCATGTCA TAACTTTCAG GAAACCCTTT 3600

TCGGTCCATA ATTCCCTGTT CTTcTAACAC TTTATTGGGA TACAAAAATC CATCCGTkGT 3660

AATGaGTTGA ACGTTTCTTC GTTTAAACGT TCGGGCCAGA ATTCTTTGCA AAAGACGAGC 3720

GGTGGTGCTT TTACCAACGG CGACACTGCC GGCAATTCCA ATGATGAATG GTGGTACAGA 3780

GACATATTCA TGTAGGAATA ATCCTTTGCT AAGTGTTAGC GATTCAAATT CTTTCATGTA 3840

CAAATGAATA AGATGCGTTA AGGGTACATA AATTTCTTGT ACATCTTTTA GTGAAATTTG 3900

GTCGTTCACA CTTTTAATGT TATCCAATTC TGCTTCGGTT AACGGTGCTT TGCCGTCATG 3960

ATAAAACCCA TGCCATTCTT CTCTGGAAAT TGGGTAGTAA TTCATTTTAT CGTCCATTGT 4020

TCCTCCACC 4029

(2) INFORMATION FOR SEQ ID NO: 299:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1136 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

AACCGGTGTG ATTACATCAT ACAATACTGA TTCAGATTTG TAAACCGCTT TCTAAAAAAA 60

ATAAATAAAC ATAGTCCATA AAATGGACTA TGTTTACGGC CTTAAAAAAA GGTAGTTATA 120

TTTGTAAAAA TTTACTCACT AATAATCTCA ATCTGATAAA AATCAGAAAG TATGAAACTT 180

TGACTATATT CAATCGGAAC CCCATCGTTT CCTCTAATCA AACGTTTAAT ATTGAATAAT 240

GAAAAGCTAT CTTCTTCACC AAAACTTTTT CTTATTATTC TCAACGGCTC TAGTTCATCT 300

GAATCCTTCG TTAATAAAAT CGTGCTCACC TCCATCGTCA TCGGAAGATT AGCCATATTT 360

CCTTGAAGAC CCaATTCATC GGATAAACTA GCTCCTTCAA TCAGGCGTTT TACTGAATCA 420

TCCCTAAGTT TATCTGCAAA GAGATAGCGA TTTTGAATTT TCCATGGAAA ATCATTCGTT 480

AATGCAATTT GGGTTAGCTT AAGCAACGCT TTGCTGCCTG ATTTGTCGCC TAATATCTGT 540

GCGATCTCTT TATCAGTAAA TTCTTCCGAA ACTAAAGTGA TCGTTTCCTC AACCGACTTT 600

TGTTTTTCGT TATTGGACAT CGATAACTTC TCTGAGAAGA AAACTTTTCT ATGCTTCAAA 660

TTTTTTCTGA CAAAGGTTCC CTCGCCCtGA CGCCGAATTA AATACCCTTC ATTTaCCAGA 720

TCGTTCAAAG CCTTTACTAC TGTTGTATTA CTGACATTAT ACTTTCTTTT TAAATCACCT 780

TCTGAATAAA CTTTCTCACC CGTACTAAAC TCTCCAGATT CAATCTTCTC TATGATATCC 840

CGCTTTATAA TCTCATATTT AGGCAACATA TGATCCTCTC CTAACACGTt AACCGTTAAT 900

CTTCATTGTA AAGAACTTTT ATATTTATGT AAATAGAATG AGTAATAATT ATTTGCTCAA 960

ATTTCGCACA CCAAAATCAT AGCTAAAAGC TTGGCATGAA AGGGCGGAAG TTTTTGGGAA 1020

TTGATTTCAA TAAATTTGAG TTCTTTTTTA GTATAATTAG CTTATAAAAA TACTGACCGG 1080

ATTAAATGAA TGATGrAGCG AATGrCTATC AAAAAACAAn CACCTACAAA AATTTA 1136

(2) INFORMATION FOR SEQ ID NO: 300:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3914 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

AGATCGGTCC nTAATAATTT ATTTTTGCCC TCCAGCGTCG nTAATGTCTC CTGATGAGGA 60

TGTTTAAAAC GATTATTCAG TCCAGCTGrA AyCAAgCyTC TCGCGGATTA AGCTGTTTGA 120

TAAATCTTTT AGTAGACGAC GTTTTACTAC CATGATGACC GACTTTCAAG ATGTCGATTT 180

TCAACTGGGG ATAGCGAGCC AACACTTGTT CTTCTCCTTC GGCTTCCAAG TCACCAGTAA 240

ATAAAAAATG CCGTCCGCCT AATTGTGTGG AAAACACTAT CGAGTCATTA TTGCCCCCAA 300

TACTGCTCTT ATTTGGTGAT AAAATTTGAA AAGAAACGGG CCCATTAATT CGCTTGGGCG 360

CTAAAATTGG CCGACAGTCT GTCCCGCTTT TTTGGAGTTG TTTCAAGACT TGGCGAAACG 420

AAAAATTGGC TTCAGCTGCT CCTTTAGGGA AATAAAGAAC AGAAATTGGA ACCTTCTGCG 480

CAATGGCTAA TAAATCGCCA AAATGGTCTT CATGTGCGTG TGTGATGATC ACTTCATCTA 540

AACGTTTGAC TCCTTGACTT TTGAGAAAAG GAATCACTGA ATATTCTGCT CCTGATCGAC 600

TATTTTTCCG TTGCCGCCAA GCTTCTTTGC CAAAAGACAA CCGCCCCCCT GTATCGATCA 660

GTGTATTTTT TCTGTGAAAA GGCGCTTGGA TAAACAAACT ATCCCCTTGG CCTACATCGA 720

TCATTGCAAC AATGCCAATT GGCGAAAGGT ATTTACTATT TATCAAAACA ATTAGTAGTC 780

CGATTGAGAA AAAAATTTTT TTAGACTTGT CTGCTAAATA AATCATTGAA AATCCGATTA 840

GTAATAACAA ACTAATTTGT ATCCAAGAAT TTAAATAGCC GATGGTCAGC TTCGGATTTA 900

AGTGAACTAG CACCCACTCA AATAAGCCAT TCATTGTTAA AAATAGTGGT TCGAGCAAGG 960

CTGCTAAATA ACGTGCTGGC ATAATAATCA CTAGAAAAAA GCTCAACGTC AAAAGCGGTA 1020

ACAAGAACCG TTCAAAAATT GGTAACAAGA GAAATGTTAA GAAACTTCCT AACAAGGGCC 1080

ATTCATAAAA TGATTGTACT AAAATAGGTA AGGAAGCCAC AGCCAATAGG AAAGAGCCTA 1140

AACAATTTTT AATGAAAGGC GAAGAATAAC CAACCAATGT TTGCTCTAAA AACAGCAGAA 1200

CAAAGGAAAG TAAATAACTA AATTGACCGC CGACTGAAAA AAAGAGGGTT GGCTGAACAA 1260

TCAAGTGCAC TAAGAGAGTT AAACTCCAAC AATCGAGGGC CGGAACCTGC CAATTAAACC 1320

GTCGATTGAA CGTATTGAGG TTTGTTTGTA AAAGCGCCCG CATGACACTA ACTGAGAATC 1380

CTGTTAAACC AGCATATATC AGAGAAAAAC CGCATTGCCA CAAAAAATAA CTTTCTTGCG 1440

TTCGGCGAGC AATTCTTAGC ATTAGCAAGC GAAATAAACG AAGGAAAAAA GAAACATGCA 1500

TCCCTGACAA ACTAAAAAGA TGCAAAATGC CTAATGCTGA AAATTTATCG CTTATTTCaG 1560

CAAATTCTTC aTTACGAAAA CCTAATAGCA ACGCATTCAT ATACATTGCA CTAGTCTCAG 1620

CAAAATTTTT CTGAATATGT ACCATCGCCC ACTTCCGCCA GCCGCTAATG ATTTCTTTAA 1680

ACCGCCAAGA AAAAAAGGAA CGCGGTTGTC CTATCTCCTC AATGAATAAA ATAGCTTCGA 1740

TTCCTTGATT CTTTAAAAAT AGTTGATAAT TGAAGCCATT TTTGTTGGTT TGACCTTGGG 1800

GCGAGTCTAA AATACCTACA GCGGTTGCCT GAATGGTTGT GGTTAGTTGT TGCCATTGCT 1860

GTTTTTCTTT CACACTTTTT AATTGATAAA ATGCCATCAC TTTCTGCCTT TTGCCTTTTA 1920

CGATCCAATT GGCAGGAAAT TTTACTTGAT CACCATCGAC TTGTATTTCA TCTGGTAATA 1980

TCGTTAGCTG TCCTGACAAG CGGACTTCTT TAGGCAACAA AGAGACTGGC TTTAACCACT 2040

GGAAAATCAA TATGCCCCCT ACTATTAAAA AACTGCACAA AAAACCTCCC AATATAACAG 2100

GAAAACTTTT TGTACAGACA ATACGGATAA CAGTGATAGA AAAAAGAAGA GCTCCTAATA 2160

GACTAGGTGC GAAAAAACAC CACAGCCCAC AACTAGTCAG CAGAGCTATG AAGAACCAAT 2220

TATTCCTCAG CTGAGGAAGG ATTGTCCTCA AGTGTGGCAA ATTCTTCTCC CCATTGTAAT 2280

TCGGCAAAAT ATTTTGGTAC TAAAGTCACT TTTTCAACCG TGGCTCCTAC TTGTTCAATT 2340

AAATTTAGGG CGTACGCATC ATTACGGTAA TCCTTTAAAT AATAAATCTT TTTAATCCCT 2400

GCTTGTAAAA TCATCTTTGT ACATTGCAGG CAAGGGAAAT GTGTCACATA AATTTCTGCC 2460

CCTTCTGTAG GGACGCCAAA CTTCGCACAT TGTAAAATTG CATTCATTTC TGCGTGGATT 2520

GTTCGAACGC AATGATTATC AACAACATAA CATCCTTCAT CAATACAATG CGTTCCGCCA 2580

CTGACTGAAC CATTGTAACC TCCTGCAATA ATTCGTTTGT CGCGAACAAT GGTTGCTCCG 2640

ACAGTTAACC GTGTGCAAGT ACTTCTTAAT GATAATAAGA CGCTTTGTCC CATAAAATAT 2700

TGGTCCCATG GAATTCGTTC CATAATAAAC CTCCTTTATT TTTAATGACT CCTGTGTTCG 2760

ATTTGCCAAA TCATTTAATC TTATTACAAA TAGTATGTTA TCCGCATCCC ATAGTCAAGA 2820

ATTCTCAGGA ATTTTCTGTT AAAAGTTAGC CAGTCACTGT CAACATGTCT TTTAAACGTT 2880

CTACTGTnTT CTCTCCAATa CCTGGgAACA TTTTTCaGTT CCTCGACGGT TTTAAACGGT 2940

CCTTGTTCGT CACGAAmACG AATAATCTCT TGTGCTTTTy TCGCACCGAT TCCTGAmATC 3000

GTTTGTAGCT CTGCTTCTGT TGCCGTATTA AGGTTGATTT TTTCCTCCTG ATTTTGCTGA 3060

GCAGGTGCTG ATTCCTGCAG CGTTTCTAAT GATTGTGCCA CTGGnTCTCC TTTTTTGGGC 3120

ACGTAAATAA TCATTTGATC TTTtACTTTT tGrGCGTAAT tGACCTGGGT AGTATCAGCT 3180

TCCTcGtCAC GCCTCCAnTA AAGtAAAGCA TCCCAAATCC GCTGCTGATT TTTTAGTTGA 3240

TAGATTCCTG GTACTTTCAC TGCGCCTTTA ATATCAcATA GATCATTGTT TCTGCCCTTT 3300

CTTTTGTGGC ATCTGTGGCG ACCTCTGTCG TAGTGGTCAA CTTGTTTCTw CmcACATTGT 3360

GGTGTCTACT TGTAGCTCCT TATTAACCAT CAAATAAATG CCCACTAATA GAGTTATCAA 3420

AATGACCATG AACACGCCAC TACCTAGCAA TAGCCACTTT TTAGGTAACT GTTTCAACCA 3480

ATCCATCGTT TGCCCTCCCT TCAAAATATA ATTACGCATT TTTTAACAAA AAAATGCGCC 3540

TGGACCAAAA ATCACTAGTA ATTTTTGGTC CAGGCCTACT TTTATTTTAA TTTCTCTAGA 3600

TAGTTCAAGG CATCTTGAAC CgTTyTTACT GGGACAATTT TCATTTTTGT CCCAATTTTT 3660

TTCGCAGCTT CTTGTGCTTC TTGATAATTC GATTTGATTT TTGGCTCAAC TTTTTTCATT 3720

TCTGAAGTTA TTTCATCATC AGGTGCAAAG AAAATTTCTG CGCCGTTCTC ACTGGCAGTT 3780

ACAACTTTTT TGTCAATGCC TCCGATTCTA CCAACGATTC CTTGACTGTT CATCGTACCT 3840

GTTCCAGCAA TTTCATGACC TTTACGTAAA TCTTTATGAC TTAATTGTTC ATAGGTCTGC 3900

AAAGAAACAT TAAA 3914

(2) INFORMATION FOR SEQ ID NO: 301:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17764 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 301:

TCGTGGCAGG AAAAATGTTG GnACTTTATC ATTACGGATT AAAAAAATCn ATGnTCCTGA 60

AAAAGCTGTC ATGACTTCAG GTGGTAATTT TTTAATCGTT CTGGTAACCC ATTGAAACCA 120

CTCCTTCTTT TTTATCATAC CACAACATCC ATTTCTCTAA AAGCGCAGCA CTTACTACTG 180

AATATCTGCT AAAATTTGCC TGATAATCTG CTTTTCAGAC AGTTGTTCAG ATAACCGCTG 240

TTCTCCCGGA AACCCTAAAT AATCTTCTGG TAGCCACCAA CGTTTCATTT CCACGACACC 300

AAATTCTTGA TTCTTGTTTC TCTGGCTATG TCTTTTCAGC GTTTCTTCAA AAGAAATATC 360

AAAATAATAA ACCTGAACGT TTCCTTCAAA GAGATGGATT ATTTCCTGAA AAAAGGAATG 420

ATAAATGGCT TTTTGAAAAA TCCCTTCAAC AATCACATAC TGGCATTCTT GATACCCAAA 480

```
AGCTACCAAC TGTTTTAACA AAGCAATACT TAGATTACCT GTTTCATCTT TAACACGAAG     540
CATTTCTCTG CGCACCACAT CTTGAGAAAT CAATAACGAT TGTTCTTTAA GACACTGATG     600
AAGCGCGTTA gcTATCGTTG TTTTACCACT ACCAGAATTA CCACGTAAAA TAATTAATTT     660
CGGTTGCATC TGTTTCCCTC CATTTTCTTG ATTGTTGAAA ACCAGCACAA AAGAGCAAAA     720
AACTTGGAAA AAACTCTCCA AGTTTTGCTG TTATTTTGAT GAATGTTTCG TTGAATCTCG     780
CTGAATGAAT TCCGTTCCGA CAATAATTTT TCTAGGAACT GGGGGGGGAT tCAAGCATAA     840
CTCTTTCATT GTCTCAACGG CTAGCTCACC CATCCATTCA GTATGAACTT aACTGTkGTT     900
AATGTTGGTG AAACATATTT GCGACACTAA TATCCtTAAA ACCAATGATA GAAATATCTT     960
CTGGCACACG AATGGCTGCT TCTTGTAACG CCCGCATTAC ACCAATCGCC AATGTATCAC    1020
TTGcAGCGAA AAAAGCGGTT GGACGAGTGT CCGCTTCTTC AATAAAGTCT TTCATCGCTT    1080
GATAGCCTTC TTCAACAGAA AAAGCTGCTT CGACAACCCA TTCTTGATGA TAAAGGTCCA    1140
GATCTTTTAA GGTTTCTTTA AAAGCGACAA AACGTGGATC AACCAATGCC ACGTGGGACT    1200
CTTTCGTATA TTCTTGCCCA GAAAGCATAC CAATTTTCTG ATGCCCTTGC GCCAAAAATT    1260
CTTGCACAAT TGTTTTAACA CTTCCAGCAA AATCAACAAC AATGGAGTTT TGGCCAAAAG    1320
CCATTGCATC AAAGTCTACA AATAGAAGAG GTGCGCCTAA GGCGGCTAAG GCCGCCACTT    1380
GCTTTTCATC AAATTTTCCT AGGGCAATTG TCCCATCAAA TCTTTTTTGA GACAGGTTTG    1440
TCAGAGATTC TTTGACTAAA AGAAAACCCA ATTCCTCTGC CCGCTTTTCA ACACCTAAAC    1500
GAATCGATAA ATAGTAAATA TCATCCAGTT CTTCTGGATC GCTGAACCAT TGGACAAACC    1560
GAATTTTCTT TTTGCCTGCT TGTTTTTTTC TGGCTCGCTT TTTATAATTT AACGTTTCTG    1620
CTGTTTCAAA TATTTTCTTC TTCGTTTCAT CACTAACTGA TAATGTTTCA TCATAGTTCA    1680
AAACTCTCGA AACTGTCGCC GAAGAAACAC CGACTTGCTC TGCTATATCT TTAATGGTTA    1740
CGATGTTCCT CACCTCAAAC TAATCATTTT CTATACATTG AATAAATCGT TCAAAAGCAC    1800
GACGACCTTC TTCGTCACGT TTAAAGACGC CGGCATCTGT TAAAATTTGA ACAAAAAATTG    1860
TCCCAATTCC TTGTTGGATC ACTTCTTTGA CATTGGCTTC AGTGATGGTT TGCTGACTGA    1920
GTTCATCCGC CCAAGCTCGA TGACTTTCAG CAATTTGATT GGGTTGCTTT AATAAATACT    1980
TTTCGACTTC TCTTAATTCT TGTGCCAGCC TTGGTGGCAA AATAGCTAGT CCCATTACTT    2040
CAATTAAGCC AATATTTTCT TGCTTAATAT GTTGTACTTC TGGATGTGGA TGAAAAATGC    2100
CATCTGGAAA TTCCTCCGAA ACATTATTAT CTCGTAAAAC TAAATCCAAT TCAAATAATG    2160
ACCCTTTGCG TCGGGCAATT GGCGTAATCG TATGATGCGG TGTGCCATCT TTTGAATAGG    2220
CCCGAACTGA GACACGTTCA TCAGAATAAT TGCGCCATTT GGTCAAAATT AATTCGGCCG    2280
CCTCAACTAG TTCCTCTTGA CGATCACTTT GAAGACGAAT CACAGACATC GGCCATTTTA    2340
AAATACCAGC AGTTATCTCA GGATATTTTT TCAAAGAAAA TCTTTTTTCC ATTGGTGCTA    2400
AATCCATTGG AAATGTATGA CGTCCTGCTT GGTAATGATC ATGAGATAAA ATAGAACCAC    2460
```

CAACAATTGG CAAATCTGCA TTAGAGCCCA CAAAATATTC AGGGAATAAC TCCGTAATTT 2520

TAAGGAGTCG CTGGAACGTT TTTTTCTCGA TTTTCATTTC TCGATGTTGT TGGTCCAAGA 2580

AAATCGCATG CTCATTATAG TAAGCATATG GCGAATATTG AAAGCCCCAA GCTTCGTCCT 2640

CTAAATTTAG TCGGATAATC CGATGGTTTG TACGTGCTGG ATAATTTACA CGGCCATAAT 2700

AGCCTTCATT CTCCATACAT AACATACACG CTGGATATTT TTTTTGACTC ATTTTCTTTT 2760

CAGCCAAAAT TTTTCGAGCA TCCTTTTCTG GCTTCGATAA GTTAATGGTA ATTTCTAGTT 2820

TACCATAGGG TGAGTCCATT GGGAAAATAA TATTTTTGGC GATTGCTTTC GTTTTAATGT 2880

AGTTATTCGT TTTACATAAC TGATAAAAAT AATCCGTTGC TTTTCGTGGA GACTCTTGAT 2940

AATAATCCAT AAACTTCTGA TTAATGATGG TTGGGGTCGG CGTAATAAAA TCCATTAACT 3000

CCGCTTCTAG TTGTTCTGCT TCCGTAATTG AGTCATCGAT TTGGTGATTG GCTTGTGCAA 3060

TATTTAAAAG ATGTTCCACC ACACGCAACG GTTCTTGCTC ACAAGCAACC GCTTCTGGTT 3120

GTTCAAAAAA ATCACAGCCA ACTAGTTTGG CAATCCGATT CGTCAAATAA TAGCGGTCTG 3180

TTTCCGTCCA ATCGCCATTT TCTATGATTA TCTCTACAAA TGCAGCAATC CATTTACTAC 3240

TTTCTTTCAT CATCTGATGT TCTCCTATTC ATTATAGCCT TGAGGATGTT TCACATGCCA 3300

TTGCCAAGCG GTGGCGATAA TATCCTTGAC CTCAGTTACT TCTGGTTGCC AGCCTAAAAC 3360

TCGTTTGGCT TTTTCACTAG AAGCAATTAA CGTACTAGGA TCCCCTGCTC GGCGTGGGGC 3420

GATTGTCGCA GGAATTTCTT GACCTGTCAC TTCGCGAGCA GCATCTAACA TTTCTTTAAC 3480

AGAATAGCCG TTGTTGCTAC CCAGGTTAAA GACGTCACTT TCGCCGCCAT TTTTCAAGTA 3540

TTCTAAAGCC AAAATATGTG CTGCAATCAA GTCTTCAATG TAAACATAAT CTCGAATGCA 3600

CGTGCCATCT GGTGTATCAT AATCATCCCC AAAAATACTT AGCTCTGCTC GTTGGCCCAA 3660

TGCCACTTGT AAAATAATTG GCACAATATG CGTTTCTGGC GTGTGATCCT CACCAATTGA 3720

GGCATCTTTT TTTGCTCCTG CAACATTAAA ATAACGCAAA GCAACATATT TCATTTCATA 3780

AGCGTTGTCG CACCATTTCA TGATTTTTTC CATCATTAAT TTACTTTCCC CATACGGrTT 3840

TTTAGGATTC GTTGGGGTTT CTTCTGTAAT AGGCATTGCT TTTGGTTCGC CATAAGTGGC 3900

CGCTGTGGAA GAAAAAACAA TGTGTTTCAC CCCGAATTCT TGCATGACTT CCAACGCAAT 3960

TTGAGTGCCG TGAACATTGT TATTGAAATA CATTAACGGT TTCTCCACGG ATTCTCCTAC 4020

TAAAGAATTG GCCGCAAAGT GCAACACCCC TTCAATTGAT TCCTTTTCAA AGACGCTACG 4080

TAAAAAAGCT TTATCACGAA TATCGCCTTC ATAGAAAGTC GCTTGTTCAT GAACTGCTGA 4140

ACGATATCCC GTTAATAAAT TATCAACAAC AATAACGGCA TATCCTTTTG AAATTAATTG 4200

ATCGACGGCA TGTGAGCCAA TGTAGCCTGC GCCACCTAAT ACTAAAATAG ACATAAATAC 4260

GCTCCTTTTA CTCCCTAATT TCTTTTTCTA TTGTACCATT AAATAAGGTC TACTTCTGAA 4320

TTATTTCATT TTGTAAACAG GATTTAATTT GGCCATTTCT CCTTATTCAT TTTTCCTATA 4380

AAAATACAGA AAAGACTTGG TAAATTTTAA CAAATCTCTA AATTAGTCGT AGAGTTTATG 4440

```
AAAAATCATT GTAACTATTG GATTTTTTTG CTTTTTTCCG TCATAATAGA AACAAATACA      4500
ATTTGAAAGA AGGAATGAAC GATGAAAAAA TTTGTATCAG GTATGTTAGT CGGTACTGCT      4560
ATCACAGTTG CTGCTTTAGC CGGTGTGGCT ACAACTATTA AAAAaACAGT CATTGATCCA      4620
ATTGAAGAAA AAGAAGATAT GATTGAAGAA AACCGTAAAA AAGCGATGCG TAAGCGTATT      4680
GCACGTTAAA AAATAGCTGT GACGTTGAGA TAAAAAATCT AACGCCAAGA AATCAACAGA      4740
AGCGCTTCTG GCTCTGTTGA TTCTCTTAAG mATAACAAGT GGTCATCCAT CGCTTTGATG      4800
GATGACCACT TGTTATGCTT TCTTTTTTAC CAAAGGAACT GACTATTTTT CTTAACTATT      4860
AAAAATAGTG GCAGCAATCC GATCCATATC GCGCATATTC CAAGGATCCA CACCAGGACG      4920
AGCTGCAGCG GCTGTTAACC AATGAATACC TTCTGTTGGA ATACCACAAC TAAAAAGGCC      4980
TGTCGCCACC TGTCCATCTT TTGTCATGAC TTGATTCGTG GCAGGCACAA TCAAGAGTGC      5040
ACCTGTTTTG ACTATTTGAC CATCGGCTAA TGCCAGTTGA TGGAGCGTAg cTAGCTTATC      5100
GGCTAACAAC TGCCTCGTTA AAGGATTCAA ACTTTTTTCA TTATTGACTT TGGGAATGCG      5160
CGCTTCAATT AAAAAGTGAC TATTATACCG ATTTTCCGGA AACAAATTAG CATAAGTAAT      5220
AAATTGACCC GCTTTCATTT CGACACGCAT TTCTGGCCCC AGAATCGTGA CTAATCCCAG      5280
TTCAACTAAC GCCTGTAGCT CTTCTGTTCT TTCAACAGAA GGTCCCACTG ACAAAAAGGT      5340
ATTCAACGGG GTAAACCAAT GCCACAACCA CTTTTGACTC TCGTCATTCG TTAATAATCC      5400
TTCATCCAAA ACAAAGCGAA TCTGATCTCT TAAATCTTTT AAACTGTCTA AAGCAGTGGT      5460
CAAAGGACCG CTAAGATTAC CTTTCTCTGC TTCTTGAATA TCCCATTTTA AATACTCTAG      5520
TAAAAACGAA CGAAAACTTT CATGAGTAGT TTTCTTTGTT TCAGGATGTT CTAACGCTTG      5580
CCAATCAAAC CAATCTTGTT TTTCAAAGTG GGCTAGGCTT TCCGGCGCTC CCTTGGTTTG      5640
AATAAACAGC TGTTGAAAGG CCTCTTTAGA AAGATGCGGA TATTTTTCTT CGATTAAGCA      5700
AGTGTAATAC ACATATTCCA CTTCTTTTTT TAAATACCCA AAAAACTCTG CGCCTGTAAC      5760
CTGTCCTTTT TTACGACATT CAGCTAAATA CGTTTCGGTT AAAAAAACAG GCTGATAGGT      5820
TTCCCCGTAG CCTTTTTGAT TATTCCCTCG AGGATGAGAA GGAAAACCGC GACCAGAACC      5880
AGCAATAATA ATAGGCTCTC TTCCAGAAGG TTCATAACGT AAGCGGCCAT TTTGACGAGA      5940
AAAAACACCC CCGCGGCCAA TCGTCAAGGC AGATAAATAG TCAAAAAATG CCAAACCTAA      6000
GCCTTTTAAC ACAACGGGTT GTTTTGCAGG TATATTGGCA AGTGCCAATA AAGCATCCGC      6060
CGCATTTTTA GGAGAAGAAT AAAACAAACG ATGCTCCTTG GCGTAGTCCG CAAGCTCTTG      6120
TTCAACAGTC GTTAATTCAT TTTCTTGGTG ACCCAAAGCT AAAATTACTT TGTCTGCCGT      6180
AAACTCAATG GATTGAGTAT ACACATAATA GTTATCCTCT CCTGCCCGCA CAGCAGTTAC      6240
TGGCTCTTTA AAAAACGTCA GCGATGTTTG TTCGTTACAC CGCGTGGCTA AATACGTATA      6300
AAACCATTTT TGGTACATGC CATAAAAAGC CCTCGAGCAG TGATCATTCG GTCCAAGGCG      6360
ATGACATTCT TTCAACAACT CAAGATAATT ATAGGGTTGT TGCTGTTGAA TGTACTGCGC      6420
```

```
AgcTTCTGTT TGGGTCCATT CATAGAGAGT GGGGCCCTTG ACAATTGGTC CTTCGCCACT    6480
AAAAGATTCA TCTGTAAATA AAGTGACTTG ACTAATCACG GTGTTCATTA ATAAATAATT    6540
AGGTTGCTCT GCTCGCCAAA TTTTCCCACC TGGACCGTAT GGATCAAACA ACGTGAGCTG    6600
AACATTTATT TCCGTTTTAC GAGCCCATTC AATGACTCGT TCTGCAGCTG ATAAGCCTCT    6660
GGGTCCAGCA CCTACAATTG CAATTTTCAT CCAATCACCT ACTCTATTTT TCTAATAATT    6720
CTTTTAAGAC AACCGCTTGA TTATGCACTG TATCTTTCGC CCCATAGACC AGCGTGACTA    6780
CTGGCTGATT TTTTACAATC GTCTGTAACT GCTTAAACGC TGTGGCAGTT GGCTCATGTT    6840
TCAATTCTTC TTGGTATTTT TGTTTAAACT CGGGATATTT ATTTGCTTCA TGATTAAACC    6900
ATTTCCGTAG CTCTGTACTT GGGGCCACTT CTTTGAGCCA TTCATCTAAC TTTTCAGATT    6960
CCTTGGATTT TCCTCGTGGC CAAACACGGT CAACTAAAAC ACGGTAGCCA TCTTGCTCTG    7020
TCGGTGCTTC ATATGCTCGC TTGATTTGAA TCATCTGCCT TCACCTCTTC TCTTTATTGT    7080
AGCTCAAAAA GTGAGGACTG ACAAAAAGCA ACATCTCTAT TAAGCTTTCC TATATAAAAA    7140
GAGTTGGTCT CAATGAAGAC CAACTCTTTT TGTTATTAAC TATTTCTCTT TTTCACTTTC    7200
CCAGTCCAGC CATCATAGCC GCCTTTTAAA ATATAAATAT CAGTATAGCC ATTTTTGCGT    7260
AATAGATTAG CTGCTCGGAT ACTTAAAGCT TTCTTCTGAT CATAAAGATA AACAGGTTGA    7320
TCCTTACGAA GAGAACCAAT TGTTGTTTTT AACATGCTGT ATGGCATGCT GCGCGCCCCT    7380
AAAATATGCC CAGCATCAAA TGTATCTTTC TCACGAACGT CAATCACTTG TGCTTTACGC    7440
ATTGTTTCTT TAAATTCTTC TTCCGTCAGC ATTTTTGCTG AACGTTTCAC CATAATTTTC    7500
AAATATAATT CATTGAATAC CATCGCTAGT ACAATAACTA GCAAAATGCC ATTGATTACC    7560
CATAAAATAC TCATTCTAAA AAGCCCTCTC TATTTCTCTT TTGAAAATTG TAATGCTAGT    7620
GAGGCGGCAC CAATGACGCC GGCCTCATTA CCTAGCTCTG CTAATTTGAT TTTTGTACTA    7680
TTTCGAACTT GCGGGAAGGT AAATTCTTGG AAATATTTTT CGACACGACT ACGTAAAAAT    7740
TCTCCTGCTG CAGAAACACC TCCACCGATA ACCACGCTAT CTGGATTTAA CGTGTTACCT    7800
AAGTTTCCTG TGGCTAAGCC TAAATAGAAA CAAACACGAT CCACAACCAT TAAAGCAAAA    7860
TGATCGCCTT TTTCGGCAAA TTCAAAAACG TCTTTACTCG AAACATCTTG TCCATCATCA    7920
ATCGCTTGTT TCAGTTCAGA GTCACCTGCA AATTCTTCTG ATAAATGACG TGCAACACGA    7980
ACGACACCAG TGGCACTAGA AACAGTTTCT AAACAGCCAC GTTTCCCACA GGTACAGTCA    8040
AAGCCGTTAG GATCAACCGT CACATGCCCA ACTTCACCAG CACAGCCAGC AACACCGTGG    8100
AGCAATTTTC CCGCTGCAAC AATGCCTCCG CCGACACCTG TTCCTAAAGT AATAAAAATC    8160
ACGTCGGGAT TGTTTTCCCC CGCGCCCTTC CAACGCTCAC CTAAAGCGGC AACATTGGCA    8220
TCATTATCTA AGGCAAACGG AATTCCTAAA GCAGATTCGA TTTGTTCTTT GACAGGCTGT    8280
ACAGTCGTCC AGTTTAAATT ATAGGCACCT ACAACAGTTC CTTTTTCAAT GTCGACACTT    8340
CCTGGTGTTC CCATACCAAT TCCCACAAAA TCTTCTTTTT TCATATTATA AAGATCTATA    8400
```

```
CGGTGACGAA TACTTTCAAT AATTGATGGC ACAATATGCT TGCCATCTTC TAAAATATTC     8460
GTTTCAATAC TCCATTTTTG TTGAACAACG CCATCGGTTG TTAAAATAGC AAATTTGATG     8520
GTCGTTCCAC CTAAATCAAT CCCAATAATC TTCTTATCCA TTTAGTCGAT TCCTTTCTGC     8580
TTCCCTCTAG ATTCTTCTAT TCGATGCTCT CTTTGCAAAA TATGCCGGGC CTGTAAATAC     8640
GTATCTCGGT CTACCAATTG ACCTTCATAC AATTTTTTTA ATTCAATCAT CATTAACTCA     8700
ATATCATAAA TCCGTGCACC GACATAAATA TACATCCCAA ACTGTTTAAA AAGTTGTTGT     8760
ACATCATATA ACGTTTCCAT ACGAAAACAC TTCCTTATTA TACAGGTAAA CCGTATTTTT     8820
TAAAACCCAA TAACAGACAA ATTACGACmA GAAAAATAAa AATTAACGTG GCCATAATTC     8880
GCTCGTGCTT ATTAAATGCT TCTTCACGAT TTGGAACAGC AAAAGCCGTC GCGACTAATA     8940
GTCCACCAAT TAGACCGCCA ACATGACCCA TCATATCTAC AGAGCTACTA AACAAGTTAA     9000
ATAACAAATT AATCACAATA AACATTGAAT AGCGTTGCAC CATATACATA ATCGCTGGAT     9060
TGTCTCTAAA ATGGCGTCCT AAAATAACAA ATGCGCCGAA CAAACCAAAA AGCGCTGTAC     9120
TAGCACCCGC TGAGACACTG TTGGGTGTCC CAAAGGCAAA ACTGGCGATA TTTCCAGCAA     9180
TCCCACTCAA TAAATAAATT CCCAAATAGC GCCAATGCCC ATAAATTGCC TCAACCTGTG     9240
CGCCGATATA ATAAAGAGTA ACCATATTCA AAATAATATG CATAAATCCA ATGTGAAGAA     9300
ACATTGGCGT GATAAACCGC CAGTATTCAT GATTTTGAGC TACCAGTGGA CGAACCATTG     9360
CTCCCCAATT AACTAAGTTT AAAATATTTT CTGATCCTCC AGTTAATTCC AATCCAAGGA     9420
AAACAATGAT TGAAATACCT AAAAGGGCAT ACGTAATGAA CGGCTGATTT TTTAGCCGTT     9480
TCAATTTCAT TTCTGTTTGA TAATTCATTC CATTCCTCCT ATTGTATAAA TAATTGTTCG     9540
ATTGGCTGAT CAAACGCTTC TGGTTGCCAT TGCTCATGGA GTTGTTCACT AAAAACTAAA     9600
CTACATGAAT GGCCCGAAAA ATGTTCCAAA TAACGATCAT AAAAACCACC GCCGAAGCCT     9660
ATTCGATAGC CAGCACGATT GAAAACAATC CCTGGCACAA TcAATAAATC AATCGCTGTA     9720
GCTGTTATTT CTGCCGCTGT TAACGGAGGT tCTTCCACGC CAAATGCACT GGTATCATAA     9780
ACCGTCTCTG GAAAGACTTG ATAAAAGTGC ATTTTGCCTC CTTTAAATGT tCGCGGCACC     9840
GsCAcTTGcT tGCTCtCTTG CATGGsCCGk TCGAAAATTG GCTGTGTTGA TAATTCTAGC     9900
GGTAACGACC GAATCATTCC GATCGTTTGT GCTTGTTGCC ATTGCTGACT AGcAAcAACT     9960
GTTCATAGAG ACTTGATTCT TTATGTTGTT TAATTTCTGG ATGTTGTTTT AGsCATTCCA    10020
GGGCAGTTAA GCCCACTTTT CTCAACTTGT TTTTCTCCAT GTACACCTTC AACCCCTTTT    10080
CTGTTAAGTT TATAGAAAGT GTTCTTAAAA AACAAGCTGA TTCCCCTTTT CTCTTTGAAA    10140
AATCCCTTTA TTTTTCTAAC GTTTTTGACG CTTCTTTAAA AAGTCAACA ACCAAATCCA    10200
TTGTTTCTCC TTTCACTAAA TGCCGCTCAC CGATTTCCGT TATAAATTGC GTATTTCTCG    10260
CATACGTTTT CCCAGACACT AATTGTTCAA AATCAGCCAT GTCTTCATAG GGAATTTTTT    10320
CATCTTCCGT TCCATGCCAA AACAGCATCG GTCGTTCTTG TAGTCTTTCT GGCGTTTGTG    10380
```

ATAAATCGTA TTGATGCACC CAACTTAGCA ACAGTGGCAA GTCTTTAGGT ACAAAAATAG 10440

CAAATTCTTT CGCCCGTTGA ATCACTCGCT GAATATAGCG CGCCGGATAG GGTGTTCCCA 10500

TTAAGCAAGC AGCAGCTTGA ATCTCAGGAT GTTGGGTCAT AAGAGCGCTG GTAGTAATAC 10560

CGCCCATGGA TACACCACCC ACGCCAATCC AATTTTCTTT AATCAACCCT AATTTATGAA 10620

AATGATGGaT AATTACTGGA AATTCAATTA AATTATATtG AATGCTACTC CAAAATGTAA 10680

TCGAAGGaAT CGTGGaAATA gGTtCTGnTT TTCtTTCTCC ATGGtTCATG GcATCTGGTA 10740

AAATGACCCG AAACCCCTCG TGTGCTAATT TTCTTGCTTG TGTTAAAGAT AATTCTTTAG 10800

CTGaTTGCCA ACCATGATAG TAAACAATTA AAGGTAATGC TTCGTTTTTT TGCTCTTCTG 10860

TTgTCACTTC TAATGCTGGA ATTTTTGCTA TGTAGCGATG TCTAATACTG ATTTTCATTC 10920

ATGATTCCTC CATTATAAAA TATCTTCTCT TATTTTACCG TTTTTTGATA AAACCTGCTC 10980

TTTATATGAT TTATCAAAAA AATCACAAAA TATGCGAATT CGCATTTTCC TATTTTATTA 11040

CTTTCCGTTA TAATCATTTT ATTAAATACA AAGGGGTGTT TTTATGGTTA AAATTGAGGA 11100

AAAAGGTAAA GTAACATTTG GACAAGCTTT TAAAGATTAT TTTCGTGGAT ATGTTGATTT 11160

TAAAGGTCGG ACCACACGTG CTGGTTATTG GTGGATGACA TTGGTATTAA GCATTTTAGC 11220

ACTTATTTTT TATATTGCCA TTGTCGGCAA AGCTGtGTCA GCAATTTTAG CAGCTGAATA 11280

TTTTGAAACC TATGATTTTG GGAACTTATT ACCGTTGATG CTTTTgCACT TGTTTTATGG 11340

TTAGCTTTAT TATTAcCAAC GTGGGCAtGT GTGTcCGTCG TTATCGTGAT GCTGGGATGA 11400

CAGGTTGGGG CGTGCTTGTT TTATATCTAC TTTCTATTGC CTGTAGCTAT ACACAAGTTT 11460

TCTCTGTTAT GTCAACACTG AAATATGATG TTCAAACAGA TACTGTGATA ACAGGTGGCA 11520

GTCCCGTCTT CTTATTTTTC ACTCTTGTGA TTAGTCTGTT TTTCTTCTTA CTAACAGTCT 11580

TACCGACAGA TAAACTAACA ACAACTAGCC AAAATAGCGT CTTACGCTTC TTCTTCCGTT 11640

ACAAAGAAGT GAAGTAAGGA AAAGTTTTAA GCCTTTCTTT TAAAAGAAGG GCTTTTTTCT 11700

TATTTTAAAA TCTTTATGAT AAAATAAGAA CTAGTCTAGC TAAAAAAATG GAAGGAGAAA 11760

ATATGGCTAC TTTTAAACGG ATTCTGAATC TAGAAAATTA TATTTTTACT TTTTTTATTT 11820

TATTTGTCTT TTACGGCGCA ATTAACTCTC CCTATTCGCT GGTAAAAATT CTTTGGGAAA 11880

AGCCTAAGCA CTTGCTTGTT TTACTTCCTC TATTAATTAT TCTATTGTTG ATTATTTATA 11940

AAAACAAAAA ATTTCTTTAC ACTAAGATTC AAGCGCTCTG GCACTTAATA AAAACCCATG 12000

ATCGTGTGGT TATTATTGGT AGTTTGATCA GTTTATTTTT ACTACAACTT TTACTCCTGA 12060

CGCAGATTAC TGTGCCGATT GGCTGGGATG TTTTTGACAA CTTTCATAGT ATTACCACTG 12120

AAAATAAGGA TTATTCAAAA ATCGTACTTT CTTTAAACCC CAATAACGAA TTTTTCTTTT 12180

TCATGATGTA TTACCTTAAT AAATTTTTGC GTTTCATTGA TGTAACAGGG AGCTGGAGCA 12240

ATACTTGGTT TAGTTGGCAA GTGGTTAATT GCTTGTTTAT TAATAGCAGT TTATTTCTCT 12300

TTTATCATGC CTCTAAGCGT GTCTTTAATC CATTAACGGC CTTCGTTGCT TACAGTTTAT 12360

TTTTTCTTTC TTTCGGATTG TCTCCATGGT TACTCACGCC TTATACAGAT ACGGCCGTCT 12420

TATTGTTCAT CAACCTAGTC TTTTTTGCAT ATAGCCTTTT TGACCAAGTT TCCCCTCCCT 12480

TCGTAAAATA TTGTTTATTA CTGTTTATTG GGATTGGCCT AGCTTGGTGT TTTTTAATGA 12540

AACCTTCGTC TATTATCTTT TTTATTGCAT TTAGTTGTAT CAAAGTGCTG CAACTTCTTT 12600

TAGTAAATCG GAACAAACAG TCTATCGTAA AATTAACCGT GGTTGCTCTC TTTTTACTGA 12660

CTGGTTTCGC CAGCGCATAT TATAGTTTTC AATTTTTTGT GGAAAAACAA ACGATTACGG 12720

AAATCGATAA AGAACAAGCC AAACCTTGGA CACTTTTTGT AATGATGGGT TTGACTGGAA 12780

CTGGCGGTTA CAATGATGCA GATACACAAG CAGTAAATCA GTTGCCTACA CAAGAAGCCA 12840

AAAAAGCTTA TACTATAAAA ATGATTCAAG ACCGTTTAAA AAATAAAGGA TTCTTCGGTT 12900

ATCTCCGTTT TCTGGCTCAA AAAAATCGTC ATAATACTGC CAATGGTGAT TTTGACTGGG 12960

GTTGGGATGG CGGTGATTTA ATCCCAGAAA CACCTTCTAA AAATCGTTGG CAAGAACACT 13020

TGCGGTCACT GTATTATCCG CAAAATCAAA AAAGTAACTA TCTACGAATT TATATGCATT 13080

TTTTCTATTT aCTcACATTa CTaGGTcTAT TGTTCAGTAT CCCTCTAAAA GATTCAAAAA 13140

ATAACTACGC GATTTTAAAA TTAGCTTTCA TTGGAGCTAT TCTGTATTTA TTGCTCTTTG 13200

AAGGCGGCCG CTCTCGCTAT CTAATCCAGT TTATGCCGTT TTGGTACTTG TTATCAGCGA 13260

GTGGTTGGCT GGGATTGAGA GAGATTCGAA GATACAAAAA AATAGTAAAA TAGAAAAGAG 13320

GATTGAGTTA AAAATGAAAA AAAGTTTGCT GGGGTTGATT GTTTTTTTAT CCTTACTGAC 13380

CCTTACTTCT TGCAGAAATA AAGTGACCAC AAAAGAGCTG ATGGCAAACG AGTGGGcCGT 13440

AAACTCCAAC GTTGATGAAG TAGTAATGAT TGTATCATTC aGCGAaGATA CCGCTACTTT 13500

CAAAATCAAT ACAGnTGAAC ACaCATCAAC TGCAAAAAAT GAGTTGGnAA AAGCAGGCGA 13560

AGAATTAGGT AAACAAATTG CAAATAAAAT AGAATACAAA GTCAAATACC ATCTAAAAAA 13620

CAATCAAATT CGTTGGGAAA ATGAAGGAAA AGAAGTAGCT TACAAGATAA AAAAAGAAAA 13680

GCAAGATCTA CTTTTCACTC CTACTAAGAC AAACAATTCT GATAACCAAA CAAAACTAGT 13740

TTTGAAACCT TACACAAAGA AAAGTATTGA TTCTTCTACT CAAAAAGATA AAACGGAAGA 13800

AACCAGCTCT AATTATCAAA ACGTCTCATC TGAAACAAAC CAATCTACCT CTTCATCTAC 13860

TACTAAAGAA CCGCTACCAC AAGTTAGCTT AGCTGATTTT ATAGGCGGTT GGGGTATTCC 13920

TCAAAGTGAT AACTTATTTT TTATAAATGC TGACGGAACA CTCACTAGCA TAACTCAATC 13980

GAATGTTCCT CTTCAAAATG TAAGTTTTTC TGTGGATGAG AATGGTAATC AAATAATGAC 14040

GTTTCTTTTG AATAATACGC CCCGAACAGT AACGAAAAAT AATGATGGTA CTTTAACTGT 14100

TAATGGACAA ATATACACTT ATCTAGGTAA TATTACGTTG GAACAATTAA TTGAAAGAAA 14160

TAATCAAACT CAACAAGTTT TTGAACAATC TGAACAGCAA CCACCACAAA ACTCTGATTC 14220

TCGTGAACAA ATACAAAATT CTAAATCaGA CCAACCTATA TACGATACGG TACGAAGTGG 14280

TGAAGGTGGG CGACAGTTAG CCGAAAGAAA TGGTTTAACC TTAGAAGAAT TATTAGCATT 14340

AAATCCAGGC ATTGAAACTT CTGTTTTTA TCCTGGTCAG TCATTACGAA TTAAATAGAA 14400

ATTTAGGTAT ACTAATAGAT AATAGTCTAT AAAAGTAGAA CTGGACCATC TACCATTTAC 14460

TTGGGATACC AGTTGATGGC AATAGTTTTG ACGAAGTGGA ACTAGACCTT GTACCTACTG 14520

GCTCAGCAAA AATTGCGACA TTTGATATAG CAGTTGGGCA TTGATTAACA GCATGCATAT 14580

AATTTGATAA TATGGAATAT ACACAAAAAT GTCTAAGGAG CCTAACATTC TTTCAAGAAA 14640

GAATGTTAGG CtCCTTTTAT TCAGACAATT TCAGCCACTC CTTGCGATTG CAATAGGACA 14700

TAGTTTTTAC AAAGACAAAA TGATAAAGCA AACCGTAAGC TAGCCCTAAC ATATTTAACA 14760

AAATTCTTAA GTAAAGACTT GTTTCTAGCC CAAATATCAT AGAAGCTGCC ATGATAGCAC 14820

CAACACAATT AAAAATCGTT TTACTTTTAT AAGATGTGCA TAAGCCTAAA CAAAGTCCTC 14880

CTAAAATACC AAGTAGGTTC GTTGGTATAT AAAATAACAA GATAGCCGAT ACTAGAGAAC 14940

CAATAACGAC CCCTATTATT CGTTCTTTTG CTCGTTCAGA CAATTTAAAC GTCTCATACC 15000

CAGAAAATAA TGACGAACTA GCAAATGTTG CCCACATAAA GCGGTCAATC TGAAGGTGCG 15060

TTCCTATAAA AAGTAATAAG CTAATGCCTA AAGCGTAATA ACCAAACCAA ATATTTCTTT 15120

GATTAAAAAA ACCATTTTCT GTAACCATCT GTATAAAAGT AATCTCTTGA TCCAACTTTT 15180

TATGTTTCAC ATGATAAACA AAAGCTAAAA GTAAATAGGC AAACACTAAT ACAAAGAAAG 15240

TTTGTTCTAA TTGCTGAAAC GATTGATAGT GAACCGTGCC AACTAAATAT AAGTAGGAGA 15300

ACGTATACAA GCCCGGATTA CCCATTTTAG GGTTTTTACC AGTTAAGAAA AACAATGCCA 15360

GCAAGCAGAT AAAATGAAGA CCCCATTGTA AAAATGATAC AGAAATCAGT GAAATTAGTG 15420

GACTAACACC TAGAATTGTT AACACGATTC CTAAACTAAG CAAGGCCTGT TTTTCCCTGT 15480

AGCCATATGA GACGAAACGG ATACTCAACA GCAAACAAAA TAATACAATT GAAAAAGGCG 15540

CTACAGCTTT TCCAAAGAAA AAAGTAATCG TCGACACCCA TAAAATCGCA AACGAAACTA 15600

ATAAAATATC CCTTATTAAC AAGGCGCGCC AAAAATATCT CCGCTGTTTT TTAGTGTCCG 15660

CTTGATGAAT TTTTTGTTTT AAAATAAATG GATCTAACTG CAATAATTGA TAAAAAGTCA 15720

CCTAAAATCA TCTCCTAAAA ATCGACTCTT CTCAATTCAT TATTCCAACT TTTCCytCtA 15780

TGaAATCaTA AGaAAAgaAG TkGTGACGAA CATTTTAcTA CaACTTTTTT TCTGAATCAA 15840

TACATGCTTT ATTTCTTATT CCaATACCtT TCCCgATACG ATTTGCTTAC TTCATAAAAA 15900

ATAAGCCTAG AAAAAATCAA AACGATTAAT TCTAGGCTCA AAATTAAATA ACTGTTGAAC 15960

AAAAGctAGC GCTTTTTTAC ATAAACTTTT CCAACAGATC CATTTCTTTT TCTAATTCTT 16020

TTTCATCATA CTTGTCATAC TTACCAGCTT CGTGTGCAAT TTTTTTAATC TCATGAATGG 16080

CTTTTTTTTC AACAATCCAT TTCTTCATAC TATGTTTTTT TGTATCCTCA TCTAGCGAGC 16140

TCAACTCTTC TAAACTTGCA TCCAATTTGT TCAAAACGTC GCAATTTTTG CTAATGCTTG 16200

CGCTTCTTTT TCTTCATAGT TTGACATAAT AGACACCTCT TCTTAAGTAT TTTTCTACTT 16260

TAAGTGTACG TCTCTGTCGT TTTATTTGCA ACTAATAACT CAACTGCTTC TCGGTACGGA 16320

GGATAAGGGA TTCGAACCCT TGCACGATGT TACTCGCcTA ACGGTTTTCG AGACCGTCCC 16380

CTTCAGCCAA ACTTGGGTAA TCCTCCCTCT ACATTCCATC TACAAATGAA CATGTAATGA 16440

ATCTATAGAT AAAAAGCCTA GACCCTTACA CAACAAGGAT TCTAGGCTCT ATCTAACTAA 16500

ATTATTTAGT TTCACGGTGT AATTTTTCTG CAACACCATA TTTAGACAGA CTAGAACTCA 16560

CAAGGGTTTG CGTAATATCA TTTTAATTCA AATTAGCAAA ATAGACTATA ACAACAAGAT 16620

AAAATCTACA TCAAAAAATT AAAAAAGGCA AGCGAATAGA TATATAAATC ATACCGAATA 16680

AACAAAAAAA GCTCTACTTC CTGTGATTGA GAAGTAGAGC TTTTAATTTA TCTATTCAGC 16740

ACCGTTAACA CTAAGGTGCA GATACCTAAT AAAGTTAAAC TAAATGATAA TAGTTCTTTC 16800

TTTTGCAAGC TCATGTTTTT CTTCATCTCG TTAATTCCCT TTAACACTAT TATAGCGACG 16860

ACAATAAAAA GTATATAGTA CGTTCGCATA CCTGAAGCCC AAACAGTAGG AGAAAAACCC 16920

ATAATTAATC TTGAACAAAA ACCCATCATC ATTAACACTA CTAAAAAAAC AGCAGATTTT 16980

TTATCATCAA AACTTATATA AATTCCTATA ATTAAACATA ATAATAATCC CAAAACTATA 17040

AAAGTAGCTA CCCAAGTTCC AGGGTGATAA ATTGATAGTT TAGTTCCAAA TCTTGTAAAG 17100

ATATTCCCCA GATTGTTGCT ATTCCAAATC ATTGATCTTT TATTTCCATT GACATAAGTA 17160

AAACTTTGTC CCATTGTGTT TCCGAAAAAA GTAATCAAAT TTAGAAAAAG CGGCAATGCG 17220

GTTnATATAC GTATATAATA GTTCTGACAT TTCTTATAAG ATAGAAAGAA TAATATAAAA 17280

AAGAGTAATA AAAATAAAAT ATTTGTGTCT AAAAAAAACG GTTTTCCAAA TGATGATAGG 17340

GAGCGTCAAT AATTTTGTGT AAATAAATTG TCCTCCTGCA AAATAATTAG TTACTCAGTA 17400

AACATTGAAA CTAATGTATC GGTTACCTGT TGAAAACCTT TATGGCTTCT GTTTAGAAAT 17460

TTTTGATTGT ATGTATCAAA AATGCTGACT AGAAAGCGTT CTAGTGATTC TTCATTTTGA 17520

AACTGCTCTT TTCTACGGCT GTATCTTTTA ATTTGCTTAT TGAAAGACTC GATTAGATTG 17580

GTTGAGTAAA TGGTTCTACG AATGCyAGGT GGrAAATCaT AAAAaGTTAA TAAGTCyTGG 17640

TTTTCTATGA GTGACTGCGT CACTTTAGGA TAGTTTnnCT TCCATnTCTC AATCATGCCG 17700

GATAAGAAGG TATTCGCTTC TTCAnTGAGT TAGCTTGGAT AAACAGCCCT AAAGTCATCA 17760

CAGA 17764

(2) INFORMATION FOR SEQ ID NO: 302:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3327 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

GGGATTAACT TGCCTCnTAA GTTTAGTGCA GATTnCGATA CAAAATTAAG TGCTGAAGAA 60

ATAGCTACTC TTGAAAAAAC GGCnCTAGAA ATGAACAAGA ATTTTCCAAC GAGCAAAGAA 120

```
GATGAAAAAA ATAAAGACGT GATGTGGGAT ATTCAACATT TAAGTGCAGA TCAGAAAAAA    180

GAATTAAGTG TATACACAAC AGAATTATTA AATGATGTGC GAAAAAAATT GGGTCTTTCT    240

CAATTAAGTG TATCGGACCA ATCAATCAAA TTTGCTTGGG ATATTGCGAA ATATTCTGAT    300

ACCGGAGAAT ACATGCATGA TGTGATAGCC ATTAATAAAG CAGCTAAGGA AAATGGCTTT    360

AAAGAATATC CTGGAATGAA TTACTATGAA AACCTAGGTG GGGGGTATTA CGAAACTGAA    420

AATGGTAAAG TTTCTAAATA TACTCTCCAA GAAAGTATCC GCAAAATGCT TGTCAATATG    480

TTGTTTGATG ATGGCCGCTT AGGTTATAGC CATTTACATT CTTTATTACA AGACGGAAAA    540

ACAGCATTAG GCGTTTCACT ATCTGGTGAA AAAAATTCTA TTTCTCCTAA GATTCATATT    600

ATTTCTTATG GTAAAGAAAA ACTGGAGGAT AGTAGTCAAT ACCAAAACGG CGAAGTGGCA    660

AGCATGAAAT CTAAAGAAGA ATTACAACAA GAGATAGCTA GCAACCAAGA AAAATTAGCT    720

ACCGCACAAC AAGCAGAATC AGACGCACAA CAAGCAAGAA GTGCAAGTCA GCAAGCCTTA    780

AATACAGCCA AAACAACACA AGCAACAGCA GAAAAAGAAC TATCTGTTCA TAAAGCGACA    840

TTGGCTAATC TTCAAGCAGT TGCGACTAAr AGTACGACAA ACTATGAAGA AAAAGTACGA    900

CAAACTGCrr CGGCAGAAAC AAACCTTCAA CAAACAAAAG ATCAATTGGC AACAATCAAT    960

GAGTTAATTC AGAATCGAGC TGTTGTTTTA GAAAAAGCGA AAACAAAAGT TGCAGACGCA   1020

CAAGCGATTG AACAAACGTC TGCTAAAGTG CTGAAAGAAA AACAAGAAGC CCAAAAAGCA   1080

GAAGAAAACA CATTGAATAG CTTGAAGGAA GTATTGGATT TAGCAAAAGA AAACTTAAAT   1140

CAAAAACAAG TTGCGCTTAA AACAAGTACA CGTTCATTGG CTCGTTTAGA AAATGCTCAA   1200

CCAACATACG AAAAAGCAGT GAACGAGTTA AACAAAGCAG AAGCGGCAGT GGTCCAAGCA   1260

AAAGAAGCCT ATGAAAATTC TCTGAAATCA TTGGAGGAAC TCAAAGAACA ACAAGCCGTT   1320

GCTACACTTG CTTATACACA AGCACAAGAA GACCTTTCTA ATGCGAAGTT AGAGCTACAG   1380

CAGTACCAAG GCATATTAAG AGACTTAGAG GCACAACAAG CCGAACAGCG CCGACAAGAA   1440

GCGTTGCAAG AGCAAGTAGC AAAAGAGCAA CAACGCCTTG AACGAGAAGC AAAGCAACAA   1500

CAAATGTTAG TAGCAAGTGC TACTTCAGCA GATAAAACAC CTGGTCTCCA ACAGTTATCT   1560

TTTTCTAAAC AAAAAGAACA GCCAAAAGCA CAAACACTAA CACATTCAGA ACCTCGTAAG   1620

ACGAAACAAG TAGCAAAAGC CCAAGAATCC TTACCACATA CAGGAGAACA AAAAAGTATC   1680

TGGTTGACTA TTTTTGGATT ATTCATGGCA GTAGGTGCGA TCAGTTTCAA GAACAAAAGA   1740

CGGAAAAATA GTTAAAAGTA TAATAGGTGA GGAGCGAAAT TTTCGCTCCT TTTTATTTTA   1800

AATACATGAC ATGAAAGGAG AAAACATTGG CAAAGAAGAT AAACCTTGTT TCCATATTTT   1860

GTTCTCACGA AGCACTAGAA CAAATTTATA TTCTATCCTA CCCGTATTGT TTAATTAAGC   1920

AATATCTATT CAAGAGAATC GAAGATATTT TTGGTTCTTT TTTTATTGGA GAAATGACAT   1980

TGCTGTGGCT GAATAGAAGT CGCTTAAAAT AAAGCAGTAG AATCCCCTTA TTGTTTTATT   2040

TGATTATGCA AGGTGCGAAT CCTTGCCAGC AAGTAGGGTT TACGGTAAAT GATCCCGACT   2100
```

GCCGGTTAGC GAAAACCGTA TAGAGATTGA GCGGTAGTGT GGCAAGCAAC AGCTCTTATA 2160

CCATTCATAA GCCTACCGCA AAAAGAATGG CAATCTCAAA AAAACAGGAA ACTGAGAGGT 2220

CGCTCCTTTT AGGTTGGCGT GTAgCATCTG GGATGCAACG GACATGAAGG AAGATAAGAG 2280

GTAGGTTCGA CTCCTACCAC GCCAGTAAGT AGCTTTGCTA CGAGAGAAAA AGAATAATAG 2340

GAGGAATAAC AATGAAAAAT TATGATCCAA ATATCCGATG GGGCTTGCAT ACAATCAAAG 2400

TAAGTTTCCA ACGAGGTGTA TACAAAGGAT TTGTTACCTT TGTAAAAAGT GGAAATTGTA 2460

AAGGGTTAGA CGTATTAGGC ATAGACGAAG AAGATTTATA TGACATGAAA TTTAAAGAAA 2520

ATCCAATCAA TTTTAGACTG CTTGGTGAAG ATGATAATGG CGATGAATGG TTCGCAATGA 2580

CTTTGAAGAA TGACAAAAAA GATGAATTGT TAGTTGAAGA TGTGTGGGAA GAGTTAAGTG 2640

AGTACATTGT GAGGATTGAA ATTATTGATT TTGAGGAGGA AAAATAGAGG AGCGAGAAAA 2700

TACTCGCTCC TCTTTTATAT ATAGAGAGGC TACCCAGCAA TGATAACTAG GTAGCCACAG 2760

ACATAAAATT ACTATTAATC TGGGAATAAA TTATTTATAC CATGAAAAGG GTTACTTATC 2820

AAGAAAAAAT ACCAATGAAT AGGAGAATGT GGAATGAAAT AAAGTAATAC ATCTATAAAA 2880

AAGAGAGGAG GAAAAAGTAT GACTTGTGTA CGTGCGCCAT GTCGGAGAAA TCGTTGTTTG 2940

TAAAGAAAAG AATAATAGGA ATTAAAAAAT ATAATTTACG AGGAGAAGGA TTCATGAAGC 3000

AACAAACAGA AGTAAAGAAA CGTTTTAAAA TGTATAAGGC AAAGAAGCAT TGGGTGGTAG 3060

CCCCTATTCT TTTTATAGGT GTGTTAGGAG TTGTAGGATT AGCTACTGAT GATGTACAAG 3120

CTGCGGAATT AGATACGCAA CCAGGAACAA CGACGGTGCA ACCCGATAAC CCCGATCCGC 3180

AGgTAGGTAG TACAACACCT AAGACAGCAG TAACTGAAGA AGCAACAGTA CmAAAAGACA 3240

CTACTTCTCA ACCGACCAAa GTAGAAGAAG TAGCGTCTGA AAAAAATGGA GCTGAACAGA 3300

GTTCAGCTAC TCCAAATGAT ACCACAA 3327

(2) INFORMATION FOR SEQ ID NO: 303:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 510 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

TTCGCTGGTA CGTTTTTtCA ATACTTAATA ATGCACTAAC GTTATATAAA ATAAACCGAG 60

TTGCATCATC TACACTGATG CCCTCTTGTA AGCAATTGGT TTCAGGTTCA CTTTCAAAAA 120

TAGCCGCAAT ATTATTGATA TAGGTTTTTT TGAAAATTTG ATACAGCGTA CGATTGTAAT 180

CAGGGCCCAT TTCTTGAAAC ATATCGTTCA ACATTAAGAA AAGATTGGTC GGATGTTCAT 240

CTAACAAAAC TTTGATTAAT TGATAAAGTT TTTCTTCTAC TGATGCATTT CCAGAAACAA 300

TTGGGACTAA TTCTTCCTGA ACACGGGACG TTAATTCTTC AATCACAGCT AAATACAGTT 360

CTTTTTTATT TTTAAAATGG TGATAGAGAT TCGGTTGGGT AATGTtCGCT TTTAAAGCAA 420

TTTCACGAGT GGAGGtATTT TgAACCCCTT tGTCATGAAT AATTCAGAAG CGACAGATAA 480

AATAATTTGC CGTGTACGTC TAnATCAGCT 510

(2) INFORMATION FOR SEQ ID NO: 304:

(i) SEQUENCE CHARACTERISTICS:
(A) LENCTH: 2047 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

GACATAAGTC GTTTTGACTT ATGTCCTATT TTCGTTTTTT ATTTCCAAAC ATTTTACGAG 60

AATATTTTAA AATTTAGCGT ATACTTAATA TAAGCAAGTT AGGAGGGCTT ATTTATGAAA 120

CCATATGTTT ATGTAATCAG TGATATTCAT GGACAAGCAG ATTTGTTTGA TGCCTTACTC 180

ACAGACTATG ACCCAGTCGA GCATCAGCTA GTTTTAATTG GTGATTTAAA CGACCGTGGT 240

TCGGATTCTA AAGCCTGTTT TCTCAAAGGA AAAGAACTTG TCGAGCAACA CGGCGCTGTT 300

TATTTACGAG GcAATCATGA AGAATATTTT TTGCAGTTTC TTCAAAATCC AGAGGACTGG 360

TTTGCAGGAT ATGTTCGCAA TGGTGGCAAG GAAACAATTG AAAGCCTCTT GCATCCAGGA 420

GCTACTGCAG AATACTCACC AACAGAAATG GCCTTAATGA TCCGTAGTCG CTATCCAGAG 480

TTAATTGATT TTTTAACAAA ACGACCACTT TATTTTGAAT GGCAGCATTA CTTATTTGTG 540

CATGCTGGGG TTGATTTAAC GATGGAAGAT TGGCGGCAAA CAGCACCTAA GGATTTTTTG 600

TGGATTCGCG AACCCTTTCA TCAAGGGAAA AATAATACAG GCAAGACCAT TGTTTTTGGT 660

CATACCATTA CTCCGATGTT ACATGGTGAT ATGCAGACCA CGGACCTTTG GCAAAGTGAT 720

GGAAAAATTG GGATAGATGG TGGTGCTATT TTTGGTGGCT CTGTGCATGG TGTGATTTTT 780

AATGAAAAAG GCATTGTCCA AGATATCGAA TATCAAAAAC GCACCCcTGC GTGGCAACCA 840

GAATTTTGAT TCCTTAATCG TATGAATCAG TTACCTCACT TCTATATTTA GAAGCAAAAA 900

TATGCTAAAA TAGATAAGTA TGCATAAGAA AAGAGGTATC TAGTTAATGG CAGAAGCTTA 960

TAACGAAGAA GTAGTAGCGC TACTTCAAAA AGAATTGACA ACGTATCGTT CAAAACAAAT 1020

AACAACGGTT TTAACTTTAT TAAATGAAGG AAATACAGTT CCTTTTATTG CCCGTTACCG 1080

GAAAGAAATG ACTGGAAGTT TAGATGAAGT CCAAATCCGT GAAATAGAAG AACGTTATCA 1140

TTATTTACAA AACTTGAAAA AGCGGAAAGA AGAAGTTCTC CGTTTAATTG AAGAGCAAGG 1200

AAAACTAACA AAAGAACTCA AAACAGATAT TCAAAAAGCA GTAAAAATGC AACAAGTGGA 1260

AGATTTgTAT CGACCATATA AACAAAAACg TCgGACCAAA GCGACTATCG CGAAAGAAAA 1320

AGGGTTGGAA CCATTAGCAG ATTGGTTGCT CTCTTTACCT GAAAACGCGG ATATCTTAGC 1380

CAAAGCGGCT ACTTTTATTA ATGAAGAAAA AGAAGTTGCA ACCGCCGAGA TAGCCTTGCA 1440

AGGTGCTCAT GAAATTTTAG CAGAACGTAT TAGCGATGAA CCAAAATACC GAATATGGTT 1500

ACGTGATTAT ATGGTTAAGC ATGCGCAATA CGTTAGCGTA GTTAAGGATG AAGAAAAAGA 1560

CGAAAAGCGG ACCTATGAAA TGTATTACGA TTTTGCAGAA CCAGTTAGCA AAATGGTGCC 1620

GCACCGTGTC TTAGCAACGA ATCGTGGTGA AAAAGAAGAC ATTTTAAAGG TTTCCCTTGT 1680

CGCTGATGAA ACaAAAATTA ATGATTATTT CCAACGTCAA CTAATCGGCA AGCAAGCAAC 1740

AAGTTTAGcT GcTCCATACA TTGAGGCTGC TTATTTAGAT AGCTATAAAC GATTCATTGG 1800

ACCAGCGATT GAACGGGAGA TTCGCAATGA ATTAAcAGAA AAAGCTGATG AACAAGCGAT 1860

TGCAATTTTC GGTGAAAATT TACGTAATTT ATTATTACAG TCACCACTTA AAGGAAAAGT 1920

TGTTCTAGGA TTTGACCCGG CTTATCGGAC AGGCTGTAAA TTAGCAGTTG TCGATGAAAC 1980

TGGTAAAGTT TTAGCAATTC CAAGTGATTT ATCCACATAA ACCAGCGACT GCTGCGAAAC 2040

GTGAAGC 2047

(2) INFORMATION FOR SEQ ID NO: 305:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5338 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

CnCCCGGGGG AAATCAnGAn GAATTTCCCG AAATTGCTCT TTAATTTTTG GAGAATTTCG 60

GCTTATTTCC GAAGGAGTTG GTTCTGTTCC CGCCGTTTAT CAGTTTTTGA GGGTGGAGCA 120

AAAATCCAAA GTGATTTTTG TCCCACGCTC ATTTTTCTTA AATGAAAGTT TTTAAAAAAT 180

TTCCATGATT TTTCAAGTAA AATGAGGTGG GCACTTTTGT TCGATTCCAT TTTTATGTAT 240

AATGAGAAGT GATGGAAAGA GGGAAGTTAA ATGGCACAAA AAAGAGCAAC AAATAAGAAA 300

AAAAGTACGA CAAAGAAAAA GACCAAAAAA CAACAGCAAC AACAAGAACG TTTAAATTAT 360

ATGTTTCTTG GCCTGATTTT TATTTTATTC GGTGTCTTTG GACTCTTTCG CTTAGGGTTT 420

TTAGGCACGT TATTAGCGAA TTGTTTACGC TTAGTTGTAG GTAATACCTT TCCATTTGCA 480

GCGATTTTAC TAATCCTGTA TGGCTTATTA GTCATGATTT ATGGGAAAGA TTTTCCGTTA 540

AAAAGAGGGC GCCCTATTTT TGGGGCTGTT CTGATTTATA TTAGTGTTTT GTTATTTTTC 600

CATGCCTTTA TGTTTCGAAA TGTTAGTGGC AGTCAACCAG ATATTTTAGG TAACACTTGG 660

GAATTTTTGC AAAGCGATCT AAAAGCTAAC CAAGTGACGC AAAACGTAGG GGGCGGTATG 720

ATTGGCGCAC TGCTTTATCA AGGAACGTAC TTTTTGGTGG CACAATTTGG TAGCTATTTA 780

ATTGCCACCT TGCTACTGTT AGCGGGTATT TTCTTGATGA GTATGTGGGA TTTTCAGCmA 840

ATTGTTGATC ATTTCCAAAG TATTCAAGAT CGCCTAAGTC ATGTTTCGGC AAAATCACAG 900

GCGCGACAAG AAGAAAAAGA GGCGAAgCGA gCAGCGAAGA AAGAAGCCAA AGCTGCAGAA 960

CGCCAAGCAA AAATTGAAGC CGCTGCCCAA CAAAAATTAC AAGAGCGGGA ACGCATGGAA 1020

CAAGCGGCAG CCGAGCGCTT AACAAAAACA CCTGTTGAAA CCCACCAACC AATGGTGGAA 1080

```
GAACCGGCCG CACCAACGCC TGTTCAAATT GATTCGTTTC AACAGCAAAA TCAGGCGATG   1140
CCTGTACCAC CAATTGCAGC AACAAAACCG CAACGAGAGC AGGAAGAAGA GGCGGCAGAT   1200
GAAGCTGGTG TTCTCGAATT TGAAATTTCT GAAGAAGCGG AAGATCGTGA TTATCAACTA   1260
CCACCAACTG ATTTGTTAGA TACAATTCAA GCAACGGATC AAAGTGGTGA ATATGAAAAA   1320
ATCGAAAAAA ATATTGGCGT CTTAGAACAA ACCTTCmAGA GTTTTGGTGT TGATGCTAAA   1380
GTAGTCAAAG CTAGTTTAGG ACCATCTGTC ACAAAATTTG AAGTTCAACC CGCTGTTGGG   1440
GTAAAAGTTA GCAAAATTGT TAACTTGACG GATGACATTG CTTtAGCGCt TGCGGCCAAA   1500
GACGTTCGTA TGGAGGCTCC GATTCCTGGG AAATCACTAA TCGGGATTGA AGTTCCTAAT   1560
AGTGCAATTA GTACAGTATC CTTTAGAGAT ATTGTTGAGG CACAACCAAG TCATCCAGAT   1620
AAACTATTAG AAGTACCTTT AGgTCGAGAT ATTTCAGGGA TGGTTCAAAC TGCAGATCTA   1680
TCAAAAATGC CGCATTTATT AATTGCGGGA TCGACAGGGa GTGGGAAATC GGTTGCTATC   1740
AATGGCATCA TCACAGGTAT TTTGATGCAA GCGAAACCTC ATGaAGTAAA ATTAATGATG   1800
ATTGACCCGA AAAtGGTAGA ATTGAATGTG TATAACGGGa TTCCTCATTT ATTAACGCCA   1860
GTGGnTTACT AATCCGAGAA AAGCGGCCCA AGCCCtGCAA AAAGTAGTTC AAGAAATGGA   1920
ATTTCGTTAT GAAAAATTTG CAGCAACAGG CGTCCGAAAT ATCACTGGTT ACAATCAACT   1980
AATTCAACAG AAAAATGCGG AAGATGGCGA AAATCGTCCA ATTCTACCAT TTATTGTTGT   2040
CATTGTTGAT GAGTTAGCTG ATTTAATGAT GGTGGCAAGT AACGAAGTAG AAGATGCGAT   2100
TATCCGTTTA GCCCAAATGG CACGGGCAGC AGGTATTCAT ATGATTTTAG CGACACAACG   2160
GCCAAGTGTG GATGTCATTA CAGGAATCAT CAAAGCAAAT GTCCCTTCTC GAATGGCTTT   2220
TGCGGTCTCA AGTGGGACAG ACTCGCGTAC CATTATTGAT ACCAACGGGG CGGAGAAACT   2280
ATTAGGTCGA GGCGATATGC TCTTCTTACC AATGGGAGAA AATAAACCAA TTCGTATCCA   2340
AGGGGCCTTT ATCTCTGATC AAGAAGTCGA ACGAGTGGTT GCGTTTGTAA CGGATCAACA   2400
AGAAGCTGAA TATCAAGAAA GCATGATGCC AACAGATGAA CCGACCACTT CAGGCGGTGG   2460
CGAAGCTCCT CAAGATGAGT TGTTTGAAGA AGCGAAAAAC TTGGTGGTTG AAATGCAAAC   2520
AGCTAGTATT TCATTGCTAC AAAGACGTTT TAGAATTGGC TATAATCGTG CGGCTCGCTT   2580
AGTAGATGAA TTGGAAGCAC ACGGAGTTAT TGGACCATCA GAAGGTAGCA AACCAAGAAA   2640
AGTCTTTCTA CAAGCGGAAT CAGAAGAGGC AGCAACCGAG ACGCCAGAAC AATAAACAAA   2700
AAGAACAAAA ATAGATAAAT TTATCTATTT TTGTTCTTTT TTTATAATTT CAGCAAAGCA   2760
TCATGTTATA ATCATCATTT TCTGGTAAAA TAGAAAGGAA TAGTGCGGTT GGGAGGGAGC   2820
GTATCATGAG AAGAAAAAGC GTGCTTTATT TAGAAGTAGC AGATCAGATT AAAGAAGATA   2880
TTTTGAGTGG CAAGTATCCA GTAGGAACGT TTTTACCAAC AGAAACAGAA TTGGAAGAGC   2940
TGTTTAACGT TAGTAAAATT ACGATTCGGC GGGCAATCGA AATGTTAGCA ACAGAAGAAT   3000
TTGTTGAGAA AAAAAGTGGT CGAGGCACAA CTGTGTTAAG CAATCGTCCC TATAATAAAC   3060
```

TATCAAAAGC AGGGACATTC ACTGAATTTT TGAATGAATC AGGTCAGAAA GTAACTAAAA 3120

AAGTTTTACA AGTTGAAAAC TTGACGTTGT CAAAAGAGAT GCCGGCTTAT CAATTTTTAG 3180

GAGAAGAAGT GGTTCATCTT TCACGGTTAT ATCTAATCGA TGAACAACCC TATATTTATT 3240

TTAATTACTA TTTGCCAGCT GCCATGAAAG ACGTATCATT GACTGAATAC AAAGAAGAAT 3300

CCTTGTATCG TTTGATGGAT CGACATGCCA TTGAAATTTA TAAATTTGAA GACCGTTTTG 3360

AAGTTGCAAC TTTGACAGAA GCAGAGCAAA AGATGTTGCA AACAACTGAA ACAACGGGCT 3420

TAAAAAGAAT TCGGCGTTCG TTAAGCCCAA CAGGTCGTTG CGTAGAATAT TCAGAAGCTA 3480

TTTATAATAC GAAGATTCAT CCTTATGTGA TTGAATATGA AGCATAACTT TTAAAAGAAT 3540

AACAGTGCAC TTACAGTAAT GTAAGTTCAT TGTTATTCTT TTCTATTGTC GTTTTTTTTC 3600

GTTTTCGATA CACTAAATGA GTGAGAAGAA TAGGAGTGGA AAATATGTTA GAAGTACAAA 3660

ATTTAAAAAA AGTTTACGGA AACGAAATTA AATATGAAGC CTTAAAAGGG ATTAATTTAA 3720

CTGTCCAAGA TGGTGAATTT ATTGGTATTA TGGGACCTTC TGGAAGTGGG AAAAGTACAC 3780

TTTTAAATTT ATTGGCAACA ATCGATAGTC CAACAGACGG TGAAATACTT TTAAATGGAA 3840

AAAATCCAAA CAATTTGAAT CAAGAACAAA TTGCTAAATT CCGTCGTACA GAATTAGGAT 3900

TTGTTTTTCA AAGCtTTAAT TtAATGCCAA CATTAACTGT TGAAGAAAAT ATTATTTTGC 3960

CATTGACACT TGATGGAGAA AAAGTCTCTG TAATGAAACG ACAATTAGCA GAATTAAGTG 4020

AGCGTTTAGG TATCAATCAT CTGTTGAAAA AACGAATTGC AGAAATTTCT GGTGGACAAG 4080

CACAACGGGT CGCCGTTGCT CGGGCAATGA TTCATCATCC ACAATTATTA TTGGCTGATG 4140

AACCAACAGG AAACTTAGAT ACGAAATCAT CTAAGGATGT GATGGGCTTG TTACAACAAT 4200

TAAATGAAGA AGAAGCGGCG ACTATTTTAA TGGTGACCCA TGATCCGCTA GCCGCAAGTT 4260

ATTGCAAGCG CATCGTCTTT ATTAAAGATG GTGAATTAAT TGATGAAATT GTCCAAAATG 4320

GCAATCAAAA AGAGTTCTAC GATTTAATCA TGGTGAAACT GGCTGAGATT GAAGGTGTCG 4380

ATAATGAATT TTAATCAGTT TGTTATAAGA AACACGATTC GAAATAAGCA TTTATACTTG 4440

GCTTATTTCC TAAGTACAAT GTTTTCTGTG ATGATCTTTT TTACGTTCAC TGGTTTTGTC 4500

TTTCACCCAG CCTTAGCTAA TGGTTTAAAT CCTAAAGCAC AAATGGGAAT GACGGCTGCG 4560

GCAATTATAA TCTATGGTTT CTCTTTCTTA TTCGTTTTAT ACTCGATGGA TGTCTTTATC 4620

CAATCTCGAA AAAAAGAGTT TGGAACCTTG ATGATTCAAG GGATGAGTCC AAAACAGTTA 4680

AAGAAAATGA TTTTTATTGA AAACTTAGTG ATTGGTTTCT TTGCTACAAT TTTTGGTAGC 4740

ATTTTAGGTG TTGGATTTTC ACAGTTTATT TTATGGATTA GCAATTTATT GATGCATCTA 4800

GGTTTAGGTT TTTATTTGCC AGTGATGCCG TTTATCATTA CTGTCATTTC ATTTGCGGTC 4860

TTGTTCTTAG TAATTTCTTT CTTTATTCAA TTCCGCTTGC CAAAAGCGAC ACTTCAAGAG 4920

CTATTAAAAG CAGGAGAAAT GGGTAAAGGT GAAATCAAAA GCTCTAAAGT TAAATCTTTC 4980

TTAGCCGTGT TACTTTTAAT TGTCGGTTAC GGAATTGCTT TAGTTGCTAA GGGACAGTTA 5040

GTTTTAATGG TGATGTTCCC AGTAATTTTC TTAGTTATTT TAGGGACAAA ATTCTTATTT 5100

GmCCAATTGA GTGTTtCTGT CATTGAaCGT TTAAAmCGGA AaCCAAAAaT CTTTTGGAAA 5160

AAAmCCAATA TGGTCGTTTT ATCAGATTTA GCATTCCGCA TGAAAGATAA TGCACGTTCG 5220

TTCTTCTTGG TTTCGGTGAT ATCAACGGTT GCTTTTGCTG CCATTGGAAC GTTATACGGC 5280

GTGAACGAGA TGATTTTTAA AGGAATAAGT TCTGTACCGT ACGAACTGAC ATTGTCTG 5338

(2) INFORMATION FOR SEQ ID NO: 306:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11854 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

TGAATTCGAA GAAGAACTCC CTGCAGAACC TATTTCCAAA TTAGGTGATA TTTATCAATT 60

AGGAAGGCAC CGTTTAATGT GCGGGGACAG TACAAATTCT TTAGAAGTAG AAAAATTAAT 120

GGGCAATAAA AAAGCCGATC TTTTGATTAC TGATCCTCCA TACAATGTAG CGTACGAAGG 180

TAAAGGAAAA GAAGCACTAA CTATTAAAAA TGATAGTAAA GAAACGAATG AATTTCATTC 240

ATTTTTATAT GAAGCTTTTA GTGCAGCCAT AAATAATATG AAATTAGGAA GTTCATTCTA 300

TGTGTGGTAT GCCTCATCAG AAGTAGTGAA TTTTCATACT GCTTTAGAAG AAGCTGGGTT 360

TTTAGTAAAA CAAGAATTGA TTTGGAATAA AAATAGTATG GTTTTAAGTC GTCAAGATTA 420

CCATTGGAAA CATGAACCTT GCTTATATGG TTGGGCTTCT GGGGGcAGTC ATTCTTGGTA 480

TTCAGATAGA AAGCAGACTA CTATTCTTAA TTTTGACAGG CCTACAGTTA ATAAGGAACA 540

CCCTACTATG AAGCCAGTTG CACTATTCGA TTATCAAATT AAAAATAGCA GTAAGCAGGG 600

AGACTGTATT TTAGATTTGT tTGGAGGtTC TGGTACTACG TTAATAGCAT GTGAGCAGAA 660

TGaAAGAGAA GCCTATTTAA TGGAGTTAGA TCCAAGATAT GTTGACGTCA TTATAGCTAG 720

ATGGGAAGCA TTCACAGGAG AAGTGGCTGT GAAGATATCA GGTAATGATA CGGCGGTGGT 780

TGATGGTGGC TAGAAATGAT AATTTAAAAC CATTTTCAGA GCGAAGCGTG GATGAAGCTr 840

GAGAGCTTGG AAGAAAAGGC GGTAAAGCAA GCGGAGAAGT TAGACGAAAA AAAGCTGACT 900

TGAAACGAGC TATTTCTATT GTTTTATCAT CAGAAGTACC AAGCTCTAAA ATGGCTAAGA 960

CGCTAAAAGA AATGGGATAT GAGAACACTA ACGAGATGGC CATGGTCTTA TCTATGACGC 1020

AAAAAGCAAT TAAGGGAGAT GTCAAAGCAG CATCCTGGAT TTCCAATATT ATTCAACCTG 1080

CAAAGGTAGA GCATGAAGTT GAAATGAGTG TGGGCGTTGA TGAGAAGCGG AAAGTAGCAG 1140

AAGAATACAT TAGGGGCCTG TTTAATAATG ACACTGGAAA TAGCACAGAA GAGAACAATT 1200

AAGCTTTTAA AAAGTTCAAC GCCCAAAGAA AAACTCAATA AGTTTGTAAA AGGTTATGTT 1260

CCAACTCACT ACAAGAGATT AAGTATATCA ATGGAAAAAG CAATAGAACT AGCCATAATT 1320

GGGGCAACCG AAAGTCTAGC ATATTATGGT GATCGATTAT ATTTTACGCA GGCTCTATTA 1380

ATGGGGGCTG TGGTAAGTGG AGAGTATGAC AACATTATTG TCGTTACCCC TTCACAGTAT 1440

GGTAAAAGCT GGTTGAGTTC GAGAATAGCT GTTTGGCTCG CTGATCACAA TCGGCGTTGT 1500

TACGTGGCGG GTGGTAAAAA AGATACCACT GATATTATCA TGCAACATGT TACAGATACA 1560

CTACAGACTG TTGATGAATC AATTGCAAGA AAATTATTAG AGCCTGTTGA TAAGCTAGAA 1620

AGACTTCAAA CGGGTTTATC CAAAAGAAAG ATTTCTTTCA GTGGCGGAGG ATCGATTGAA 1680

GGGATTTCAT TAGGTGAACA TTTCAAAGGA AACAAATCTG GAAACCAAGC GATTGGTCGT 1740

GGTGGAGATT ATATTATTGA TGAATCAGCG TTCGTCTCAA ATGAAACATA TGCTGAGCTT 1800

GGTCGTAGAA ATTTTGCAAA TGTGGATGGT AAAAACTATC TATCTTTTGA AATATCTAAC 1860

CCACATAATA AAGGTCGATT TTATGACAAA CTAACTCAAG AAAATATTCC AAAAGGCATG 1920

TTGGTTGTAT GGGCAGATGT TAGGACCGCT TTTGAAGAAG ATCGAGTTAA AAGTATTGAA 1980

CAAGTAATAA GTTCTGAATT TTTTCAAAAT AAATCTACAT GCCAACGTTA CTTTTTATGT 2040

GAGCTTCCAG ACGAAAATGA AGATGGAATG TTTGGGACAC CTCAAACAGA AGAAGAACAT 2100

ACCGAAAAAA ATTGGGAGTA TTTCCTTGGT GTAGACAGTG CTTATAAAGG AAAAGATAAA 2160

ATCAAAGCCA CGTTATCAGC ATTAGATGCA CAAGGACAAG TACATGTTAT AGACACTATA 2220

GAAATTGAAA AAGGTGACTG GCAAGACGGT GTGACTAGTA AAAAGATAAT TACTCAATTG 2280

TTGATGATTA TAGAACATTT TGAAGTTAAA GGTGTATGTG TCGACGTAGG TTACGGTGTT 2340

TATATTGTTG AAGGTTTAGC ACATATTAAC GGAGATTTCG AATTACACGG CATAAATTTC 2400

GGTGCTGGTA CAACTAAAGA AAGGGTGGAA AAAAACCACT ACTCGGCAAA ATATGGGGCA 2460

AATAAGCGTG CTGAAATGCA CATTGACTTA CAAGAAAACA TTGATAACAG AAATATATTT 2520

TTCACTGAAA AAGTATATGA AGAAGTTATA GATGAACTAG TTCTTGTAAG TAGCaAGATC 2580

aAGTCTAACG GAAAAACAGC CATTGTTCCA AAGGAAGAAA TCmAAGCTAA GTTGGGCCAC 2640

TCACCAGATA CACTTGATTC AGTTCTGCTA TCGCTACATG CGATTATTCT ATATAAATTA 2700

AACGAACGAT TCTATATCTA TTCTTGATGA AAGGAGGGAA ATGAATGTCT GAAACTGAAT 2760

TAGTTGGGAA AGATGAACTA TTAAAAGCGA TGACTATGTG TAATAGTTGT CCAGATTTTA 2820

ATTTAAACGA TTTAAGAGGT AAGGGTGATA ACAACTATCA GCTTTATGAT TGGTTAATTC 2880

ATAATTTGCC AACAGCACAA TATGTTCTTG GTAAGTTAGT AGAGCTTATT TTTTCTAATA 2940

ACTTAACTAC AGGCGATGAA AAACAAGATG AAATATTGAA TAACTTTTTA TATGGTCAAA 3000

CGAATCCAGA AGGAGTTACC AACTATCACG TACTTGTTCA ATCAATTAAG GAATCAATTG 3060

TATACGGTCG ATCTGGTTTA CGTTTTTTAT CTAAAGATGA TGGATTGATT AACGTAAAGT 3120

GTAATCATTT TGGCGTTGCT CAAATATTGA ATAAAGAACA TTACGGATAT AAAGAATTGA 3180

TTGGTTTCGT TATTGACAAA AAAGGTCGAG CTATTACAGA TGTCGATCTC CGTGAAGGAG 3240

AAATTGATTC AGAAGAATAC TTTAAAAAAG GAATATTTGT TTTTAAAAAC AATGACAATA 3300

TTTTATTGCC ACCCGAAAAA TTCGTTAATT TAAGGGTTGA TACATCTACT CCCAAAGGAT 3360

CGAGTGTCTT TGATTCGGAT ATCCAACGAG TCCTGCTTAT AGCTTCGGTA TATAAACGAC 3420

TACTATATGA TATCGAATAT GATGGCGCAG GAAGATTAAT TTTCTGGGCA GATAATGCCA 3480

ATAGTAATGA AGAAAGCTCA AATAACTTTT TGAACGACAC TGAATCGGCA ACCAAACGAC 3540

GGCAGGATAA ATATAAAAAA GAAGTTGAAG AAATAATGAA GCTCGTAAAA GACAGCAATT 3600

CAACAAGCGT TTTAGCAGTT TCAAATGCTT TTAAAAAGAT GGACCACTTG CCACGTGTAA 3660

CTAAAGCGAC GGAGTTTCTT AGTTATTTAA ATcAGGAAGG CGAGATTATG GCGCAAGTAT 3720

TCGGTGTTCC TAATGTTCTG CTTGGATTAG GTAAAATCAG TGGGAATATA TCAATGGAAA 3780

AAGTAATAGA TAATGCAATG CTTAATTCAA TTATTCCTTT ACGTGAAAAT ATTGCGACTC 3840

AAATTTCGAG CATATTAACA AATAATTTAA AAGTTCCAAA AGTCTATTTT GACAAATATG 3900

AACTAAAATC TCAGTCAGAT ATAAATGATA GGCGATTGAA AGTTTTAACT GTTGCTGAGC 3960

GACTTAAAGC ACTAGGTAAA GAGGATTTAG CTAATAAAAT TATTGAGGAG GAAATTCAGT 4020

TATGAGTATT TTAGAAGATC TAAGCAAAGC AAAAAAGAAA GCAAAACCAT ATGCAGTAGT 4080

CGGTGGACGT GAGGTTTATG ATTATGACTC TTTAGAGAAA AAAGTTGAAA TCGATCAAGC 4140

CGAAGCGAAA GTTGGTGGTG GCCAAGTAGA TTTAGGGAAA ATGAAACCGA CAAAAGATGG 4200

ATGGGGGTAT ACCGACATGG GAAATAGTTT CTCGGCAATT CCTCAAGATA TCTTATTTAT 4260

CaATCGTTAT AAAAAAGAAA aTGATGCTTA TTTAATCGTA ACAGACTATC GTGcCATTAA 4320

AGAACAGTCT AGTGGTCAAa TTTATGCATC AAGTGTTCAA GCGCTAGTTA TTAAAAATaA 4380

AGGCAAAAAA GGTGaAGAAG AGATGTATCT CGAAAGTATT CGAAATGTAT CTGATACTGA 4440

ATTTATCAAC GATTTTACTG GAGAATTGTC AAACATTTCA ATGGCAAAAG TTTTTGAAGC 4500

TATCGATAAA GACAAGGTGA AAGAAGTAAG CAAAGATGAA ATCAGTTTCT AATATTGCTT 4560

TAGAGTACAC CACAACAAGC AATGAACAAA AGCAACGTGA ACTATTATTG CTATTAGTTT 4620

CTTACTTCCT AACTCTTTAT GATATGGAAA AAGAAAACTT TGCTGATGAA CTTGGTATTT 4680

CCAGTGAGTT TGTAGAAGAT GCTCAAATTA AGTTTGATTA CATGAAGAAA ATTGAACTGA 4740

CATTGCAAAA TATGCGTGAG ACTGTACAGA GTGAAAAAGA CAAAGACAGT GAAGCTCTAA 4800

CTGCTCTATA TTTTAATCGA ATTTTAAATA CAGATGGAAA AAAAGCTAAA GAATTAGCTC 4860

AAATCGAAAC TGCAAAACAT TTAACTCGAC TTGATCGTTC TAAGAAAATT AGAAAGAGAT 4920

GGAAAGCATT TTCAGGATGC TGTGATGTGT GTAGAGCAAT GGATGGAGTA ACAGTAGCTC 4980

TTGATGAGCC ATTCATGTAT CAAGGTCAAG TTGTTGAATT GAGCAGTGGT GAACGTTATA 5040

TCAACAATTA CGCAGCGATG GACACGCCAA ATGCTCATCC AAACGATAAA TGCTCTATTG 5100

AATTTATTAT CGAATAATTG AAGGGGGGTG TAAACTATGG CGAAAATTAA GTTACCAATC 5160

GTAACAATTA ACGCAAAAGG TGGCGCTCAA TATGTTGTTG AAGGCGTCGA AGCAGTAAAT 5220

GCGTTAACTC AATTAGAAAA TTATCCTGAA CGTGGTATTC GTATCAAAGT TGATGGGAAA 5280

ATGACAGTAA TTACAGAAGG TTGTTTATGT TCTGCAGCTG TTACTGGAGA rGTTGAAGTA 5340

GATGTTCCTG AAATCGTTTG TGAAGAAGTA GAGTGTGAAC CAATTTCCAA TATTTTACCA 5400

AATCCAAAAA ATCCAACTGA TCCAGGGGAA AATTCAAAAC CAGAAGAAAA AGCTGGCGAA 5460

AAATAAAACT AGGAAGTGTC AGTATGAAAC AATTTAAAAA CAATTTAGAG ATGTTTCGTT 5520

ACATTGATGC CATTAATAAA GAGCCTAGTA GAGAATTTAA AATTGAGCCA TTACGAGAAG 5580

AGATTATTAT TGAACCTATC GAAGAAGACG TAGAATCAAA AGCTAATAAA AAAGCTAAAA 5640

ATAAGAAGGG GGCGAAGGAA TGAACTACTT TAAATTAGTT GACGGTATTC GTTCCCCTCA 5700

GTCTATCGAT GTAGTACGTT CAGAAAACGG ATATAAAAAA TTTGGTTGGA TACGTGTCCT 5760

TCCAGATGAA CGATATCCGT TGGGAGACGA TGAAGCATTT ATTCAATCAT TAGAAAATGC 5820

TAGTGTTGAA AAGCTTTACT CCGACAAATT AGTTACTGAG CTCGAAAATA ACGGAATTCA 5880

ATTTGAAGTC TTTAACGGCG GATGTTGTGG CGGAAAAATC AAAAAAGTAA GTTATAAGAT 5940

CATCGATATT GTTAGAGATG AGGTGTAACA TGTTTGATTT TATCAAGAAA CGTGATGCTC 6000

AAACGCAAGC ACGAAAACGT ATGAAAAAGG ATTTAGACGA AATTTTCGAA TTTAATAATG 6060

AAGAAAAACA GCAACATACA TTGAACCACA TTGTTCAACT AGCTAATAAT GATTTGAAGG 6120

AAATTGGCTG GGTTCGATTA CTGGATGAAG GAACAGTATT ATACGGTGAT GGAAGTATTC 6180

GAAGTTATAT AAAACGTGGA ACTATAAAAG AATTTTATGA CTCTTTAGAA GATGATTATA 6240

TTGGCTATAT CAATATTGGA CACATTAATT TTGCTACGTT ACCAATATTT GTTGGTCAGT 6300

GGACTAAAAA TGATTTGCGC TTAGTAGATA TTGGAGAAAA CCGTCAAGCA TTAGAAGTCA 6360

ATATGAAAAT TGATGAAAGC TTATCTGTTA TCCAAGATTT GCAAAAAATG CCATATACAA 6420

TTGGTATTTC AGCCGAATTT ATGGCAAGTT ATGACGAAGA ATTnTCTTAC GAATATGAAT 6480

TTCCAGTTAT TGAACATCTT TTTATCATGG GATTTGGCAT CGTTGGAGAT GTAGGTAATG 6540

TGAATAGTAG CGGTATAAAT TTATCCGCAG AGGAGGCAGA CAAAATGGCT TTAGCAGATT 6600

TATTTGGCAA GAAAAAAGAA AATGATCAAG AAACGATCAA AGAACCTGAA ACTAAACAAA 6660

CAGAATCAAA AGAAGAGCCC AAAGAACCTG AAGTGGCAAC AAAAAAATCA GAGGAACAGA 6720

ATGTTGAAGA AAATGACAAA AAAGAATCTG AAGCGGAACT TTCCGAAAAA GAAGACCATA 6780

CATTCGAACA ACTTTATAAT TTATCTATGG AGCAAAATGa AAAAATGGTT GTTGAGCTTG 6840

AACAGTTACG TGCTGaAAAT aAATCTTTaA AAGAAGaAAA ATTGgAAGCA AGAACAAACA 6900

AATGaAAAAG CTGTTCAACG CTTGGAAAAA TTGATGAATC GTATTGAAGT TTCAGCGCTG 6960

CCACAAGCAA CTGGATCTAC TAAAAATAAA TGGGGGGAAT AACAGATGGT ATTAGAAGTT 7020

AATCAAGCCA TCGTTAATGA ATCGTTACAA ACAGAGCAAG TTATGGAACA ACTAAGCTCT 7080

ACTGTAGATG TCATTGTCGA TAATATCGAA AAATATACaA ATGCTGCTCG TTATGGTAAA 7140

GGAAATAATT ATGCTTTAAA TAACTTGCGA ACAGCAGTTC AAAACCAATT TCCTTTAGTT 7200

GATTGGTTAT TAAGCACAAG CCTATCTCAA ATGATTGAGA ACGCTTGGCA AAATGGCTCA 7260

TTGcCAACTG TAACTGATGA AGATGGAAAT GTTTATTTAA AAGCACCTTT AAATGTATTT 7320

ACGACACCAC CAAAAGATAC TAAAGGTGAA TGCTGTTGGT TACCTTTTGA TATTGCAGCA 7380

TGTGGGGGCA AAGCACCAAT TAATATTCTA TGTTTAAAAG ATTGTGTGGA TATGTTAAAT 7440

CACTTATTGG ACCGTAAATT AAAAGTTCAA TCGAATGATT TAATTGGTTT CTTTAAGCAA 7500

GCAGGGCAGA CTTATGAAGA AGTTCGTGAT TTCATGAATC GTGAAAGTAT GGCTTTCTAT 7560

ACTGCCAATA CAATTGTTAA TGGACAACTA GATGTTACAA CACCTATCTT AAAAAAATTC 7620

CATGGATTAA TGGAAATTTT AAAACGCCCA GAAGTTTTCA AAATGCAAGG AACAAATATT 7680

TTAGCAGCAT TTGATTCTAT TGGATATCGC ATGGACGTTC TAGGTGGATC ATTCATTTTT 7740

GCAGCTCATC CATTAACTGT AGCAAGCATT AAAAATGCCA TTCGTCCTAA TCGTTACGGT 7800

ATTCTTCCAG ACGGCTGGAC CATTAACGGA GAAAGTATTT TCTATAAAGG AGCACAAGTT 7860

TTACCTGATA AAAcTGTACC AGTTGATGTT GAAAATGGCA CAGGCGTTAT TTGGCAATTG 7920

TCTGGTGAAT CCGTAGTGTA TTCTTGGGGA CAACTTtGCG TCCTGCTGAA GACTATATTA 7980

TTCGCAATCa ATTTACAACA ACAAACGATG TAAATAAAGG ATGCGCAACA GAATGTGATA 8040

TTTATTACAA CTTAGGAAGT GTAGTTACAA ATAATGTAGC TCGATTAGCT GTAATTACAG 8100

ATGTTCCTTT GGCTGCTGGA tCAACGCCTA CTCATTAGCt AATTTATTAG ATCGaGTATA 8160

TGTAGAnACA TTAGCTCCCT AGAAAGTAAG tGCTtACCTT ATGGATGAAA TCATTGAGCA 8220

ACTGAAAGAA TTTtGCGATt GTTTTCCTTG TGATGTAGAG GACAACAAGT TAGAGAAAAC 8280

AGTTAAAGAA GCAATTCATT TAATAAGTTT ATTAACTTGT TGGaCTCAAC GTCCTTGCGA 8340

AACATTTTTA ATGAGTGAGA GACAAGAAGT ATTTGATATG GACAATTACT TACCTTGTTC 8400

GTGTGATGAT GGAATTATGG AATTAGATTT ATTCTATGCA CCATTTGCTT TAGCATCATT 8460

CCGAGTTTTT TCTGTACATC GAGAAGGTGT AAAAGAAATC ATTAGAGAAT TAGATGAAAA 8520

AGAATTTGGT TATAGCGTTG TAAAAGACAA ATTACTTGTT GATATTCGAA ACTATGCAAA 8580

TAGGGAAAAC GGATGTTGTG TTTGTAAAAA AGAACACCAG TTGCTAGTCT TATATGATGC 8640

TGGATTTGAA GAGTTACCTC AATGTTTATT ACAACTTTTC TGCGATTTAA TTCATGTAAT 8700

TTACAACAAG AATAACTGTG ATTGTCATGC TTGTGCTACT TGTCAAGATA ATTCAGATAG 8760

CGGCTTTATT GCAAATGAAG CGATGACAAC AGATGAGTTA GTAGAAAGCT ACCTCAATAA 8820

ACTAGTAATA GACAGCTACC GAAAACAATT AGGTCTTATT AGCCTTTGTG GTAAAAGTCT 8880

AGAACAGATT TGGGGGATTA GAGTATGAAA GTTCGATTTT TAGGCGTTCG TGAACACATT 8940

AGCGCTACAG GTTGTTCGTC CTGTGGAGCG AAAAGGTATG CCAGTGGTTT AAAGACAGAA 9000

AAGACATTTT TCTTACCAAG CGGAAAAAGA TTAGATGTTG AGTTTAATCA TGAATATGAA 9060

ACCACTGATA CAGACGGAAT GTTTTTGCTT GAAAGTGGGC ATTTCGAGGA AGTGTTTTAA 9120

ATGGCACTAA AAAAGATAGT TATTCAATCA ATTGAAAAAG AGATAGATAA TTATGCAAAA 9180

AAAATGGAAA AAATTATAAA AGAAGAAGCT CACGTTAAAA CAGGAGCGTT GAGAGATTCG 9240

```
ATAACAATAG AAAAGGAAAG TGACGGGAGT AGATTAATAG GAGTTGATGT CGCGAAACTG    9300
AAATCCGATC CTCGCAATGT TGGAGGCTTA GATTACTCTA TTCCTTATTA CAAAGGTCAT    9360
AGTGGCTACA CGATAAGGCC AAGAAAGGCA AAGGCTTTGA GTTGGGTTGG TAAAGATGGC    9420
AAACGTCATT TTGCTAAAAG TGTTTATATA CCACCCCACG CAGGTGATCC TTTTTTGAAG    9480
CGAGCTGTGT TACGTAGACC AAAATTATAG GAGGTATATC AAATGGCGAA ACGTGCAATG    9540
AATGCATTAG CAGCAGATGG TGAAACGACT TATCAGTTAT CAGCTAAAGA CTATACAATC    9600
GGTGATTCTG AAGTCACAGG TGTGTATGAT AATGAAAAAG CTGTAACTAT TAAATTATTT    9660
GTAGACGACG TAGCAGTGGA TGAGATCACA CCTGATAATT CAAAAAATAT TTATGCTATT    9720
TCTACAAGTA AAACTACTAT TGTTAAAGAT AGTAAAGTGG AAGTTGCAGA ATATGATGCA    9780
GATAAGAATG AATTAACAAA AATTCTAGTT ACTGTAATTG ACCCAAATGG AGGAGGAAAT    9840
GAGATGGATG AAAAAGAAAA GATTGATAAA TTCATTTCAC GAAAATTAAC AGTATTAAAT    9900
GAAAAAGACG GTATTGTCTA TGAACAGCTA GCAATCCGAG TTATTCAAGT TAATCAAAAA    9960
TAATAGGAGG AATTAGACAT GGGAAAATGT AATACAGAGC AAGTTCTTTC AATGATTGGT   10020
GTTAACAAAT TAACGAAGGC ACAGGAACTG TTTTTTTCGG TTCTACAGGA CAATGACTCT   10080
TGTGTAAAAA TTAAAACCAG TGACTACTTG GAAATCATTG GAGACAAATC AGCATTCAAT   10140
AAATATTTAC GACCAGAGGA TGCATTTAAT TGCTTAGCAG AAGGATGTAG AAACACAGGT   10200
GGATTATTAA TTACAGGCAA CGAATTTCCG TTAGGGGCTA CCTTTAAAAA AGTTACTGAT   10260
GCGACTGATT TTTATGCTGG TGCAACCACT TTTTATTTAG ATTTGCCAAA AGATGGAACA   10320
TACACTATCG AATTTAAAAT TGCAGCAATT AATGACAATA GTTTTGTAAA TGCTGATGTA   10380
TATAGAAAAA AATTCACTGG AACTAAGGGC TATAATCCTA TTTTTATTGA TTTTTCAGTC   10440
GTTCCTGAAG AAGTTTTAGG CGAAGGTTGG cAAGCAAATG AACGTGGTGT TTATGTGTCA   10500
ATTACTGTGA CAACTGAAGA AGAAATTCCA TTAAAACAAA TTCATATTTC TTCGATTAGT   10560
TTTTACAACT CAATCGAAGA ATTACAAAAT GATGAAGTTG TAACGATTGG ATGTATTACA   10620
GAATACGGTG GAGACATGAC TATGGATGTT GCCGATAGTG TATGTTTTGG TGCTAAGTAT   10680
GATCCATCAA GCGCTAGTAT TACTCGTACA TTCACAGGTG GCAAAACAAG CGGAAACTAC   10740
TGGTTACTTA ATCCATTTAT GCGTCGAGGG GATTTATCAA AAGGGTGGAC AGTTGTTAAA   10800
GAGAAAGATA AAGTTCGTGA ACTCACTATT GACGGCCGAC GTTATGGGTA TATCTTGTTA   10860
AACGGTTTAT CAAAACAAGA ATGCTCTTTC TCTAAAGCTT TAGTTGCAAG TGAATGTAAC   10920
TTTACTGaTG CAGAATTAAC TAAAGTCAAC CTCCCaGATG TAGCAGTATt AAACGAAAAA   10980
CAATACCAAA TTATTAAGCA TGGCGAATAT GATGGATATT TAATTGTTCA TGAACGTTTG   11040
ATCGGACAGC CATTGTTATA TGCATACCCT AAAGAGGTTT CTATTGAGCA ATACGTCGGT   11100
GAAGATGACG CATATGAAGG ACGCCGTGTA CGTCTATTCT TCCCAACTGT TCAGACAGAT   11160
GGTGTGAAAG TGAACTATAT TTTCAACAAT GTATTAGTTA CTTCATTCCC TACGACTTTA   11220
```

AGTAATACAG ATGAAACTAC ATTTGAATTT GAAGTATCTA TTCAAAAGGA TAACAATGGT 11280

CGATTCTTTG AAGTTCAGAA AATTATTGAA TAATTATTTA AGAAAAGGGG ATTTTTGTAA 11340

TGAAACAAAG CGATTTAACG AAAATGATTA CAAAAAATGA TGTAATTGAT ATGAAACATA 11400

AAATGGATAA AACTCAGGAT ACAAGTAAGC CATATGCAGT AATTGACAAC GATAATATTG 11460

CTGTTGTTGG TGACGCAAAC GAGATTCAAA AAGTTGAAGA AAACTATGTG ATCAAGTTTC 11520

GTGTTCCAAA AGAATTCtTT GAAGAAtTCC GTATGGAGCA aCTACTGTAG ATAAATATGT 11580

TATTTTCGAA GTTGAATACA GTAATGCAAG TGTAACAGGT ATGAATGATT TAAAAATTGT 11640

AGATGCACTT TTGAACATTC AACCATTTTT GAAGGAGTTT TACcAAAAgA CGAAGACGGC 11700

AAAGTAACTA TTGkGGAAAa GACAGaTCGA GAaGTyCTtA AAATGTtATC AACATCAGAG 11760

GACGAAGTTA TTtaTGGGTT TTATAAAGTG GTGGCAGCTT TcTTAGGTGt AGATGAAGAG 11820

TTAATCGAAT ACATGCTTCC ATTTTCAGTG ATTG 11854

(2) INFORMATION FOR SEQ ID NO: 307:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4888 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

AGTGTTTTTC ATCGAGAAAG TTAAAAAAGT TTTTTTGTTA TAAATAACAT ACTGTTTATT 60

TAAAGTGGTG TATACCACTn AATTGAGCAT TTTAAATAAA AAAATAAAAA AGGGTATACA 120

AATGCAATCG TTTGCGCTAA AATGTGnTCG AGTATTAGAA AATACACGAA AAGCGACAAT 180

TAAAGGAGAG AnTTAAATGA ACACTGCTGC AATCTACCAT CGTCCAGAAA GTGACTTTGC 240

TTATCTTgTA TACAGaAGAA ACCATGcATA TTCGTTTGCG TACGGCAAAA GGCGACATTG 300

cCTCTGTCCA ATTAGTGcAA GGAGATCCTT ATCTTTTAGG GAAAGAAAAG TGGTACCAAC 360

AGTCGTTAAC CATGGAGAAG TTAGTATCTA CGGAGTTATA TGATTACTGG TTTATTGCAT 420

TAAGTGCAAA ATTTAAACGC CTTTCATATG CTTTTACTTT AGTAGGGGCG GATGGTTTGA 480

GGGCGTTTTA TGGCGAACAC GGCATTTATC CTTTAGAAGA AAAGTATTTA GAGATGGCGA 540

ATAATTATTT TCGGATGCCG TATTTTCATG AAATCGATCG TTTTAAAGCA CCTGAATGGG 600

TAAAAGAAAC GGTTTGGTAT CAAATATTTC CAGAGCGTTT TGCAAATGGT GATCTCGCCA 660

ATGATCCTGA AGGGACGTTG CCGTGGGGAT CTAAAAGACC AGATCGACAA GATTTTTTTG 720

GTGGGGATTT GCAAGGGGTG CTAGATCATC TTGATTACAT TGAAGATTTA GGTGTTAACG 780

GCTTATATTT TTGCCCAATT TTTGAAGCGT TTTCTAATCA CAAATATGAT ACGATTGATT 840

ATTTAAAAAT TGACCCAGCT TTTGGTGATG GTGCTACCTT TAAGCGTTTA GTCGAAGAGT 900

GTCATCGACG GGGCATCAAG ATTATGTTGG ACGCCGTTTT TAACCATATG GGGGACACTT 960

CGCCTCAATG GCAAGATGTG TTAAAAAATG GCCAGGACTC AATTTATGCC GATTGGTTCC 1020

ACGTAAATGA GTTTCCAGCT AGCTATGAAG AAAGTGCTGA TTTTGAAGAA GCAAGCAACA 1080

TTACCTATGA TACTTTTGCA TTCACTCCGC ATATGCCTAA ATTAAATACA GCAAATCCAG 1140

AAGTGCAAGC CTATCTTTTA GAAATTGCTC GTTATTGGAT TGAGGAATAT GATATTGACG 1200

CCTGGCGTTT AGATGTGGCA AACGAAGGGn ATCATGCGTT CTGGAAAAAA TTCCGGCAAA 1260

CATGTGATGA CGCAAAAGAA GATTTTTATA TTTTAGGAGA AATTTGGCAC TCGTCGCAAA 1320

GCTGGCTCCA AGGAGACGAA TTTCATGCAG TGATGAATTA TGCATACACA GATGCGATTA 1380

TGGGCTACTT TGTCAAGCAA GAACTGTCCT TGGAAAAGAT GCTTTCTGAA ATGAACAACC 1440

AGTTAATGCT CTATCGACAA CAAACGAATC AAATGcAGTT TAATGTTTTA GATTCCCACG 1500

ATACACCACG CTTATTGCAT GAAACGAAAG AGGATAAAGA GCTGATGCGC CAAGTATTAG 1560

CGTTTACGTA TATTCAACCA GGTGTCCCTT GTCTTTACTA TGGTGATGAA ATTGGTATGA 1620

CAGGTGATAT GGATCCTGAT TGTCGTCGTT GTATGGTTTG GGAGGAAAAT CAGCAAGATT 1680

TGGATTTGAA GGAATTTGTA AAATCATTGA TTGCCCTTCG CAAAAAGTAT GCGGCTGTTT 1740

TTTCTGGTGG AACAATCGAT TGGACAGGAA CTTCTTTTGA AACGGGGTTG ATTAGACTCG 1800

CTTGTCGTTT GGAGGAACAG ACGATTTATG GGATTTTTAA TACTGGCCAA GAGACACAAT 1860

TTGTCGATAG AACTGATCGA GAAGTTTTGC TAAGTTCTTC TATCGAAATT GCTTCAGAAC 1920

AGATAATGGT TTTACCAAAA GGTTTTCTTT TATATCAAGC ATAATAATTA AGCAAAGGAA 1980

TGTGAAGAAG AATGGAAAAA CATTGGTGGC AAGAAGTCGT GGTTTATCAG ATTTATCCTC 2040

GCAGTTTTAA GGATrGCAAT GGAGATGGCA TTGGAGATTT ACCAGGAATT ATTGAAAAAT 2100

TGGATTATTT GGAAACGTTA GGCATTGGGG CTATTTGGTT ATCACCAGTT TATCAATCAC 2160

CCAATGATGA CAATGGTTAT GACATTTCTG ATTATGAAGC GATTATGACA GAGTTTGGGA 2220

CGATGGCGGA TATGGATCGT CTAATTGAGG AAGCGAAAAA GCGAAAGATT GAAATTATTA 2280

TGGACTTAGT GGTTAACCAT ACGTCAGATG AACATCGTTG GTTTATTGAA GCGAAAAAAA 2340

GCAAGGAGAA CCCCTACCGA GACTATTATG TATGGGCTGA TCCTGCTTCA GATGGCGGTG 2400

CGCCTAACCG ATTGAAATCT GCCTTTTCGG GTTCGGCGTG GACGTTTGAT GAAGCGAGTG 2460

GGCAATACTA TTTGCATTTG TTTAGTAAAA AACAACCAGA TTTAAATTGG GGAAACCAAC 2520

AAATGCGTCA GTCGGTTTAC GAGATGATGA ATTTTTGGAT TGATAAAGGA ATTGGCGGTT 2580

TCCGTTTAGA TGTAATTGAC TTAGTGGGTA AAATTCCAGG AGAAGAGATC ACCGCGAATG 2640

GGCCTCATTT ACATCGTTAT TTACAAGAAA TGAACGCCGC GACATTTGGC GGCAAAGAGC 2700

TGTTAACTGT AGGTGAAACA TGGGGGGCCA CACCTGAGAT TGCAAAAATG TACTCTAGTC 2760

CAGAGCGTCA TGAATTATCT ATGATTTTTC AGTTTGAACA TATGAGCTTG GATCAACAAC 2820

CAGGTAAAGA AAAATGGGAC CTTCAGCCAA TGGAAGTTGC CAAGTTGAAA CAGGTTTTTG 2880

CAAAATGGCA AACAGAATTA GGAAATGATG GCTGGAATTC TTTATTTTGG AATAACCATG 2940

ATTTACCGAG AATGATTTCA CGTTGGGGAA ATGATCAGGA ATATTGGTTG GAAAGTAGTA 3000

AGTTATTCGC TATTTTACTG CACATGATGA AAGGCACACC TTATATTTAT CAAGGGGAAG 3060

AAATTGGTAT GACCAATACG CCGATTACAG ATATTCGTGA AGCGCGGGAC ATCGAAACGA 3120

TTAACATGTA CCATGAGTAT CTAGAAAAAG GATATTCTAA AGAAGAAATT TTGCTGAAAA 3180

TTAATACTAA AGGTCGTGAT AATGCACGCC GGCCTATGCA GTGGACCGCT GAAAAAAATG 3240

CGGGCTTTAC GACTGGTACA CCTTGGATTG ATGTTAATCC CAATTATCAA ACGATTAACG 3300

TCGCTGCTGC TTTGGCAGAT AAAAATTCAC TATTCTATAC GTACCAAGAA ATGATTCGTT 3360

TGAGAAAAGA ACATCCTTTA ATCGTTTGGG GCAACTTTGA ATTGCTGGAA ACAGTGGAGG 3420

AAGTGATCAG TTTTTATCGT ACCTATGGTG AGGAACGCTG GCTTGTTGTA ACGAACTTTT 3480

CTGACAAAGT TCAGCCATTT TCTGCAGACG TTTCTGTAGA ACAAGTAATG ATTGAAAATA 3540

TGCCGACAGA TGTCACTGCT TTAGCAGACT ATTCGTTAGC GCCTTGGCAA GCATTTGTTG 3600

TGAAAGTGAG TCAATAAGAT GGACCAATGG TGGAAAAATG CTGTCGGTTA TCAAATTTAT 3660

CCTCGGAGTT TTAAAGATAG CAACGGGGAT GGCATTGGTG ATTTGCAGGG AATTATTGAG 3720

AAATTACCGT ATTTAAAAGA ACTAGGGGTC GATTTTTTAT GGTTAAATCC GATTTACACT 3780

TCGCCAAACG TTGACAATGG TTATGATATT GCGGATTATC AAGGGATTCa GCCaGAGTTT 3840

GGTACGATGG AAGATTTTCA AGAACTtTTG GACCAACGCA TCaATTAGGT TTGAAGATTA 3900

TTCTTGATTT AGTGGTCAAT CATACCAGTG ATCAACATCC GTGGTTTGTT GAAGCGAAAA 3960

AAAGTTTAGA TAATCCGTAT CGTGAGTATT ATTTATGGGC CGATGCGACA CCAGATCGTA 4020

TGCCGAATGA ATGGCAAAGC TTTTTTGGTG GCTCGACTTG GACGTATGAC GAAGGTACAA 4080

AGCAAGCATA TTtTCATGTG TTTGCAAAAG AGCAACCTGA TTTGAACTGG AAAAATCCTA 4140

AAGTTCGTGA AGAGATTTAT GCGATGATTC GCTGGTGGTT AGATTTGGGA ATTGATGGTT 4200

TCCGATTGGA TGCGATTAGT CATATCCAAA AAGAACCTTG GGATTTTAAA ATTACCACGA 4260

ATCCGTGGGC TCCATTTATG AATGTAAAAG GCATTGAAGA CTATATGCTG GATTTGAAAG 4320

TTATTTTTGC TGAGTATGAC ATCATGACAG TGGGGGAAGC TAGTGGCGTT TCAAGTAAAA 4380

AAGCGGTGGA ATGGACCAAT GATGCGGGCT ATCTGAATAT GATTTTTGAA TTAGAGCACA 4440

ATGTTCGAGA AGGAAAGCCA GGAGAAGAAC GGTTAAATAT TTTAGGCTAT AAAAAaGTAA 4500

TGGCTCGCTG GCAAAAGCAT TTAGGGACAG AAGGCTGGAA TGCATTATAT GTAGAAAATC 4560

ATGATAATCC CAGAATCAAT TCAATTTTAG GCAATGAAAC CTCTCATTCT GCTAAGGCGA 4620

TTGGAACAAT CGCCTTGCTC TTGCGAGGAA CACCGTTTAT TTATCAAGGA CAGGAAATCG 4680

GCATGGTTAA TTATCCGTTT CAACAGATTG ACGAGTTAGA TGCTAAAGAT TCTCATAACC 4740

ATTATCGTTT GCTTATmGAA AGTGGCTACG ATGCGAAcAA GCATTAmAmG AAGTGGCGCA 4800

TTGGACGCGG GACCcATTCG cGTAyGCCTa TGCAGTGGAC GAGTCaAGAG GCAAGTTCAT 4860

TCACCTCGGG GCATCCGTGG TTAGCAAT 4888

(2) INFORMATION FOR SEQ ID NO: 308:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2406 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

```
AGAAGTCAAA ATTCGAACAT TGTAATCAAA TGCTGACATA TTTTTGAGTA TAGCCCGCCC      60

AAATTGTAAA AACATTGTCT ACGAAAGTAT AATACATATT TTCTTGTAGT AAGGTATAGC     120

CTTTTTCCCA CTCGGCTTGT AAAACAATCG ATTTTAAAGT AGTACCTTCT TTTTTAATTA     180

ATAAAGATAC TAATAAACTA TCATGAATTA TTTTTGTATT GTCACCCATA TAAACGAAAT     240

CAATCACTCC GCTGGCATTG CTTGTATCGT TATCTGCAAT AATAATTGAA AACATTTCAA     300

TTGCCGTCCC ATTCCCCTGG GTCTTCTCTT GTCCAAAAAT TACTTTACCT CCAmGCTCTT     360

GAATATGAAC ATTCGGAATA ATAAAGGtTT TTTTCATCCC AAAATCnTTA TAAATATTTk     420

GATATATtTC TAGATtGTAT TGTAAAATAT TTtGGATTCG TGTGATACGC TCTACTTGTT     480

CTGGATCATT AGAATATTTT TCTTTTTTCA CTAAGTCTAG CCCTAATTGT TGGATGTATT     540

TCTCTTTTAA TTCTTGATAG AAACGTTCTT TAAATGCTTG ATCAACACGA TTCATATCAA     600

AAATAAACCA GTTTAATGTA TTAATAATTT TTAACCGATA TGCTTCAATT CGATTTGGAG     660

CATGTTGTTT TAAGTACTCC AACATTTTGT CATGTAAATT GAACTCATCA CGATAATTCA     720

TTCCATACGT AGACATCAAG CCATCTTGTC TAACGATTCG TGTCAGCAAA TATTTGTTTA     780

GTGCAACAAT TTTTTCGGTA CCTAAtAATG CTGTATAAGT GAAATAAGCA TCCTCGCCAA     840

CAAATTGTTC TAAGTTTACC AAGTTATTTT CTTTGATGTA TTCTAATAAA AATAATTTAT     900

CATATAAATA AGGGGTTACT TCATCCAATG AAATTTCTTC ACCAGACATA CGTCCGCCAT     960

TAGTTAGTAA TGGATTGTGA ACAACGTGCA CTTCTCCATT TGGTAACTTT TCTTTCATGc    1020

CAAATAACAG AATATCTGTT GCACTATCAT ATGCACGATA ATAAGCGTCT TGTAATAAAT    1080

TCGCTTCAAA ATGATCATCT GCATCCAGAA ATTGCAAATA CCGTCCTCGT GCCATCGAAA    1140

TGCCCACATT TCTTGCGTGT CCAGGATTAC TTTTGCCAAT AGAAATCCCC GTAATTCTCG    1200

GATCTTTTAC TGCAAATTCG TTAATAATAT CCATGGTATT ATCTGTAGAG CCATTCTCTA    1260

CTAAAATGAT TTCAATCTCA CGTAAAGACT GCTCACAAAC AGTTTCTAAT GTTTGTCTTA    1320

AATATTTTGA aGCATTGTGA ATCGGCATGA CAACACTAAC TTTATACAGA TTTTTCAAGT    1380

AGTCCTCATA ATAAgCTGCT GGgATCATTg CAATTTGATA AAATtGTTTG ATGTATTCcA    1440

AAATCGATAT CATnCAGCTG GTCTGTTTCT AAAGCTTGtC TAAAATCTAC TGCCcCTTTT    1500

gAATCACTTC ATAAACCGAC TCATCACGTA AATGGGTTAA TTTTTCAATC GCCCAAAAGT    1560

ATTTTTCGAA CATTTGTTTA TAAAAATAAC CTTTGTAGCG TTCAAATTTa TTTkGTTCAA    1620

TTAAATTTGA aCmAATTAGT TTAAATTGTT GAAACAAGAA AGGAmCCTTT GTTGCCACAT    1680
```

TATTTACTGT CGAATTTGGG TTGTCTTGTC TATAAAAATA CAwACAtCTT TAATAAAAAT 1740

AACATTCTGA GATGCTGAAT TGATTTTCCA AAATAGACCA GTGTCCCCAT AACTTTTAAT 1800

TTCCTCATCC CaAGTTACGT TTTCTTTCTC AATCATACTT TTTTTATAAA TGGCATTCCA 1860

TGGATAAACT TGTACAATTA AATCCGTCGT TTTTAAGTCA ATCAATTGAT TGTATTTATC 1920

GGGATATTGC CATAATATAT CTCGCTCAAC CTCTTCTTCA GAAGAAAATT CAACCCAATT 1980

ACAACGGACA ACATCTGCAT GATGACTTTT AGCCGTATTG TATAAACGCT CATAGGCATG 2040

CAATGCTACA TAATCGTCAC TTTCAAATTC AGCAATATAC TCACCTTTTG CTTCTGAAAC 2100

ACCCAAGTTA AAGGCTTTTC CAATTCCCCC ATTTTTAACA TTGAAAACGC GAAAACGTGG 2160

aTCATCTTGT GCAATTTCTT CAATAATTTT TTGACTATTA TCAGTAGAGC CATCATTAAT 2220

AATTAAGAAT TCAATTTCTT TCAATGTTTG GkTTTTCAAA CTCATTAATG CCTCCnCCAA 2280

ATAGGGTTCC ACGTTAATAG GACnGGnACA ATnGACTGGA CACTTTCACC ATTTTTCCCA 2340

TAAAACCAAT TTCCCCCCAC TTAACCATTA AAAGTGGGAA ACCCCAAAAT AACCCCGGAT 2400

TTAAAC 2406

(2) INFORMATION FOR SEQ ID NO: 309:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9179 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

TTTTGATTGG TGCCTGTTTC GGTTCTTTTT TATGCTTAGT GGCGGACGCT TACCAGTAGG 60

GCGTTCTTTG TGGTGGCcCC CTTCCCACTG TCAAGGTTGC CACCAGCCTC TACAACTCTA 120

TGAATTAATT CCAGTTGTTT CGATTTTGCT GCAGCGCTTT CGTTGTCGAA AGTGTCAACA 180

GCCAGTAGCT AAGTGTTATC TTTTAGCAGA ACTTGTCATG GGCGGGTTAA CCGCAAGCTG 240

TTTTTTTGCT GGCCTGACGA TTGATGCTTT GATTCTCTGG TGGTGGCTGA CGAGCGCTTT 300

CACATTATCC TTGATTGATT ACTGGTATTT AGTGGTTGAA CCGAAGATTC TTTATCCTAG 360

TTTTTTTGTG CTTTGCTTGC TGAAAATTGC GGGTCAGCAT TCGTTTTACT TACTGACTGG 420

TCTTTTTTGT TTTTGTTTTT TCCAAGCCGT CTTACATTAT TTTCCGGAAG CCATGGGACG 480

AGGGGATCTT TTGTTACTGG GTTTGTGGGG CTGTTTTTTA CAGGTACCTC AGTTGCTGAT 540

GCTACTATTT TTTGCCAGTA GTTATGGCCT AATCTATGGC TACAGCTGTA AATTTTTGGG 600

GTATCCAGTT GAACAGACAC TTCCCTTTGT CCCCTTTTTA AGTTTGGGCT TACTCACAAT 660

CAGTTGGTTA TAGCAAATAA AAAAAGCCGT AAGGAGCATG CTCCTTAsGG nTTTTTAGGG 720

tTTATTTATT GATGTTTTTC ATCGCATCTT CAGCAGCCAT TTGTGCTTCA ATATCTGAAA 780

TACTATAAAC GTTTTCGCTA GCTACGCCAA CTTCTTTTGG TTCCATGTTA CGGTAACGAG 840

CCATACCAGT ACCAGCTGGG ATGATTTTAC CAATGATAAC ATTTTCTTTC AATCCAAGTA 900

ATGGATCTTT TTTGCCACGG ATAGCAGCAT CTGTTAACAC GCGTGTTGTT TCTTGGAAGG 960

ATGCAGCAGA TAAGAAGCTG TTTGTTTCCA ATGACGCTTT GGTAATACCA AGTAAGACTG 1020

GACGACTAGT AGCAGGGACG CCACCAGCAA CTAATGTGTC GTAGTTACGG TCTTTAAATT 1080

CTGCGATGTC CATTAATGTT CCTGGAAGaT TTCAGTGTCA CCTGGATCCA TTACGCGGAT 1140

TTTACGTAAC ATTTGACGAA CCATTACTTC GATATGTTTG TCGCCGATTT CTACCCCTTG 1200

CATACGGTAA ACACGTTGTA CTTCACGTAA TAGGTAGTTT TCAACAGATA AAACATCACG 1260

AACTTGTAAT AATTGTTTTG GATCGATTGA ACCTTCTGTC AACGGTGCAC CACGGTGGAT 1320

GATGTCGCCT TCAGCAACTT TCATACGTGC AGTGTAAGGG ACAGTATAGG TACGTGTATC 1380

TGTTTTTCCT TTGATGGTTA CTTCTTTTTG ACGTGTGGCA GGATCTTCAG AAATATCGAT 1440

AACTTCACCG GTTACTTCTG TAATCACTGC TTGCCCTTTC GGATTACGTG CTTCAAAGAT 1500

TTCTTGGATA CGAGGTAAAC CTTGAGTGAT ATCGTCACCG GCAACCCCGC CGGTATGGAA 1560

CGTACGCATG GTTAACTGTG TACCAGGTTC CCCGATTGAT TGAGCGGCGA TTGTACCAAC 1620

TGCTTCTCCG ACTTCAACGT CAGAACCAGT TGCAAGGTTA CGTCCGTAAC AGTGCTTACA 1680

TACACCATGT TTTGTGTTAC ATGTAAAGAC AGAACGGATG GTTACTTTTT CAATACCAGC 1740

GTCCACGATT TCTCTGGCAA GATCTTCAGT GATTAATTGA TCCGCACCAA TGATGATTGC 1800

ACCAGTTTCT GGATGAACAA CTGATTTACG TGTATAACGA CCTAATAGAC GTTCGTCTAA 1860

TGGTTCGATA ATTTCGTTAC CTTCACGAAT TGCTTCGATT TCAAGACCAC GGTCAGTGCC 1920

ACAGTCGTCT TCACGAATGA TTACGTCTTG GGCAACGTCA ACTAAACGAC GTGTTAAGTA 1980

ACCTGAATCG GCTGTCTTCA AGGCCGTATC GGTCATCCCT TTACGAGCCC CGTGAGTTGA 2040

GATAAACATT TCTAAGACAG AAAGTCCTTC ACGGAAGTTT GAGATGATTG GTAATTCCAT 2100

GATTCGTCCA TTCGGAGCGG CCATCAAACC ACGCATACCA GCTAACTGTG TAAAGTTGGA 2160

GATGTTACCA CGGGCTCCAG AGTCAGACAT CATGAAGATT GGGTTACGAG CTTCCATACT 2220

TTCGATCAGT TTTTGTTGGA TTTCGTCTTT GGCACCGTTC CAAACGCCGA TTACACGTTC 2280

ATAACGTTCG TCATCAGTAA TTAAACCACG ACGGAATTGT TTAGTAATTG TTTCAACTTG 2340

TTTGTGGGCA GCATCAATGA TTGCTTGTTT TTCATGAAGA ACCATGATAT CGGCAATCCC 2400

TACTGTCATC CCAGCATAAG TTGAGTGTTT GTAACCTAAG TCTTTCATGC GGTCAAGCAT 2460

TTTAGAGGTT TCCGTAATGT GGAATCTCTT GAAGACTTCA GCAATGATAT TTCCAAGATT 2520

TTTCTTCTTG AATGGTAAAA CTAATTCTTG CTCTTTAATG TGTGCAGGAA TGTCTGTACC 2580

AGCTTCTACG AAGTATTTAT CTGGTGTTTG AACAGTTAAG TTATAGTCTG TTGGTTCGTT 2640

CAAGTAAGGG AACTCAACAG GCATGATTTC GTTAAAGATG ATCTTACCAA CTGTGGTGAT 2700

TAAGATTCGT TCTTTTTGCC AATCAGTAAA TGGTTTGTCA CCTAAAAGTG TTGTTTGAAC 2760

CCCAATCCGT GAGTGTAAAT GCACGTAGCC ATTTTGCCAT GCTAAAACAG CTTCGTTCAT 2820

GTCGCGGAAG ATCATGCCTT CCCCTTCACG GCCTTCTTCT TCCATTGTTA GGTAGTAGTT 2880

ACCTAAGACC ATATCTTGAG ATGGTGTAAC AACTGGTTTA CCATCTTTCG GGTTCAAGAT 2940

ATTTTGGGCA GCAAGCATTA ACATACGAGC TTCTGCTTGT GCTTCTTCAT TTAGAGGCAC 3000

GTGAACCGCC ATTTGGTCTC CATCGAAATC GGCATTGTAG GCTTCACATA CTAATGGGTG 3060

AAGACGAATC GCACGACCTT CAACAAGAAC AGGTTCGAAA GCTTGGATTC CTAATCTATG 3120

CAACGTAGGT GCCCGGTTAA GTAAAACTGG GTGTTCTTGG ATAACTTCTT CTAAGATATC 3180

CCAAACTTCA TCTTCGCCAC GTTCGATTTT ACGTTTCGCA TTTTTAATGT TTGTCGCAAT 3240

TTCACGTTGT ACTAATTCAC GCATTACAAA TGGTTTGAAT AATTCAATCG CCATTTCTTT 3300

TGGTAAACCA CATTGGTACA TTTTTAAGAA CGGACCAACG ACGATTACGG AACGACCAGA 3360

ATAGTCAACA CGTTTACCTA GTAAGTTTTG ACGGAAACGA CCTTGTTTCC CTTTCAACAT 3420

GTGAGAAAGA GATTTCAATG GACGGTTACC TGGTCCAGTA ACTGGACGGC CACGACGACC 3480

ATTATCGATT AAAGCATCGA CCGCTTCTTG TAACATCCGT TTTTCGTTTT GAACAATGAT 3540

GCTTGGCGCA TTTAAGTCTA ATAAACGTTT TAGACGGTTG TTACGGTTGA TAACACGGCG 3600

GTATAAGTCG TTTAAGTCAC TTGTTGCAAA ACGACCACCT TCCAATTGAA CCATCGGACG 3660

TAAATCTGGT GGAATAACTG GGATAACATC CATTACCATC CAGCTTGGTT GGTTACCAGA 3720

AGCACGGAAG GCTTCTAAGA TGTCTAAACG ACGGATGGCA CGTGTCCGTT TTTGACCAGA 3780

AGCTGTTTTT AATTCTTCTT TTAATTGCGC AACTTCGCCA TCTAAATCGA CGTTATCTAA 3840

TAATTGTTTC ACGGCTTCGG CACCCATAGC AGCTTTGAAG GCTTGGCCAT ATTGTTCACG 3900

TTTTTCACGA TATTCGCGTT CCGTTAACAA TTGTTTTTTC TCTAATGTTG TGTCACCTGG 3960

TTCGATGACA ACGTATGAAG CAAAATAAAT GACTTCTTCC AATGCACGAG GGCTCATGTC 4020

TAAAACAAGA CCCATCCGAG ATGGAATTCC TTTAAAGTAC CAGATATGTG AAACTGGTGC 4080

GGCTAATTCA ATATGTCCCA TACGTTCACG ACGTACTTTT GAACGAGTGA CTTCAACGCC 4140

ACAACGGTCA CAGACAATAC CTTTATAACG GATACGTTTA TATTTACCAC AAGCACATTC 4200

CCAGTCTTTG GTTGGACCAA AAATGCGTTC ACAGAATAAC CCTTCACGTT CGGGTTTTAA 4260

TGTACGGTAG TTAATTGTTT CGGGTTTTTT AACTTCCCCA TAAGACCAAC TTCTGATCTT 4320

TTCTGGTGAA GCAAGCCCTA TTTGCATACT TTCGAATTTA TTTACATCGA TCAAAAGGAA 4380

TTCCCTCCAT TTCCTTATTT AGAGAAGCTG AGTCAGCAAC TTGTTAGCAA ATGGCAAGCA 4440

TCACTTGAAA GAAGGTCTGA CGACACTTGG AACAAGTAAT GGCTTGCGCA GTTGCAACAA 4500

AGTCGCAGTT GAATCAACTC ATCAGAGTGG TAGCGACAAA TTAGTCTCTG TCTTCGGCTG 4560

TTTCAATTTC TTGAACAACG TCTTGTGCTT CTTCTTTCAC GACTGATTCT GCTTCTTTTT 4620

CCAGCTGTTT CGCTGATTGT TGTTCCGCAA ATTTGGTTAG TGCGTCAACG GTGATTAAAT 4680

CATCATCGTC ATCGTCCATA TCGCGTAACT CAATCTCAGC TTCCTCAATG TCCAAGACAC 4740

GCATGTCTAA ACCAAGTGAT TGTAATTCTT TCACTAATAC TCGGAATGAT TCAGGCACAC 4800

CTGGTTTTGG AATTGGTTCA CCTTTAACGA TGGCTTCGTA TGTTTTCACA CGTCCGACAA 4860

CGTCATCAGA TTTGTAAGTT AAGATTTCTT GTAATGTGTA AGCAGCACCG TAAGCTTCCA 4920

GTGCCCAAAC TTCCATTTCC CCGAAACGTT GTCCACCGAA TTGAGCTTTA CCACCCAACG 4980

GTTGTTGCGT AACAAGAGAG TAAGGTCCGA TTGAACGAGC ATGCAATTTG TCATCAACCA 5040

TGTGGGCTAA TTTAATCATA TACATGACAC CAACGGAAAT ACGGTTATCA AATGGTTCAC 5100

CTGTACGTCC ATCGTAAAGA ACTGTTTTAG CATCGCTAGC CATACCAGCT TCACGAACAG 5160

TTTCCCAAAC GTCTTCATCG GTTGCCCCAT CGAAAACAGG TGTTGCGACG TGAATGCCTA 5220

ATTGGCGAGC AGCCATACCT AAGTGTAATT CAAGTACTTG TCCGATATTC ATACGAGAAG 5280

GTACCCCTAA TGGGTTCAAC ATGATATCAA CAGGTGTTCC GTCAGGTAAG AATGGCATAT 5340

CTTCTTCCGG CATAATACGG GAAACAACCC CTTTATTTCC GTGACGTCCC GCCATTTTAT 5400

CTCCTTCGTG AATTTTACGT TTTTGAACGA TATAGACACG AACTAACATG TTGACACCTG 5460

GTGATAATTC ATCGCCAGCT TCACGAGTAA AGATTTTCAC ATCGTGAACG ATACCGCCGC 5520

CACCGTGAGG TACACGGAGA GACGTGTCAC GAACTTCGCG GGCTTTTTCC CCGAAGATTG 5580

CGTGTAATAA ACGTTCTTCT GCAGATAATT CTGTGACCCC TTTAGGTGTG ACTTTACCAA 5640

CTAGTAAGTC GCCATCTTGA ACTTCAGCAC CAATGCGGAT AATCCCCATT TCGTCTAAGT 5700

CTTTCAATGC GTCTTCCCCA ACGTTTGGAA TTTCACGGGT AATTTCTTCA GGTCCTAATT 5760

TTGTATCACG TGCTTCTGAT TCATATTCTT CAATATGCAC AGAAGTGTAA ACATCGTCTT 5820

TAACTAAACG ACGGCTCATG ATAATGGCAT CCTCGTAGTT GTAACCTTCC CATGTCATGA 5880

AGGCAACTAA GACGTTTTGC CCTAAAGCCA TTTCGCCTTC TTCCATAGAA GGTCCATCCG 5940

CTAAAGTATC GCCTTTTTCA ACTTTTTCAC CTAAGTGAAC AATTGGGCGT TGGTTGTAGC 6000

TTGTTCCTGA ATTTGAACGA CGGAATTTAG TAACCATATA TTTGTCTAAT GCGCCATTGT 6060

CGCGACGAAC GCGAATTTCT TTTGCATCGA CGAATTCTAC GACACCGTCA TGTTTACATA 6120

GTAAAGCAGC ACCTGAGTCA TGGGCTGATT TATATTCCAT CCCTGTACCT ACCCACGGAG 6180

AGCGTGGTTG AATTAACGGC ACCGCTTGAC GCTGCATGTT GGCACCCATC AAGGCACGGT 6240

TGGAGTCATC GTTTTCTAAG AAAGGAATAC ATGCTGTTGC GACTGCGACT ACTTGTTTTG 6300

GTGAAACGTC CATGTAGTCA ACTTTGTCTA CGGCAACTTC TAAGTTTTCA CTTTGTAGAC 6360

GCGCCATAAC AACATCATTG GCAAATGTGC CATCTTCATT TAAAAGTGAG TTCGCTTGCG 6420

CTACGATATA ATGGTCTTCG ATGTCTGCTG TTAAGTAATC TACTTGATCA GTAACACGGC 6480

CTGTCGCACG ATCAACACGG CGATAAGGCG TTTCGATGAA ACCAAATTTA TTCACTTTTG 6540

CGTAACTAGA TAAACTATTG ATCAACCCGA TATTTGGTCC CTCAGGCGTT TCAATTGGAC 6600

ACATACGACC ATAGTGAGAG TAGTGAACGT CACGAACTTC ATAACCGGCA CGGTCACGAG 6660

TCAAACCACC AGGCCCTAAG GCTGATAGAC GACGTTTATG GGTTAACTCA CCTAATGGGT 6720

TTGTTTGGTC CATGAACTGT GATAACTGTG AAGAACCAAA GAATTCTTTG ATACTTGCTA 6780

CCACTGGACG GATGTTGATT AATTGTTGTG GTGTCAATGT TTCTGTGTCT TGAATAGACA 6840

```
TTCTTTCACG AACCACACGT TCCATACGGG CTAAGCCAAT ACGGAATTGG TTTTGTAATA   6900
ATTCACCTAC TGAACGGATA CGACGATTAC CTAAGTGGTC GATGTCATCG ACATTGCCGA   6960
TATCTTCCAT TAAGTTGAAG AAGTAGCTCA TTGAAGCAAC GATATCTGCT GGACGAACTG   7020
TTTTTACGCT GTCGTCTGGA TAGCCGTTAC CAATCACATT TACGATACGT TCAGGATCTT   7080
TTGGTGAAAG AACTTGAATC ACTTGGATTG TCATTGGTTC AGTTACTACA GCATCTTCAC   7140
TTGGATAGTA AGTTACGCTG TTTAAGCCGT TGTCAATGTA TTCGCCTAAT GTTTCCATGA   7200
TTTGGTGTGT TAAAACCGTG CCTTTTTCGA CAATGATTTC ACCAGTTTCT GGATCAACTA   7260
GCGTTTCCGC TAAGGTTAAG TTTAATAGAC GTGTTTTTAA GTCTAATTTT TTGTTAACTT   7320
TGTAGCGACC AACGTTTGCC AAATCATAAC GTTTTGGATC AAAGAAACGT GCAGTTAACA   7380
AGCTACGTGA GCTATCTGCT GTTTTTGGTT CGCCTGGGCG AAGACGTTCA TAAATGTCTT   7440
TCAAGCCTTC TTCTGTACGA GAATCACTTG CATTTTtGTG TAAATCTTTT TCAATTGTGT   7500
TGCGTAAGCT TTCGcTGTCG CCGAAAATTT CGAAGATGGT ATCATCTGAA CCGAAACCTA   7560
AAGCACGAAC TAACACAGTT AAAGGAATTT TACGTGTGCG GTCAATCCGA ACaTAAGAAA   7620
TGTCTTTCGC ATCTGTTTCC ATTTCTAACC ATGCACCACG GTTAGGAATG ACTGTTGAGC   7680
CaAAACCTTC TTTGCCGTTT TTGTCCACTT TTCCATGGAA ATAAaCACCT GGAGAACGAA   7740
CTAATTGGGA AaCGATAACA CGTTCTGCCC CGTTGATGAT GAAGgTACCC ATTTCTGTCA   7800
TTAATGGGAA ATCGCCGAAG AATACTTCTT GGGATTTAAT TTCACCTGTT TCACGGTTGG   7860
TTAAACGTAA TGTTACATGT AATGGCGCAG AATAGTTGGC ATCATGTGCG CGGGCTTCTT   7920
CTACTGTGTA CTTTGGTTCT TTTAATTCAT AGTCAACAAA TTCTAAGGAT AAGTTTCCTT   7980
GGAAATCATC AATTGGTAAA ATGTCTTCAA ACATTTCACG AAGTCCCTCA TCTAAAAACC   8040
ATTGATAAGA GTCTGTTTGG ATTTCAATTA AATTTGGTAA TTCCAATACT TCACTGATAC   8100
GTGCGAAACT TCTACGTTCG CGATGCTTTC CGTATTTTAC TACGTGTCCA GCCAAGCTCT   8160
TCACCCCTCT AGTTATAAGT CTATCAAAAC ACCGATCCGA AAAATTACAT TTACGTTACA   8220
TATATTAAAT ATATGTAACA AAAAGCATTT TTTGCAGTTT TTAAGCAAAA AAAAACCAAA   8280
AATAGAACAT GCTTCCGAAA AATTTTCTGG CGTTTCTACT TTTGTTTTCC TATAAAAAAC   8340
CGGAACGGAC AATATTTCGC TCCTGCTTCG TCACTTTCAG ATAGTTTTTG CTCTATAGAT   8400
TCTACTACCT TTGTCTAGAA AAGTCAAGGG TTTCTTTAAG GTTTTCTTTT TCATAAAAAG   8460
GAGTTTTTTT GTAAAGGGTT TTAATGAAAA AATGCGTAAT AAATAATTAC GCATTTTCCC   8520
CTATCTTCCT CTTCACACAA AAAAACGAAT GCAAAGATGC ATTCGTTTTT GACTAACCGC   8580
CACCTAATAA AAAGTCTAGG ACATTcATGC CGCCGCTTGT TTGGCTGCGG TAAAGTTCGG   8640
TGTAGCCGCA TTGGGTACAG CTAATTGTAA TAAATTTTTT ATTTkGAmCA TCAAAAATTT   8700
TCGCAAAGTT CCCACCTGTT GCTTGAAATT GGTCCGACAC ATAGTGCATG TTGCCACATT   8760
TCTCACAAAC ATATTGTTGC TTTTCCATTC CCCTCACTCC TTCTCAATAA AACAAGTATA   8820
```

GCAAACGCCT GCTTCATTTT CCTCCATCCT AAGATGGATT TTTTCACTAG TTGACCAAGC 8880

CTTTTCCTTT CCGCTACAAT AAGCATAACG AATCGGAGGA AAAAATATGT CGACAAAAAG 8940

CCAAGTCCTA ACTTTATTGA TGAAACAAAC GCCCGCATTC CTTTCGGGCT GTCTCTTGAT 9000

CAGATCTTGA TCCCCTGCGC CATCAGATCC TTGGCGGCAA GAAAGCCATC CAGTTTACTT 9060

TGCAGGGCTT CCCAACCTTC CCAGAGGGCG CCCCAGCTGG cAATTCCGGT TCGCTTGCTG 9120

GTCCATAAAA CGGCCCAGTC TAAGnATnGC CATGTAAGCC ATGCnAGTAA CTGCTTCCC 9179

(2) INFORMATION FOR SEQ ID NO: 310:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4289 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

CATAAGTCGA AACGACTTAT GCCCCACTTC CGTTTTTTTT ATTCCTTCAA TGAATTTTGA 60

AACGTGCTGA TTTTTTCAGC CAATCTTTGA TAAGTCGTAA AAGAcGGATG CTGTAGCTCT 120

TGTTCTTCGT TAACGGATAA AGTTTGTTGT GTCGTAATGG CCTTTGTTAG TGCATCTACA 180

AAATGAATAA AAGAATTGTT GGTTTGTTTA TTGTATCCTG GTGAATAATA TTTCGATTGA 240

CGATAAAAAG GCCGCAGGCT GTTTCTGTGA TTTTTTCCTG TTCATCTGAA AAGTATTCAT 300

CATAGCTGAC ATGTTTTATT TTCTGCTCTT CTGTTATCAC TTGAAGTCGA GCGATTAGGC 360

CATTGCCTAA ATCTTCTGCG TAACCGATAT ACGTTTCTTT CGATGGCTCT TTTAGCCAAG 420

TGTCCATCTC TGCTGCTAAA GGCATTAATC CTTCACGCGC AGAAGTTGAT GAACCGCGTA 480

CCGTATAAAA ATTACCTGTA ACACGATTTT CGTCACGCAT TTGTTTTTCT ACATAAGTAA 540

TACCATTAAC AACGGTGACT AACGTATCGT CTGTACGCGT GTTGCTTGCT TGAAATTCAG 600

GATAATCCGA CAGGCGTTTG TTGACACCCG AATAGCGTGT ATCATAATAA GGATCACTAG 660

CAAACTCTGT AAACTCAACA TTCAGAATCC TTTGCGTTTc AGGTTCCGTA ATAATCTCTA 720

CAATGCCGTA ATTTCCGCCA GTAAATACAC CTTCATTTTT ATAATAGTTA CCCGTCATTA 780

ATCCCAAAGG CGGTTGATTG GACCAAAACA TTTTTTTCTT TTTTTCTTCC GGTGTATATT 840

TTGCTGGTGG GGTTGTTCCA AATGTCGTTT GTGTGGCCCC TGTAACCACT TCGTACGTGT 900

ATTCATGAGA ATCAAAAGTC AATTGATTAG CAATCAGTTC GTCATTTTTA GTCGGCGCTT 960

GTTTTTCCAC CGCAGAAGTC GTTTGATTGG TGACGGTAAC TTGCTTCACT TGCTTTGTCT 1020

CTTGCGGACC ACTTGTTGTT TGTTTCGTTG GCTCTTGATC TGGCGCACAA CCAACCAAAA 1080

TAAATAATCC GATAAAAACC AACATTCCTT TAATATGTTT CATTCACTTC TCTCCTTTTT 1140

ATAAACGCTT AATTTCCTCT TATCTCCAGT GTAGCAAACA TTTATAAACC AAAACTGACG 1200

GAAAAATGAA TAAATAATAA ACAAAAACGC TAAGTCTCCT GCAGCTTAAG AGACTTAGCG 1260

TTTTTTTATT TAATTGCTTG GCGCTTTTGG CGCAACCACT CTTCTTTAAT GTGTTCAGAT 1320

AGGCGAATCA TTCGTTTACC AACTTCACTA TATTGGTTAG TAGCTAAATT GGCTAAAGAA 1380

ATTCGAACTG ACCATTTGTC ACTCCCGAAA GCATCCGCTT TTAAAAGCAT TAATTTTyCT 1440

TTTTCTGCTA ACGAAGTTAA TACTTTTGTG ATCGTCCACG ATTTCGTTAA ATAACTTGAA 1500

AATTCTGGGC CATAACGTTT TTCTGTCCAA TCTCTAAAGT TAATTTCACA ATAGTATGCT 1560

GTATTTAAAC TTGCTAAAGG CTCTTCAATG CCTAAGGTGC GGAACAAGAG CTTTTCACGC 1620

GTATGACAAA TATCCATCAC TTCGTCTTTA TATGCTTGAC CGTCTTTTAA CAAAGCATAC 1680

AAACTAAATA AAGCCATCAT TGCTTGCTGA ATAGAAGaAA GTCCCGCTGC ATGGTTCAAC 1740

GCAATATCGC GGCTGTctGC GACTAAGCGG CTGATAAAGT TAATTTGGCT TGTGTCTGCA 1800

TTCAATGTAG CATaACGTGC CTGTAATTCC ATTTTTCGAG CCACTGGCAA TTCTTTTAAT 1860

AATTGATCAA AGATGTTCTC TTGACTGACC GCAATAGTCC CCACACGCCA ACCGGTAGCG 1920

CCAAAGTATT TAGAATAAGA ATAGATACAG GCTGTATTGT AAGGTAACTC CGTAAATAGT 1980

GAACGGAAAG CTGGCACAAA GGTTCCGTAA ACGTCATCTG TTAAAATCAT TAACTTAGGG 2040

TTATCCACTG CAGAGATTTG TTTGATTTGT TCAATGGTAG GTTTTCCCAT AGCATTAGCA 2100

GTTGGATTGC TGGGATTTAC GACAAAAACC GCTTTGATGG ACGGGTCTTT TAGTTTATCA 2160

ATTTCCTTGT TTGAGTATTG ATACGTCGTT TTTCCATCAA TGATCATTTG CTCTGCTTTG 2220

ATTTTAACCA CGTCAAAGTC ATAACGTGGT AGTTCTGGAA TTTCTAAATA AGGAGCAAAC 2280

GTTGGTAATA ACAACGCAAT ACGATCACCT TTTTCTAGTA AATAATTATT TGCCAACGTA 2340

TCAAACAAGT AACAAATCCC TGCTGTACCA CCCTCAACGG CAAAAATATC AAACGGTTGT 2400

GCTTCGCTGG AAAATAATTC TTGATTTAAA TATGCTTTAA TCGGTTGTTC ACACGCCTTC 2460

AAGCAACGAA CAGGATTCGG ATAATTATCC CCAATCACAT AGTCTAACAT GGCATCTAAC 2520

CACATTTCCT TATCCATCCC AAAATAGTTA TGCTCTGCTG TCCAAATTTC tTGTAAGAAG 2580

GTTGCCCCTT TGCTTGGATT TTCTGCTAAA AAGTTTAAGA AGCGTTGCGT TCTACCAAAA 2640

CTTGGCTGAA TCATACCAGC CGTTTGTTCG CTCCCCTCTc TTAACGTTTC TTTTGTAGCA 2700

AATTGACCTA ATAAAAAAAA GGCTTCCCTT GGTGTTGGCG CCGTCCAATT AGGATTGCCT 2760

CGCCCTGCAT TTAACAATGT TTCTAAATTA TGATTTTCTT CATGTAATTT CTTCTCTAAA 2820

AATAGACTTA ATTCAAACGG ACTAAGGTTT TCTGTCAATT CATCACTTGT TATCATTTTA 2880

TCCTCCTGAA GCACACACAG TCATTTAGTA TTTAGTATGT AGTCTATACT ACTCTTCTAC 2940

CTCCAGATGT CAACCATTTA AGCGGAATAT TCTGACAATT AATAAAAAAG AAACAATCCT 3000

TTATAAAAAA TAAAGGATTG TTTCTTTTTT GCTGATTAAC CAAAGAACCG TTGCATCTCA 3060

ATTTTTGCTG ATTGAACAGA ATCAGAAGCA TGGATTAAAT TTTTGGTGAC AGACAACCCA 3120

TAATCCGCTC GTAAGGTGCC ACTTTCAGCA TCTAACGGAT TCGTTGCTCC AACTAATTTA 3180

CGCACTTTTT TAACCACATT GTCCCCTACA ACGACCATCA CAACCACTGG GCCACTTGTC 3240

ATATAGGCTA ATAATTCAGG GAAAAATAGT TCATCTTTTA AATGCTCATA ATGTTCCTTC 3300

ACTGTTTCTA CTGTTAGTCG TTCGAATTTA AGGTGCAATA AGTCAATACC TTTATCTTCA 3360

AAACGTTGAA TGATTTTGCC AACTAACTTT CGtTCAACGC CATCTGGCTT AATAATTACT 3420

AAACTTGTTT CCTTCATTTT TTCCATCCCT TTCAAAACTT CCTTTTGCCC TTTCCCAAGG 3480

GGTAGCTACA GAATAAAATG AAAGCGGATA CTTTGTCAAT GATAAATTTT CTTTTCCTTT 3540

TGATTAAATA AAAAATTAAA GATTATTCAC AAGATTTGTG AATTGTTCTT TCAGTTCTGT 3600

ATTGCTTGTT TCTTTTAAAA ACAAATAAGG TTGCTCCTCT ACTGGCTGTT CTTTACTCAC 3660

ATAAACTTTA GCTTCACCAT TCACTAAAGT AAATAAATAG ACATGCTTCT GGGCAACTTG 3720

TGTCGCACTG TCAGAGTATA CTGCCACCAG ACAATAACCC CCTTCGCTGC CTGTTTCTGA 3780

CCAATTTAAC CAAATGGGAT TGTTCCCAAT AGCTGGTTGT ATTTTAGGTT CATAACTCAA 3840

TAAATCATCA GGCACTTGAA TCGTGTAAAA AGCTACAGAG TGACCTGGTT GGTACGCTTG 3900

ATAGTTCTGA TTGCGTTCTT GACCCCAGGT TTCCATGTAT TGTCCTAACT GTTTTGCTTT 3960

TTGCTGATTC CAAATCGTCC CTTCTTCTGT AACTAATTCA GTTAATTCTT GTGTTTGCGT 4020

TGTTTGCTCT AATTCCAACG CCGTGGTTTG ACTAACTTCG GTTGCTGAAA TTGCTCTTGT 4080

TTCTTCAGTA TTTACATTGT TTGTTTGTAC GGTTTGTACT TTCTGTGATT CCTCTTTCGA 4140

TTTTTGCCAA CCACAGCTAG ATAAAACGCT TCCAACAACC GCTATCATAA TAATCCCGTT 4200

TAGTTTCATT CCTTAAACTC CTGTTTTTAT TTTTAtTCnA TCGAAATAAA CCTAAAATCT 4260

CAATCCCTTT CTTTGAATAG TnAAAACCA 4289

(2) INFORMATION FOR SEQ ID NO: 311:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7925 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

AGCTCGGTAC CCCCAATTAT TATTTAAAAG CTATATTGAA TATAGCTTTT AAATAATAAT 60

AACTGAATGA AAAAAAACTG TCCAATAATA TATAGAAATT TCTGAAAAAT CATTTkTAAT 120

GTAAATTATT TTAATAAAAT TAATAAGAAA GGAAAAAAAT ATGCAACTAG AATATCCATT 180

AACGAATCAT ATGAAAAGAA ACATGGTTTT AACTAAAGAA AATCAAGTGA TTGCGTATTA 240

TCGTATTCGG TCAGAAACAG TTGGCTTAAC AGATTTTGAG AAAAAAAGGA AAACAAAGAA 300

AAAAGTAGCT CGAACGTTGA AACGTTTACA AGAGAATGAG GGTTTTGAAA TTGTTCTTTT 360

ACCTGTGAAT GCGGATATTC GAGGGAAAAT GGGAGGTATG CGCCAATTAG TGGACAAAGA 420

AAATCATAGT GTAGCGGTGG ATAAGTTGAT GAAAACTGCT CAGTATCTTG AAAATGAAAT 480

TGGCATGGTC TATGAGTATA TTTGGTTAAT TGGCGTGCCT TTAGTAAAGA AAGAGCGGTC 540

AATTGATATT AAAGAAACAT TTTCTACGGC GTTAAATAAT CTATCTGAAA AAGTGGTTAA 600

AGGACTTGGT TTAGAAGTAG GAGTAGCGGA GGATTGGGAA GAAGCTTATA AAGACCAAGA 660

ACAAGAAGTA TATCAGAATT TGTCAGAATT GTTAGTGGAA CGATTGACAG AAGACGAGTT 720

GTATTACTAC CAAGCCTATC AATTTTTAAA TAATATTCAA CATGAGAAAA AAGAACTTTT 780

GACTAGTCAG AACTTAGATA ATTTACTAGC GAACAAAATC ACTCCTTTTC GAGGTGGGTT 840

AGAATTGTCT AATGAGTTTG GGAAAAGTTA TATTGCTCAT TTACCATTGG GGGATTGTGG 900

GGTCACGATT GATGGTAATC ACTTGCTAGA ACTGGTTCAA AAAATGTCTT ATCCTGTTTC 960

TGTGAAGTTA CAGGCTGGTT TTGCTGAAAC AAAAGGCCAA TTAGCGTTAT CAGGACGTTC 1020

TGCTCGTGCG AGAACACGAA CAAAGAATAT CATGGAAGAA GCACATTTAG CAGGTTCAAA 1080

ACAAAAAAGA AAAATTGTTG AGGGGCAACA TTCACTTGAT GATTTAGATC AAAAAATTGA 1140

TGATGATATT CCGATTATTG ATTGGAAAGC AGTATTAGTG ATTAGTGGGA ATTCAAAAAA 1200

ACAATTACGT GAGCGAAAGA AAAATTTGAT GAATCGACTA GATAGCTTAG GAATTCCGTT 1260

AATCCGTGCA ACGTTTGATA ATGTCTATCT TTTTCAAGCA ACGTTATTAG GCAACTTTCA 1320

ACGTTTTTCA ACCAGTAATT GGCAACATAC GTCGACGTTA GAAACATTTT CAGAGTTGAA 1380

TTTCTTTACT TCGTTACACG CAGGTACAAA GACAGGGTTT TATTTGGGAC GTGTGGACGC 1440

CACGTTAGAA GAAAAAGAAA GTCGGAAACA AATTGTTTCT GGCTCTAAAA ATATTGTCTA 1500

TATGAATCTA TTGCTTGCAA ACAAACAGAA TATTGAAGGA AAGAAAACAA ATAATCCCCA 1560

TTGGTTGGTA TCGGGAGATA CAGGTAACGG GAAATCAGTA TTTTCTAAAT GGTTATTTTT 1620

GTATTCTTCT CTATTAGATG TGAAAGTGTT ATACATTGAT CCAAAAAAAG AAGTTCGCCA 1680

ACAGTTTATG CGAACGATTA ATGATCCAGA ATATCAAAGA AAATACCCAC TTGATGTGGC 1740

ATTTATTAAG ACGTTTAATT TTGTGACACT TGATGTTCGT AAAAAAGAAA ACCATGGGGT 1800

ACTTGATCCA ATTGTATTAT TTGATGAAAC AGAAGCAATC GCAACGGCAA AAGCTATGTT 1860

AAATAATATT AACGAGGATA AGTGGAAAAT GCCACATAAG ACGGCGATTA ATGAAACAGT 1920

AGCAGAAGTG GTGGCTGAAC GGAAAGCAGG CAAACAAGTT GGATTTTGGC ATGTAATTGA 1980

ACGTTTAATT AGTCATTCAG AGAAAGACGT TCATGAAATG GGTCGTTTTT TATTATCTAC 2040

GATTAAAGGA TCGATTTTAG AGTTAGCCTT TTCTCATGGT GAAGTAGAAG GACTTTCTTT 2100

TGAGAAGAAA GTGACCATTT TAGAAATTGA AGATTTAGAT TTACCCACAG ATCGGCACGA 2160

TGAATTAGAC GAGAATCAAC GGTTATCTGT GACTTTAATG TTTGCATTAG GCACGTTTTG 2220

TTCAAAATTT GGGTCACGCA ATCGGAAAGA AGAAACAGTC ACTTTCTTTG ATGAAGCGTG 2280

GATTTTTCAA TCGTCACCAG AAGGACAAAA GATTCTAAAA TCGATGAAAC GGATCGGACG 2340

TTCGTTTAAT AATTTTATGG TATTAATTAC GCAATCGGTC AACGATGTGT TAGATAATGG 2400

CGATGGTACA GGCTTTGGAA CGGTGGTTTG TTTCGATGAA GTGGATAATC GAGAAGGGAT 2460

TTTACGTTAT TTAAAATTAC CAGTCAATGA GATTAATTTA AAATGGGTAT CGAATATGAT 2520

TCAAGGACAA TGTCTCATGA AAGATATGTT TAACCAAGTG AATCGAATCG TGATAGACGT 2580

TATTTTTGAA GAATGGTTAG AACTCTTTGA AACAGTTGAT GATACAGAAG CGTCCGTTAT 2640

GGAAAACGAT TATGTGGCGA GGTGGTGAGA AAGTATGAAA TATCTCATTG ATAGTTATAC 2700

AGCCTATATG GTGAATGGTG GTTTATTAGA TAAAGGATCA GAAGTCGTCA ATTGGTTCTT 2760

TATTGAATTG CCTTTTTTCT TTTTGCGTAT GGGGGCAATG ATTTTCTTAA TTATGGAGAA 2820

TGTCATGAAT CAGTCAGATT ATTTTGTTGG TAAACAACAA GAAGCATACG ATTATTCTTT 2880

GGATATTTTA AAAGGATTTG GTGGAACTGG GATTGTCAAA GGGAGTTTGC TTGGATTGGC 2940

AATTATCCTA AGTGCCTATT ATCTATTGTA TAGCTTTTTT TCTAATCGGC GTAATTTCAT 3000

GAAAAGTTTG TTGCATTATT TTGCGGTGTT TGCCTTGTTT ATTTGTTGGT TTGGACAAGT 3060

AAAAACAATT GATGGTAAAA CGCAAAATGG GGCAATTTTT TTGATTAGTT CGGTTAGTGA 3120

GATGACAAAA CAAGTACAAG GCAAGTTTAC TTCTAATGTA AACTTTGGAG GGGATACAAG 3180

CCAAGAAGTA GACGATGGAA AGAAAAAAGT GTATCAAAGC CCAATGTTTG ACGCAACGGT 3240

TTTACAGACG TTTAACTATG TCAATAGTGG CTCGATTGAT GGCAAAATGG CAAATGGGAA 3300

GAAGCTTGAT TATGATAAAC TATTAGAAAA GCCTAACCTG AATGATAAAG AGAAAAAGGA 3360

TTTTAGAGAA AAAAGAAATG CTTATATTGA ACAGCAGGAG AAAGACAATC CTTATTTTTC 3420

GCAAGATACC TTAAAAACTA TGGAAAAAAG TTTTGCGGTA TGGACGGGTG CAACAAATTT 3480

ATTTATTCTA GCAATTCCTG TGTTATATAT TAATTTAATG TTGTCGTTGA TACAGCTGTT 3540

AGTTGTCTTT TTAATTCTTA TATTTCCAGT TGTTTTGTTG GCTTCCTTTT TTCCACGGTG 3600

TCAAATGTTA CTATTCAAGT TCTTTCAAGG TTTAATTGGT GCGTTATTCA CGCCAATTAT 3660

TTATGGTATT TTTTTATCGG TGCTGTTTTG GaTTAACAAa CTGATTGATG GGGCATTTTT 3720

AGGTGTCGCA AAAAAAGTCA GTGGCAGCTT GCTTGAATTG ATTTCAGGTA GTACGGTTTA 3780

TATGATCGTT TTATTTGTAG CaGTGGTAGT GAAGATTGTa GTTTTACGAA AAGTTTGGAA 3840

AAATAAATAT GCCATTCTTG CTTACTTTTC AAATGGGCAA GTGTCACAGC CTGTGTTTGA 3900

ACAAAAAGTG GAACAGTTGT CTGAACGAAC GAAAGAAGTA GCAACTGGTG GAGCGCAAGT 3960

TGCCCTAGGT GCTTATACAG GCAATCTGGG CATGGTCGCA AACGGTGCAG GTAATATTCt 4020

ACCGAATCAA GATAAAGCCA TTGAAATGGG CCAAGAACAT TTTATTGATG ATAATGGTTC 4080

ATTTACTGGG GTAAAAGCGG GTTTACAGTC TATTTTTAGG CAACCAGTGA CAGAAGAAGA 4140

ACAAGGGGAG CTGAACGAGG AACAGTTTTC AGAAGTGGAA ACGGAATCGA TAGATCATGA 4200

GTTAACAGAT ATTGAAGAAA TAGGACAGAC AGAGTTAGAA TCTGATAGTT TAGATAACTC 4260

TATTATTGAA TTAGAGCCTA TGACGAGTGA AGATTTAGTA GTGGATGTCG ATAATATGAG 4320

TTCTGATATA GATGTGGAGC TAGCATCAGT AGAAGAAGTT GGGGAAGAGC CACCAATGAA 4380

CGCAGAGGAA TATTCAGGTT ATGATAATGT GATTGTGACC AACTTTGATG AATTGGCTAT 4440

GGCAAGAGAA GAACGTGATT ATTTTGACGG TGGCAAAGAA CAAGAGCTAG TCGAGACAAA 4500

TGGAGTTGAA TTCGTTAATA ATATTTATCC GTTTGTTCAA ACAGAAGAGT ATCTAGCTAA 4560

TCTTGAACAG ACAGAAGCCC AGTTTTTTGG AACAGATGAG TTTGAGAAAG AATTAGAACA 4620

GATAGAGGAG TGGGCATAAC CACTTCTCTT TTTATGTTAG GTGGTGCAAA TGAAAAAAAT 4680

CTTTTTAGGA ATCCTTGGCT tGTTTAGCTT TTTGTTTTTA TTGCCTTTTT TGTTATTTTT 4740

GGGCACCATG TCAACAGAAA TTGGAAACAA TACACAATTT CAAGCAACAA CGCCACAAGA 4800

AAAAGTAGCG CTAGAGGTAT CCAATTTTGT TACCAAAAAT GGTGGATCGT TACAATTTGC 4860

CTCTGCATGG CTAGGAAATA TGGAACATGA GAGCGGATTA AATCCTGCTA GAATACAAAG 4920

TGATTTAACG TTTAATTCAG CGTGGGCGTT TAATCCTTCG ACCAATGGTT ATGCGTTAGG 4980

TTTAGCAATG ATGGACGGTG AGCGCCGTGT AAACTTGTTA AATTTTGCCA AAGAACAAAA 5040

GAAAGATTGG CAGGCTGTCC CAGTACAACT CGAGTATATG TGGAATCATG ACGGCTCAGA 5100

CAGTGCCTTG CTGAAACGTA TGTCGAAAAG CTCTGATGTG AATCAATTAG CTGTAGATAT 5160

TTTGGTACAT TGGGAACGTG CAGGCACTAA AAATGATCCC AACGAACAAA TCAAACGAAA 5220

AACAAGTGCG AATAATTGGT ATAAGAGACT GTCTACAGGT TCTATGGGGG CAGGTTCAGC 5280

CAATATTGGT GGTGGCAAAA TTGATGTGTT AGAACAAATG TTAGGGCAAA CAGTCAATGG 5340

AGGTCAGTGT TATGGGGGGA CTTCTTATTA TGTTGAAAAG ATGGGCTTTC AATCTTTAAT 5400

GAATACAGGG CATATGTTTG CCAGTGAAAT TGGCAATGAT TATGCGTGGG AACAATCTGG 5460

CTGGCAGGTA ATCAAAAATC CAAACTATTC TGATGTCAAA GCaGGTGATG TGATTAATTT 5520

CGCAATGGGC GGTTATGCGA CCAGCGTTTA CGGTCATACT GGAATTGTAG cTAGTGTTGA 5580

AGGTAATGGA AAGCTTGTTT TATATGAACA GAACGCTGAA AAAGGACAAA TAATTGCTAA 5640

ATATTTTCGA CAATGGGGAG CAGAATATCC TAACGTGACA AGTATTGTAA GGAAGAAATA 5700

GAGGAGAGAT TTTTGATGAA AAAAGAAGAA ATGGACGTTG TTTCATTGCG GATTATAAAG 5760

CTTTATTTTT TAGGCATTAG GGAATTAAAT TTCCCAGATT ATAATAAAAG GTTTCAGCAA 5820

AAAGATATGG AATTGTTTAT TCAACTAGCT GATATGATGG AAAATTTACC AAATCTTGAT 5880

GAACAACTAA TTTATGAATT AGAAGAGTTG AAAGATTATT TATTCTATGT CAAAACCGAA 5940

AAGTATTCTT TAACAGTTCA TGATATGTTT TTGGAAATGA AATCCGAGCT TGAAAAAATA 6000

ATTTAGAAGA AATGTAGTGT CAAAAAAGAG AGGGTGTAGT GTCAAAAATG TTTAGAAAGA 6060

AAAATGGTTT AACAAATCGG CAAGTATTAG CGATTTGTAC AAAATTTCAG TGTAAAAAGA 6120

ATGAGTTTGT CGCCCATCGA TTAAGTGATG GGTATTTAGT AGCAGTAAAT AATAAAGAAT 6180

ACCGAGTAAA ATTTTCGGAA GGCTTTTTTT CAAAAATTGT ATATGCAAAA GAAGTGCAAC 6240

GGAAAGGAGC TCGGAAATCG TGAATAAGCT ACCTTTACTT ATTTTATTGT TAGGCGGAGT 6300

GTTGCTTGTT AGTGGCTGTC AAAGCCATAA GGaAGAAAAC AAGTCTAGTA AAGTATCGAC 6360

AGAAGAAACG ACAGTGATTG AAACAGTAGC AAGGGAACAA TCGAAGGAAT CGTTTACGAG 6420

TGAAGCAACT AAAAAACAGA CAGAAACAAC GAAATTAGAA GAACCAGATC ATGTAAAACT 6480

TCTAGAAGCT TATGGAAATG CGTATGCGAA CTTTACAAGT ATTAATGATC GCAATGAAAA 6540

GCTAAAGCCC CTCATGACTG AAAAATGTAT CAAAAAAAAT GGAATTGATG TTAAAACTGG 6600

AGTAGCGTTA GTTTCCGTAG GAAAGGTTAC AACGATTTAT AAAAATGATC AACATGAATA 6660

TGCTTTACTT TTGGATTGTG AACAAAATGG AACGCAGACA CGAGTGTTAC TTTTGGCTAA 6720

GGTGAAGAAC AATAAAATTT CTGAAATGAC CTATAATTCA GTTAAGCAAG AGTATTAGAA 6780

ACTCTTGATA TTAATAGATT CCATGCTAGA TTTATGGTAC AATTACACAT AATAAATAAA 6840

GGAGAAAACC AATGAAAAAA AGTATTGTAA CGtGTTATTT TTGATGTTGG GGATAGTTGT 6900

ATTTACTAGT TGTGGGTCTA AAATGGATAA GACAGCAGAA GAGTTTAAGA ATAAGATTAT 6960

AAGTCATGTC GGTGATCTTT ATGATGAAAA ATATCAAGAA AAAGAATTTA GCTTTTTAAT 7020

TTACAAAGAT AAGGATACAA ATGAATATTT AGCTAAAGCT TTGGTTCCTT ATGAAGGTGA 7080

GCCTAATAGT GTAGAATCTA AGTATTTTTA TAATGTAAAT AAAGAATTAG AAACTATTGA 7140

ACCATTTAGT GGAGGTAGAA CGTTTGATTA TGCTAAGTCT CATGGTAACT ATGAGGTAAT 7200

ATATAAATCA GGAAAATTTA AATAGAATTT GAACTTAAGA TAAATTAAAT TTATGTTTGA 7260

GAGCAGTGCT TAGAGCATTG CTCTTTTTTG TTTTAGGAGG GAAAACAATG TTAATTGTAA 7320

CAAAAGAAGA CCTAATGAAT TTAGGATATG GAAAATATCA AGCGGAAGAT ATAATTAGGA 7380

AAGCAAAGGC AGTAATGGTG TCTAAGGGAT ACGCATATTA CCTAAATAAA CGATTAGGAA 7440

GAGTTCCTGT TGGAGCTGTC GAAAGTGTAT TAGGAATAGA ACTAGAATCT ATAGCAAAAG 7500

AGGTTGAAAC GGTTGGTTAA GTTTAAAAAA GATAAGAGAG GAAAGTATTA TTTTACTGCT 7560

AGTTTGGGTT TTGATGAAGT AACTGGTAAG AGGGTTCAAA AAATGCGTTC AGGGTTTTCC 7620

ACGATTAAAG AAGCAAGGGA AGCTTATGCG GAAATAATTA ATAATTTCGG TAAAGAAGCA 7680

TATTCTTCAA ATAGCACAAT GCTGTATGAA GAATTTTTTT ATACTATTTT TTTGCCATAT 7740

TATAAGGGAC GGGTTAAAGA AAGAACTTAT AATAATAGAC TTAGTATGAT AAAAAATACT 7800

TTTGCCTATT TTTTTAAAAT GAAATTGAAA AGTATAGCTC CTGTTCATAT ACAGAAGTGG 7860

CCGAATGAGC TGTTGGAACn GTATGAAAAC ACGTATGTAA GAAATATTTA TGGACTGTTT 7920

CnGAT 7925

(2) INFORMATION FOR SEQ ID NO: 312:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2594 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

TAAAATAwTw TTTATAATAG AaAGTTATCG AAAGGATACC TATGTTTAAA TTATTAAAGA 60

CTTTTCAAAT TGTGTATGAG CAAATGAACT TTTCCAAAGC AGCTACCTGC CTATATATAT 120

CACAACCAGC CGTTTCTAAT CAAATCAAAC AATTAGAAGA AGAATTAGGC TGTTCATTGT 180

TTTTAAGAAA TGGTCGACAA GATGTAGTCC CTACTAGGCA AGCGGAAGTC TTGTATAATC 240

GTTTGCTAAA TTTAGCTGAT GATTGGCAAG AAACATTAAC AGCTCTTCAT CAGGCGAGAT 300

TACCTAAAGA AACTTGCCGA ATTGCTGCGT CTAATACTTT TGCAGTTTAT TATTTGCCGC 360

AACTGATGCA GCATTTACAG ACACAGTATG CAACGATCAA TTTTGTTTTA GAAATGAACA 420

ATTCAGAAGA AGTAGTGGAA AAAGTTGAAA AACATCAAGT TGATTTTGGC TTTATCGAAA 480

AACCATTAAT TACGAAAGGA GCAACTCGTG AAGAAATTAT TCAGGATCAA TTGGTTTTAG 540

CTGGGGACCC TgCCAATcAG AATTGGTTGG TACGTGAAAA GGATTCTGGG GTCTTTCATT 600

ATACCCAACA GTATTTAGAA GAAAGTAATC AAAGCCCAAC ACTCATGACG GTTAAAAACA 660

ATGAAATGAT TGTTAAAATG TTGGAATTGG GCATGGGGCA ATCGTTGCTT TCAAGAAAGG 720

CGATTACTGA AAAAATTCCT TTCCAAACGT TAGGTGAAAA GTATTGGCGT TCCTTTTATT 780

TTTTGACACG GGGACATTTA AAATCCTCCT TGCTTCAAGA AGTAAAACAA GCAATTTATC 840

GATTCTATCA GACGGAAATG AATAAATACT AAAACGTTCT CTTTGTTTAT TTATCAATCT 900

TTGACTACAC TAAAGAAAAC GAAGGAGGGA ACCACAATGC CTAAATTTTC ACCATGTTTA 960

TGGTTTAACA CACAAGCGGA AGAAGCTGCT AACTTCTATA CAACAATCTT TGAAAAAGGT 1020

GCGATTCTAA GTAAAACGAA CTACGTTAAT GAGGAGCATC AACCTCaAGG aACATCATTA 1080

ATGACCATTG AATTTACCTT GGCGAACCAA ACAATCATTG GTTTAAATGG CGGACCAGAA 1140

TTTTCCTTTA CGCCAGCTAG TTCATTTTTT GTTGAATGTA CAACATTACC ACAAACAGAA 1200

ACGTTGTGGA AAAATTTAAC AGCAGATGGA CAGATATTGA TGCCTTTTGG GGAATATCCT 1260

TTTAGTCCTT TATATGGTTG GGTAGTGGAC AAATTTGGTG TCTCTTGGCA AGTGTCTTTT 1320

TCAGGAAAAG AACAAACAAT TGTACCAACC TTTATGTTTG CTAATGAAAA ATATGGCGAA 1380

gcAGCCAAAG CTCTATCAGA ATGGTTGGCG ATTTTCGGTC CAGGCGAAAT AATTGAAAAA 1440

GTTGAATATG AAGATGGGAA CATTGCGCAA CACTTTTCAC ATTACAGGAG CAACCGTTTC 1500

GAGTAATGGA TGCACGAGAT AAACATGATT TTGATTTTAC CATGGCCTTC TCCATTTATA 1560

TTGATTGTGA AAATCAGGAG GAAATAGATC GGCTATGGCA ACAAGTGACT GCTAAAGGGA 1620

AAGAATGGCC GTGTGGCTGG ATGGAAGATC AATTTGGAGT CAGTTGGCAA ACAGGCAATC 1680

CTGAACTGAA ACGTTATTTA TCTGACTCAA ATCCAGCTAG AGCAAATGAA GTGACCAAAA 1740

AATTGTACCA AATGAAAAAA ATAGATTTAA ACCAGTTAAA AGAAGTGTAT GACAAATATA 1800

ATCATTAGAA AAACTGAGGT TGTCTGATCC ACTACAcTTC GTTACGTTGG TATCAATTCC 1860

AATACATGGA ACATACTAGA GTTGAAGGAC TTGTCTTTTA AGGATCAAGA CCTTCAACTC 1920

TTAAAAAAGG CTCGATTCTT GCGAGTTTCA AAGAAGCCAC TTCTATTCTG TTGTTTATCT 1980

TAAGGGTTAG GACTCTGCCG ATTTTTTTAT AAGATTATGC TATAATTAAG AGTAGCAATT 2040

ATTGAAGGAA AAATGAAAAG TCCAAAATAG GAAGTssTAA AATGATTGAG TTAATTGTGT 2100

TGTTAACAGT TTTTACTTAT GGAACAAACT TTGTTTTATA TCTTATTTTG AAAAATAAAA 2160

AGTCCATGCC AATTGTCGAA AAAGCTTCTA TTGTATTAGG AATTAATATG TCTGTGTTAT 2220

TATTAGATGG TATTTTTGCT TTCGTTGGTA AATTCATCAT GTTATCAGAA TGGGAATTAT 2280

TAGCGCCTAT TTTTGAAATA GTGTAAAAAT CATCATAAAG TCCTATATCC AAAAAGAATG 2340

AAGTTTGTTA GGATAAGAAA GATGAAAATT TAGAAATAAA AGAGGTGGGA TTGCATGGCT 2400

AAAATCATGA TTGTCGAAGA TGAAACAACC ATTCGAGAAT TAATTAGCGA GGAATTACAA 2460

AAATGGCAAT TTGAAACGAT TGGAACAACC GATTTTAATG ATGTATTGGA TGATTTnCAA 2520

GAGGAGAATC CTCAATTAGT CTTGATGGAT ATTAATTTAC CTGTTTACGA TGGGTACTAT 2580

nGGTGTCAGA AAAT 2594

(2) INFORMATION FOR SEQ ID NO: 313:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4234 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

TGTGCTAAGC GTTTnTTAAC TTTTTTTAnG TAGTTTGGCA GGTTCATTAG GAGnTGGCGC 60

TCTTTTATAT AGTGTCATAA TAGGTGTTnT AGGATTTGCG GACTATCAAA AGATGTTTTA 120

TCGTGTCGAA CCTATTTATC CAGACGACTT AAAAATGATC ACAGAAGTTA GTTTATTAAA 180

AGAAATGGTT GGTTTATGGC CGTTTATTTT TGTGGTAGCG TTAGGTTGTG TAGCACTTTT 240

TTTCTTAGGA AAAGCCTTTT ATAAAAGTTT CTTTTTATCG AAGAAAAAAC AAACGATTCG 300

TGTGTTAAGT CTTGTTCTTT CGATTGGTCT ATTTTCATAT ATTGGTGATT TTAACAATCC 360

GCAAAATTTG TTACGTAAAG CCTATAATAA AACCGCGTTA TGGATTCCTT ACAGCCAAAA 420

AATGAATTAC TACAATACAG GATTyATTGG TGGTTTyCTA TATAATTTAA AAGTAGAGGC 480

TATGGACCAA CCAAAGGGCT ATTCAAAAGA AGCTATTTCA GAAATTACTC AGAAGTATCA 540

AAAGTTGGCC GAGGCGAAAA CACTGCCGAG CAATGAGCAA CCAAATATTG TGTTTGTCAT 600

GAGTGAAAGT TTTTCAAATC CAGATCATCT GAAAGGCTTG TCCATTATTG GTAATCCTTT 660

AGCTGATTAT CAGCAAATCG CTGATAAAAC TTACAGTGGC CAGATGCTTT CACAAAACTA 720

TGGGGGCGGA ACTGCGAATA TTGAATTTGA AGCCTTAACA AGTTTTcGAT GGAATTAATG 780

AATGCTCAAA TGACTACACC TTATACGATG ATGATTCCTA AGATGACgGA ATTGCCTTCC 840

ATTGTTTCAT TACTTAAGGA ACAAAAGTAC CAAACAACGG CTATTCATCC GTACAATACA 900

TCTATGTATA AACGAAAAGA TGTGTATAAA ATTTTAGGTT TTGATCAATT TATCAATGAA 960

AAAACAATGA ACTATACGGA TAAACTAGAA AATAATCCTT ACATTTCAGA TCAAGCAGCG 1020

TATCAGGAAG TGACGACACT GCTTAAAGAA AAAGAGCAAC CGCAATTCAT TCATTTAGTT 1080

ACTATGCAAA CACATATGCC CTATGAAAAC AAATACACAA CGTCCAATTA TTTTTCTCAA 1140

GGAAACGGCA ATCAAGTCGC GCTCAATAAT TACTTGCAAG ATATTGCCTA CAGTAGCGAA 1200 yTTTGcaAAA ATTTCCTACA AGAAATCGAA AAAATGCCCA AACGCACATT AGTCGTTTTT 1260

TGGGGAGATC ACTTACCAAG TATTTATGAT GAAAACATTC AAGCGTTGAA TGAAGGGGCA 1320

GCCCTTCACC AGACAGAnTT TCTAATGTAT GATACGGCGA ACCAGTTAAC AGrAAAAAAT 1380

CAGCATCAAG CAATCATTAG TCCATTTTAT TTnGCTCCCT CACTTTTCCA mCAAAGTGGC 1440

TTACCGCAAT CGGGATTTTA CGCTATGCTA AATGAGGTAC AAGAACAATT GCCTGCTTTT 1500

GAAAAGGGGA ATTACTATCT CGGTGGGGAA TGGAAAAAAA CAGTAGAAAT GAATAAAAAG 1560

CAAGAACAAC TATATGAGGA ATACCGACTG ATTCAATATG ACATTGTTTC AGGGAAGCAA 1620

TATAGTTTAG AAAATCAATT TTTTAGCTAA AAAGrGCCTG GGACaAAAAT CACTTTGGAT 1680

TTTTGTCCCA GGCTCAAAAC CTGATAAACG GCGGGAACAG AACCAACTCC TTCGGAAATA 1740

AGCCGAAATT CTCCAAAAAT TAAAGAGCAA TTTTCGGAAA TTTCTTCTTA TTTCTCGGAG 1800

CTAAACGGTT CTGTCCCsAC CTCTTGAGTT TTATTTTAAA GTTTTACGAT AAACAATGAC 1860

ATAATAGCTG AAAGCTAACA GACAAGCAAA ACTTAAAAAG TAAAGCGGCA AGACCCAAAT 1920

GGTACCATTG TGAAAGTAAC TAATCAAATA GGTAAAGAAA CTAACAAACA AGTAATAGAC 1980

AAAACTAAAA ATCCCACTGG CAGTACCAAT TACAGATTCA TAGCCTTTTA AGGCCATACT 2040

TAAGGCATTG GGCAAAGTAA TATTAATCCC TAAGAAAAAG AGAAAAATAA ACAAAATAAG 2100

ACCAATGATT GTTTCAAATT GTGTCATTAT TAATAAACAG CAAGCCGCAA AGATACTTAG 2160

CAGCAAACCA GCAAAGGCAA TTTTTTCCGC TGACCAATAT CCTAGCAACC AATTGATAAG 2220

AACCGCAGCA AGAATACTAG ACACAGCTAA AACTAAACCA AGAAAGCCAT ATTGGACACT 2280

GCTAAAGCCA AAGTGATTGA TGAAAATAAA CGGAGCTTCT GCATAGTAGC TAAACAATAA 2340

ACCATTAATG CCCCCAATCA ATAAGCCGTA GCACCAAACT TTCGAGTCGA AGGAAAGGCG 2400

TTTCGCCACG GACCAAAAAG GCGCTTTCGT TAAGGTATTT CCTAAGTGAA AGGTTTCTGG 2460

TAAACGATAA AAACTATAGA AAAGTAATAG AACAGCCATG ACAATTAAGG TGCTAAAGAC 2520

ATGCTGATAG CCAAAATAGG TTTGAACAAG TCCGCCGACT AAGGGTCCGA TGGCAGGAGA 2580

TAGCGATAGG GCAGCGCTTG TTTGAGCAAA AATTTTCCCT CTTTGAAGAT CGTTAAAGCT 2640

TTCCCGCATC ATCGATTGGG TAATGACAGA ACCAACGCTT GCACCAAATG CTTGAATTAA 2700

ACGGAAGAAT AACAGCCAAG GAAATGTCGG ACTCAAAAAA AGGCCAATAT TTCCAATTAA 2760

ATAGCATAAA ATTCCTAATA ACATAGCTGG GCGACGTCCA ATCGTATCGG ATAGAATTCC 2820

CCAGAATAAA ACGCCAACGG CAAAGGCAAC AAAATAAATA CTCATGGTTA ATTGTGCCAC 2880

GCTATCGTGA ATGTTAAAAG CAGAACTAAG CGCTGGCAAA ATGGGTGTGA AAATCGACTC 2940

GCTGATTTGT GGAAATCCCA CAAGCAAAAT AATAATTAGT AATGATGGGA CGGTGTTTCG 3000

GGTGTTTTTC ATGATTTTTC TCCTTAAAAG TTTAATCTAA TCAAAAACAC CCATTGTCTT 3060

TTACAACCGC CAAATTTAAA ATAAGTAAAA TAATGTCACC AACTTTCCTT GAAATTTTTT 3120

ATTATACAAA GTCTAAATGA AGCAGCGCTT CTTCTTGACT AAATTCATAA GTGTAGCGAC 3180

CGGATAAACC GGCAAGCTCA CCTGATGCTT GAAATAGTTT ACCTGGGCAG TCTAAAAGAC 3240

CCTGTGTAAa CGTCCCTTTT TCGCAAGCAG TAAAGGTGCC TTTTTTTCCT AAATAAGTTC 3300

CTTCGAAATG GAGGAAACCG ACCACTTGAG CCACAGCTTC AAGCCCATCT GCCATTGAAG 3360

AGCCTTGATA AGTCATTAAG TATTCCACAG AAGcTGTCCC AGATAAGGCG CCTGACAATT 3420

GATAAGTAGC GCTCGCTTGA TTTACTGGGA AATTTTCATT AAGATTGTCA ATCGGTTGTT 3480

CTTGCCAACT AGATACAGTA AATGATGCTT GCTCCATTGC TGTCCCTTCT TTCTTATTAA 3540

TAAGAATAGG GTAGCGGGAA GCGCACTATT TGTCAACAAT CCACCACTTT AACTCTTCAT 3600

TTTTTGCACA AAAGAATGTA ACGCTAAGCG GTTGATGTGA TAAGCAGTTC CTTGTTTTTG 3660

AATAAAGCCG CGCTCTCGAA ATTGTTTGAA GGTATAAATA AGATGGCGAT AACTAACCCC 3720

TAAGTATTCG GCAATCTCAG TATGTTTTTC TTTATAAACA TCTTCGTGAT TAGCGGCAAC 3780

CAATAGCAAT TCCGCTAGCC GATAACTTAG TTCATAGGTC TGATTCGTTG TAAAGTGGTC 3840

CACCCGTTGA AGCATTTTTT TCCCTAGATA ATGGCTGAGT TGCTGCATAA AAAGCACCTT 3900

GTTTTTTAAA ACAGTTTCTG CGTAAGCTAA AGGTACAGCC AAACAACAGG TTTCACCTAA 3960

GGCGATAACG TCCTTGGTTA TCGTTTCCGC TTGAACTAAT GTTAAATCGC CAATTAAATC 4020

ACCTGCCGTT AAAAACTGAA CAATGGATTG GCGGCCATTG CGTTCTGTTT TTAAAATTTT 4080

CGCTTTTCCT TTGATTAAAT AAAAAAGATA ATTTAATGGC TCTGCTTGGT GATGAATGAA 4140

TTCTTGATTT TCAAAATCGA TAATAAAACT GTGTTTTAAA TCAGTTTCAT TAAAAAAAGG 4200

CGTATTTTTA GCTAATTCTT GTTTCGGTGA TCGT 4234

(2) INFORMATION FOR SEQ ID NO: 314:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2642 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

TAAGGTTGCA AAGCCAATTT TTTTATTCTT ATCCGTAAAA TATTTTTTGC GTCGATGGTT 60

CACTGAATGA ACTTTGTCTT CAAGGGTTCG TAAAGATAAG CTAATCTTTT GAGAATATTC 120

ATCAACATCA ATAACTTGAA CGGTTACTTT TTGTCCAACG GTTAAGAGTC CATGGATATT 180

TTTGGTAAAG CCAGCTTGTA CTTCCGAAAC ATGAATCAGT CCTTGTGTTT CTTCGTCTAA 240

AGAAACAAAT GCGCCATATG GTTGAAGTCC GGTAATTGTG CCAGAAAGTA CCATACCAAT 300

TTTGTACTTC ATCAAATTTC CTCTTTTCAT ACGATAAATT TCTTTTTCCG TTTCTACTCA 360

TTTAAAATAA TAGCATTTTT TCAACGTTTG TGCATTATTT CTTCCCTAGT TGTTAAAACT 420

ATTGTTTCAT CTTTTAAGCA AAATCTACTA TACTATAGGA AGTAAGACTT TATATAGAAG 480

AAAAGGAGTT ATCTTATGTT TAATTTTAAT GAAACCATTG AGCGTCGCCA CACGAATTGT 540

GTTAAATGGG ACACCGTTGA AGCGAGTTAC CACGAAAAAG ATTTACTACC TTTATGGGTT 600

GCTGATATGG ATTTCAAAGT ACATCAGCCT ATTCTTGATG CCTTGTCACA AGTTATTGAA 660

CAGGGCATTC TAGGATATGC AGTTGCACCT AATGAACTGT ATCAAGCAAT TCAAGATTGG 720

CAACGGCAAC ACCATCAATT GATTGTTGAA AAAGAAGAAA TTTTGTTTAA TAGTGGCGTG 780

GTACCCAGTT TGGCCACAAC CGTTCAAGCG TATACAGCAC CTGCTGATTC TGTCATGATT 840

TGTGATCCAG TCTACCCCCC TTTTGCAGAT GTGGTAAAAC AAAATGAGCG TAGACTCGTG 900

CGCCATTCTT TACTAGAAGT TAACGGACAT TATGAAGTTG ATTTGGTTAA AATGGAACAA 960

CAGATAATCG AAGAAAAGGT GAAACTGTTA CTTTTTTGTA ATCCCCATAA TCCTGGTGGA 1020

CGTGTTTGGA CAAAAGAAGA ATTGCTGGCA ATTGGCCGTT TATGCCAAAA ACACCAAGTA 1080

ACTGTGGTTA GCGATGAAAT TCATCAGGAT TTGATTTTCA AACCGCATAC CTTCACTTCC 1140

TTCACTGTCG CTGATGATGC ATTTAAGGAA TTCACTGTTA CCTTAACCGC AGCAACGAAA 1200

ACCTTTAATT TAGCGGGGAT TAAAAATTCA ATGCTGTTCA TTCCTAATGA AAAACTACGG 1260

CAATCTTTCG TATCTTTACA AGACnAAAAT CATCAAGGCG GCATCaATAC TTTTGGCTAC 1320

GTTGGAACGG CAGCTGCCTA TCAAACAGGG GAGGAATGGC TAACGGCATT ACTTGATTAT 1380

TTAAAAGAAA ATATTGATTT TGCCCTTTCT TTTTTCCGTG AAGAGCTGCC TAGCGTTCGT 1440

GTAATGGAGC CTGAAGGGAC CTATCTTTTA TGGTTAGACT TTAGTTCTTA TTCACTAACG 1500

GATCGAGAAC TTCGCGATAC CTTAATTCAT AAGGGGAAAG TGGTTCTAAA TCCAGGAATT 1560

AGTTTTGGTC CCCAAGGTTC TCAACACATG CGTTTAAACT TAGCTTGTTC AAAAGAAACT 1620

CTCGAAGAAG GCCTTTTACG TATCAAAAAA GCCTTTAATT AAAAACAGCT GTCCTAGTTT 1680

TTAGGACAGC TGTTTTTAAT TCTTTCTTTT GCGTTTAGCT AAATAAACTC TGACGAACCA 1740

GATACCTGAA AATAGAaGTA AAATTTCTCC AATTATGATT GTATAAAATC CTGGTTCATG 1800

AACGATTCCT TGCTGATCTA CATACGTAAA AAACGTTCCT ATAAATGGGA GTATTACTGC 1860

AAACAAAGCC AAGGCCATTG GATATCTGTA TTTTTCTAAC AAATTTGCTT CCTCTCCTTC 1920

TATTCTTTAA CTCAGTATCC TATAAAGACT TAATTAAAAG AAGAGATTTG GATGCTTTTA 1980

AAAAAAAGTT AGTCGTTTTA AGAAAwAAAT AAAGAGTACC AACGAATTTG TCCCTCGTTG 2040

GTACTCTCTT TTTCAATTGG TTAAACTTAT TCACCAATGA TTTCAACTGT GTCCATCACG 2100

ACGTCAAATG TTGGACGGTC TTGGGCATCC CGTTGGACTC CACCAATTTC ATCCACTACA 2160

TCCAtACCaT CAACAACATG ACCAAAAACA GTATGACGGA AGTCTAACCA AGGTGTTCCG 2220

CCTTGtTTGT AAGCTTCAAT AATTTCTTCT GGAAAACCAG CATCTTCTAA CTGACTCATC 2280

ATTTGAGCTG GAACATTTTG ATTAGTAACA ACAAAGAATT GGCTACCATT TGTGTTAGGA 2340

CCAGCATTTG CCATTGACAA TGCGCCACGC AAATTGAAAA CATCCCGAGA AAATTCATCC 2400

TCAAATGCTT CTCCATAAAT ACTTTCGCCG CCCATTCCTG TTCCGGTTGG GTCACCGCCT 2460

TGAATCATAA AATCAGGAAT CACACGATGG AAAATAACGC CATCATAGTA GCCCTTTTTC 2520

GCTAATTCAA CAAAATTTTT AACCGTTTTA GGCGCTTGTT CTGGGAATAA TTGAACAGTG 2580

ATGTCGCCTc GGTTTGTTTT AATGACCGCT TTTGGGCCTT TAGCATTTTC TAAATCTAAT 2640

TG 2642

(2) INFORMATION FOR SEQ ID NO: 315:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4605 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

CAGCTTAATT TTCCTTCATT TACATAAAAT TGAAACATAG TGTGACAAGG CGGCAACGCC 60

ATTGAAGGCA CATCTTCTGG ATTCCAAGCG GAAACAATTA AACGGCGAGA ATCAGGATTG 120

GTTTTAATCA TTTCAATCAC ATTAGCTAAT TGATCGATAA AGCTACCATC TTTGGTTTCC 180

CAGTGACGCC ATTGTGCACC ATAAATATTC CCTAATTCAC CATATTTTTC TGCAAACTCT 240

GCATCATTTA AAATGGCATC ACAGAATTTT TGATGTTCTT CTTTGTATTG TTCTGCAAAC 300

GCTGGATCTT GTAAGACACG ATGCCCAAAA TCCGTCATGT CTGGTCCTTG ATAATCTGCG 360

CTTTTTACGT AACGTTCGAA CGCCCATTCA TCCCAAATAT GATTATTTCG CTCTAACAAA 420

TAGCGAATAT TCGTGTCGCC TTTTAAAAAc CACAATAACT CACTTTTAAT CaAGcCAAAC 480

GGTACTCGTT TGGTTGtTAG TAAGGGAAAT CCTTTGGCCA AATCAAAACG CATTTGATAG 540

CCAAATAAAC TATACGTCCC aGTTCCTGTA CGGTCTTCTT TAAAATGACC TTCTTCTAAA 600

ATCTTTTTTC CTAGCGCTAA GTATGCTTCT TCCATCTTAT TCCCTCCGTC TCTTAATCCG 660

CGTATTCACT CAAGTATTCC CAACGATTCA TTTTTTCTTC TAACAATTGT TCCGTTGCTG 720

TTACTGCTTC TTGCAATTCT TGTAAACGAA CAAAATCATC ACCCTGatGA TTCATTTCTT 780

CAGTTAATGT TTCTAATTTC GTTtCTAATT GTGTGATGTC TTCTTCAATT GTtGCCCACT 840

CTTTTTGTTC CATATAAGTC AGCTTTGTTT TttCTTtCTT TtCTGTCGAt GCTCTTTTAC 900

tGGTTtAtGA gsGGTTyTTc TTTGACtTTT CTTGTTCTTT TtGGTTCGCC AAATATTCAC 960

TCATCGTGCC ATAAAATGAT AGGAkCTGAC CATTTCCTTG AAAAACTAAC AATTTAkCCA 1020

TTGTTTTATC CAAGAAATAA CGGTCATGAG AAACCGCAAT TACCGCTCCT TTAAACGTTT 1080

GGATATAATC TTCTAAAATG GTCAACGTAT CAATGTCCAA ATCATTGGTT GGCTCATCCA 1140

ATAATAACAC GTTAGGCTGA CTAATTAATA ATTTTAAGAG ATATAACCGG CGTTTTTCTC 1200

CACCAGACAG TTTGCCAATC AGTGTGCCAT GCATAAAACG AGGAAATAAA AAGCGCTCTA 1260

ACATTTCAGC GACCCCAACG CTGCTTCCAT CACTGCGTTT CACTTCTTCT GCTGCTTCTT 1320

GCAAATAAGC AATCATCCGT TGATTAGGAT CCATTTCTTC GTTTTGCTGG GTATAATAGC 1380

CAATTCGGAC GGtTTCACCG ATACTATATA GACCACTCTC TAAAGGAATG CGTCCAGCTA 1440

AAATATTTAA TAATGTGGAT TTGCCTGCAC CGTTTTTACC AGTAATGCCA ATTCGATCGT 1500

TTGCTTGAAT GAGTAAATTA AAGTCTTTCA AAAGCTGTTT ATGATCAATG GCATAATTTC 1560

CTTCCtTGAT TTCCAGAACT TTTTTCCCTA AACGCTGCGT GGCAAAGTCT GCTTCTAAGG 1620

TGCCATTGGT TTGTACCTGA TGCAAGTtTT CTTTTAAATC CTGAAACCGA TCAATTCTTG 1680

CTTGTTGTTT CGTtCCTCGT GCTTGAACGC cTGCTCGCAT CCaCGCAAGT tCTTGTTTGA 1740

AGAGGCGTTT TCTTTTTTCT TCTTGTTCTA CAGCGACACG TTCACGTTCG GCTTTTTCCA 1800

TCACATACGT TTCATAGTTT CCTTTGTATT CATACAGTTT TCCAAAAGAA AGTTCGAAAA 1860

TCCGATTGGT TACTCGATCT AAAAAGTAGC GGTCATGGGT CACCATTAAA ACCGCGCCAC 1920

GATAGTTATT TAAAAAGTTT TCCAGCCATT CAATCGTTTC ATAATCCAGA TGGkTCGTTG 1980

GTTCATCCAA TAACAGTAAA TCAGGCGCTT CAATTAAGAC TTGCGCCAAG CTTACTCGTT 2040

TTAACTGCCC ACCAGAAAGC TCACCTATTT TCTTTTCCAA TGTTGGAATG CCTAATTTTT 2100

GTAAAATAAT TTTTGCATTG GTATCTGCTG TCCAAGCATC TTGTTTATTC ATGGCTTCTT 2160

CTGCTTGCGT GTAACGTCTT TGAACCCCTT CGTCTAAACC ATTCTCACTT AAAGCAGTAA 2220

CGCTAATTCA TAATTTTTGA CCGCTTTTAT GATCGGGCTG TCTCCCTGAA ACACGGCTTC 2280

TGCAATCGTA ATATTCTCGT CAAATGTTTG ATTCTGGGAA AGGTACCCAA TTTGATAATC 2340

ACTTGGCTGC TGAATCGCAG CAAGATCACC ATCACCGCTA TCTACGCCAG CAATAATATT 2400

TAATAAACTG GTTTTCCCTG TGCCGTTAAT CCCAATTAAA CCAATGCGAT CTTTTTCATG 2460

AATCAAAAAA GAAAGATGAT CAAATAGTGT TTTTTCCCCA TATGTTTTGG TTAAATCTGT 2520

TACTTTTAAT TCTTTCACCG ACACTACCAA CTCCTTCTAT TTTGGTCTTG CTTATTGTAT 2580

CAAAAATTGG ATTATAAAAA AAGTGCATGA GATAAAAATC GTCTTGGTTT CATCTCATAC 2640

ACTATTACTT ATTCTTTATT AAAATTTTTA ACAAATCGTT TCAGCGGAAC ATTATTCACT 2700

TCGCCGCTAA TCGAATAACT CTCTAATAAA TTGACATTTT TTCGTAACTC GTTCGCTTCG 2760

ATAATTGAAA TGCGTTCAGA TTGTTCGTAA TGATCAATCA TCCGTCGTTC TTCTTGGAAG 2820

CCTAATAATA GGTCTTTGAC ATAGTCTGTC TCCGTATGAA CAATATACGA ATCAATAAAT 2880

TCGCCTTTTT CAAACTTATG GGCTAAATCC AAGCGTCCGT CAATGAAGAA ATCAATTAAA 2940

TCATCATCAT AGACACCGCG CAAATTATCT AAACTCTTCA GAATAACTTG AGAATTTCTT 3000

AGAAAAACTT GGCGAATATT TTCTAGTTCT TCATTTCGTA ATTCTTTTTC TTCTTCAGAA 3060

GCTTCACCAC GATTGAAAAA TAGCTTAGGA TGTAAAAGAA TCGAAATGAT TTGTCGAGCA 3120

AATAATAACC AAAAGCCTAC AATCGATAAT AATTGTTTGG CAATCGACCG TTGCATCCGT 3180

GTAATCAAGC GTTGATACAT ACGATAACTT CTGACACTAA TTTCTCCTTG TTTCAGACTT 3240

TCTTCTAAGC CGTCTCGTTC AATGGACACA ATTAACGTCC GCAACTCTTG CACTTCTTGT 3300

TTTCGTTCTG CaCTTAGCTG cTCGGTATAC AATTCTTTTA AACGATCTTG ATAATTTTCA 3360

ATGACCGCAT TCATTTCACC ACGAGGATTC TCTTCATTTT GCTGTTyTAA GACATCAATG 3420

ACATCACGTA ATAAATCAAT TTCGACTGAC TCTTCATTTT CCACTTCCGT CTCTGAATCA 3480

GTTAAAAATG GTAAAACTAG AATCCCGCCA ATCAGAGTCA CCAAAATAAC ACAAGCCGTA 3540

ATAAATAGTA ATAACGGTCG TTCTTTAAAA TGCTCGCCGT TAATCGCTGT TGGCAAAATA 3600

AAGATGGTGG CTAAACTCAC TGTTCCTTTA ACACCGCCAA AAGTTAAAAT CaGCATTTCA 3660

TTCaTTGCGT TTTTAACTTT CTTAAGTTCa TGTTGAAGCG TGTAAACACT GAAAATGGCT 3720

AAGAAACGAA TCAAAAAGAC AACGGCGTTA ATAATTAGAA CAACCAGTAG TAGGAACCAG 3780

TTTGAATATT GTTGGTCAAT CCAAATTGGC GAAAACACTT GAGrAAGTTC AATTCCTAAG 3840

AATAAGAAAa CTAAGGCATT TAAAGTGAaC TCAATCAAaC TCCAAATACT TTCGCCAACA 3900

TTTGACAGCT CAGCTTCAAA AAGCGAAACT TTTTTCAAAC TAGACGCTTG CATAATTCCA 3960

GCAATCACAG CAGCAATAAT TCCAGAAACG CCAATCAACT CTGAAATTAA ATAAGCAATA 4020

AATGGCAAGA GTAATTCAAT CAACAAGTAG CCTGTAACAT CTTTTGCCAC TGCAGTTTCA 4080

AGTAAATGAA CAATTTGCTG TTTTAATAGA ACTAACACGA TTCCGACTAC CGCTCCACCA 4140

ATGCTGGAAA TAACCAAGGT CCCTGCCGCT TCCACCGGCG AAAATGACCC CGTTAATAAA 4200

GCCGCTAATG CAAATTGAAA GGCTGTAACG CCTGATGCAT CATTAATGAG TCCTTCACCA 4260

GATAAAATGT TCATCGAACG TTCAGGAATG TCAATTCGAC CTGATAACGA ACCTACCGCC 4320

ACCGCATCCG TTGGCCCTAA AGCCGCTCCA AAAGCGAAGC AAGCAGCTAA TGGTAAAGTA 4380

GGCAAAATCA TATGCAGGGT CCAGCCGACT CCAACTAAGG TAATTAACAC ACCAAAAAAG 4440

GCCAAGAAAA GAATCATCCC GAAATTTTTC ATCATTGCTT TCAGATTGAT TAGCTCCCCT 4500

TCtCTAAATA AAAGCGGTGC AATAATCATT ACTAAAAACA CTTCAGGCTC AAATTCAATT 4560

TCCTGCCCCA TTTTAGTTaA ACCAATTAGA ACGCCTACAA TAATT 4605

(2) INFORMATION FOR SEQ ID NO: 316:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2726 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

TnTACTTCTA CTGTCACnTG CGGTAnTAnT TTCCTTTGTG GATTGTATTT GTTTATGCAC 60

CTATTTGCTT TTCGTTATGG GGACATGGAA ATTGGCTAGG GAAAATGGGA GTACTGGATT 120

ATTCAGGCGG ACTAGTAGTC CATACTACCG CTGGAATAGG AAGTCTTGTT TTAGCAATAA 180

CATCACTAAT ACGTTTAAAG AATTCATTGA TATTTAAATC ACAAGAAATG ATAGCTTTTG 240

TAGGAATGCT GTTTATTACC TTAGGATGGT TCGGATTTAA TATGGCACCT TCTGGAAAAA 300

TCGGTGAAGA AGCCATTCAA GTTTGGTTAA ATACACTTAT TTCAATTTTA GGCGGTAGCA 360

TTAGCTGGAC TTTTACCTAA TGGATTTTAA TAAAAAAAGT ATCTATCTAT TCTATAATGA 420

ATGGAATAAT CGGCGGATTA GTAGGTAGTA CTTGCTCGGT TGGATATGTA wCTCCTGCAT 480

```
TTAGTTTATT AATTnCGGTA AtTGkATGtA CCCTTTGtCC AATTGTwATt CACATAATGC     540
ATCTAAGCAT AGCAAATTTT GATGACGCTG CAGATTCTTT TGGGATGAAT GCAGTGGGTG     600
GAATTGCAGG ATCGATTTTA ACTGGTGTCA TGGCAGAAAA AGGGGACTTT TTTCTTCAAT     660
TATTTGGGAC TTTTTTAATC AGTATTTGGA GCCTTAGTTT AAGTCTTTTG ATTCATTATT     720
TATTGAAAAA AATAGTAAAT GATTGTTATT CTCAATGTAT TAAATAAAAA AATTAAAATT     780
ATTCCTAGTG TGTATGAATA TCAAAATAAT AGTCAATAAT AACACTCAGG AAGAAATTTG     840
AAGACTTTAA ATTTTAGTGG AGGTAAATAA TGGAAAATAA AGAAGTTATA ATTATTGGAG     900
CAGGCGCTGC TGGCGTAGGT ATGGGGGTTG CTTTAAAAGA TTTTGGCATA AATAATTTTG     960
CTATTTTAGA GAGAAAGCaG GTAGGTAATA GTTTTATTAA ATGGCCTGAG CAAACAAGAT    1020
TTATCACTCC TTCTTTTACA AGTAATGGAT TTGGGATGCC TGATTTAAAT GCAATAGCAA    1080
TTGATACATC TCCTTCATAT ACTTTAGGTG AGGAACGTTT ATCTGGAAAA GATTATGCTA    1140
AATACCTACA ATTAGTTTCA GAAGAATATA AGTTACCTAT TAAGACAAAT TGTAAAGTTC    1200
AATCTATTAA AAAAGAAAAA ACTGGTTATT TGCTTGAAAC AACAAAAGGA TTTATTTATG    1260
CAGAATATGT AATTTTTGCG ATGGGAGAGT TCTCTTTTCC AAATAAATCT AGTATTAAAG    1320
GTGCTTATAA GAATAGTTTA CATTATGGCG AGATTAATTC TTGGATTGAA ATTAAAGGGG    1380
ATAAGCAGAC GATTATCGGA GGAAATGAGA GTGCGATTGA TGCAGCTATT GAATTAGCTA    1440
AATTAGGTAA ACGAGTTACA ATTTATACAG ATACTTTGGG GTTAAATATA AGAGATGCTG    1500
ACCCGAGCAA AAGgTTAGCA CCACGCACAA GACAGAGATT TTTTGATTTA CGTGTAAACC    1560
AAAAAAAATT AGATAGTATT AAAATATACA CAACTACCAA AGTTAAAAAA ATTGAAAAAA    1620
AGTCTACTAG TTATATATTG ATTACAGAAA ATGGGAAATC ATTACCTGTA GAAAATATAC    1680
CAATTCTTTG TACAGGTTTT AAAAATGGCA CAAAATCTAT TACTGCATCA TTGTTTAAGT    1740
ATAAGGAAAA TGGAGAGGTT CTTTTAAATG ATTTTGATGA ATCGACAATA GCTAAAAATA    1800
TTTTCTTAAC CGGTCCAAAT GTACGTAAAG GTAATACCAT TTTTTGTTAT ATATATAAAT    1860
TTCGCCAACG TTTTGCCGTA ATCGCTAATG AAATTGCCCG ACGAAAACAT ATTACTATTG    1920
ACGAAAAAAA ATTGTCATAC TATAAAAATC AATCTTTTTA TTTAGATGGT TGCTCTGGGT    1980
GTGAAGTAAG ATGTAGTTGT TGACAAATAA GTAATTGTTA CAACTTATTT TATAAATTAA    2040
AAGATTTTTA AATAATTTAA GTCAAGTCCA GACTCCTGTG TAAAATGCTA TACAATGTTT    2100
TTACCATTTC TACTTATCAA AATTGATGTA TTTTCTTGAA GAATAAATCC ATTCATCATG    2160
TAGGTCCATA AGAACGGCTC CAATTAAGCG ATTGGCTGAT GTTTGATTGG GGAAGATGCG    2220
AATAATCTTT TCTCTTCTGC GTACTTCTTG ATTCAGTCGT TCAATTAGAT TGGTACTCTT    2280
TAGTCGATTG TGGGAATTTC CTTGTACGGT ATATTGAAAG GCGTCTTCGA ATCCATCATC    2340
CAATGATGCG CAAGCTTTTG AATATTTTGG TTGATCGATA TAATCATGAA TCAATCGATT    2400
TTTAGCCTCA CGCGCTAAGT TAATATCTGT GAACTTAAAA ATTCCTTTAA CAGCTTCTCT    2460
```

GAAAGATTTT GAATTTTTTT TAGGAATGGT GGTAAAGATA TTTCTTAGGA AGTGAACTTG 2520

GCATCyTTGc CAmCTTACGT TGGTGAAGGA TTTTCTAATG GCAGAGACTA ATCCTTTGTG 2580

CGCATCAGAA ATAACGAGTT CCGTAcCTTG TAAACCGCGT CCTTTAgGAw TCmAAAAATG 2640

TkGTCCAGGc TCTCGCTTCG cCACTTGAAT CAGaAGCgAT AATTCACGGC GCAnCTTGGT 2700

ATCCAACGGC AAAGAAGCTT TGAGnG 2726

(2) INFORMATION FOR SEQ ID NO: 317:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 741 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

ATAAAAAACT ACAATTTTTT GGTACAAATA GTAATGTTTT TTTAAGTTTA TTCAATACTG 60

TTGGGACAGT AATATTGATA GCTATAGGCA GCTTATTAGT ATTAAAGAAT AGTATAAATT 120

TAGGAAATTT AACAGTATTA ATTTCATTAT CAGGAATAGT AGTATCATCC CTAAATCAAT 180

TTGGGTCGTT TCAAGCGCAA TATGAATCGA TGCAGGTGGC TAGTCATAGG CTTGAATCTA 240

TATTAATAAA TATGGAAAAT GAGAATGTAT GCGGGGAAAt AATTTTAGAC AAGAAAATAG 300

AATCAATAAG ATGTAAACGT GTTTCCATAA AAAAAGGAGA TACTTTATTA CTCGACACTG 360

TTAACTGTGA AATCTATAGA GGGAAAAATC TTTCAATACG AGGTGAAAAT GGTTCAGGTA 420

AATCTACATT AATTAAGTCT TTAGTAAGAT TAGATGATGA CTATAGGGGA CAGATTCTAA 480

TTAATAATAT TGACATAAAA AAAATCAACT TGGACTGTCT ACGATCAAAA TTAGTGTTTG 540

TAGAGCCCAA TCCTAAATTT TTAGAAGGTA CAATTAGAGA TAATTTATTA TTGGGACATA 600

AAGTTCCGAA TAGTATATTT AATAAGTTGA TAAGAGATTT TGAAATAAAC AAAATATTAG 660

ACGATCTACC TTTAGGTATA AACTTtCCTG GAGAAgCTGC AATAAAATGC TTATCTTCTG 720

GACAAAAACA AAAGCTAGCC T 741

(2) INFORMATION FOR SEQ ID NO: 318:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6157 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

AGTAAATAAA TGAATACATT GCAATTATTT AAGCAGGCCT TTTTTTCAGT TTCAGAATTA 60

CGGCATGCGA AAAAAATGGC GTTTTGGAAA GTAATTCTCT ATGCTGTTTT TTTAAGTGCT 120

ATTTTAGCAC TACCTATTAC CAAACAAGTC TTTTCTGTGC TACAAGAGGT GCAACAAGAT 180

GGCCAAAAAA TTGCAGAAAA GTTGCCTGAC TTTTCTATTG AAAATGGCAC GTTGCAAACA 240

AAGGCCAAAG AATCTGGCTT TATTTATCAA ACCGACTCGA TTATTTTTAC GTTTGATCCA 300

GATGGAAAAC GTACAGCCGC TGATGTCCAA AAGGATTTAA TTGGTAATGC GTTTGGTTTG 360

GCTTTTTTAC AAGATGAATT TGTAGTCGCC TTACCAAATT CAGGCGCAAC TGAATCAATT 420

CTAGGCACCG ACCAATTTAT CTTCCCCTAT AGTAAAGGAA CCTTAGATGG CGTTAACGCT 480

CAGAGTATCA AGACTGCGCT AAGTGAGGCA GCGATCCCCT GGTGGACAAA ACTCATTGTC 540

TTTTTAGTGG CTATCTACCC TGTTCTAATC GGTTTAGTTT TAGACTTACT GATTGCCGCT 600

ATCGGTGCAT CTCTCTATTC AAAATTACGT TTTTATTCTT TACGTTTGAT TGATTGTTTA 660

AAAATTATTA CGTATTGTGC TACCGTGCCA GTTATCTTAA GTGCTATTTT AAACTTCGTA 720

AATCCGATGT TTAACGATGG CATCCTAGTT ATTTTACTAA GTCTGTTCTT TTTCTTTGTC 780

ACAACCAAAA ACGAACCACG GAATATCCCT GGTTCTTAAA AAAAGGTGTT GTCTTAAAGA 840

CAACACCTTT TTACTATTCT TCAAATTCTT CTTCATCTGT TAATTCAGGT AAGTCAATCG 900

TGGTCACCGC AGATTGGCCA ATTTGCGTTA AATATGCAGC TAATTCATCC ACATTTTCCA 960

CAGGTGAGAC GGCTGCTTCA ATTGCCATTG CTAAATTAAG TCCAGCGACT ACTCGCAATT 1020

GCGGATAAGT GCCCATGGCC ATCATAGCCA CATTACAAGG TGTTCCTCCT TTTAAATCAG 1080

CTAATACAAG CGTCGGGACA TTGCCAGCTT CTTTTAAGAT TGCTGCCAGT TTTGCTTGCG 1140

TGCCGCTCAA GCCATCTTCT GCGGTCATTG ACACAATCGC AGCATCTGCT AATTCCCCGA 1200

CAATCaTTTG aGTAGAAGCT AAGGTTtCTT CTGCCATCCG TCCATGGcTC ATTAAAATTA 1260

ATTTTGGTTT CATCATATCC TCCTAAAAAT ACTGCTGATA AATTTTTGTT CGTGCTTTCa 1320

GAAAGCCACG TTCTTCACGA GTGGGaTGAA TCCGATTCGT TAACAAAATT AGGGCTTCTT 1380

TGGTCTCTTT TTTCATACCA ATCGCCGTTC CAGTAAAGCC AGTATGATAC AAATAGTCTG 1440

CCCCTGCTTG CGTCTTTTTC ACTTCCCAAC CAAATGTACG TCCCCCTTGA TTCATAGACC 1500

AGAGAGCAGT GAACATTTTT TCAGAAAATA ATGGTTTGCC AGACGGGTAG CGATTGTTCA 1560

TAAAACACTG aACAAAAGTG AGTAAATCTG GTAAAGTTGC AAATAAACCT GCACTGCCAA 1620

TTGGCGCTTT TAGTTGATAG GCTTTCGAAT CGTGGACCAC TCCCTGAATA ACGCCGCGTT 1680

GTTTTGTAAT CTCTGTGGGC ACTGCCTGGG CAAAATCGCT GACCGTAAAA CTTGTATGCT 1740

GTAGGTTCAT TGGTTGAAAT ACGTATGTTT GAAAGCTCTC TTCTAATGAG CAGCGATCCA 1800

ATACTTGAAT AATTTCTCCC AATAAGTAAT ACCCTAAATC AGAATAAGTG GTTTTCCCTC 1860

GTTCAGATAA CGAATGCGTT AAAATGATTT CTTGCAAGCT TTTTTTTGTG ACATTTTTCT 1920

TATCGACAAC ATCAGCTGGC AGTCCACTCT GATGCAGTAG AAGCTCCCCA ATCGAACAAG 1980

ATAGTCCTTG ATAATCAGGT AGGATTTCCC CCACAGTGGT TGAAAATGTA AGCTGATTTG 2040

TATCGATTAA TTGTAACATT CGATTTGTTG TCCCAATAAC TTTCGTTAAG GACGCTAAAT 2100

CATACAGGCT AGCTGATGAA ATCTTTTGTC TTGCAAATGG ACCAGTATTA CCTAAAAAGC 2160

CCGTATACTG TTCTGATAAG CTATCTTCTG TGATCGTCCC CCAAGCGACA CCCGGAACGA 2220

CTTGCTGATT AATCAGTCTC TGAATTGTTT CTTGAATTGC GATTTGATTC ATTGTTCACC 2280

ACTTCTATGC TAAAACTTCT GAAATCCGTC CGTCATGTTC TGCTAATAGT TGTTTGGCTT 2340

TTTGAAAATC AACGTGGGCT TTCGCCATTA CAATGGCTGG TGCCACTTCT AGCTGTGCTG 2400

CCTCCAAGTA TTCTTTGGCT TGCGATTCCT CGATTTCTGC TGCTTCTTTA ATGATATTCG 2460

TGGCACGTTG AATTAATTTT TCGTTCGTTG GTTGTACATT GACCATCAAA TTTTGATAGA 2520

TATTCCCTAC TTTAACCATA ATTCCTGTTG AAAACATATT TAACACCATT TTCTGTGCTG 2580

AACCCGCTTT CATTCTAGTT GACCCTGTAA TGACTTCCGG TCCTACAATT GGCGCAATCC 2640

CAATTTCCGC CAATTGATTC ATCGGACTTT GATTATTACA GGTAACTGAA ATGGTTAGAG 2700

CGCCCACTTT TTTTCCGTAT TCAATTGCGG AAACAGCATA TGGTGTCCGA CCACTAGCAG 2760

CAATCGCAAT TACGACATCT CGGGCAGTCA ATTGATGTTG CGTTAAATCT TCGATAGCTA 2820

ATTCTTTCGA GTCTTCAGCG CCTTCAATTG CTTGATACAT TGCTTTTTCA CCACCAGCTA 2880

AAATACCAAA TGCGCGTTCT GGCGACACAC TATATGTGGG TGTTAATTCA ATCGCATCCA 2940

AAGCACCTAA ACGTCCAGAC GTTCCTGCAC CACAATAGAT TAAACGGCCC CCTTTTTTAA 3000

ATCGTTCTGC CGCTGCATCA ATTGCTGCAG CAATCTGCGG AAGCACCTTT TCAATTGCTT 3060

GTGCTACCTT TTGGTCTTCT TGATTTATTA AAGTTACCAT TTCTAATGTT GACACTTGGT 3120

CAATCTTTTT AGTCGCTTCA TTTCTGGCTT CTGTCGTTAA TCCTTCTAAG TTCATTTGCT 3180

TGCACACCCC TACTTCTCTT TTAGACGGCT GACAAGTTGA CTCGTTTCAA TCATATTGCG 3240

TTCAATTTCT GTGTCTAAAT CATCTTGAAT TGAACCTAAA TACAGAACAT CGCCAATTGC 3300

CATTGACGAA GCAATCGATG AAATAGCCCC AACTCGAATG AGGTGCTCAT TATTGGGGAC 3360

TAAAAGAACT TCATCGCTCA ATGATCGCAC GTTTTCTCCA TCGTTACTGG TAATAAAAAT 3420

AACCGTTGCG CCTCTATTTT TGGCAATTTC ACAGGCGATT AAGGCTTCTT TTGTCAAACC 3480

ACTATAGGAG ACCGTGATCA AAATGTCTTG TGCCGTTGCA TAATTTAAAA ACTCCAACAT 3540

CATGTGACTA TCATAATTAA AAATTGCTTG GCGACCAGCG CGGTTAAACT TGTGATAAAG 3600

ATTATAAGTC GTTAAAGAAG AGGCACCAAT TCCAACCAAA TGAATTTTTT CCGCTTTTTT 3660

AACATGCGCG ATTGCTCGTT CCAAGGCTTG TTTATCCACT GTATAAAATA AATCCTCAAT 3720

CGTAGCATTT AATAAAGATT CCACTTTCAT ACAAATAGTA TCCACCGAAT CATCCACCGC 3780

TACAATTGGA TCAATATCTT TTTTACCGTT ACTGGCAGCT TGTACCGCTG CAATGGATAG 3840

TTTTAACTCT TCCCAGCTGT CAAACCCTAA TTGTTTAGAA AAACGGGTAA CTGACGCCGG 3900

AGAAGTTTCG CTATTGCTAG CAATTTCGTT TGCTGTTAAT GAGAGAAGAC GGGTCGGATC 3960

CTTCAGTATA TAGTCTGCGA TTCTCTTATT TACTTTGGAA AAGTCCTTAT ATTTTTGTCT 4020

GATTGAACGG ATTGGCTCCA TCTTCTCATC TCCATTTCTA TTTTATGAAA GTTAAAATTG 4080

CTTCGTTTGG TTCAGAAGGG ATCATCTGTG CGGTAATAGT CACCCCATTA TCTGTTAAAC 4140

GACGAACGGT TTGAATATCT TCATCGGTTA ATGCAACAGA TTTTTTAATT GCTCGACTGC 4200

CTTCTTTTTG AGAAATATTG CCGACATTGA AGCTCTGCAA CGGAACACCG CCATCAATTA 4260

AAGCCGCCAT ATCTGCAATG TCTTTCGTAA TTAAAAAGAC TTTTTCTTCA TCATAATTGC 4320

CTGCTAAAAT TTTTTCAATC GCTTTACGTT TTGATAAAAT GGATAATTTA ACTCCTGCaG 4380

GCTTAGCCAT TTTTAAGGCA CCAATTTGTA ATTGATCTTT GACAGCGGCA TCATTTACTA 4440

CCATGATTCT TTGTGCGCCA ACCGTATTCG TCCAAACGGT TGCAACTTGT CCGTGAATTA 4500

AACGCTCATC TACTCTTGCA TGTACAATTG TCATAATAAA TCTCTCCTAT TATATACTAT 4560

TTTTTTAAAA TAAACCGCCA ACCCAAGCCA ACATTGGTTG TAAATTGCCT AAAACCATGC 4620

CAATtAAAAT TAAAACAAAG ATTAGCCGTG TTGAATTCAT TTTCTTTTTC CCTAGTAACC 4680

AATAGGATAA AAGAACAATC CCTAAAGGAA TTAATGCGGG TAAAATTTTA TCCAACATTT 4740

CTTGAACAGA TAATGTTACA TCTCCCATTT TATAAGTTAA ATCTAATTTG TAAGTAATTA 4800

CCGAAGGAAT CAAACCGCCG ACCACTGTTA ACCCTAAAAT GGCTGCrCCT TGCGTAAACA 4860

ATTTCATTTT ATCCGCAAAT TCTGTCGCAA TTTTTTTCCC TTGATGGTAA CCAATCCAAG 4920

TGAATTTATA GCGAACCCAT AGGACCACTG CTCCGGCAAT TAACGGGATC AGTAACCCAA 4980

CAACCTGGCC CCCTTGCGCC ATGTAAGCTG CAATCGAAAA AACAATCGsA CGrTAAATnG 5040

CGATAAAAAT GGTATCACCA ACACCAGCAA ACGGACCCAT TAGACCTGTT TTCAAACCAG 5100

TAATGGCCGT TTCATCTTCA ATGCCGACTT CTTCTTCCAA AGCCATATCA GCTCCGACAA 5160

TTAAGTGCGA CATGGCTGGC GTTGTATTCA TAAAGCCAAG ATGTGTCTTC AAGGCTTTCT 5220

TCATAGCTAC TGGATCTTCA CTGTATAGTC GCTTTAATGC AGGCATCATC ACCCAAGCGT 5280

AACCAAGTCC TTGCATTTTT TCATAGTTCC AGCCCCACTG ATAGCCAAAA AGATTACGGA 5340

CGTAAACTTT ATTAATATCC TTTTTTGTTA ATTTGTTAGT CATCAATCTC AACCTCCTCA 5400

TCATCTTCGT AAATAACCGT GCCACTTGTT GTAGTGCGCC GCCTTTGTTA TGATTCATCA 5460

CATAAATCGC TGCTAAAGCC AAGCCAACCA GTGCGACACC TAAAACTGAG AAGAATTGGG 5520

CCCCATAAGC AAGTAACACG AACCCAnTAA TAAAaTAAGG CCAGTATGTT TTAATTGGCA 5580

AGTAACGCAT TAAAATAGCA ATCCCCaTCG CTGGTAAAAT AGCGCCTGCT GCTTTTAACC 5640

CATTCATCAC CCAAATAGGA aTCCAATCAT TGATAGCGTT TACAACTTGT TCACCAAAAG 5700

CTAGGCCAAT AAAAACTGGG ATGACACGAG AAATCGTCCA AGGAAAAATT CCTAAAACAT 5760

TGCAGCGCTC GACCCCTTTA TAATTTCCTT CTTCCGCATA ACGATCTGCT TTGTGTTGAA 5820

AAAACGTATT AGCCATTCTG CCTAAAATAT CTAGTTGTGT TAATAATAAC CCAATTGGAA 5880

CCGCAACTCC GATACCATAC TCTGCCCCTT TCCCTGAAAT GATTGCATAA GCCGTTCCCA 5940

TAATGGCCCC TGAAAGAAAG TCTGGGACTG TTGCTCCTCC ATAAGTCGCT ACACCTAATG 6000

TCATCAATTG AAGCGTTGCC CCAATCATCA AACCTGTTTG AACATnAaCC TAAGACTAGG 6060

CCAGTAAACG TTGCAGCAAA CAATGGTGTA TATGGACCAA TCTGAATAGA AATTTGGTCA 6120

AGTACACCAA CAACTGTGTA AAGACATAAA ATCAAAA 6157

(2) INFORMATION FOR SEQ ID NO: 319:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2378 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 319:

CAATCTCAAA TAGCAAtCAG GCTTCAGTGG ATATTTTAAA GACAACTGCT CATAGGGGAT 60

TAACCCCAAG CAACATAAGT ATCGATCCTT CACCCAGTTT TTCGACGAAT GGGGATCAAA 120

TAATGTTATC TAGTGTGACT AGAGAATCAA TGATAAACAA ATCAATTAGG TTTTCTTATG 180

CACCTCAAGA TTTAATAGAA GGTTCTACAA CAGGAACTAA ATCATTTAGT GGCACAATCA 240

ATTTTTTTGA AGTTAGCAAT TTAACTAGTG ATTTAAAACC TAATTTAAAA AAAGTATATT 300

CAGGAGATAA AGTAGAATGG ACTTGGGAAA TAATTAATAC ATCTTCTAAA CAAGCGTCGA 360

ATATTTATTC GGAATTAAAT TTACCTGTTG GATTAATTAT TGATCAAAAC AGTATTATAA 420

AGAACGGAAA TGTAGCAACG ATAAATGATC TTAACGGAAC GAATAACCTA GGAGCATTAG 480

ATTCTAATCA AACAATAAAA TTTACTTTTG AAACTATAGT ATCTGGAAAT TCAGGAGAAT 540

GGCTGGAATT AAACGGAAAA GTAAATTGGG AAGATCATGG GGCTCGTTTT GTGACAAGTT 600

CAAATCAAGT GAAAATAATA GATGATGAAC AAAAGGATAA AGCTGAAGAG ACAGATGAGT 660

TAGAGATTCT TTCAGTACCA CAATCTTTTA AATATGGCAT CCTAAATAAA AGTAATGTCG 720

AACAAATCAT TCATTTGAAC ACnAATAATT ACCAGACAAA TACTTCGGTG GTTACGGATG 780

GATTTTATAC ACGTATGAGA GATGAGCGTC CTAGTCTTAG TGGrTGGrAA CTAACTGCAC 840

AGTTATCyGr TTTTAAAGAk GTTmCAAtTy CAcGGTGTTA aCGAATtCTG GGrTAGCyTT 900

AAAAATGGaA AATATGAAAt TGAATCTATA AAAAATCGAG ATAsTCCACA AGAGAGTATT 960

GATCAAAATC CAACAGGAAC ACCTTCAACT GTTAGCACsA aTGAAACGCT GATATCAGGA 1020

CAAGCAGCAA AAACATTGAT CAATGCACAG GCAAATGAAG GTCACGGTAC GTGGCAACTA 1080

AGGATTCCtT TTGATAAAGT GTCCTTGACT GTACCAGCGA ATACAGGGGA AATAAATAAA 1140

AATTATACAG CGACTTTGAC ATGGTCATTA GATGATACAC CATAGGAGGA GAGTTCTGTT 1200

TATACGAGTA TGGATTGATT CTTGAATCGT ATTCTTGCCT AACTTTCCTC TATTTTGTTA 1260

TTAAGGTATG GTCCATAGGG TGTGGACATT GGTGGGAAAT CAACAGAATT ATTAGTAGAC 1320

TTTGGTTTGA TGTTTATAAC TTAGTGTATG tAGTTGAAGC TGCTTATTAT ATCTTGATTA 1380

TGTTAAGCAG CTTAATATTT yTATCTTATT AAAAAAGGGG GAAGCTCAAT GACAAGAGTA 1440

CTGATAATTG CAAGAAGTCC TTTAGCGGAA CAAGAGTTGG AATGTACACT TCAGCGGTCA 1500

TTCGATGAGG TCTTTTGTTC TTCTGAATTG ATGAAAGAGG TCGAAAAATA CTCGTTAGTT 1560

ACTCAATATT TTTCAGTGGT CATTTTTAGT GATACAATCT CCACCAGTGA AATGGCCACA 1620

TACTTACCAT ATTTTAAGAA ATTGGGCTTG TCTATTCTGC GAAAAGGACA GAAAGAACAA 1680

TTGAAAACAA CAGAATATAC GTATCTAATA GATGAAATTG ACGATTGGTT TGATGAACAA 1740

ACATCACCGA ACGTGCTGAT AGAGAAAATT GTTAAACTTG ATAATAAATC TCAAAAAAAC 1800

AGGATACCTA ACCTTCGTAT AATTGAATCG GTACAAAAAT CAATAGTTAA ACGAGATCAG 1860

GTTATCATTT ATTTTACGAA AAATGAACGT AAATTATTGT ACTATCTATA TAAATCTAAG 1920

GGTAAATGTG TTTCAAGGAG AGAACTGTGT CAATTAATGT GGGGAGAGGA AATAAGCAAT 1980

TCTTCCCTGA GTCAATTATC TACATTAGTA GCACATGTGC GAGAAAAATT aACCCATGTC 2040

GGGTTTGAGG AGTTATGGAT AAAAACGATT TGGGGGAGAG GCTATATTTT AAGTGATGAA 2100

TTTATCGACT ATTTGTCTCA AAATAATACT TtcAGTAAAG GTcGTGTGTc TAAATTaTTG 2160

TAGATTTGGA AATACATTTC GTTAAAAAAG ATTTAAAACA ATATAATAAG TGTTGTGCTA 2220

CATAGTTAGT AATGCAGCGA AAAATTTAAC CGAAGCTAGC GATTGCTAAC TTTGGTTAAA 2280

TTTCAAAAAT ACAACTGGAT TAGAAACTTC TGGATAGTCn ACTTTCnGGT AAGTATTTAT 2340

ATACCAAATC CTTTTATTTT AGnAATGAAT AGGGGATC 2378

(2) INFORMATION FOR SEQ ID NO: 320:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4480 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 320:

GATTAAAGCT ATTAATCTCT TGCTGCAGTT CAATAGTTTC TTGAAATTCT TCCGAGTTTT 60

CTCTCAAAGT ATCGTAATAT TTAGTATTGG CTTTCGTTnG GGATTTCTTT GATAACAAGA 120

TTAAGACAAA ACTAAATAAC GTTGGCAAAA TAACAGCTAA TCCCATTTTG ACATTACCAA 180

CCAAAAGCAT CACTGAAATA AAAGGGAAAA ACAGAGCCAT ACCGCCGGAT TTAGGTATTG 240

CATGACTCAT CGCATGCTCA ATACCTTCAA CGTCAGACAT GATAGTTTGA GATAAATCTG 300

ATAAATTATG TTTTGAAAAA TAGGATAATG GTAAATTTGA CAAATCGTCT GCAATTTGCA 360

CTCTTAAATG CGCACTTTCT TTATAAGTAC TATTATATAA GTtTTCATAC TCCCTATCTA 420

ACAACCAATA CATGAAAAGC AAAGTAACAA ACGATACTGC AAAATAAAGC CAACGGGGCT 480

TtGCATTTTC TAAAACTAAT TCGTCTATCA GCATCATTAT AATGAATGCT GGTACCATAT 540

TAATGCAGTA TACAAAAAAA CTAGAAATAG AAGCTTTAGT TAAAGCTTCT GCGCCCTGAT 600

CAGTTAAAGC AAATTTTTTT TTATAAAAAT TTTTCATTGT CAAGCCTCCA ATCATTTGCA 660

TTCTCATAAT TAGTGACCAA TTGTTTATAC ACTCGATGTT TAGCCATTAA TTCATTATGA 720

TTTCCACGTT CTACTATTTT TCCTTCTGAC AAGACAAGAA TTTCATCTAC AGATTTAATG 780

CTGGATAATC GATGAGCTAT CATAATAACT GTTTTTTCTT TCATCAAATA TTGAAAAGCC 840

TTTTGTAGTT CATACTCATT ATCTGGATCA ATAGATGCAC TAGCTTCATC CATAATAACA 900

ATCTTTGCAT CTTTTAAAAT TGCTCTGGcA ATAGCTATAC GTTGTTTTTC TCCTCCTGAC 960

AGATAAACTC CTTTACTACC AATTAAGGTA TGCTCTCTTT CGGGAAATTT TTTTAAAATC 1020

gATTCGCAAC CGGCTAATCT TAGCGCGTCA AATACTTtAC TTGCtGGCGC TgAGGATTTG 1080

CTATAGAAcG TTATCGTAAA TACTCTGATT GAATAGTTTC GTATCTTGAA AGACAAAAGC 1140

AATTGTGTTA ATAATTGTGT CTTTTGAATA GTGATTTAAA GGCTTACCAC CAATTTTAAT 1200

CTGTCCCTTA TCCACATTGT AGAAACCTGA AATTAGTTTA GCTATCGTAC TCTTTCCGCT 1260

TCCTGATTTT CCTACTAATG CATAAATTCT TTTTTGTTCT AAAACTAAGT TGAAATTTTC 1320

TAAAACACAC TCTTCTTTGT ATGAAAAGCT GACATTTTCA AATGCAATGT CACAATTTTT 1380

AATCTGTGTC TCGTTCCCAT ATTTTAGTTG GTCTTGCTTC ATTTCTAAAT AAAGGGAGTC 1440

TAGTGTTTCG ACTGCATATC TTGCTTCAAA CAAGTGCATA CTTACATACA TTATTCTCAT 1500

GAAAGAAACA AAAATAATTC CACTTAAAAA AAACAACATC AACAAGTCAA CAGCTAATAT 1560

CTTGGGAGTG GTAAGAGAAG TTATAAAAAA GGTTATAGGA ACTATCAAAA TGCAAACTAT 1620

GCCAAAAAAT AGCCATTGAA AAAGAACAAA AGGTGTTTTC CCACTCAAAG AATACTTATA 1680

TGCAAACTTA GCATAATCGT TAATTGCTTT ATTCAAGGCT TTAAATGATG TTACTTTAGC 1740

ACCAAATATC TTTATAACTT GAATTCCGCG GATATATTCT ACCGTTTCAG AGCTTAACAC 1800

ATCTAACGAT TCTTGATATT TTTGAATAAA ATTTGTATTC CCTGTCATTT TCTTTAACAA 1860

GAATACACCA ATTACTAACA TGGCGATTAG AACCAATCCA ATTTTAAAGT TAATGAAAAA 1920

GCCTAAGGCA ACTGATAAAA CAGGTGTTAC TATGGCTTGA GCACTATCTG GAATCAAGTG 1980

AGCAATTGCC ATATGTsTTT TAGCCGCGTT GTCATCAATT GTCTTTCGAA TATATCCCGA 2040

AGAATTTAAA TCAAAAAAAC GAAAACTTGC TTTTGTTAAa CCGCTAATTC CCTTTTytCT 2100

CAAATTAGTC TCyAAGCGAA ACCCAAGGAT ATGAGAAAAT AAGCCAGAAA AAATATAAAA 2160

TAAAGCTCCT AAAGTTAACC AAGCAACAGT TTGAACCGCA TAACTCTTAG CTAAATATTC 2220

ATCTCCTACA ATAATTAGAG ACTGAAGAAA TTTGTAAATA CTGTAATATC CATAGACTGT 2280

TAATACGACA GAAAAATCAG AAAATAAAGT GGCTAACCAT CCGTAGTATT TCATTTTAGG 2340

GACATAACTA AATAATTTTT TGTATACTTT CATTTTTTCC TCCCCCGCAT TCAATTTCTT 2400

CTCTTATTCG TTTACAACTG TTAATGCTTT CCGTAATGGC GTAAATTAGA GATGTACAAA 2460

CAGATAGAAG GTAAAmTACG TTCTATTTTT TGTATTTTTT TAAGACTTTT CTTCCAATCA 2520

ATACACCAAT CATAGATAAA CTTGCTACAA CTAACACTGA AATAATTCCC CATATACCCA 2580

CCGCGTAGTG GATCATTGCA TCCATTGTTG CAGCAGATTG TCCTCTATCT ATCCAGTCTT 2640

GACGAAAACG ATCTGTAAAG AACCATATGG GAATCCACTG TCCGGCAGCA CTGCCTAGTT 2700

GCATTAGAAC CCAAGACACA ACTTGAGGAA TTGCTCGATC ATATCCAAAT ACATATAGAA 2760

TAATATCAGC TAAGATCGCC ATACTAATCG AACTGATAAA CCAAGGTAAG TACGCCCCGC 2820

CCATTAACGT AAAAAATAAC AATAAAACAA TTGTGTAAAT AGTAAAAATT CCTTTTCTAG 2880

GGTAGTTATA GCTAATATAA ACAAATATGA TTCCAGATAA TAATCCAAAA ATCCCTGGGG 2940

ATATGCTGTG ACCAAACGCC CCACCAATAC TACTAACCAT TATTGAAAAA AAGTAAATTA 3000

CTAGGGAAAT AACTGTAGCT AAACCAATAA AAATATAATC TTTTGTTTTC ATATATATTT 3060

TCCTCAATTC TATTTCGTTA TGTCATTAAA AATCTGTTCT ATCAAAGTTG AATCTTCTAC 3120

TCTAAGTTTT ATATCTTTTT TTATTCCACC ATTTTTTAAA TAAATAATTC TATTGCATAT 3180

TTtATAAAGA AATTCCGGAT CATGAGTGGC TATGACTACT GCACCCTTTT CTTTTAAATC 3240

TTTAATAAGG TCAGCAATCC TTTCCATATT GATATAATCC AACCCACTTG TAGGTTCATC 3300

AAATAGAAAT AAATTAGTTT TTGACAATGa AGCTATAGAA ACCAATAATC TTTGTtTTTC 3360

TCCACCGGAT AAATCAAACG GTAATkGATC TTTCATTTTA TAAAGTTCCa TTTTTTTAAA 3420

ATTTTTTCTT GCTGATCTTT ACTAACTATT TCATTCCCCA GGGCTAACTC GCTCCTAACT 3480

GATTCTGTGA AAAATTGGTA ATCCcATTTC TTGCATTACT AAAAACGGCG ATGACAAATT 3540

TTTnnnATTT TCGATTATCT TTCCTTTGGT AGGGCGAATA ATCTTCATGA GTATTCTTAG 3600

TAAGGTTGTT TTCCCCaCAC CGTTTTTCCC TACCAAACCA ATAACATCTC CaACTTTGaC 3660

ATCAAAATTT ATGTCTTTTA GAATATTTTC ATAAGTAATT TCTTCAACTT TTAATACATC 3720

TGCACTTGCC TTTTCGATAT AGGGAACTTC AATCGAACTA TTAAAAATGT CAAATTTCCT 3780

TAATGCGTTA TTGTTTTCAG TTTTTAACAA CTGAATAGTT GTAGTTTTTA CAGTTCTATT 3840

TTTAACAAAT AGAACTTGAT CAATTAATTC TTTCAAATAA TAAAAACGAT GTTCAGCTAT 3900

AACAATAGTA AATCCTTTAT TTTTTAAATC ACAAAGTAAG TCAGCAAGTT TCTTAGTCGC 3960

TTGGAAGTCC AAATTTGAAG ACGGTTCATC CAATATAATT AATTTAGAGT TTAGCGATAA 4020

ACTACATCCA ATAGAAATTT TCTGCCGTTC ACCACTTGAT AGCGAAAAAA TGTTTTTATC 4080

TAATATATTT GAAAGATCTA ATAATTCAGC TAATTCATTG ATTCTTTTAT CCATATGCTC 4140

TTTTGTAAAT CCAAAGTTTT CCATCGAAAA TACTAACTCA CTCGTTGTAT TTGTTGTAAA 4200

GAATTGTCCT CTAGGATCTT GACTTACGTA GCCAATATTT TTTGATATAC TTTCTTGTGT 4260

TAAGGATTGT AGCGAGTTAT TTTCAAATAA AATTTCCCCA ACCAATTTTC CGTCATAAAG 4320

TCTTGGAATA AGTCCATTCA TTAACTTTAA TAGTGTGGAT TTCCCACTCC CAGAAGAACC 4380

AGTAATCAAT GTAAGACTAC CTTTTTTTAT TGAAAAATTA AGGTCAACAA GCGCTTGTTC 4440

CTTAACATCC TTATATGAAA AAGAAACATT GTTAAATTCA 4480

(2) INFORMATION FOR SEQ ID NO: 321:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6720 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

TACAGATTAC TTGAAAATCT TTACGTTCCA GTTTTATTCT nTCTAAAAAA CGTACTCACT 60

TTTTTTCTTA GACTTCTTTT TTTGATTTTC CTCTTTTAGG TCTTGTCCTT TTTGTCTTCA 120

AAATATTTAA ATAAATCCCT TTCAAAAGTA TTTATAGGTT TTGGACAATA AGGAATATCT 180

TTTTTCTTCA TATAAACACC TCATTAAACT ATTTTATATC ATCACATTAT CAAATTAACA 240

AATGTATTAA AAGAAAGTTT TTTATTAAAA cTTTCTTGGC CTTTTGGGAG ATTTCAGTGC 300

GATTTCTTTT TATTCCAATA CTTTTTTGTT GGTCACTTTC GTATACCCAT TCTAAGCGAT 360

AATCGGCGAC AATGTTTTTA CAACGAATGG TAGGACAAAA CAATGGGAGA AAAAACTTTT 420

TCTACAGCAA TCCCTTGTTT TCACCTACCA CATGTTGTGA ACCCAAAGCA CGCCGACGAA 480

CGCTTAGTTA TACGAAAATG ACCTACCAAA AACAAAAAAA GTGAGGTGAA TAGAAACAAA 540

AAGAAATTGA AATCTCCCAA AAGGTCGTCC ACCTCGGGGG CAACTTTCTG CCGAAAATAG 600

TAGTTGGGCC TATGAACAAA CTAACGTTAC TTTTTGGTAA ATTATTAAAT TAACCGAAAA 660

AACAGCATGG TTTCTCCCCA TCAAAAAGCA AACTGTAAAA ACATATACTT CTATGAGTGG 720

AGAAACTATG CCAATGCTAA ATCAATTTAT CgrTTATCCm CTTgcACGaA CCAAAGACAG 780

CATTGTTAGA AAACAAATTA ACCCgCCcTT ATATACCAAT AAATTATACG GTAATCGTAG 840

TTGGGATACA TTTTACTACT ATGATAATCA ACTGGTTTAT GAAaGCAGTT aTTACTAAGA 900

GGGAaTTTTC TTCTTAGTTA GCGTTAAAAA TAACAAAAAT AAAAGGAACA AGAACAATGA 960

AAACAATAAC TCTCTATGAA ATCTTAGGTG AAGAAGGGTC TGACAAATAT ATTTCTGTAT 1020

TCAAATCTTT AATTGCTTCC AATTTTACAG AGAATGAAGA CATTTTATCA AGGTGTCAGC 1080

TATTTATACA CAATAATGTC AGTGCTGAAG AAAATTTATA AAGATGAACC AGAAAAACTG 1140

AAAGAGACGG CAAGCGCAAA CTTAGAATTT TATAACACGA ATCCACAAAT GTTGACATTT 1200

ATTACCAGCA TGCAGTTAGC GATGTATGAT AATGATCAAA GTGTCTCAGA TACACGCAGC 1260

ATTAAGATGG CACTTATGGG ACTCCTATCA GGAATCGGAA ACTCGATTGC TCGATTTGGG 1320

ATTGCCTCCT TATTTTCAAC AATATTTGCA GGATTAGCGA TGAATGGTCT GGGATTTGCT 1380

CTGATGTTTT TTTGGTTATC GATGTTAATC TCAATGCTAG TCATCAAATT ATTAATGGGA 1440

GGGATATTTA GGGTTTAAAC TTGGAACGAG TGTGATCGAA ACCTTAAGTG ATAAAATTGG 1500

TAAAATTTCT TCAGCAGCAA ATATTGGCGG AGTGGCTGTT ATTGCGGCAT TAAATACTTC 1560

TTTTGTTAAG GCAAATATCG CCATTAAGTA TTCACATAAT GTAGAATCTG GTGAAAAACA 1620

AGTAATTGCT ATTCAAGATA TTTTAGATAA GATTACGCCC AAAATGCTTC CAGTCATTTT 1680

AACTATTTTA GTTTTTTATT TGATCAAAAA AAGAAAATGG AACACGTATC AATTGTTATT 1740

TTTTTTATTC GTGATTGGGA TTCTAGCCTC TGACTTAGGG ATTTTAGCAT AGTAAGTGAG 1800

TGTAGTAAGG GACAGGGTTT TCGATTTCGA GTAAACCTTG TTTCCTTATT TATAAGTTTT 1860

CTGAGTGATG GGCATAATAT TTTTTCAGTT AACGGAAATA AAATTTGAAA GCGAGAACAC 1920

GATGAATCTT CAAAAAATCG AAAACTACCA ACTAAAGTTT TATCAGAAAG ATTTACTTTC 1980

AGGATATTTG GAAAAACATT CAAAACTTTT GGAACCACTT TTTGAACGTA CTTATTTTCT 2040

ATTAAAAGAT CAGATTATTT ACAATGATGC GAkGGATATG GAAGCCTGTT CTATTCCGTA 2100

TTCTCTAAAA GAAAATACAC GGAATCGATA ACATGGAGAT GACCCAGAAT gGCTCTTtAT 2160

GTTGAATTGA CAAAGTTTtC TTCTAGATTT ATCACAGGCA TACGCGTTAA TAAAAGAAAA 2220

ATGCTATTTA CAAAAATGGT GCAGTCTGCT TCTTGATTTT ATTCAAGAGG AAGGCGAACC 2280

AAATTCTACT AATAGGAATG TTTGGCGACA CCTTGATGTT GGAATTCGAG TGATGAACTG 2340

GATGAAGAGT TTGACTTATA TTTCTATTGC TGATTATAAA CAATTGGGAA TTGATAAAGT 2400

GTTGCGAAAC GCTTTGCTGG TTAATCTCGA GTATTTGGAG AGGAGCTATA TCGATAAGTA 2460

CAGACTTAGT AATTGGGGAG TATTGGTAAC GGGTGGAATG GCTGCTATGG ATCTATTTCT 2520

TCCAGAACTA GTGAACAGAG TAAACTAATA TGGGACCGGT TAACTGACTT ACAATTTTAT 2580

TCAGATGGCA TTCATTGGGA ACAGAGCGCA TTGTATCAAT ATGAAGTTTT GATGACGTAT 2640

GTATATCTAT TACAGATTTC TGAATATCTT GAAATATCAT TGCCATTGGA TCTTCGAACG 2700

AAATTGAAGA TACCTATCTT ATCTACCTAC TATATAACGG ATAATCAAGA TGTTCTGAAT 2760

CCTATCAACG ATAGCGATCA TGTCAATTTT CGATATGTAT ATGATAGCTA TCGTAATATG 2820

AAGAAAGAAC TTGGAAAACA CTGCAGCCAG AGAAACTTTT TTCGTGGTGA ATCAAGTGGT 2880

TTAGTGTTTT ATAAAACAGA GGATATTTAT TTTACGCTTT TTAATGGATT GCATGGAAGC 2940

TCACATGGTC ATGCTTCTAC AGGGAGTTTT ACGCTACAAT TACAGGGTGA TGATTTGATA 3000

TCTGATAGTG GTTGTTACAG TTATGTCAAC AAAGCAGAGT GGCTTCAACC GAAGGAGTGT 3060

GATTCGCATA ATACAATGTT TATTGCAGAG AATTCTCATA CTTTAGTACT GATACATGGG 3120

GCTACGGGAA ATTACCAACC CCATTATTTC AGCGAATAAA AGAAATTCCT GTCGGTTTTT 3180

TTGCGGAATG CGGCTGGATA GATAAAGATG ATCAGAACTT AATGATTTTT AAGCGTAGTT 3240

TTATCTACTT GAAGGCCATT AATTCAGTAA TTATTATTGA TAGTTTTATA GGTCAAAAAG 3300

AGACTGAAAT TACGAGTACT TATAATTTAG CACCTACGCT AAATTGTCAA AAAAGCAGAT 3360

CAATTCAATT TAACTACAAA TTCACAAAAA TATAAACTCC TTATAGTATC AGGTCAAACG 3420

GATCAAACTA CTGTTGAAAG ATCAGAGATT TATAATCAGC TGATCGAACT ATATACGCTT 3480

GACCAATAAA TTTCATTATA AAACGGGGAA AGAAGTTCAG GATACCGTGA TTTCCCCTTT 3540

GGAAGATATC CAGATTATGC CAGTAAAAGT GAACCAAGCG GGAGAAGACG AGCCATTCTG 3600

TCAGGCCAAA GGATTTTGCG TGATCACTGG AGAGAAAAAA TTTTATATTT TTGTTATATG 3660

TGGAGATATT GTTAAAGGGA ATAAATTATT AGTGAGTATG GTCAATTTTT TATGGCCGAC 3720

TTGTTTTGAT TGATCAAAAT GAAGCGATTA TTCGTATTAA ATAAAGGAGG AGAAGAATGA 3780

ACTTTTTAAT AACTGGACAT AGTGAATTTG CTATTGGAAT GCAAAGCGCC TTAACGATGA 3840

TTGCAGGAGA AATAGAAGAT GTGATGTTTA TCCCTTTTTT AGCCAATGAA ACGACAGAAG 3900

AATATCGTCA AAAAATTGAA TCTTCAATTG AGATGTATGA ACATGTCGTT TGTTTTACCG 3960

```
ATTTATTAGG CGGTACCTCA TTCAAAAAAT GTATTGAATT CTCTATGACA AAAAAACAAG      4020
TTTTTGTTAT AAGTGGTACA AACATAGCGA TGTTATTGGA AGGATTGATG ATGCGGAATT      4080
CTGGTATTGA GCCGTTTGAA CTTGCCAAAT CAGTGGTAAA AGTTGGAAAA AGTAATATTG      4140
TGCTTTTTGA AAAAAATAAT GAAGCTGAAG AACTATCAGA AGAGGTAATA TAAAATAATA      4200
GTTGTATAGC GATTATCTAA AGTGACTGGT CAAACGTTTC TAATCGTATA TCAGTCATTC      4260
AGTGAGTATA ATTTAAACCA TAACCTACCA ATTGCCAAGG AAATATAGCG TATTTTACTC      4320
TTGTCAATTG ACAGTTTATA GTTTAGAGAT GGCTATGCCA AACAACTAAT ATATGAAAAA      4380
GGAAACTAAA CATCATTCCT CTATTGGAGC AATTTCTTCT TTGATGGCTT ATAAAATGTA      4440
CTAATACTTC ATTTGAGGCA GTAGAAAGTT GAGAAATAGT TGTTTCTTGT GCTGTTTTTG      4500
TCGTTATTAA GTTTGCCAAA AAGCAACTTT CAACCAACAT TTTTGTTTCT TATCATATCA      4560
AGAACCTGTT TTTTATTTTT CTATAAAAAA TGTCGAAATT ATTAAGTTAT CGACATGATT      4620
TTGTTATATT TAAGATAAAT TAGTCAAAAT TTTTGCCAAT ATTTTTTCCA AATTTTTCTT      4680
TATTTTTTCT TTCCTTACCT TTTTCAATTA ATATCATTTG GTGTTAATAT ATCTCTGATA      4740
ATACAAAAAA ACATTGACCT ATCATAAGAT AAAGAGTGAT AGTTTAATAC TAAAAGAGGT      4800
ATATTAATTT TATTTAAAAA TTGACATTAA TGACTCTTAA TCAGAGGGTC TAGGGTTCGA      4860
ATCCCTATGC CCCCATTGGG TGCCAAACCC ACGAGAACGG TATTCTGTCG GCGTCTTAGG      4920
ACGTTTCGAA GaCGGAGTAC GCGTTCTCTT TTTTTACATA AAATGTCGTG TGTCGTGAGA      4980
GATACACTAG TTAAATCGGT TAAAGAAtnT ACGCTTAGAA GAGCGAGAAA GTACGTTATA      5040
TAAAAAAAGT TAAAAAAATT TCAGGCTAAA AAAGCGAATT TAACAAAAAG CGCACTTTTT      5100
GGTTCCTATG AAACGAAAAA GTACGTACTA TACAGCATGT ACGGCACGGT ATATGATGAT      5160
ATCAGAAAAC GTTAAAGGAG CGAATCAAAT GATTACATCA CTAGCAGATA CAGTAACATT      5220
AAACAACGGC ACTAAAATTC CAGGAATGGG TCTCGGGGTA TTCCAAATAC CAGATGAAGA      5280
GACAGCAAAA GTGGTCGAAG AAGGGATTAT CAATGGTTAT CGTTTAATCG ATACAGCGCA      5340
AATCTATGGC AATGAATCTG GCACAGGAGC GGGAATTAAA GCAGGCTTAG CGGCAACTGG      5400
TTTGAACCGT GAGGACTTAT TTGTTACCTC GAAAGTTTGG AACGCGCATA TTTCCTATGA      5460
TGAAACAATT CAAGCTTTCA ATGATAGTTT GGAACGTTTA GGTTTAGATT ATTTAGATTT      5520
GTATTTAATT CATTGGCCAG GAAATAATTC TTACAAAGAA TCATGGCAAG CTTTAGAAAC      5580
TTTATATGCT GAAGGAAAAG TGAAAGCAAT TGGTGTCAGC AACTTCCAAG TTCATCATTT      5640
AGAAGATTTA CTTTCTTATG CAAAAGTTGT TCCGGTAATT AACCAAGTGG AACTACATCC      5700
AAAATTGGAT CAAAAAGAAG TGCGTGATTT CTGTGAAAAA CATGATATCA AAGTCCAAGC      5760
TTGGTCACCA TTGATGCAAG GCCAATTATT ATCTAATGAA ACCATTTTGG CGATTGCTGA      5820
GAACCATAAT AAATCAGCAG CCCAAGTGAT TTTACGTTGG GATATTCAAC AAGATATCTT      5880
ATTAGCGGTG AAATCTGTCC ACAAAGAACG GATGATTAGC AATGCAGCGG TCTTTGATTT      5940
```

TGAACTAAGT GCAGAAGAAA TGGCACAAAT TAACCAATTG AATGAATCGT TACGTGTGGG 6000

ACCAGATCCA GACACGTTTG ATTTCTAAAA ATTAACTAGG AGGAGATACA ATGAGTGAAG 6060

CAGTTAAAGG TTTTTCAGTA GTCAACCACG TGGGGATTAC AGTCAGTAAT TTGGATGAAT 6120

CAGTGAAGTT TTATGAAGCA TTAACAGGAA CAAAAATCGC AAATCGTGAT GAAATTGGTG 6180

GAAAACGGAT GGCCCAAACA CAAGGGTTAG ATGATACATT AATCAAGTAT GCCAATGTGC 6240

ATCTAGAAAA CATCAATATT GATTTATTAG AATATGTTAA ACCAAAATCA GAAAAAGCTT 6300

CTTATTCCAA CGAACAAATT AGTGCAATGC ATATGTGTTT CGAAGTGGAA GATATTGATG 6360

CAGCTGTTGC ACGTTTAAGA GCGATTGGCG TAGAACCTGA TGGAGAGCCG ATTACATTCC 6420

AAAAAGAAGA TGGCTTGAAG AGCGGTTTTG GTACAGGGGT TGCTTACTTT ACCGATCCAG 6480

ATGGTACAAA TTTAGAATTG ATTGCTCCTA AAGGTCCATT TACACGAAAT TAAGTTAAAA 6540

GACTTTCAAA CmAACATCAA AAAGACCCCC TCATTCTAGT ATCACTAACT AGAATGGGGG 6600

CyCTTtTATT TTTTgAGACA GAAGTCGCTA AAGAAAAGAT ATTCCAAATC CGTAATTGAT 6660

GGTATACTAA AACGTAACAA ATAAAGGAGG nnAGACAATT GATGCTTAGC GTCTTAACAG 6720

(2) INFORMATION FOR SEQ ID NO: 322:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5069 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 322:

GTAAGAGGGG AGTATGAAAA TATGGGAAAA GAATATCATG TTGCCGTGGT AGGAGCAACA 60

GGTGCTGTCG GGACAAGAAT GATTGAATTG TTAGCAGAAA CAAGTTTGCC AATTGCCAGT 120

GTGAGGTTAT TGGCGTCTAA GCGTTCTGCT GGGACAAGAG ATTCAGTTTA AAGGGCAACC 180

TCTTATCGTG GAAGAATTAG TCCCGGAATC ATTTGAAGGA ATCGATTTAG CATTATTTAG 240

TGCTGGTGGG GAGTATTTCC GAAAAGTTTG CACCAGAAGC mGTTaAGCGC GGACTGTCGT 300

TGTTGACAAC ACTAGCCATT TTAGAATGGC AAAAGATGTT CCTTTAGTTG TTCCGGAAGT 360

GAACCCGGAA GCCTTGAAAA ATCATCAGGG AATCATCGCA AATCCCAATT GTTCTACGAT 420

TCAAATGATG GTGGCCTTAG AACCAATTCG GCAAGCGTTT GGCTTAGAAC GAATCATCGT 480

TTCAACGTAT CAAGCTGTGA GTGGCGCAGG TATCCGTGCA ATTAATGAGC TAAAAGAACA 540

AGCCGAAGAT TTTTTAGAGG GAAAACCTGT CAGCGAATGG TCAGCTGAAA TTTTACCAGC 600

TGGCAGTGAC GAAAAACATT ATCCAATTGC TTTTAATGCG TTGCCTCAAA TTGATGTTTT 660

TTCAGAAGCA GGGTATACCT ACGAAGAATG GAAAATGATG AATGAAACAA AAAAAATCAT 720

GGGAGACAaC AAGCTAAAAG TCGTtGCCAC CTGTGTCCGT ATTCCGGTAA TTTCAGGACA 780

TTCTGAATCT GTTTATATAG AAGTGAAACA AGCAGATGCA TCGGTCTCAG CCATCCAAGA 840

ATTAATGGCG CAAgcTCCAG GTGTAATTTT GCAAGATGAG CCAAGTGAAC AACTATATCC 900

ACAAGCGCTA AATAGCGTCA ACCAGAAAGA AACATTTGTC GGCCGTATTC GTCAAGACGT 960

TGACGTGAAG AACGGTTTCC ATTTATGGAT TGTGTCAGAC AACCTATTAA AAGGAGCTGC 1020

TTGGAACTCG GTCCAAATTG CTGAAACGCT ACACGAAATG AATTTGCTTC GTGTACCTGA 1080

AAACGACTAG AAGGTTTGTA TTGTATACTC GGGAGGGAAT AGGATGGATT TAACGAATGC 1140

AACAATTATT ACGGCAATGG TTACGCCGTT TCAAGAATCG GGAGAAATAG ATTTTGATAA 1200

ACTACCGCAA TTGGTAGACT ATTTATTGGC CAATCATACA GAAGGCGTTA TTTTAGCAGG 1260

AACAACTGGT GAATCACCCA CGTTAACTCA TGAAGAAGAA TTACAATTAT TCCAGCGAAT 1320

TATTGaACTG ATCGATGGAC GTATTCCGAT TATTTGTGGT GTAGGTACGA ATGATACACG 1380

TGACTCAGTC GCTTTTGTGa AAGAGTTAGC GACAATCGCT GGtATCGATG CTGTTTTGGc 1440

GGTTGTTCCT TATTATAATA AACCAAATCA AGAAGGTATG TATCAACATT TTAAAACCaT 1500

CGCAGAAGCA AGTGAACTGC CAATTATTTT GTACAATGTT CCTGGTAGAA CGGCTGCATG 1560

TCTAGAAGTG GAAACTACTT TGCGcTTAGC ACAATTrGaA AAAATTGTGG cAATTaAAGA 1620

ATGTGCaGGG TTAGATGCAA TTACTGAGTT GATTGAAAGA GCGCCAAAAG ACTTTCTAGT 1680

GTATACTGGA GAAGACGGGT TGGCTTTTGC CACCAAGGCA CTTGGCGGAC AAGGGGTTAT 1740

TTCGGTAGCC AGTCATGTCT TCGGTTCGTC AATGTATGAG ATGTATCAAG CGCTGGAGCA 1800

GGGGAACTTG CCGGAAGCGG CTAAaATTCA ACGACAGTTA TTACCAAAAA TGAACGCATT 1860

GTTTTCAGTG CCTTCGCCGG CACCCGTGAA AGCGGCCTTA AATCATTTAG GTATTCCTGT 1920

TGGAAATTTA CGTTTACCTC TAGTTGCTTG TACACCAGAA GAAGAGCAAA GAATTATTCG 1980

CACGTTAGAG ATTTAAAGAA CGTGTAATCA TAGAGAGGTG AAAAGAGTGA GTACAATAAA 2040

AATCGTTCCA TTAGGCGGCG TTCGTGAAAA TGGAAAAAAT ATGTACGTTG CTGAAGTGGA 2100

AGATGAAATT TTTGTACTGG ATTGTGGATT ACAATATCCA GAGAACGAAC TTTTAGGAAT 2160

TGACGTCGTC ATTCCGGATT TTACGTATTT AGAAGAAAAT AGTGAACGAA TTGCAGGGGT 2220

CTTTTTAACA CATGGCCATG CTGATGCAAT TGGGGCCTTA CCGTATTTGC TTTCAAAAAT 2280

TCAAGTGCCT GTTTTTGGGA CAGAATTAAC GATTGAACTA GCAAAACTAA GTGTCAACAG 2340

TCATGCGGAA ACGAAGAAAT TTAAAGATTT CCATGTGATT GACCCGCATA CAGAAATTGA 2400

TTTTGGTCAA GCAACGGTGA GCTTCTTTAA AACAACGCAC ACCATTCCAG ATTCGATTGG 2460

CGTTAgCATC AAAACAAAAG AAGGCAACAT TGTTTATACT GGTGATTTTA AATTTGATCA 2520

AAGTGCGATT GAGATGTATC AAACCGATTA TGGCCGCTTG GCTGAAATTG GGAAAGAAGG 2580

CGTCTTGGCG CTACTTAGTG ATTCATCAAA TGCTGAAAAC CCAGCCCAAG TTGCATCGGA 2640

AGCCCAAATC GCTGATGAAG TGTTCGACAC GATTCGCTAT TGGGAAGGAC GAATCATCGT 2700

GGCATGTGTG GCAAGTAATT TGCAACGGGT CCAACAAGTA TTGAATGCAG CGGATCGTTC 2760

TGGTCGTAAA GTCGTTTTGA CAGGCCAAGA TTTTGAACGG ATTATTCGGA CGGCGATGAA 2820

ATTGGAAAAA TTACAATTGC CAAGTGAAGA TCTATTAGTG AAACCAAAAG AAATGAAAAA 2880

ATATGCACCC GAACAGTTAT TAATTTTAGA AACAGGCCGG ATGGGAGAAC CTATTAAGTC 2940

ACTTCAAAAA ATGGCTAATA ATACACATGG TGTCGTTCGA ATTGAAGAAG GCGATTTAGT 3000

TTATATTACG ACAACGCCGA CAACGGCCAT GGAAACAACG GTTGCCAAAA CAGAGGATAT 3060

TGTCTATCGG GCCGGTGCTA CGGTGAAGCA AATTTCCGAC AATTTACGTG TGTCTGGTCA 3120

CGCTAATCCG AATGACTTGC AATTGATGCT GAACTTAATG AAACCAAAAT ATTTCATTCC 3180

TGTCCAAGGT GAATACCGCC AGTTAGCTGC GCATGCGGAc TTAGCACACG AAATTGGCAT 3240

GCCTTATAAA GATATCTTTA TTACAGGTCG TGGGGACATT CTAGAGTATA CAAAAGGTAG 3300

AATGTCGGTA GCTGGCAGCA CAACGGCCGA AAATATTATG ATTGACGGGA TTGGTGTCGG 3360

CGATATTGGT AATATCGTTT TACGGGATCG TCGTATTCTT TCCGAAGATG GTATTTTTGT 3420

GGCTGTTGTG ACAATTAATC GTCGAGAAAA ACGAATTGTT TCACCAGCTA AAATTACTTC 3480

TAGAGGTTTT GTTTACGTTA AAACAAGTAA AGACCTAATG AAAGAAAGTA GCAATATTGT 3540

TACAGAAATT GTTGAAAAGC ATTTAGAAAG TGACGATTTC GAATGGAGCA AACTGAAACA 3600

AGAAATCCGC GAAAACTTAA GCCGTTATTT ATTTGAACAA ACCAAACGTC GTCCTGTGAT 3660

CTTACCAGTT ATCATGGAAG CGACGCAACG CAAGCGTCCC AAAAATAACG CATAGTTCGT 3720

CCTGATAAAA GCAGGTGATT TTAGGAGAAA TTCTTCTAAA GTCATCTGTT CTTTTTTGTA 3780

ACTTTCTTCT ATAAATATGG TACTATTAAT CAGAGCGCTG AATAAAGGAG GCGATTTTAG 3840

TGGAAGAGAT TGTGTTGTTC CATACCAATG ATCTTCATTC TCACTTTGAA AATTGGCCCA 3900

AAATACGGCG TCTGGTTAAA GCCAAACGTT CGCTGTACCA AAAAGAGGGC AAAACAGTCG 3960

TAACCATTGA TTTAGGTGAC TTTTCAGATC GGTGTCATCC TTTGACCGAA GCAACAGATG 4020

GTCGAGCAAA TGTAGCTATA ATGAATACAT TAGCGTATGA CCTGGTTACA ATTGGAAATA 4080

ATGAAGGAAT CGGCAATTCA AAGAAACAAT TAGAGCATTT ATATGATCAG GCAACATTTG 4140

AAGTGGTCTT AGCTAATTTA GAAGAACCAA AAACCCAAAC GTTACCAGAT TTTTGTCAAG 4200

CATATAAAAT TATGACAACT AAAGAAGGCA CAAAGTTAGG GTTTATTGGC TTAACCGCAC 4260

CATTTCCTTT AACGTATAAT CCCAATGGCT GGACAATCAA ACAAGTGGAA GCAGTCCTTC 4320

CTCAACTGAT TACAGAAGTT GCGCCTCAAT GTGATGTTCT CATTTTGCTT AGTCACTTGG 4380

GGATAGATAC TGATTTTATG ATTGCGGCGA ATTATCCTGA AATTCAAGTG ATTTTAGGTT 4440

CCCATACACA TCATTTATTT AAAGATGGTG AAAAAATTAA TCATGTGCAA CTAGCAGCTG 4500

CTGGCAAATA TGGACAGTAT ATTGGTGAAG TGCATCTGTT TGTGGATGCC GACACGAAGC 4560

AAGTCACAAG TTATGCGAAA ACCATTGAAA CAGCAAGTTT AGAAGAGCAA GCAACGGATG 4620

CCAAGGAAAT TGCGGGTTAC TTAACGGAAG GTCATCGACT ATTGCAAGCG CAGcAAATTG 4680

CACAAATCCC TGAAACCTTG TCAACGGACT TaCGGCAACC GCATGCCTTT ATAACTGTTG 4740

CCTTGAAGGC GTTAAAAGAA GCGGGGCAGA CCGAAGCGGC TGTGTTGAAC AACGGGTTAT 4800

TTTTAGCAGA TTTACCAGAA GGAATAATCA ATGCAGACCA ATTACATGAG GCCTTGCCTC 4860

ATCCAATGCA TTTAATCAAG GTTACCTTGA AAGGGTCAGA CATGAGCCGT TTGATTCGTG 4920

AAATGGAAAA AAGTCGTCAA TTTCTGCGGA AATTTCCGAT TCGCGGGATG GGcTTCCGCG 4980

GGAAAATTTT TGGCGAATTA TGTTATGATG GGATTCGTTT TGAGCCAAAT AGCCAGACCG 5040

TTTTTTGGCA AGGGAAACCT ATTCAGCCA 5069

(2) INFORMATION FOR SEQ ID NO: 323:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6279 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

ATTTTAGGAT TATGGTTAAA TCAAACAGAA TCTAAAAATC GCTGGATGCA AGTATTTGAT 60

GAACTGAAAG CACGTGGTGT CGAAGATGTC TTTTTCATTT CGATGGACGG TGTTTCTGGC 120

CTCGAAGAAG GCGCAAAAGC GATCTTTCCA TCGGTAATTG TTCAGCGCTG TATTGTTCAT 180

CTTGTTCGAA ATGCATTGCG CTATATTCCA AGTAAGGACT ATAAGGAAGT CTGCCGAGAT 240

ATGAAAAAGT TCTACGGTGC TTCTTCTCTA AACGCTGCAC ATGCTGCTTT TGGCAGTTTT 300

CAAAATCGGT GGTCTCATTA TTCTGGTGCT GTAGATGTCT GGAAACGGAA CTTTGCACAT 360

GTTGAACAAC TATTTGATTA TGGTAGTGCC ATTCGAAAAA TTATGTACAC TACCAATGCG 420

GTTGAAAGTG TCCACTCCAG CTTCAGAAAA GTCACAAAAA AAGGAGCGTT CTCTAATGAG 480

AATGCACTCC TAAAACTACT TTATTTACGC ACGAAAGAAT TACACGCGAA GTGGTCAGGT 540

GGTCGAATTC AAAACTGGGC TATGGTGCTT AAGCAGTTAA TGATCAACGA GACCTTCTCC 600

TCTAGAATGA AAAGTATGCG ATTTACCTTC CGTAACAACG GTAGGTTTAT TCAACTATGT 660

GTCCGATCAA AATTTATCTG ATTTCAAAAT AGCCGAGAGA ATTATGGTGG TTAAAAACTC 720

TTAACTATTT AACTTCCACA GTTTAGTTGA CAAACCCCAA TTAGTTCTCT ATTTTCACAT 780

ACTATTTTCA ATAGTGTGTT TTTTATTTAT ATTCTTAGTC AAGCTAACTG TATGAGTATT 840

TAGTAAAGTT GTTTCTTTCa AAAATTCAGA AGTCTTATCm CCTTGTATAC CTAACCAAAT 900

CAGTGTCCAA GGTTCCTTTT TATCTGTATA ATACTTCACT TTTTGACCTT GAAAAAAAAG 960

AAAATATCGT GTTCTATAAG TTCAAAATGT TGAGCTTCTA CTTCTAAGAA TCCCTTACCA 1020

GATAAAACAT AATGAAATAT CCATTTTTGT TGAACTATAT AATTACGAAT AGAAGAATTG 1080

GGATGATAAG ATTCATATCC GCATTGGGAA AGGGAAATTG GAGCAGAAAC TGACTTATTA 1140

TGATACCAAA AATCCTCCAT TATTATCCTC CTTTATAGAA CTAACTATGA AGAAAATACC 1200

ATGTTTTAAG TATAGATTAT CATGTTCAAA TGGTCAAAAA AATCTATACT ATAATTAGAT 1260

AGCGCTTTCT AACTATTTTT TGAAGGAGTC AATCATAATG AGTGaGAAAA TAATGAATAC 1320

AAyCCAGAAT AAAGTTTTAC CTATAGCTAC TAAGATTGGG AATCAACGAT TTTTAGTAGC 1380

GTTAAGAGAT TCATTTATGG GTACGATGCC TGTAATAATG ACaGGtTCGG TTGCAATATT 1440

ATTAAATGCT TTTTTAGTAG ATTTTCCTAT GCAATTTGGC TATGAGAAGA TTACTGATTA 1500

TTTTCaATGG TTAGTAGATA TAAATAATTT GaTATCaAAA GGAAGCATTT cTATTGTTTC 1560

CTTACTGTTT ATTTATTGTT TAGGAGTAAA TATTGCTAAA ATATATAAAA CAGATACTTT 1620

GTCTTCAGGA CTAGTCGCCC TGTCTTCATT TATTATTTCT ATAAGTAATT CTATGACTTC 1680

AACTTATAAT TTATCGAATA ATAATAATAT TGACTTAACA ACACTATTTA CTGATGTTGA 1740

GGGAATAACA GTTACAGGTA ACAGTTTAAA TGTGACTATA TCAGGATTGT TGCCTGGTAC 1800

ACAAATTAAT TCAAATGGTT ATTTTACCGC AATAATCATA GGTTTTTTAT CTTCAATTAT 1860

TTTTTGTAAA CTCATGAATA AAAATTGGAC AATTAAATTA CCTGATACAG TTCCTCCAGC 1920

TATCGCTAAA CCATTTTTAT CAATTATTCC GGCTTTAGTT TCTCTTTATA TAATTGCTAT 1980

TTTGACATTT TTGTTAAATA GAATTACAGG AAAAATTTTA ACAGATGTAA TTTACGAAAT 2040

ATTACAACGC CCTATGTTAG GACTTTCCCA AAGTTTTTTT GTTGTTATAT TAGTTGCTTT 2100

CTTAGTTCAA TTTTTCTGGT TTTTCGGTAT TCATGGTGGA AATGTTATGG CACCTATTAT 2160

GGAAGGTGTA TTTGGAGTAG CGCTTTTAGT TAATTTAGAA GCTTATCAAA ACAATGAAAC 2220

TATTCCTTAT GTTTGGACAA GTGTATCATA CGGATCGTTT GTCTGGTATG CAACTTTAGG 2280

TTTATTAATT GCGATTTTTT GGGTATCAAA AAATAGTCAT TATAAAGAAG TTGCAAAACT 2340

TGGAATTATG CCTGTTTTGT TTAATATTGG TGAACCAGTA ATGTATGGAT tACCTACTGT 2400

ATTAAATCCT ATTTtATTTA TTCCATTkkT GTtATGTCCT GCTGTAATGt CTAGTGTGGC 2460

TTATTTAGTA ACAGATTTTA AGTTGGGTTT CTCCAGTAAC ACAAAATGTA ACTTGGGTAA 2520

TACCACCGGT CTTATATGGA TTCTTTTCGA CAGGATTTGA TTGGAGATCA ATTATTTTAT 2580

CAATTATTAA TTTATTTTTA ACCACGTTAA TTTATTTGCC ATTTGTTAGA ATGGCAAATA 2640

AGAAAATTGG ATAATTTTTA AGTAAGTCGT AGTTAGTGTG TATTAGAAGT ATGATAAGCC 2700

TCATTTTTGA TATATTAAGG AAGGrGAAGA CaTGCAAaCG TTTGAGATAA ArGArGAATT 2760

TTTGGTAGAT GGAAAACCAA CTAAATTGAT TAGTGGCgCA ATCCATTATT TTCGCATGAC 2820

ACCGGCACAA TGGGAAGATA GTTTATACAA TTTGAAAGCA CTGGGTGCGA ATACCATAGA 2880

GACTTATATT CCTTGGAATC TGCATGAGCC AGTGGAAGGC GTCTATGACT TTGAGGGCAT 2940

GAAAGACATC GTTGCTTTTG TGTCATTGGC GCAAGAACTT GGCTTGATGG TGATTTTGCG 3000

GCCATCTGTC TATATTTGTG CGGAATGGGA ATTCGGCGGG CTACCTGCTT GGTTGCTGAA 3060

AGAGCACGTG CGATTGCGTT CAACTGATCC ACGATTTATT GCAAAAGTCC GTACTTATTT 3120

CAGTGTGCTT TTACCAAAAT TAGTTCCTTT GCAAGTGACA CATGGCGGAC CAGTCATCAT 3180

GATGCAAGTC GAAAATGAAT ATGGTTCTTA TGGCATGGAA AAAGAGTATT TGCGGCAAAC 3240

AAAACAGGTG ATGGAAGAAT TTGGCATCGA TGTTCCGTTA TTTACGTCAG ATGGTGCTTG 3300

GGAAGAAGTG CTGGATGTGG GCACACTGAT TGAAGAGGAT GTTTTCGTGA CAGGAAATTT 3360

CGGCAGTCAT TCCAAAGAGA ACGCTACAGT GATGAAAGCA TTTATGGCAA AACACGACAA 3420
AAAGTGGCCG ATTATGTGCA TGGAATACTG GGACGGTTGG TTCAATCGTT GGGGAGAACC 3480
GATTATCAAG CGTGATGGCC AAGATTTAGC TAATGAAGTC AAGGATATGC TAGCCTTAGG 3540
TTCATTGAAT CTCTATATGT TCCATGGCGG AACCAATTTT GGTTTTTATA ATGGTTGTTC 3600
TGCTCGTGGT GTGTTGGATT TGCCACAAGT GACTAGCTAT GACTACGATG CGCTATTGAC 3660
GGAAGCTGGC GAGCCGACTG AGAAATACTT CCATGTACAA AGAGCCATTA AAGAAGTATG 3720
TCCTGAAGTG TGGCAAGCCG AGCCCCGCAG GAAAACATTT GGTTCTTTAG GAACATTCCC 3780
TGTGCAAAAC AGCGTCTCAC TGTTGGCGGT AAAAGACCAG ATGATGACGG CmCAAGAAAC 3840
GATGTATCCA ATAACGATGG AAGAAGCAGA ATCTGGTTAT GGGTATATGC TTTATTCAGT 3900
AAACTTAAAA AATTATCATC ACGAAAATAA GCTCAAAGTT GTGGAGGCCA GTGACCGATT 3960
GCATCTTTTT GCAGACGGTA GCCTGCAGAC AATTCAATAT CAAGAAAACC TTAGGGAAGA 4020
AGTGATGATC AAGGGTACGC CAGAGaAAGa ATGGATTGaA CtAGATGTTT TAgTAGAGAA 4080
CTTAGGCCgA GTGAAcTATG GCTTTAAATT GAATGGGCCG ACACAAGTGA AAGGCATTCG 4140
TGGGGGAATC ATGCaAGACA TCCcATTTTC aTCaAGGCTA TCgGCAATAT GCaTTGaCAT 4200
TGTCAGCAGA TCAACTCaAA AAAATCGATT ATACTGCCGG TAAAAATCCg GCCCAGCCTT 4260
CTTTTTATCA AGCTGAATTC ACATTGACTG ACCTCGCGGA TACATTTATC GATTGCCGCA 4320
GCTATGGcAA AGGGGTGGTC ATCGTCAACG GTATCAATTT AGGTCGTTAC TGGCAAAGAG 4380
GACCGATTCA TTCCTTGTAT TGTCCGAAAG AGTTTTTGAA GAAAGGCACA AATGAGATCG 4440
TGATTTTTGA AACGGAAGGG ATAGAAATTA ATGAGCTCAT CTTTTGTGGA CAACCGATAG 4500
TCAAGAAACT TTTGACTAAT GATTTTTCTG AAATAGGTTC AAATATCCAT AACCATATAG 4560
ACAGCTAGTT TAAATTTATT GCAGCAAACC AATCATTGTA TGTGTCGCGA TTTTTCCAAA 4620
TCTCAATACT TATTCTAACG CATCACAAAT ATTTATCTCA ACAAGTCTCG AATGATTTTT 4680
ACTTTTCATT TATGGTTCAA GCAGTGAATA ATTTACACTT TCTTTTGAGA TTGACTTATA 4740
TTCGCGATAA ACATCCTAAA TGATAGAGAT AATTTAAGCA TTTGATCGTT GATAAAAGGT 4800
ATAATTATAG ATTCATCACA GATACATTGC TTTTTTCGAA GAAACTATAA AAAATGAGAG 4860
TCTAGGTTAT TTGAGAATAT CAATTGAGCA AGTACAAATA ACTCCAAATT TAGTAATATT 4920
CCTAATTGAT TATTGGTGAG TAATTAAAAA GGAATACTCA AGCTACATGT ATAGCATAGA 4980
AAATCAGCAT TCAATCAGAC AATTAGGATT GAATGCTGAC TTTTTTGTTA TATTGATATT 5040
TTCAATTATC TTCTTTATTT TCTTCTAGTT TTTCGAAAAA TAGGTATTCT CTGATCGTTG 5100
TATTAGTGGA AAGGGGATAA AGGAGCATAT TGATCGTGCA ACAAAACAAC AATGAGATAT 5160
CTACACAACA CCAATTTGAT GCCTACTGTA AAAAAGTATT ACGCAATGAA GCTAAGTCTA 5220
TTCGAAAACG AAACGATAAA ATAAAAGAAA ATGAGGAACC ATTAAATGAT TTAAATGAAG 5280
GTAAGTATTC ACAAACTCAT TTAGATGAGC AAGATGTTTA TTTCTTATTT GGAATGGAGA 5340

TATTAATTTC AGATCAGAAG TTATCAGCAG CAATTGATCA GTTATCGGAT ACAAGGAAAA 5400

AGATTGTTTT ACTTTATTAT TTTGCAGGTT TTAATGATAC TGAGATCGGT AAAATTTTCA 5460

ACATGAGTAC AAGTGGTATC TGGTATCAAC GAACGAAAGC GGTTGAACAA TTGAAAATGG 5520

AATATGGCTT ATGGTAGACA AACAAAGGAA CCAACATTAC CCATTAGTTC CTATATCTAT 5580

TATCTTATCT GCTTTGAAAG AAGACCCAAT TTCAATCAAA ATGATACTTA AACATTACCA 5640

CGGTTATATA ACTAAGTTAT GTTTAAAAAA TGGtTTTAAT GaGGTGGGGC AATTTaTTAC 5700

CTATGTAGAT GAATaTrTGC TAAGACAGCT AGAAATTAAA TTGATAGAGG CTATATTGAA 5760

ATTTAAAGTT AACTAATTAA AAAGTCCCCT TTCCACTTTT GCTATCTAAA AGAATCTTCT 5820

AATATTTGAG AGATATGTTT TTTAGTTAGA AAACCTATTG ATCATTGAAA ATTAAATAAA 5880

TTCTATCAGT ACGTTAACAC TTCAAGGAGC aTATTTTAGT TGCGCCaTGA TCTCTTATTG 5940

TAAAGAAGGA GCGATTAAAA TTCTGAAATA GTTTTTATAG TCAATAAAAA GCGATGATTT 6000

GAAGTTGAAA TAATGATACT TCCGTAAGGC TCGCCGTAAA CTTGGGGAGA GGTGAAATTC 6060

CTATGTACTT TTGACAAAAG TGGCTGATAG ATTAATTATT GTTAATAGAA AGGAAGAGAC 6120

AATCATGTCG AAGAATAACA ATAAAAGTAA TTCATCAGAG TTAGAAAAAT ATTTGTTTCA 6180

TCCAGAAAAG ATAGAGGATT CTAATAATAA AATGAAGAGG ATAAATCAAC TACCATTAGT 6240

TTTAACGAAT AAAGATTTAA AAGAACAATT TCAAATAAG 6279

(2) INFORMATION FOR SEQ ID NO: 324:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5330 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 324:

GCGTTCATGG AAnCCAGAAG AGACTTGtTT CAGCAACAGA TAAAACAGGT CAAGATGTCC 60

CATTCGArAA AATCACTGTT TCAGGTCAAG TTGATACTAG CAAAGCAGGC GTTTATCCAA 120

TCATATACAG TTACGAAGGT AAAGAAGAAA CAGCCCaTGT GGCCGTCAAA CCCGACCAAT 180

CTAAGTTAGA GGTTAAAGAT ACAACGATTT ATGTTGGTGA TTCGTGGAAA CCAGAAGATA 240

ATTTCGTTTC AGCGACAGAT CGTGATGGTC ATGCTATATC CTTTGATAAG GTTCAAGTAA 300

AAGGGGAAGT TGATACAAAG AAAACAGGAG AGTACCAGAT TTCCTATACC ACTGAGCCAG 360

TTAATGAAAC CAAACCAGCT GTGCAGTCAC GACTTTTCTC AATGTTTTCA AACGAAACAC 420

CTCGTCAATT AACGACAGTA GCAACTGTTC ATGTGATTGA TCGTAATCCT ACACCGCTAC 480

CAGATAAAAA TGAGAACAAC CAGACTAGTT CTTCAACGAA TCAAACCACG ATAAAAAGTA 540

GTCAGTATGT CACACACATT GTGAAACCTG ACAAGCAAGG ACGTtATCCC AAAACGGGAG 600

aACAAACTAA TGGTCTATAT CGTGTGCTAG GACTTGTTGT ATTACTTATT GTGATTATTA 660

GTGGTATAGT CATCAAGAAG AAACGAAAAT AATTAAATGA GCGTGGGGCA AAAGTTTAAA 720

GTAACTTTTG TTCCACGCTT AAAAGATGAA TAAGTGCATG AATGAGGGTA GCTGTTTCCT 780

TGGAAAATAG GAAGTTTTAC AAGTGACTTT TATTTTAAGA ACACCTATTA GCAGGTTTTT 840

CTGTTAAGAT AGTTATTTTT ATGAAATAGT TTGGAGGAAT AATAAACGAG ATGAACCAGC 900

GAACTATTTA TTATGGTAGT ATCGGCACAG CCTTATTTCT TTCTTCAGGA ATGTTGATTG 960

CTTCAGTGAA TAGCTATGCA AACGAAGCGA CAACACCAGG CAAATACACC GTGACTTATG 1020

AAGTGGATGG CAAAATTGAA ACAGCAACCA TTACGGTCAA AGAAAATCAA ACGGAACTTA 1080

AATTAAAAGA CACGTTTTTA GGAGTTAATC AAGCATGGTC ACCATTAGAC AATGTCGTTT 1140

TTCTAAAAGA TAAAGAGGGT CGCCCTTTAA ATTTAGAAAA TGTAATTATA GAAAATAATG 1200

TAAATACTCA AAAAGTAGGC CTTTATTTTG TAACCTTCCG TTACGAGTCA TACCTTTCTA 1260

TTGCGAAkGT GTATGTCAAC CCTACTAATT TTGCTGCaGA GGGGaATaTG cTTGCaGaGc 1320

TTCCtTATAG TCAAGCAATG ATTGGCCCAG AAAGATCaAT TATTGCCtCG AGACaGGACA 1380

AGGAGCAACA AACAAAACAG GAGAAAAAGA AAAATAAAAT TGAACAGGAA CGAGTAAATA 1440

AAAAGGTGCT GTTAAATGAT GAGTTGTTGA AAAAAATAGA TACGATAGCC CCAGTGGAAA 1500

TTACTGGAGG AGCGGCAAGT GGGTCCTCTC AATTTGGCAC CACTTTAAGC TTCCTGTCAG 1560

GCTCACTTTT GTATGGAACA AGGAGAGTTT AAATGGAATT CTTGTTATTA GATGATCGGG 1620

AAAATTTAAA ATTAGCGATT ATTAGAAAGT TAGAGCAACA ATATTCATTT TCTGAACGAA 1680

AAGATCGTTT ATGTAAGCAA TTAGCTATTT CTCCTTATTT GCTAGAACGG AACATAACAG 1740

AAATCAATGC AGATTTGCAA CGCTTTGGTC TGATTAATGA GATGGAAATC AGTGAAAAAA 1800

ATAATGAAAT TTTGTTATCT CAAAGCCTTC GTATTTCTTC AAGTATTATT GAAGAATATT 1860

ATTTAAAAGA CTCGTTAGAA TTTAATCTAT TAAAAACGAT TTTCTTTCAT CAGTTTACCT 1920

CAATCAAAAA ATACGGTGAA AAGCACGGGA TGAGTCGAAC GGTTGTATAT AAAATTGTTG 1980

ACCATATCCG GAAAGAATTA GGTCAATACG GTATTAAGTT ATCCAAAAAT TTACAACTTA 2040

GCGGCAATGA AATGGCGATT AGACAATATT TTACTATGCT TTATTACCGA ATTTACAAAG 2100

ATTCTGAAGA ATTATATAAT CAAACAGATC TTCsCsCAGT GAACCAATTA TTGGCTCAAT 2160

TGAAAGGCTC TTATGAAAAT ATCACTAATT TTCATTTATT TAAACATTAT GTATTGGTGG 2220

CGTTAGAACG AACGCAAAGA AAAGCTAATT ATTTTTTATC CCAAGAAGAA AATCCATTCG 2280

CTTTTGATGA AGAAAGTTCG ATTtATCAAG AAATTCAAAG TTGGATCAAC GAAGgAnTGA 2340

AGGCGACTCA TGCAGAGAAA AACGCAGAAA TTCAAGgAaT TATCGGCAAT TTAgGTCTAT 2400

CAtGAACTTA TTTCGacATT TGCTGAGTcT CACAATGAGG CAATTACGGC TACGAAAACA 2460

TTATTCTTTT CTTATATGCC GTTTACAATT TCGGACGAAG AATTTTATCA AGAAATTGTT 2520

CCGATTATTT ATCAACATCG TTTTATTACG CCGTTCATCG ATATCACCTT ACGAATTATG 2580

GATTTAGAAT TTTTTCAGGA ACGCTACCCA ATTGTTTTTA ATAGCTGTCG TCAGTTTTTA 2640

TTTGCGTTGG ATTGCTCGGC GTTTGAGTTT AGTAAGCTTT CTTTATTTTT TGATTTGCTA 2700

TTAGTGTTAT CCCGTCTTTA TGACCAAAGG AATGAAAAAA GCACGATCAA TCTTTATGTG 2760

AATTTTACGC AAGGAGAAAA ATATACTCAG TTTATCAAAG AACAAATCAA AATATTCGAA 2820

TCTTTCAGCA TTCACTTTCA TTCAGCGATT CGACCAGATA CTGATTTAGT AGTTTCGGAT 2880

TATTTACCAA AAACGTTATT TTCAGTAAAa TGTTTAATTT GGCTAGCGCC CCCGCGAGCC 2940

AGTGACTGGC AAAATTTTGG CAACGAAATT GTCCGAATTA ATAAAGAGCT ACAACAAACG 3000

AAACAAAGAA AAAGCGAATA AAAACAGTCG CAACGTGTTT TAGTTGCGAC TGTTTTTTGT 3060

TACTGAATGT TTATAGTAAA CTTTCTCATT ATCACTGATG TAAAAAATTT TAAATAAAAG 3120

CCAAATAGGA TTGAAGAAAA AGAAAAAATT GCTAGTGAAG GACCaATTCT CTAAAGTTTT 3180

TCTCAAATAT AGTAGAATTC CTTTCTTTTC TTCTGTGAGA AAGATATAAT AAAAAACGTA 3240

TGAACAAAGG AAAAGGGATG ACGAATGAAT AGAAAAGAGA AAACAAATTT AGATAGTCGA 3300

ATTGATTACG GCGTGATATT ACCAGTTTTT CTTTTATCAC TAATTGGTAT GCTATCGCTT 3360

TACGTAGCGC TTTACAATGA TCCAAGCAAA CCTAAAATTG GTAGCTTGTT AATGAAACAA 3420

GGACTGTGGT ATTTAGTCGG AGGATTAAGT ATTGTCATTA TTATGCATTT TAGCTCAAAA 3480

TTACTCTGGC GGTTAACGCC CGTTTTTTAT GCACTTGGAT TAGTATTAAT GGGATTAmTA 3540

CTGAAATTCT ATGACCCCGT GCTAGCAGAG CAAACAGGTT CTAAAAACTG GATTCGCTTT 3600

GGTGGCACCA CATTCCAACC ATCGGAATTG ATGAAGATTG CGTTTATTTT GATGTTGGCC 3660

TATATAGTGA CCATGCATAA TGTTAAATAT GTGGACCGAA CACTCAAAAG CGACTTTTGG 3720

CTAATCGCTA AAATGCTATT AGTGGCGATT CCGGTAATTG TTTTAGTACT TTTACAAAAA 3780

GACTTCGGGA CGATGTTAGT TTTCTTAGCC ATTTTTGGGG GAGTTTTCTT AATGTCTGGC 3840

ATTACTTGGA AAATTATTGT GCCAGTTTTT ATTTTAGCTG CTCTAGTGGG GGCAGGAACT 3900

ATTTACTTGA TTACAACAGA AACAGGTAGA GATTTATTGT CTAAGTTAGG GGTCGAAGCC 3960

TATAAATTTG ATCGGATTGA TTTATGGCTA AATCCTTTCC ATACAGATCC AGACCGCTCA 4020

TTCCAACCAG CGCTAGCATT AACAGCAATT GGTTCTGGGG GCCTTTTCGG AAAAGGCTTT 4080

AACGTGAGTG ATGTGTATGT CCCTGTTCGT GAGTCAGATA TGATTTTTAC AGTTGTTGGT 4140

GAAAACTTTG GCTTTATTGG TGGCTGTTTC ATTATTTTGT TGTATTTTAT TTTAATCTAC 4200

CGCATGATTC GCGTTTGTTT TGATACAAAT AATGAATTCT ATGCCTACAT TGCCACAGGA 4260

ATTATCATGA TGATTTTGTT CCATGTTTTT GAAAACATTG GAGCGAATAT TGGTTTACTA 4320

CCATTAACTG GGATTCCATT ACCGTTTATT AGTCAAGGTG GTTCTTCCAT TTTAGGTAAC 4380

ATGATTGGGG TGGGCTTAAT TATGTCCATG CGTTACCAAC AAGAAACGGT TCGTACCAGA 4440

AGTGGTCGTT AAAGAAAGAA GAAAGAAGGG GTAGAAATGT ACACACTTTA TGAATATCCT 4500

AAATGCTCAA CTTGTAAAAA AGCGAAAnCG wGGTTAGACC AACAAGGCGT TAAATACCAA 4560

GCAATTGATA TTAAAGCAAC ACCGCCTAGT AGTGAACAAT TAGCAAAATG GATGAAAGAG 4620

ACCGGCTTGC CTGTTCGTCG TTTTTTTAAC ACAAGCGGGG TTTTATATAG AGAACAAGGC 4680

TTAAAAGATC TTGTTGACAG TTTTTCTATT GAAGAAGCAA GTCAGCGTTT AGCTGCGGAT 4740

GGCATGTTGA TTAAGCGTCC AATTCTCTTA AAAGACAATA CATTTCTAAC GAATGGTTTT 4800

AAAGAAGCAG ATTACGAAGG AGTGCTTGGA AAATGAGTGA AATGAAACAA ATCGATCAAT 4860

TATGGATCTT ACCAACAGAA GCAGGTTATA AAATTGGTTT AACCGCCGAA GCCCAAGAGG 4920

ATTTAGGGAA GATTACTTTT GCGACATTTC CCAAAGTCGG ACAAACATTA GCTAAAGGTG 4980

ATTCTTTAAT TGAATTAGAA GCAGAAAAAG CAGTCAGTGA ATTCAGTAGT CCATTATCTG 5040

GTAAGGTAGT TGCGATTAAT GAAGCGGCTG ATCAAGAGCC AAGTGCATTA GATGAAGAAA 5100

ATGCCTGGAT TGCTGTTTTA GCAGATGTTG AACCAACAGA ATTTGAAAAT CTTTAGTCTT 5160

ACGTTGACAA GGCCAATGaT TATTGcTATA ATTTTCTCAA TCTAATCAAA AGAAAGCTTT 5220

GaAAAGAAGA GTACATGTTa AGGACGTTTG AAGAGAGCCC kGTTTGCTGA GAaTGGGtAA 5280

CAGATTTaAG ATGGrAGATG GtCyTTGrGc aGAGTAAGTG AGCGAaGCAG 5330

(2) INFORMATION FOR SEQ ID NO: 325:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 547 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 325:

TGACAGrGaC AATTkGTkGA AgCTTTCAGG AATATTGGCT ACACTTGCTC CGAATAATTG 60

TAAAAAAGTA TCGTTAAAAA TAACACCTTT TAACCCTAGG AACAATCCTG TTGCTGCAAT 120

TACTGGAACA ATTGGAATGA ATATTTCACT TAAAATGCGC ATTGTTTTGC GCATGCCTGT 180

TTCATTATTT TTTAAAATTG CTGTTTGCTC TGCTTTTGaC ACAACATTGa TTCCAAGTTT 240

TTCTATTTCA TCATACACTT TAGTCACAAT ACCAGTTCCT AAAATCACTT GGTATTGACC 300

ACCATTATAA AAAACCCCCT TGACTAAATC GATTTTTTCT AGTTGTTTAC CATCTACTTG 360

TTCATGGTCC TTCAAAATAA AACGTAACCG CGTTTGACAA tGGGTGACCG ATATAATGTT 420

ATCCATACCG ATTAATTGAA TAATGTCTTT GGCTGCCTGC TGATAATTTT CTGCCATGTk 480

ATACTCCTCC TAGCTACTCA TTGATAGTAG nCAGTAGTCC TTGATAGCTn TTGAATAATG 540

nCCCnGC 547

(2) INFORMATION FOR SEQ ID NO: 326:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5354 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 326:

CCTTGTTCAT TTGACACCGA CGTATTTAAG TCGTTTATTA AAAGAGGAGT TTGGTATTCC 60

```
CTTTAGTCAA TTTTATCGAC AATTAAAAAT CACATGGGCT AAACAGCTTC TTTTGGAGAC    120
AGATAAAACA ATTGCCCAAA TCAGTGAAGA TTTAGGGTTT GTTGAAGATA GCTATTTTGT    180
CCGGATCTTT AAACAAGAAA CGGGTGAAAC ACCCCTTAAA TACCGTAAAA GAAAGAACAG    240
AGCATAAAAA AACAACGACT CCTAGAAGTC GTTGTTTTTT TACACCTTTT GCTCTTGCTG    300
TGGCTAATGC AACAAGTGCA AAAGGCGCAT TATATCTGTC GTGCGACAAA GTGATACTAT    360
TAAATATGAA TTGGTAAACC TAAAGCCAAT TCAGCTGTAT CCATAGTAAT TTCGCCCAAT    420
GATGGGTGTG GGTGGATTGT TAAAGCAATG TCTTCTGCAT TCATGCCAGA TTCAATAGCT    480
AAAGCTAATT CAGAAATCAT GTCACTTGCA CCGACACCGG CAATTTGTGC ACCGATGATG    540
ACATTGTCTT CTACAGTTGT AACTAAACGC ATGAATCCTT CAGTTTTATC TAAAGAGATT    600
GCACGGCCGT TACCAGCAAA TGGGAATTTG TAGCCTTTTG CTTCGATTCC CGCTTCTTTT    660
GCTTCTGCAA CAGTCATACC AACGCTTGCC AATTCTGGAT CAGTAAAGGC AACAGCTGGC    720
ATTGCTTTGT AATCAACTGC AACTTTCTTA CCAGAAATTG CTTCAGCAGC AATTTTTGCT    780
TCGTAGCTTG CTTTATGCGC TAACGCAGCA CCTGGTACGA TGTCGCCGAT TGCGAAGATG    840
TTTTTCACGT TAGTACGTCC TTGGTTGTCA ACTGGGATTA AACCACGTTC GCCAATTTCA    900
ACGCCCGCTT GTTCTAAGCC TAAGTCGTCT GTGTTTGGAC GACGTCCAAC AGTGACCATT    960
ACGTAATCTG CTTCAACACT TTCTTCTTTT CCGTTAACTT CATATTTAAC AGTGACGCTA   1020
TCGCCATTGT CAACAGCTTC TTTAGCCATT GCAGAAGTCA CGATTGTTAC GTTTTTCTTC   1080
TTGAAGTGGT CTGTGACAAC TTTAACCATA TCTTTTTCAT AAGTTGGTAA AATTGATGGG   1140
CTACCTTCTA AAATTGTTAC TTCTGAACCT AAGTTAGCAT AAGCGCCACC TAATTCAGCA   1200
CCGATGACAC CGCCACCGAT AATAACGAAT TTTTTAGGAA CTTCTTTTAA GTTTAAACCG   1260
CCTGTAGAAT CTAACACGCG TCCGCCAAAT TTGAATCCTG GGATTTCAAT TGGACGAGAA   1320
CCTGTTGCTA CAATAGCATT ATTGAATGAG TAAGTTTGTG CTGAGTCTGG GTGAATAACA   1380
CGCAATGTAT TTTCGTCAAC GAAGAATGcT TCGCCTTCAA TAATTTCTAC TTTGTGTTTT   1440
TTCAATAACA TGCCAACGCC GCCTGTTAAT GATTTTACAA CTGTGTTATC TTTCCAGTCT   1500
TGTGTTTTTG CAAAGTCTAA TTCGACTCCT TTAGCTGTTA CACCAAAAGT TGAAGAATCT   1560
TGTGCTTCTT GGTAATGATG TCCAGCAGCA ATTAAAGCTT TTGAAGGAAT ACATCCAACG   1620
TTTAAACAAA CGCCTCCGAT GTATTCACGT TCGATAATCG CAACTTTTTG ACCCATTTCT   1680
GCGGCACGAA TTGCGGCAAC GTATCCTCCA GGACCAGCTC CGATTACGAC TGTATCTAGT   1740
TCAATGGCGA AATCTCCTAC TACCATTTGT TAATCATCCT TCCATTAATA GTAATTCTGG   1800
ATCAGCTAAT AAGCGTTTAA TGTTGTTCAT TGCTTTTTGA GCAGTTGCGC CGTCAACGAT   1860
ACGGTGGTCA AAGCTTAATG ATAATTTCAT CATGCGTCCC ACAACGATTT CGCCGTCTGC   1920
ATTAACAACT GGTTCTTGTG CAATTGTACC AACGCCTAAA ATAGCAACTT CAGGGTAGTT   1980
GATTACTGGT GTAAACCAGC CGCCACCGAC TGAACCAATG TTACTAATTG TGATTGTACC   2040
```

```
ATCACGCATA TCTTGTGCAG TTAATTTACC TTCGATAGCC AATGCTGCTT TTTCGTTGAT    2100
TTCATCAGCG ATAGCAAACA TGCTCTTCGT ATTAGCATTT TTAACATTTG GTACATATAA    2160
GCCATGATCT GTATCAGTAG CGATACCAAT GTTAAAGTAA TTTTTGTAAA CAATTTCTTG    2220
TGCTGCGTCA TCGATTGATG CATTCAAGAT TGGGAATTTT TGAACAGTTG AAGTCAATGC    2280
TTTTACAACG TATGGTAAGA ATGTTAATTT TGTACCATTT GCAGCAGCAA CATCTTTAAA    2340
TTTCTTACGG TGATCCCATA ATTTAGAAAC TTCTACTTCA TCATGTAATG TTACGTGAGG    2400
AGCAGTGTGT TTGCTGTTAA CCATTGCTTT AGCAATTGCT TTACGTGTTG GTGTCATTTT    2460
TTCACGTGTT TCCATTTCGC CTAAATCAGA AGTAAAGGCT TTTGGCGCTG CTTTAGGTGC    2520
AGCTGCTTCC GCTTTAGGTG CTGCTTCTGT TGCGGCAGCT TCAGTAGCTG GGGCTGCTTG    2580
AGAACCACCA GAGACAAAGG CATCAATATC CGCTTTAATG ACACGGCCAC CTTTACCAGT    2640
TGCAGTTACT TGTGTAATAT CAACGTCTTT TTCACGCGCA TACTGACGAA CAGATGGCAT    2700
TGCTAAAACG CGTTTGTTTG GATCAGCGGC TGCAACTACG CCTGTTGAAG CGGCTGGAGC    2760
TGATGCTTCC GCTTTAGGAG CGTCAGTAGC TGGTGCTGCG ACTGACGGTG CTGCTGAATT    2820
ATGACCAGGT GCGTCAATTT CAACTAACAC GTCACCCACA TTGGCAACTG TTCCTTCTGG    2880
TACAACGATA TTTTTTACAG TACCTGTTAC TGGTGATGGA ATTTCTTCTA CTGATTTGTC    2940
ATTTTGkACT TCTAATAATG AATCATCTTC ATwAATTGTG tCGCCCGCTT TAACGAACCA    3000
TTTAACGATt TnCnTTCGGC AATTCCTTCA CCGATGTCTG GTAATTgAAT TGGAAAAATC    3060
CGCCAGCTGC TTCGGTTGTT GGTACAGCAG CAGGTTGTGC TGGTGTTTGT TCTTGTGCAG    3120
GAGCTGCTGG TGCTGCATCG TTATCTTCGT GACCAGGTGC gTCGATTtCG AyTAACACGT    3180
CACCAACGTT TGCAACTGTT CCTTCTGGTA CAACGATATT TTTTACAGTA CCTGTTACTG    3240
GTGATGGAAT TTCTTCCACT GATTTGTCAT TTTGTACTTC TAATAACGTA TCGTCTTCGT    3300
TGATTGTATC GCCAGGTTTT ACAAACCATT TAACGATTtT CCTTCGGCAA TCCCTTCACC    3360
GATATCCGGT AATTTAAACT GATAAGCCAT TTTTTAAGTC TTCCCTTCTT CGTTGGATCT    3420
GTCTATTTTC TTAGGAGCAA AACGCGGGGA AACGACTATA GTAGCTGAAC ATTTCATGCC    3480
TATAGCCGTA ATCCACCTGT TTAATTCTTC GTTTATCTAC TTCGAATAAC TATTAAAATT    3540
CGACGATTTC TCTAGCTTTT GCTTCGATAT CTTTCGCATT TGGTAACCAG ATATTTTCTG    3600
CTTGTCCGAA TGGGAAGATT GTATCTGGAG CAGATACACG TCCGATTGGT GCTTCTAATG    3660
ATAATACGGC ACGTTCAGAA ATTTCAGAAA CAACCATAGC GCCAACGCCA GCTTGTTTTT    3720
GTGCTTCTTG AaCGACAACC ACACGGCCAG TTTTTTCAAC AGAGTTAATA ATTGTTTCCA    3780
CATCTAAAGG AGCCACTGTA CGTAAGTCAA TGATTTCTGC TGAAATATTG TCTTTCGCTA    3840
AGCTATCAGC TGCTTTAATC GCTTCACGAA CCATAGCACC GTAAGTGATG ATTGATACGT    3900
CTGTTCCTTC ACGAGTTACA GCCGCTTTAT CTAAAGGCAC TTCATAAGCT TCGTCTGGCA    3960
CTTCCTCACG GAATGAACGG TATAATTTCA TGTGCTCTAA GTAAACAACT GGATCGTTGC    4020
```

TACGAATAGA TGAAATTAAT AGTCCTTTTG CATCGTAAGG GTTTGATGGA ATAACAACAC 4080

GAACACCTGG TGATTGTGCG ATTAATCCTT CTAAGTTATC TGAGTGAAGT TCTGGTGTAT 4140

GAACACCACC ACCAAATGGG GCACGAACAG TAATTGGCAT ATTACGAGTT CCACCCATAC 4200

GGTAACGCGT ACGAGCCATT TGACCAACGA TTTCGTCAAA TACTTCAAAA ACGAAACCAA 4260

AGAATTGGAT TTCAGGAACT GGACGGTAAc CTTGCAAGGC AAGACCGAAA GCCAATCCAC 4320

CGATACCAGA TTCAGCTAAA GGTGTATCGA AGACGCGGTC TTCACCAAAT TTTTCTTGTA 4380

AACCTTCAGT tGCACGGAAA ACCCCACCGT TGTTACCAAC GTCTTCACCG AAGATTAAGA 4440

CATTTTCGTC TTTCTCTAAT TCAAGAGCTA AGGCATCTGT AATTGCTTGG ATCATAGTTT 4500

TTTGTGCCAT TTTTATTTCG ACTCCTTCGC TTCATAAAAT GCAATTTGTT CTTTAATTGT 4560

TTGAGGTTGA ACTTCAAACA TATTTTTCAA GAAATCAGAA ACTTTTTGTT TTGGCGCTTT 4620

ATCCGCTTCT GCAATCGCTA CTTTGATTTC TTCTTTTGTT TTTTCAATAA TTTCTTCTTC 4680

TTTTGCTTCA GACCATAAGC CTTTATCTGT TAGATATTTA CGGAAACGAG TCAATGGATC 4740

TTTTTGTACC CATTCGTCAT CCATTTCTTT TGAACGGTAA CGTGTTGGAT CGTCTCCAGA 4800

TAAAGTATGT GGACCATAAC GATAAGTTAA TGTTTCAATT AAAACTGGAC CGTTTCCTGC 4860

AGCTGACCAA TCGCGTGCTT CTTTTGCAAT TGCGTAAACT GCTAATGGAT CCATACCATC 4920

AACTTGAATA CCAGGAATTC CTGCTGCAAC AGCTTTTTGA GCTAAAGTTT TAGCCGCTGT 4980

TTGTTTTTCA CGAGGTGTAG AAATCGCAAA ACCATTGTTT TGGATAATGA AGACACCGTT 5040

TGCTTGGTAA GCACCAGCAA AGTTAATTGC TTCATAGAAG TCCCCTTGTG AAGAACCGCC 5100

GTCACCAGTA TAAGTGAAGA CAACATTTTC TTTTCCACGT TTTTTCAATC CTAAAGCAAC 5160

ACCAGCTGCT TGGATGTATT GAGCACCAAT GATAATTTGT GGTGGTAATG CATTTAAATC 5220

TTCCGCGTAG TAGTTCCCTG cTACGTGACC ACGAGACCAT AAGAAAGCTT CTCTTAATGG 5280

TAAACCATGT TGTACTAATT GAGGTACATC cACGGTAACC TGGTAGTAAG TAGTCTTCTT 5340

TTTCCATTGC AAAT 5354

(2) INFORMATION FOR SEQ ID NO: 327:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6005 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 327:

CGTAATCTTT TAAAATGACA AGTTTAGAAC AGAGAGATGA GTGCGGACAC TTATTTTTCT 60

GTTCTTTTTC TTTCTGmAAA AGTCTTTGAA ACTTATCATT TCGGAAGGTT TTCTTTTGAG 120

AATATGATAG AATGAAGCTG AGGTGAGCGA akgGnCAAAA AAAATCAATT TACCGAaGCC 180

GTTATTTTCA GAAAAAGGAA AACGGGCGGT CCTTTTATTA CACGCTTATT CTGGCAGCAG 240

```
TAATGATGTA CGGATGTTAA gCAGACGTTT AGAAAnAGAA AATTATACTG TCTATTCGCC     300
TAATTTTTCA GGACATGGCA CGCTAGTTCC TGAAGATATT TTAGATCAAA CAACGGAACA     360
ATGGTGGCAA GATACACAAG AAGCAGTCGC TTTTTTGAGA GAAAGAGGCT ACCAAGAAAT     420
TGCCATTTTT GGTTTATCAA TGGGCGGGAT TTTCAGTATG CATGCATTAA CGGAACAATT     480
AACAGGATTA ATTGGTGGTG GATTTTTCTG TTCACCGATC TACCCTGTCG AAAATAAAGT     540
CCCaGAAAaT TTTGTATTAt ACGCAGAAAA CGTGATGAAA ATCGCAGAAG TTCCTGCAGA     600
GGAACAACAA AGCCGGTTGC AATCGATTGA GCAACGAGTC AAACAACAAT TAGGAGCGAT     660
TGAAACGATT GCTAGTCAGA CGGCTGACAA ACTAAACAAA ATCCACGCAC CTCTTTTCTT     720
AGCGCAAgcT GGcAAAGATG AAATGATTGA GCCAATGGGC GTCTATCAAA CGGCCCAAGC     780
ACTAACACAA GCACGTGTGA CCTTACAATG GTACCCAAAC AGCGGACATG TCATTaCGGT     840
GAGTGGAGAA CACAAACAAT TAGAGCAAGA TGTGGTGCAA TTTTTAGCAA CACTACCTTG     900
GAATGAGGAG AAATAATGAC AAAACAAACA ATTAAAGAGC AAATTTTACA TTTCATGGAA     960
AGCCAAAAAA AGAAAAGCTT TTCAATGGAA GAAATTGCAC AAGGCTTGAA TTTAGAAAAA    1020
AGTTCAGATT TTAAAATTTT AGTGCAAACC ATTGCACAAA TGGAACGAGA AAAGTCAGTT    1080
AGTTTCAACA AAAAAGGCAA AGTCCTGTTA CCAATGAAAG ACTTATTAAT AGAAGGAACG    1140
TTTCGTGCAA ATGAACGAGG CTTTGGTTTT GTAACCATCG ATCCTGAAGA ACCAGATGTT    1200
TATATTCCGA AAGAGGCAAC GAACTTTGCA ATGGATGGAG ATACGGTTTT AATCGACGTG    1260
ATCCAACATG CGGATCCTTT TTCAGATCGC GGCGCAGAAG GTAAAGTCAA AGAAATTAAA    1320
GAGCGAGCAG TGAGCCAAGT TGTCGGAGAA TTTGTGGCAT ATAGTGAAGA AGAAATGGCA    1380
GAAATGGGAC TGTATGGCTA CATGATTCCC AAAGATAAGA AATTGAATCA GTATACTGTA    1440
TCAATTGCAC CTGAAGGGAT TAAGCCAGTA GATGGTAGCA TTGTCATTGC TGAAATTACC    1500
TATTATCCAG ATCAAGAATA TCCaACGAGT ATGGAAGGAC TAGTCAAACA AGTGATTGGT    1560
CATAAAAATG ATCCAGGAAT GGATATTTTA TCAATCGTGG TGGCTCATGG AATTCCCACA    1620
GCATTTCCTG ATGAAGTTTT GGCTGAAGCG GACCAAGTAC CAGAAACTAT TGCAGAAAGC    1680
GATTTAGTCG GTCGTCGGGA TTTACGTGAT CAGTTGATTG TGACGATTGA TGGAGAAGAT    1740
GCGAAAGATT TAGATGATGC AGTAACGGTA CAAAAGTTAG CAAATGGCAA TTTCTTTTTA    1800
GGGGTGCATA TTGCAGATGT TTCTTATTAT GTAACTGAAG GAAGCCAATT GGATATGGAA    1860
GCGTATGAAC GTGGTACAAG TGTCTATTTG ACAGACCGAG TTGTGCCAAT GATTCCGCAG    1920
CGATTATCAA ATGGGATTTG TTCGCTAAAC CCACATGTTC CACGTTTAAC CATGAGTTGT    1980
GAGATGGAAA TTACACCAGA AGGAGAAGTT ATTTCGCATG AAATTTTCCA AAGTGTAATC    2040
CAAACAACGG AGCGAATGAC CTATACAGCA GTCAATGAAA TTTTAGAAGA GCAAAAACCT    2100
GAAACGTTAG AACGCTACAA AGAACTAGTT CCCATGTTTA AAGAGATGGG TGAGCTGCAC    2160
CATATATTAG AAGAGATGCG TATGCGCCGT GGTGCGATTT CTTTTGAAGA TCGTGAAGCC    2220
```

AAAGTCCTAG TTGATGAAAA TGGTCATCCC AAAGATATTC TTTTACGCAC GCGTGGTGTG 2280

GGTGAACGAT TAATTGAATC CTTTATGTTG GCAGCGAATG AAACTGTGGC GCGTCATTAT 2340

CATGACTTAA AACTGCCATT TATTTATCGG ATTCATGAAC AACCAAAAGA AGAAAAAATG 2400

CAACGTTTCT TTGACTTTGC TGCAGTACTT GGCATTCTTG TTAAAGGAAC AAAAGAAAAC 2460

ATTTCACCGA AAGATTTACA AAAAGTTTTA GAGCAAGTTG AAAATAAGCC GGAAGAAGTG 2520

GTTATCAATA CGATGTTGCT AAGAAGTATG CAACAAGCGA AATACTCAGA AGACAACTAC 2580

GGACACTATG GGTTGGCTGC TGAATATTAT ACGCATTTCA CTTCACCAAT ACGTCGTTAT 2640

CCAGATTTAA TTGTGCATCG TTTGATTCGC AGCTATAGCC AAGATCAATC TGAAAAAAAT 2700

CAAGAAAAAT GGAACGAAGC ATTACCAGAA ATTGCCAATC ATAGTTCAAG TATGGAACGT 2760

CGCGCAGTTG ATGCAGAGCG TGAAGTGGAC GCCATGAAGA AAGCCGAATT TATGGTAGAT 2820

AAAGTGGGAG AAACGTATGA CGGGATTATC AGCTCAGTCA CAAAATTTGG TATCTTTGTG 2880

GAACTGCCTA ATACAATAGA AGGCTTGATT CACGTGAACA ACTTAAAACA AGATTATTTC 2940

CACTTTATTG AAAATCATAT GGCGTTAGTT GGTGAACGGA CGGGAATGAC TTTGAAAATT 3000

GGTCAGAAAG TTCAAATCCG TGTTGAAAAA GCTGATCCAG AAACAAGAGA AGTTGATTTT 3060

GAATTGATTT CAGCTGAGGA AGTCGCGCCA GTCGAAGGAC CAAAAGGACG TAAAAAAGGC 3120

AAAGCCAATT CTTCAACTCG TTCAAATAAT CmACGAAGAA ATAAAAAAGA TGAATCATTT 3180

GATGGTAAGA AAAAGAaAAA TAAGAAAAAA GGCAAAGGCA AAAAACAaCC TTTTTATAAA 3240

GAAGCAATGA AACAAAAAAA TAAAAAAGGG AAAAAGAAGA AGTAGAGGAG GTCGCTTATG 3300

CCAAAAGGGG AAGGAAAATT AATCGCACAA AACCGCAAAG CTAGACATGA TTATTCAATT 3360

ATCGATACGG TTGAAGCTGG GCTTGTTTTA CAAGGAACAG AAATCAAATC TATTCGAAAT 3420

GGTCGTATCA ATTTAAAAGA TGGTTTTGCA CGAATTCGAA ACGGAGAAGC CTTTTTGTAT 3480

AATGTTCATA TTAGTCCATA TGAACAAGGC AATATTTTTA ATCATGACCC ATTACGTACC 3540

AGAAAATTAT TATTACACAA AAAACAAATT AATAAGTTAA TCGGCGAAAC CAAAAATACA 3600

GGTATCACAT TAGTACCTTT GAAGGTCTAT ATTAAAGACG GCTATGCCAA AGTTTTGATT 3660

GGTTTAGCTA AAGGGAAAAA GCAGTACGAC AAGCGTGAAG ACTTGAAACG TAAGGAAGTT 3720

GACCGACAAA TCAGCCGAAC ATTAAAAAAT AACCGACGTT AAAAATGTAA ATTTATTCTT 3780

TTTCGCATAT TTGCTTGCCT TTCTTCGGAT TAGGTGTAAA ATCTGGTAAG TGTTGAGTAA 3840

ATGAATGTAA ATGATTCATT TTTTAATCGC TTCGGAAAGC AGTACCACCC TGCATTAGAT 3900

TGGTTTATGG AAAAGGTTTT CCAACCAATT TTCAACACAT GAAATTTTAA AAGTAAACTA 3960

TTTGTAAATT ATCAACAAGT GGTGGTACGA TAGCTTTGCC AGGGATTATC ACTCAGGAGT 4020

AGGATGAAAC TTCTGAAAGA TTGTTGGAAA AGAGGTGAAT TGATTTGGAA GAGAAGAAAT 4080

TACTCTTCAA TATCGGACCA ATTTGGTTTG ATGGAACCAT TGTTTTGATG GTGCTGTTGA 4140

CGTGTATCAT TGTTTTCGCC TTTGTCTATG CTTGTACGAG AAATATGAAG CTACGACCTA 4200

AAGGCAAACA AACGGTCATC GAATGGTTAG TGGATTTTAT TCGTGGAATC ATTACAGACA 4260

ATTTACCTCG TAAAGAAGTG AGCAATTTTC ATTTAATGGC GTTTACCCTC TTCATGTTTG 4320

TACTTGTTTC AAACATCTTG GGTCTCGTTA CAAAAATCGT TGTGGGCGAC GATCTCAGCG 4380

TTTGGAAAAG TCCAACCGCC GATCCAATCG TAACCTTAAC GTTAGCGATG ATGATGATTG 4440

TCTTAACACA CTTCTTTGGT ATGAAACGCT TTGGTTTTAA AGGCTATCTA GTTAATAGTT 4500

ATTTAAGACC CGTTGGTTTC TTACTACCGG TAAAATTAAT GGAAGAATTC ACGAATCTAT 4560

TAACATTAGG CTTGCGTTTG TACGGAAACA TTTTTGCCGG CGAAGTTCTA TTGGGCTTAA 4620

TTGCTGGAAC CGTAGCAAGC GTCGGATTAT GGGTTATCCC ACTAGCAATT CCGCTAGAAA 4680

TGATTTGGGT AGCATTCTCA ATCTTTATTG GATGTATCCA aGCGTTTATC TTTGTGACAC 4740

TTTCAATGGT TTATATGTCA CACAAAATTG AAACAGAAGA ATAATCAAGA AATACGAAGG 4800

ACAAAAATTA TCTTTTAGGA GGATTTTACA CAATGGAAGG ATTAAATTTT ATCGCAGCAG 4860

CAATTGCAAT TTTCGGTTCA GCTATCGGAG CCGCTATCGG TAACGGACAA GTTATTTCAA 4920

AAACAATCGA GTCAATGACT CGTCAACCAG AAATGTCTGG TCAATTAAGA ACAACAATGT 4980

TTATCGGGGT CGCTTTAATC GAAGCTGTGC CAATTTTAGG TGTCGTTGTT TCTTTATTAC 5040

TTTTATTCAG ATAAGCAATA AAATGGAGGC AGGACGCTGA TTATTATCGG GACCGTCACT 5100

GCTTTCTTCT ATCTTTAGCC TTTTCAATAG CTGGTGATAA AGCAGTTGAA GAAGCTAAAT 5160

CGATTTTTGC AGGAAAGGAA GTTGAGACTT CATGCTACTT ACTACATTGG TAGTCGGTGA 5220

AaCGGCTCCT AGTACCACAC TAGGCACGAT GATTGTTGTC AGTGGTGCAT TTTTAATCTT 5280

GATGCTCTTA TTAAAGAAAT ATGCATGGGG AGCAATCGTA GATATTTTGA CACAACGTGA 5340

AGAAAAAATT GCCAACGATT TAGACTCTGC AGAACAATCT CGCGTAgcAG CAGCGAAAAT 5400

GGAAAAAGAA CGTCAACAAC AATTACTTTC TTCCAGATCA GAAGCAGCTG AAATTATTAA 5460

AAACGCGAAA GAAAGTGGAG AACAGACTCG CCAAAAAACA TTAAAAGAAA CAACTGCGGA 5520

AGTTAcTCgc TTACGCGAAA AAGCTCGGAC AGATATTTCT CAAGAACGTG AAGAAGCATT 5580

GTCTTCAGTA AAAAATGAAG TGGCAGATCT TTCTCTTCAA ATTGCAGCCA AAATTTTAAA 5640

TAAAGAATTA ACACCAGATG CACATGAAGC ATTGATTGAT TCTTATATTG AAAGTTTAGG 5700

TAAAGCAAAT GAAACTAGAT AAATATACAG TTGGTAAACG CTACGGCAAG GCTTTGTTTG 5760

AGTTGGCAGT TGAAAAGaAC CAAGCTGaAG CAmTTTATCA AGAATTATTG ACGCTTCGcG 5820

AAGTCTATCA CCAAGTACCT GGTATTGGCG ATATTTTAAG TGATGATCGC TTAGAGCCTT 5880

ATGAAAAAGA CAGCATCATG GnAAAACTAG TAACAGGTTT TAGTGAAATG ATGCCAAAnT 5940

TTTTGCGAGT GGTTTACGAA TATCGTnGCA TGTACGATCT GCTCTTGATG ATTGAnGGAT 6000

ATGAA 6005

(2) INFORMATION FOR SEQ ID NO: 328:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 643 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 328:

TTCGTTGCAA CAGAGTTTTC ATTTTTCTTA GGAATTCCAG TAATGTTTGG AGCAAGTTTT 60

TTAAAAATCT TTAAATTCTT GGCCAAAGGA AATACGTTTG GGTCAGAAGA AATCATTATT 120

TTAATAACTG GCTCTATTGT TGCCTTCGTT GTTTCCATCA TTGCGATTAA ATTCTTATTG 180

AATTACTTGA AGAAAAACGA TTTTACAGTT TTCGGTTGGT ATCGCGTAAT TTTAGGTGCG 240

ATTTTAATTG GCTACTGGTT ATTCTCATAA AAATAAAACC GTTGCTTTCA ACAGTGAAAG 300

CAACGGTTTT ATTTTTTTAA ATCACCAAAC GTTATTCGCT CAATCCCCTG TTCTTCAAAC 360

CACTGACGAA CACGAGGATC ACAAATCGTT GCTAATTCTT TGCAGCGTAA AGTGGTATAT 420

GAAGAATTAT CTAACACAAA TTGATCCAAA TACCCAACAT CAAAATGCAT CTCAGCAATT 480

GCCGATTCCG GCATCCGTAA TAAACCAAAG TCATCTTCTA AAAAATTCTC CACTGATACG 540

CCTTCGTGaA TaTAAGATAG CGCCCCAGAC TCTAACAAAT GTGTTACCTG GAAATAACCT 600

TGTTGcGGkG GaTAwTCtTT TTcCCCCACT nATAATTnCA GGG 643

(2) INFORMATION FOR SEQ ID NO: 329:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3384 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 329:

TTTGGCAAAC TCTTCATTAT ACATGTTAAT AAGCGCATTA ATATGCGGCT GAATAATTTC 60

TTCAAATAAC GCTTCCTTAG AAGTAAAATG TTTATAAAAA GCACCCGTAG TTACGCCTGC 120

GGTCAAGCAC AGTTGCCTTA TCTTAGTTTG CTGAAATCCT TCTTTTTCAA AATGGCTTAG 180

TCCACTTTCT ATTAATTTTT GGTGAGTTGT AATATAATCT CTGGTTTCTA TGGGTACTTC 240

CTCCTAATAA GATGAATTAA CTATATTATC TTACACAGGA CTTGAAGTTT CAATCGTTAA 300

CGTCTATTAC TGAACTCATT CCTTAGAGCT AAACTAATGA GAAAGAGCAA AATAACTGAT 360

ACAAAAAGCA ATCCTACAAG AATTAATAGA CTCATCTAAA TCATTCACCT TTTCTAAAAT 420

AATCCGAGAA AGATTTAAAA AAATTATATC CAGCTATTAG GATAAGTCCT ATCATAATCA 480

CTCCAATAGC CATTTTCTTA TCCTCCCTAA AAATAAAGTC TTTTACTTAA AACCTTTTTT 540

GTAATCGTCA ATAAATTTTT TCCCTTTATC AGTTAATTGA AATCCTATAT CTAAATAGTC 600

CTCATATCTA AGTTTAGCCA GCATCTCACT TACAGATGCT GGACTAACAC TTAGAACTTT 660

TGCTAAATCA TTATTTTTTA ATTTCTTTTT TCTTTTTGAT AATTCCTGAA TTGCTAGTAA 720

ATAAATTAGT ATAGTAGTTG TTTTAATTTG TACCATACCT ACCCTCCAAA TTATTTCGAA 780

AGGCCTTCAG CAATTTTATC CAGGTTCCAT TTCATCATCG CATAGTAACT ATCACCATCT 840

TGTCCTTTTT CTGCAATTGA ATCAGTAAAA ATCGTTGAAT AGATAGGAAT ATTGGTATCT 900

TTTGATACTG TTTTCATCGG TCTATCGTCC ACACTACTTT CTACGAATAA GGAGGGAACT 960

TTTGTTGTGC GTAATTTTTC AACTAAGTGT TTTATTTGAT CTGGTGTTCC TTCTTCTTCA 1020

GTGTTGaTTT CCCAAATGTA AGCAGAAGGC ACATTATACG CTTTCGAGAA ATATTTaAAG 1080

CATCCTTCAC TTGTAACAAT CATTTTTTTA TCATTCGGAA TTGAAGCAAA TTTAGATTTA 1140

GCTTCTTTGT CTAGAGAATC CAACTTTTCA ATATACTTAT CTAGATTTTC TTTATAGAAT 1200

TTTTTATTAT CAGGATCTTT TTCCGCTAAC CATTTTTCAA TATTTTTAGC GTAAATAATA 1260

CCGTTTTCTA AATTTAACCA AGCATGGGGA TCTTCCTTCC CTTTCTCACT CTGACCCTCT 1320

AAGTAAATAA CATCTATGCC ATCACTTGCT GCAAAATAGT CTTTGTTTTC CTCTTTGTTC 1380

GCATkTTTTA CTAATTTTGT AAACCAAGCA TTTCCTCCAG TkTCCAAGTT AACACCGTTA 1440

TAAAAAATCA AATCTGCTTT TGAAGTTTTT TGAACATCTT CAGGCAAaGG TTCATATTCG 1500

TGGGGATCTT cCCAATAGGT ACGATACTGT GTAAATCAAT TTTATCTTTT GCTATATTTT 1560

CAGTAATATC TGCTAAAATC GAATTAGTAA CTACTACCTT TAATTTTTCG TTCGAAGCCA 1620

CTGTATCTTT GTCTTTACTA TTTGTTGAAC ATGCCCCGAA TATAAAAAGA GCAATCAACA 1680

TTCCCAGACT TCCCACTACT ACTTTCCATT TTCTCATCAT AACGATCTCT CCTTCTTTCT 1740

TAAAAATTTT TGTTTTGGTG AGATTAGAAA ACTAATTACA AATAAGGATG CCGCTGTTAA 1800

AACAATACTT GATCCTGCGG CTATGTTAAA ACTATAACCA ATAAATAATC CaAGCAATGA 1860

TGAACTTGCT CCGACTGTTG CAGAAATAAA TATCATAGTT TTCAAGCTAT TTGCATATAA 1920

ATATGCTGTA GCTGCTGGCG TAATTAACAT TGCAACAATT AATATAGTTC CTACACTTTG 1980

CATAGCTGTT ACTGCTACTA AAGTTAATAG AACCATCAAT AGATAATGAT AAAAATTGAC 2040

ATTCATCCCC ATTGCTTTTG CCATTAATGG ATCAAAAGAA GTGATTAATA ATTCTTTAAA 2100

AAATAGAACT ACTATAACCA ACACTAAAAC AGAAACGCCA ATTGTAATCC ACTTATCAAT 2160

ATCTTGTACA GCTAAAATAT TTCCAAACAA GATATGAAAT AAATCGGTTG AGCTATTTGC 2220

AACACCTATA AGGATAACTC CTAGTGCTAA AAAAGAACTG AAAGTAATAC CAATTGCTGT 2280

ATCTCCTTTT ATTACACTAT TGCTTTTTaT AAAaGTAATA ATGnTAGrAG CTAATAGACC 2340

AAATACAATC GCACCAATAA AAAAGTTAAT GCCTAAAATA TAAGAAATTG CTACTCCTGG 2400

CAAAACAGCA TGAGAAATAG CtCGCCCATA AGAGACATTC CTCTTAAAAT AATAAAACAA 2460

CCTACTGTTC CAGAAACAAT CCCTATAACT ACAGATGTAA TTAGAGCATT CTGTAAAAAG 2520

TGAAAGTCAT ACAGACCATT AATAAAATTT TGAAGCATAT TATTCCCCCT CTCCAATAAA 2580

TATAGTATCT CCATAGGCTT TTTTTAAATT CTCTTCATTA AATACAGATT CTACACTTCC 2640

GTGAGCAATT AGATTTCGAT TCAAAAGAAT AATATTATCA AAATACTtCT TTACCyTGcT 2700

TAAwTCATGA TGCACAATTA GCACTGTTTT TCCCTGATGT TTCAATTCTT TTAACGTATC 2760

CATAATTATT CGTTCACTAA CTAAGTCTAT CCCTACAAAT GGTTCATCTA AAAATATAAA 2820

ATCCGCATTT TGAGCAAGAC ATCTTGCTAA AAGTACACGT TGAAATTGTC CTCCAGATAA 2880

TTCTCCTATT TGATGATTAG AATATTTTCC CATATCGACC TTTTCCAAAG CCTTTGAAAC 2940

TTTTTGCCAT TCTGCATTTT TAATTCTTTG AAAAACTTTC ATTCCGGCAT ATGTTCCCAT 3000

AGATACACAT TCTTTTACTT TTATTGGAAA AGTAAAATCA ATATCAGCTT TTTGTTCTAC 3060

ATATGCAATC TTTACCAGTT TTTTATTGAT ATCTTcATTA TTAATGGAGA CCTCTCCCTG 3120

ATGAGGGATA ATATTTAACA TTGCTTTAAG CAAAGTAGAT TTTCCAGCAC CATTTGGACC 3180

AAAAATACCT GTTAAAGCAC CTTCTTGTAT TTTTAACGAA ATCTCGTTCA ATGCTAAAAA 3240

ATCGTTATAA CAAACmnGTA AGCTTTTTCA CTTCTAACAT AACTTTCTCC TTCTAAATGA 3300

TAATGATTAT CACGATTGTA TGGtACACaA AAATAATAtT TaGtCAaGCC TAAAATTTTa 3360

AATTgGGATT TTnCgGTgGT TGGA 3384

(2) INFORMATION FOR SEQ ID NO: 330:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7213 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 330:

AnAATATAAA TCCTGAGTAG TTGCACTGAG AAAGCCAAGG ATTAAAACTT CAAAGGAAAA 60

GCCTGGGACA TCAATCATTT TGGATTTTTG TTCCAGGCTT AAAAATGATA AAGGGTGGGA 120

GCAGAAGCAA CTACTAACAA ATAAGACTTT TCTTTACGAC AAATAATGCG GATCGGTCTT 180

AGTCGTGCTA ATGTTGTCGC GGATTTTTTT CAAGGAATCG GCATTTTGAT CAAAATTGTG 240

ATACATCGTA TTTGTATATA AAATATCCAT TACCGTTAAG TAACTGATCC GTGAAGCCAA 300

GGATTCCGCA CGAAACAAGC CTTCTTCGGT GACGACAATA ATTGCTTCGT CGGACAAGCC 360

TGCCAACTCT GAACCGCTGT TACCAGTTAA CGTAATCATT TTTGCGTTGG TTTTTTTAAC 420

TTGTCGTGCT AGGTTGATTG ATTCTTTGGT TTTGCCTGAA TGGGAAAAAA TAAACACGCA 480

ATCTTCACTG GTTAATTTTG TGACAAAACT TAACTGCATG TGGTAATCAA AAATGTAGTT 540

GCACCGATAG TTAGTCCGAA TGAATTTATG GAAACTATCA AAAGCCACAA TCGAGGACCC 600

GCCTTGCCCA AAAAAATGAA GCGTTTTGGC AGAGTACATC AAAGCCAACA CGTTATCTAA 660

CAGCTCCTTT GTTAAGAACT GTGTGGCGTT ACTGAGGGAA TAAATATTGG CTTGTAAGAC 720

TTTTTTAGCA ATGGTAATAC TATCATCATC TGGTGAAATA TCATCGGCGC CATTCACGCT 780

ATTGACGTTT TGAAAAGTCG GCTGATAATT AGAATTTCTG GCGATATCAA TTTTAAATTC 840

TTGGAACCCA GAATAGCCAA TCTTTTTTAC AAATTGATAC ACGGaAGATT GGGaGACGCC 900

AATTTCGTTG GCAAGACTTT CTAACGTCAT TTGTGTCAAT AATGaTTCAG AATGAACAAA 960

GTAATCGGAA ATTTTCTTTT CTGTGAGGCT CAAATTCGGT TTTTTGaTCC GAATCCGTTG 1020

ATCCAAATAG TGGkTTTTCa TTAACACGTT CCTCCAATCT TATTTTTTTA AAAATAAAAT 1080

GAATGCCTTT TCATTGTTGG TATAGACTGA TTATAGTGGA AAAAGAGGAC AAATACAATT 1140

TCTCTTTTTC TCCACAAAAA AGAAATTCTT TAAAAAAAGC CAAAAAAAAA TAAATAATCT 1200

TTTTTCTGAA AAATGAAGGG GTCTTGAAAA GACTTCTCTC AACTAAAGAA AGTAGCCAGT 1260

AAAAAGAACG TGATACCAAG AGATTTACTT GTGTTTTTTG TATGAAAGCT CCTTAAAAAT 1320

AAAAACGTTT TACCTTTTGA GTAGGCAAGA TTTTTGCTAA AAAATAAAAA ATATTTTTTA 1380

ATTAGTGTAA TCGTTTACAT TATTTTTTTT ATCCTTTATG aTTGaAGTAT CAAAAGAGGC 1440

AGGTGaAAGA AAAATGGACA TTGGCTTTAT TGGACTtGGG AAAATGGGTT tGAACATGGC 1500

ATTAAATGTT CACGAACAAG GCTGGCCCAT TATTGGCTTT GATGTAACAA AGGAAGCAAG 1560

AGCGACTGCT AGAGAACAAG GTCTTTCCGT TGTAGATTCG TTGTCAGAAT TATTGAAAGC 1620

GCTTAACAAA AGGAAGGTGA TTTTTTTAAG TACCCCCGCT GGTCAGATTA CCAATCAATT 1680

AGTGGCAGAA TTGGTAGAGC AATTAGCACC AGAGGATATC ATTGTGGATA GTGGCAATTC 1740

AAATTTTCAT GATAGTGTTG CGAATGCTCA GCTTGCTAAG GAAAAAGGAA TTTATTTTAT 1800

TGATTGTGGT ACTTCTGGTG GCATCAAAGG AGCGCGGGAA GGAGCGTGTT TAATGGTTGG 1860

CGGTGCGCcT GAAGCTGTCA AAGTGTTAAC CCCATTTTTT GAAGATTTAG CTTGTGAACA 1920

AGGGTATCTT TATGCTGGGA AATCAGGAGC TGGTCACTAT TTGAAAATGG TGCATAACGG 1980

GATTGAGTAC GTCATGATGC AAGCCATGGG CGAAGGATTC AATTTATTAG AAGCTGCCGA 2040

ATATGATTTT GCGTTGGAGA AAGTAGCCGA TGTGTGGAAT CACGGCTCGA TTATTGAAGC 2100

TCGCTTGATG GGCTTAGCGA AAGAGGTTTT TGCGGAAAAT CCTACGTTAG CCAATCTAGA 2160

AGGGAAAGTG GCAGCAAATG GCGAAgcnAA ATGGATGATT GAAGAAGCAT TGCGTTTGGA 2220

GATGCCCGTC CCAACTACGG CGTTGTCCTT ATTTACTCGA AATGAAAGTA TGCTAGCAAA 2280

TCATTTTTCA AACAAAGTGG TCGCTTCTTT GCGACAAGGA TTTGGCGGGC ATGAAGTCGT 2340

AAAAAGTCAG TAAAGGAGTT TAGTCATGTT AAAAAAACAA AAAATTGCCA TAGTGAATTC 2400

CAGTAGTTTT GGTCAAATTT TTCCTGAACA CTTGACGCGC TTAGAAAAAA TAGGCACCGT 2460

GAAGCATTTT ACAGTCGACA GTGAAATTGG TGGCAAAGAG TTAGCGGAAC GTTTACAAGG 2520

ATATACGATT ATTATTGCTA GCGTGACGCC ATTTTTCACG AAAGAATTTT TTGAACATAA 2580

AGATGAATTG TTGCTGATTT CTCGCCATGG TATTGGGTAT AACAACATCG ATTTAGACGC 2640

TGCAAAACAG CACGACACAA TTGTTTCCAT CATCCCAGCT TTAGTAGAAC GAGATGCCGT 2700

GGCGGAAAAC AATGTCACAA ATTTGCTAGC TGTTTTACGA CAAACCGTGG CTGCAGATGC 2760

CAGCGTTAAA GCGGATCAAT GGGAAAAACG AGCGAATTTT ATCGGTCGGA CGTTATTTAA 2820

CAAAACTGTG GGTGTCATTG GTGTTGGGAA CACAGGAAGT TGTGTCGTGG AGACGTTACG 2880

AAACGGATTT CGCTGTGACG TGTTAGCGTA TGATCCCTAT AAATCCGCCA CTTATCTGCA 2940

```
AAGTTATGGA GCCAAAAAAG TAGACTTAGA TACGTTGCTA GCTTCAGCAG ACATCATTTG    3000
TTTATGTGCC AATTTAACTG AAGAAAGTTA TCACATGATT GGCTCGGCGG AAATTGCCAA    3060
GATGAAAGAC GGCGTCTATC TTTCCAATAG CGCACGAGGC GCCTTGATTG ATGAAGAAGC    3120
CATGATCGCC GGTTTACAAT CGGGCAAAAT TGCTGGGCTT GGGACAGATG TTTTAGAAGA    3180
GGAACCAGGT CGGAAAAATC ATCCTTATCT TGCGTTTGAA AATGTAGTAA TGACGCCGCA    3240
CACCTCCGCT TATACCATGG AATGCCTGCA AGCGATGGGA GAAAAATGCG TGCAAGATGT    3300
GGAAGACGTG GTGCAAGGTA TTTTACCTCA ACGGGCAGTG CAAGAAGTCA GTCGTTACGT    3360
TAGTTAAGAA GGAGGAGGAG CATGTTAACT GATTTAAAAG TGAAAGTTGG ACCGCAATTT    3420
TATCGTTACC ACGAAGGCGC CTTAGCTTCG GTGCCGTCAC TTTTCAAGGA ATATCACGCC    3480
CAACGAATTT TGGTGGTGCA TGGCACCGTA TCCTTTGAAA AAGCTCAGCC ATTTTTACCA    3540
TTCTTGGCAG ATTCAGAGTA TCAATTCTTC TATCACACTT ACACAGGAGA ATGTAGTTAT    3600
TTTGGCGCCG AACAAATTTC ACAGCAAATA AAGGAACACC AGATTGACTT TCTTCTGGGA    3660
GTCGGCGGCG GTAAGTTAGC CGATTTAGTG GGCTACAGCG CCCATTTAAA CAATCTGAAT    3720
TTTGGCTTGG TCCCCACTTT GGCTAGTAAC TGTGCGCCGT GGACACCGTT AGCTGTGATG    3780
TATCAAGAAA ATGGCGCGGC CGAAGGCAAA ACGGAACATT TCTTTCGCCA AGCTGCGTTT    3840
TTAATCACCG ATCCTAAGTT GCTTCTGGAT GCGCCCCGTG ATTATTTTGT CGCTGGTTTA    3900
GCTGATACGT TAGCGAAATG GTATGAATCA GAGACCATTT TACGCCAAGC CCATTTGCAA    3960
GGGGAACCCT TTTTACAATT AGCGGGAGCG ACCGCCAAGC TTTCGCAAGA AGCGATTATG    4020
CGGGACTCAA AAGCAGCCTT AGCAGCGATG GATGAGGGCA AACTCACGCC TGAATTTGTC    4080
CATCTTTCAG AAATTGTTTT CGCTGTTTCT GGATTAGTCG GCGGCTTCGG CGACAAATAC    4140
GCCCGCAATG CAGCAGCACA CGCCATGCAT GATGCTATGA GTAAATTTTT ACCCAAAAGT    4200
CATGACTATT TACACGGCGA AAAAGTAGCT TACGGCATCT TCTATCAATT AGCATTGGAA    4260
AAAAGGTGGG CCATCATAGA CGCATTAATC CCTTTTTATC AAGAGCTAAA CTTGCCCATG    4320
TCGTTACGAC AAATGGGCCT TTACCCAAAA GAAGAAGCAG TACTCGATGC AATGGTTCAA    4380
TTTATTGATT CAAAAGAAAA GGTTCATTTA ATTCCTATTG AGATTTCGGA AGAACGTTTA    4440
CGTCAAGGGA TTAAAGAGCT AGAAAAGTAT ATCCAAAACC AAGGATAAGA AGGAGGAACT    4500
ATGGAAACGT TAAGTGTTTT ACAAAGTTTG ATAATGGCGC TTTGGGTGGC AGCAATTATG    4560
TCCAGATGGC TTGGAGGAGG AGCAACGTTA ACGTTGCGTT TCTCGCCCTT AATGACAGGG    4620
TTAGTAGCAG GTTTAGTGAT GGGCGATGTC CCGAAAGCGA TGATTGTCAC CGCAGCATTA    4680
CAAATGATTT ACATGGGTGT TTTTTCACCA GGTGGCTCTA TGCCAGCAGA ACCATCAATT    4740
GCGGCAGCGA TTGCTGTGCC AGTTGCGTTG CTAGGGAATT TGAAGCCGGA AGCGGCTATT    4800
GCCGTAGCGG TTCCTGTCGG TCTTTTAGGA AGTTATTTAT ATCAATTCCG TTTCTTTATT    4860
AACACGTTTT TAGGGAAGTA TACCGATCGT GCGGTGGCTG AACTAAATTC TAAAAAAATT    4920
```

GCCCGCTCAA TTATCTGGTA TCCAACGATT GCCTCATTTA TTTTATTCGT TCCATTAGTC 4980

TTTTTTGCTC TTTATTTAGG GGCACCAGTC ATTGCCGATA TTATTAAAGC ATTAGAAGGA 5040

ACCGTTGTGA TTCACGTGTT AGAAGTAGTC GGTGGTGGTT TAGCCGCAAT TGGGATTGCG 5100

ACAACGGTCT ATGTCATTGG TCGAAAAGAT TTCTTGGTTT TCTTTTTCCT TGCATACTTC 5160

ATGAGTATCG TCTTTAAGTC ATTGGAAATT ACCATGGTAA CGTATGCCAT CTTTGGCGTA 5220

ATTATTGCGT TAATTTTTGT GCAAGTCCAA AAAGGCAAAC CTGTAGCAGA AAGTGCTGGT 5280

TCCGCCAGTG CCACCACTGA TTTTGATGAT GATGACGATT ACGATGACGG GTTTTAAAAG 5340

AAAAAGGAGG AACTGAAATG ACGGAAACAA CAGAAAAAGT CGTGCCGAAA ACAACGAATT 5400

TAGCACCAGA AGAAATTACC AATAAAGATG TTACAAAAGC CTATCTCCGT TGGCATTTTG 5460

CCAATGAAAT TCCCCATTCC TTTGAACGTT ACTTGGCGCC TTCGTTATTG TACGCAATGA 5520

TGCCACTTTT AAAGAAATTA TATAAAGACG ATGAACAATT AAAAGCCGCT TATATGCGTC 5580

AACTGCTGTT TTTCAATACG CAATTAAGTT GGGGCGGCGG GGTCATTACT GGGCTGATGG 5640

CTTCCATGGA ACAAGAGCGG GCCAAAGAAG AACATGAAGG CCGCGAAATC ATGATGCAAG 5700

ATGATTTAAT GTATAACACC AAAGCTGGCT TGATGGGCGC GTTAGCGGGG ATTGGTGATG 5760

CGATTGACTC AGGTACGGTT CAATATATTT TTATTGCGAT TGCGGTGCCG TGGGCGCAAC 5820

AAGGTAGCGC ATTAGGTGCG ATTTTCCCAT TTGTGGCGTT TGCCCTTTAT CAAGTACTGT 5880

TAGGTGTGTT CTTTGCGAGA CAATCCTTTA AAATGGGGCG GAATGCCACA GGCTTAATGC 5940

AAAGTGCGGG GATTCAAAAA GCCATTGAAA TGTTATCTGT TTTAGGGTTG TTCATGATGG 6000

GGATTTTAGC TGGTAATTAT GTCAAAGTTT CTTCAACCCT TCAATTTAAA CTTTCTGGCC 6060

GTGAGTTTGT GGTTCAAGAT ATCCTTGATC AGATTGTCCC AGGACTCTTA CCGCTAGCTG 6120

TTGTAATGGG CGTGTATTGG TTCTATACGA AAAAAGGCTT GAAAGTTACG CAAGCATTAT 6180

TGTGGTTAAC AGGTATCTTA ATTGTCTTAG CAACAGTCGG TATTTTATAA AAAGAAAAGG 6240

ACGTGAATCA GATGGGCGTA GTAAATTTAG CAAGAGTAGA TGAACGATTA ATTCATGGAC 6300

AAGTGATGGT GACCTTGTCC CAAAAAAGTG GCGTTAATTC CATTTTCGTT GTCGATGAAG 6360

TGGTGGCAAA GGATAAATTT ATGCGTGATT TATACAAAAG TGCAGGCAGT CGGACAGGAC 6420

AAAAAACCAT TGTCATTACG CCGGAAAAAG CTAAATTTTA TTGGGATGAA TACCAATTCA 6480

AAGAGTACAA CTGTATTTTA ATCGCTAAAA CCGTCTCTGT AATTTATGAC TTAGTAAAAC 6540

ACGGAGTCCC TATGAAGGAG TTAAACATCG GCGGTATTGC ACAGAAGAAT CCAGAAAAAG 6600

ATTTATTAGT AACCAAATCA GTTTATTTAA ACAAAGAAGA TGCCGAAAAA TTAAAAGAAC 6660

TACATGAGGT GTATGGCGTA GAAGATATTT ATTTCCAAGC AACGCCTTCT GCACCAAAAA 6720

CTAGTTTAAA AGAGGTTTTA GGCAAATTTA ATTTATAAAA AAGGCCAGTC GCTTGAGTCA 6780

TTGCATGACT GTTTGAAGTA ATGACTCAAG CCACTGAACT GAAGGAGGAA CTTTTATGTT 6840

AGGGATTGTG ATTGCTACAC ATGGTGCACT AAGTGACGGT GCGAAAGACG CAGCTACCGT 6900

GATTATGGGC GCAACCGAAA ATATCGAAAC CGTTAATCTA AATAGTGGAG ACGATGTCCA 6960

AGCACTAGGC GGACAAATTA AAACAGCGAT TGAGAACGTC CAACAAGGGG ATGGCGTGTT 7020

GGTCATGGTT GACTTATTAA GTGCGAGTCC TTACAATCAG GCAGTTTTGG TGATTAACGA 7080

ACTCGAACCA gCTCTACAAA AGAAAATCTT tGTGGTGAGT GGTACCAACC TACCAATGGT 7140

CTTAGAAGCG ATAAACCATC AATTATTGGG CACACCGATT GCTGAAnCGG CGCAAGCGAT 7200

TGTTGCGCAA GGC 7213

(2) INFORMATION FOR SEQ ID NO: 331:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 564 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 331:

ATCCAAAAAG AAAGGAAGTG CCTTATGGCT AAAAAATCAA AAATTGCTAA AGCAAAAAAA 60

CAACAAGCAA TGATTGCAAA ATATGCGCCT ATTCGCCAAG CCTTAAAAGA AGCTGGCGAT 120

TATGAAGGTC TTAGCAAATT GCCAAAAGAT GCTCATCCAA GTCGCTTAAA ACTCCGCGAT 180

CAAACAGATG GGCGCCCACG TGGCTACATG CGTAAATTTG GCATGTCTCG GATTCGTTTC 240

CGTGAATTGG CTCACCAAGG CTTGATTCCT GGCGTCAAAA AAGCAAGCTG GTAACTCATT 300

ATTTCGATAA AGACAACTAG CGAGGACAGT CGCCTTTGGG CTGACCTCGC AGCCCCCTTT 360

ATCTTTTATA CGCAATTAAA AAAATCAAAT TGGAGGACGG TaCTATGCGC CAAACCATTA 420

CGTTAGCTTG CGCTGAAACT GGCGAGCGCC TTTaCTTAAC ArGTAAAAAC ArACGAAATA 480

CACCAGAAAr AtTACAGCTG AAAAAAtATT CaCCAAAACT aCGGcGcCGs GsACTTTTcA 540

CkGAAGTCnA ATraCCGGGG GGAA 564

(2) INFORMATION FOR SEQ ID NO: 332:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 921 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 332:

CAAGAATCGT GTGGAACAAA AAGCGGAAGA CGGCGACGCA AAAGCTCAAA AATTGCTTAA 60

GGTATTGCAA GACCCCAATA ACTTTTTATC CACAATCCAA GTAGGGATTA CTTTGGTTAA 120

TATCCTATCT GGTGCTTCAT TAGCAGAGAC TTTATCTAGT CGTTTGGCCC CTGTATTAGG 180

TGGCGGCGCg CAGCTAAGAG TCTTGCCAGT ATCATTATCT TAGCTTTGTT AACGTACGTT 240

TCTATTGTAT TTGGTGAATT ATACCCTAAA CGAATTGCTA TGAATAAATC AGAAGAAGTC 300

GCTCAATTGA CTTCTGGTGC AGTCCGCTTT TTAGGAGTTA TCGCGCGTCC CTTCGTTTGG 360

TTGCTATCGG CTTCGACCGA TTTATTATCC AAAATCACAC CAATGACTTT CGATGATGCT 420

GATTCGAAGA TGACACGAGA TGAAATGCGC TACATGCTTG AAACAGAAGG TGTTTTGGAA 480

AACGAAGAAC TAGAAATGTT ACAAGGTGTC TTTTCATTAG ATACAAAAGT GGCACGAGAA 540

GTAATGGTTC CTCGGACAGA CGCATTCATG GTTGATATCC AAGATGATGT CCAAGAAAAC 600

ATCAATTTAA TTTTAGGAGA AAATTATTCT CGGATTCCTG TTTaTAGTGA AGACAAAGAT 660

AAAATTGTTG GTATTTTACA TACCAAAACG TTACTAAAAG CTGCACGTAA TTTAGGCTTT 720

GAAAACATTG AATTAGGCGC GATTATCCAA GAACCATTAT TTGTCCCAGA AACCATTTTT 780

ATTGATGATT TACTTTATGA GTTAAAACGA ACACAAAACC AAATGGCGAT TTTACTAGAT 840

GAATATGGCG GTGTTGTCGG TTTAGCAACA CTGGAAGATT TACTAGAAGA AATCGTTGGT 900

GAAATTGACG ATGAAACAGA T 921

(2) INFORMATION FOR SEQ ID NO: 333:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1534 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 333:

TGAACGTCCT ATTAGTTATA GACGAATTTG AGGACCAAAA AGTTTCAATT TTTTCAATTT 60

AGTTAGTTGT TTATTTAATA AAAAATAAAA GGTTGCTAGA GGATGAAAGG AGTATAATTT 120

ATTTATCTAG AAGAGTAATC GGAATGAGAA AGGGCTGTGG AAATGGATCA TCAACAGGTG 180

TTGAAAAATG AAATTAAACA ATATTTAGTG AGTAAAAACT GTGAGAAGAC GTATTATCAT 240

TGTATGGAAG TTGGAGAGTA TGCCTATCmA CTCGGTGAAA AATATTTAAC TAGTCCaGAA 300

AAAGTAAGTA TTGCmGGaTa TTTACATGAT ATCTCTGCGG TTTATCCAAA TAATCAGrGA 360

ATTAGTGTCG CACAAAAATA TGGGATTGAA TTAAACGAAA CAGAAATGGC CTTTCCAATG 420

ATAATTCATC AAAAAATTTC TAAGTCTATT GCGAAAATGG ATTTTGGTAT TGAAGACAAT 480

GAAATTTTAT CAGCTATTGA ATGCCATACA ACATTGAAGA AAAACTATTC GGACATAGAT 540

TTAGTTTTGT TTGTCGCTGA TAAAATAAAA TGGGATCAAG AAGGGAAGCC ACCATACTTA 600

GATGGTTTAT TACAAGCTTT AAATTGTTCT TTAGAAAATG CCGCATATTT CTATATCGAT 660

TATATTTTGA AaCACGATAT TaAAGTCGTG CATCCTTGGT TATGGGATGC TTACAATCAA 720

CTTAATTTGA TAATTAAATA ACAAATAAAT ATAGCCGTTT TTAGTCGAGT TTTGCTGGTG 780

AGAAAAGGTG AAAAGTAGTA CAGTATTCTT GTCAGAAAGA AGGGATTCCC AATGAAAAAA 840

ATAATCTTGA TAGGTCTTTT ATTGAGTATT TTATCAGGAT GTTCTTCCAA CAAAAGTGGT 900

GATAGCACTG CAAGCAAATC AAGCAAGCCG ACAGCAACAT CAGAAATGAA AGCACAACAG 960

TAGCCAAGCG AGTCAAATGA GCACGGAGTC AAAATCTAGT ACAACTCAAA GTAATTCTTC 1020

TGCTTCTACT AAAACTGAAA ATCAGTTGTG GTCTGTGCAA AAAAAAGCAC AGTTACAACA 1080

ATTTATGGCT GACTGGGGGC AACGGATGAA CCAAACGTAT AAAGAATATA TACCAGGAAA 1140

TTTAATCAAT TTTAATGGCG TCCAAGCGCC TGATGGTCTA GTAGGAAATG TTACTATGCA 1200

ACCAGCGATT GGTCAGCATC CCATTTATTT AGAATGGTCT GAAAaTGGAC AAGTGAAGGA 1260

TGGATATGCG TTAGTTGCTG TTTACTCTGA TGCAGAAACG CAACCAGAAA TGCAAAAGCA 1320

TTTGTATTTA TTCACCATCG TTAACAATCA GCCCTTAGTT TTAGTGACGA TGCAAAATCA 1380

AGGAGATCCT TATGGCTATC TGTATTTTGG AGCAACAGAT AATGCAGAGT TGAGGGCTGG 1440

CTTTGAAAAA ATTGTAGGTG CCCCGTCAAT TACTAAAGAA CAAATCCCAA tATTTCAGTC 1500

AATCCgTGkc tCaAAgAGGa GCGGnTGnAT TTTA 1534

(2) INFORMATION FOR SEQ ID NO: 334:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9212 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 334:

TTGATTTGGC AAATATTTCA GGAAAAGCGG GTAAAGAGCT TGGTAATTTT GGCATGATGA 60

TTCAAGACTT AACGAAAACA GTGCCCTATT TAACGATTAC AGAATTGGTG AAAGAAACTT 120

TACAACGTAG TGGTTATCGG GAAGCTTTGA TGGCGCAAAA TAATTTAGAG TCTCAAGCTC 180

GCTTAGAAAA CTTGGACGAA TTTTTATCTG TGACACAAGA ATTTGATAAA CGTTTTGAAG 240

CGCAAAACAA CGACGATCCA AATGGCGAAG AAACAAAATT GGCGGATTTC TTAACAGATT 300

TAGCCTTGGT TTCTGATTTA GATAATTTAG AAGAGAGCTC TTCGCAGGTA ACCTTAATGA 360

CGTTGCATGC GGCCAAAGGG CTGGAATTTC CAGTGGTCTT TTTAATGGGC TTAGAAGAAG 420

GGGTTTTCCC ATTATCACGT GCCATGCTAG AAGAAAGTGA ACTGGAAGAA GAACGGCGCT 480

TGGcTTATGT TGGtATCACA CGAGCGGAAG AAGCCTTGTT CTTAACGAAT GCTTATTCGA 540

GAACATTATA TGGTCGCACG CAATATAATC GACCAAGTCG CTTTTTAGAA GAAATTGAGG 600

CGGAgCGTTT AATGAAACAA GGGGCGGCAg CCAAtTTACA GAAAGCTCCA GCGCGCACCT 660

TTGATCCTAA AGTATTTAAG CCCAGCGCAA GTAAACCAGC TTATACACAA CCAGCTACAA 720

AATCAGTTAG CAACAAAGTG GCAAGTGGTG GCGAAAAAGT TTCTTGGCAA GCAGGTGATA 780

AGGTCCAACA TAAAGCATGG GGCGTTGGTA CAGTCGTTCG TGTAGGCGGT TCGGCCAAAG 840

ACTTAGAGTT GGACATTGCT TTCCCAGAAA AAGGCATTAA ACGGTTGTTA GCGGCATTTG 900

CACCAATTGA AAAAATTTAA CATGCAAAAT AGGTCAGCGA TTTAATGAGA AATCATCATT 960

AGTTAAATTC TTTTTATTTT CTATCTTAGT GTAAATGTGT TAAAAAGACG ATAGGAGCAT 1020

TACGCTTATG AGGAGGAGTT CTACGATGGA ACAACAACCA TTAACATTAA CGGCGGCGAC 1080

AACTCGTGCG CAAGAATTAA GAAAACAACT CAATCAATAT TCCCATGAAT ATTATGTCAA 1140

AGATCAACCA TCCGTAGAAG ATTATGTTTA CGATCGTTTG TATAAAGAGT TAGTCGATAT 1200

TGAAACGGAG TTTCCTGATT TAATCACGCC AGATTCGCCG ACACAACGTG TAGGTGGCAA 1260

AGTTTTATCT GGTTTTGAAA AAgcTCCGCA TGACATTCCA ATGTATAGTT TAAATGATGG 1320

GTTTAGCAAG GAAGACATTT TTGCTTTTGA TGAACGTGTT CGTAAGGCGA TTGGTAAACC 1380

CGTTGCTTAC TGTTGTGAGT TAAAAATTGA TGGGTTAGCA ATTTCATTAC GTTATGAAAA 1440

CGGGGTGTTT GTACGCGGAg CAACACGAGG CGATGGAACG GTTGGTGAAA ACATTACAGA 1500

GAACTTACGA ACGGTTCGCT CTGTACCGAT GCGTTTAACG GAGCCCATTT CAGTGGAAGT 1560

TCGTGGCGAG TGCTACATGC CTAAACAATC CTTTGTCGCA TTGAATGAAG AACGTGAAGA 1620

AAATGGCCAA GATATTTTTG CCAATCCTCG GAATGCAGCG GCGGGAAGCT TGCGCCAACT 1680

AGATACAAAA ATTGTGGCTA AACGGAACTT AAATACGTTT TTATACACAG TTGCTGACTT 1740

TGGACCAATG AAAGCGAAGA CGCAGTTTGA GGCATTAGAA GAGCTTTCTG CAATTGGTTT 1800

TAGAACCAAT CCCGAACGGC AATTATGTCA GTCAATTGAT GAAGTTTGGG CCTACATTGA 1860

AGAATACCAT GAAAAGCGGA gcACTTTGCC ATACGAAATC GATGGAATTG TGATTAAAGT 1920

GAACGAATTT GCGTTGCAAG ATGAATTAGG TTTCACTGTC AAAGCACCTC GCTGGGCGAT 1980

TGCCTACAAA TTTCCACCAG AAGAAGCAGA AACAGTAGTG GAAGATATTG AATGGACAAT 2040

TGGGCGCACG GGGGTGGTCA CTCCTACTGC GGTCATGGCT CCTGTACGAG TCGCGGGAAC 2100

CACGGTTAGT CGGGCAAGTT TGCATAATGC AGATTTTATT CAAATGAAAG ATATCCGTTT 2160

AAATGATCAT GTTATTATTT ACAAAGCCGG CGATATTwTC CCaGAAGTAG CGCAAGTGCT 2220

AGTTGAAAAA CGTGCAGCCG ATAGTCAGCC TTATGAAaTG CCGACACaTT GTCCaATTTG 2280

TCaTAGTGAG TTGGTGCATT TAGATGAGGA AGTGGCGTTA CGTTGTATTA ATCCAAAATG 2340

CCCTGCGCAA ATTAAAGAAG GGCTGAATCA CTTTGTCTCA CGAAACGCAA TGAACATTGA 2400

TGGCTTGGGG CCACGAGTGT TAGCACAAAT GTATGACAAA GGATTAGTGA AAGACGTAGC 2460

GGATCTGTAT TTCTTAACAG AAGAACAATT AATGACTTTG GATAAAATCA AAGAAAAATC 2520

AGCAAACAAC ATTTATACTG CCATTCAAGG GAGTAAAGAA AACTCTGTCG AACGTTTGAT 2580

TTTTGGCTTA GGGATTCGCC ATGTGGGGGC GAAAGCGGCG AAAATTTTAG CGGAGCATTT 2640

TGGTGATTTA CCAACCTTAA GTCGTGCAAC CGCGGAAGAA ATTGTGGCTT TGGATTCAAT 2700

TGGCGAAACA ATCGCAGACA GCGTTGTGAC GTATTTTGAA AATGAGGAAG TTCATGAATT 2760

AATGGCTGAA CTGGAAAAAG CCCAAGTGAA CTTGACATAT AAAGGTCTCC GGACAGAACA 2820

ATTAGCAGAA GTGGAGTCAC CTTTCAAAGA CAAAACAGTT GTTTTAACAG GAAAATTAGC 2880

GCAGTACACT CGCGAAGAAG CCAAAGAAAA AATTGAAAAT CTTGGTGGAA AAGTTACAGG 2940

GAGTGTTTCG AAAAAAACTG ACATTGTGGT TGCTGGCGAA GATGCAGGCA GTAAATTAAC 3000

CAAAGCTGAA AGCTTAGGCG TGACCGTTTG GAATGAACAA GAAATGGTCG ATGCGTTAGA 3060

TGCAAGTCAT TTTTAAGAAG AAAAATTAAC GTTGTCAGAT TTTATGTAGT AATTTGATGG 3120

TATGCTTATT TTAGTCATAA AAGTTATATG AACGAGGACT TTCTGAGTTG AGAAAGGAGG 3180

TTGATGCTTA TGCGCAATTC TTCTTATCTT AATGAAAGGA GCAATGTTTT TGTCCGTTGA 3240

AGCAGCGTTG GGACTGATGA TTGGTTTTGC AACACTTGTT GTGACCATTA TCTTCGTTAT 3300

CTTAGCGCTT GTCTTAGACA ACAAAAATAA CCGTTCATAA GCTTTGAGGA GCTGAACGGT 3360

TAAGCGTTTG AAATAAAACC AATAACTCAG TAAGCCGCTA GTTCTTATAA CTAGTAGTGG 3420

CTACTAGAGA GGTATGTTCC TGCATATCTC TCTTTTCATT TTCATAATAA CATGTTTCCG 3480

TCATTTTTGC GAGCGACAAA TTATCAATGA AATTCAAGGT TTCATCTAAA ATATCCGACC 3540

ATTTTATGGT AGAATAGGAA CAGAGTGAAC AGCTGTGCTA AGCTGTTTAA AGAAATACAA 3600

TCCAATTGAT TAATCTGAAA ATGAAAATTT CAGTAAATGA TAGAAAGAAG GGTATCTATG 3660

GCAATTACAG AAGAACAAGT CAAACATGTC GCTAAATTAT CAAAATTATC TTTCTCCGAA 3720

GAAGAATTAG CGGACTTTAC CAATCAACTA GATAAAATTA TTGATATGGT GGAATTACTA 3780

GAAGAAGTAG ATACAACCGG GGTTCCTTTT ACTTCAAATG TGAATGAATC AATCAATGTC 3840

ATGCGTGAAG ACGTAGCGAC ACCAGGAATG GACCGTAAAG AATTAATGCG GAATGTTCCG 3900

GAATCAGAGA ATGGTTACAT TAAAGTGCCT GCAATTATGG ATAACGGGGA GGCAGGAGCC 3960

TAATGGAAAA ATTATACGAC AAATCGTTAA CAGAGTTACA TGATTTATTG GTCTCAAAAG 4020

AAATCACTGC AGTTGATTTA ACCGAAGAAA CATTAAATCG TATTCAAGAT ACAGAAGAAC 4080

AACTAGGTTC TTTTATCACA GTTAGTGAAG AAAAAGCAAT GGCATTAGCC AAAGCGATTG 4140

ACTTAAAAGG CATCACTGAA AGTAATCCAT TAGCTGGGAT TCCTATTGGG ATTAAAGACA 4200

ACATTGTCAC AAAAGATATT TTAACAACAG CTGGcTCAAA AATGTTACAC AATTTTGACC 4260

CAATTTATGA TGCGACCGTG ATGGATaAAG TCTATCAAGC AGATATGATT CCTGTCGGAA 4320

AATTGAAcAT gGATGAAtTT GCCaTGGGCG GAAGTACGGA AACGTCTTAC TTCAAGAAAA 4380

CAAAAAATGC TTGGGATCAA ACCAAAGTTC CTGGCGGTTC TTCTGGTGGT TCAGCTTCtG 4440

CTGTCGCAGC CGGACAAGTC CCAGTTTCTT TAGGGAGCGA TACAGGAGGT AGTATCCGCC 4500

AACCAGCCGC ATTCAACGGT ATTGTGGGcT TGAAAcCAAC TTATGGACGT GTGTCTCGTT 4560

TTGGTTTGAT TGCTTTTGCT TCAAGTTTAG ACCAAATTGG TCCaTTGACA CGAAATGTTA 4620

AAGACAATGC CTTAGCTTTA AATGCAATTA GTGGTTATGA CGAAAAAGAC GGCACTTCTG 4680

CAGGCGTTTC TGTTCCTGAT TTTACGGCAG ATTTAACTGG CGACATCAAA GGCATGAAAA 4740

TTGCTTTACC AAAAGAATAC TTAGGTGAAG GGGTACAACC TGATGTTCGT GAAGCGGTAT 4800

TAAAAGCAGC AGAAACGTTC AAAGCTTTAG GCGCAACAGT GGAAGAAGTT AGCTTACCAC 4860

ATTCTAAATA TGGGATTGCC GCTTACTATA TTATTGCCTC TTCAGAAGCA AGTTCAAACT 4920

TACAACGCTT TGATGGCATT CGTTATGGTT ACCGTTCCGA AAATGTCCAA AATCTAGAAG 4980

ACGTCTATGT TAATTCGCGT TCCGAAGGCT TTGGTACTGA AGTCAAACGT CGGATTATGT 5040

TAGGAACCTT CTCTTTAAGT GCTGGTTATT ATGATGCGCA CTTTAAAAAG GCTGGTCAAG 5100

TCCGTACACT AATCAAACAA GATTTTGAGA ATGTCTTTGC GGATTATGAC TTAATCATTG 5160

GCCCATCAAC ACCAACTGTT GCTTTTGGTT TGGGTGAAAA TATCAATGAC CCAATCACGA 5220

TGTATATGTA TGATATTTTA ACTGTTCCTG TTAACTTAGC TGGGTTACCA GGCATGTCAA 5280

TTCCAGCAGG CTTCTCAGAA GGATTGCCAG TTGGTTTACA AATTATTGGA AAACATTTTG 5340

ACGAACATAC AATGTATAAA GCAGCCTATG CTTTTGAACA AGCGACAGAT TTCCATACGA 5400

AAAAACCAGT CATTTTAGGG GGGAACGATT AATGAACTTT GAAACTGTCA TTGGATTAGA 5460

AGTCCACGTA GAATTGAAAA CCAACTCTAA AATTTTCTCC TCTGCGCCAG CGCATTTCGG 5520

AGCAGAACCA AACAGCAACA CAAACGTGGT TGACTGGAGT TACCCAGGTG TTTTACCTGT 5580

GATGAATAAA GGTGCTTTAG AATTTGGCAT GAAAGCAGCA CTTGCCTTGA ACTGTGAAAT 5640

TTCAAAAGAA ACACATTTTG ACCGTAAAAA CTATTTTTAC CCAGATAACC CGAAAGCATA 5700

CCAAATTTCT CAATTTGATC AACCAATCGG ACATGATGGT TGGATTGAAA TTGAAGTCGA 5760

AGGCAAAAAG AAAAAAATTC GCATTGAACG TGTGCATTTA GAAGAAGATG CTGGGAAAAA 5820

TATTCACGGC ACAGATGGCT ATTCTTATGT AGACTTGAAT CGTCAAGGCA CGCCTTTAAT 5880

TGAAATCGTT TCTGAAGCAG ACATGCGTTC ACCAGAAGAA GCGTATGCTT ATTTAGAAGC 5940

GTTGCGTTCA ATTATCCAAT TTACAGAAGT TTCTGATGTG AAAATGGAAG AAGGTTCTAT 6000

GCGTTGTGAT GCCAATATTT CTTTACGTCC GTATGGTCAA GAAGAATTTG GTACTAAAGC 6060

GGAATTAAAG AACTTGAACT CTATGAACTT TGTAAAAAAA GGTTTAGCTT ATGAAGAAAA 6120

ACGCCAGGCC AAAGTCCTAT TATCTGGTGG CGAAATTCAA CAAGAAACAC GCCGCTTTGA 6180

TGAAGCAACT AGTACAACGT TGCTTATGCG TGTTAAAGAA GGCTCAAGTG ACTACCGTTA 6240

CTTCCCaGAG CCAGATGTTC CTCGTTTTTC AATTGATGAT GAATGGATTG AAAAAGTTCG 6300

CGCTAGTTTG CCAGAAaTGC CAGcCTCACG TCGTGCCCGT TATATCAGTG AATTAGGGTT 6360

ACCTGAATAC GACGCAATGG TCTTGACTTT AACGAAAGAA ATGTCAGATT TCTTTGAAGC 6420

AACTTTAGCA AACGGTGCAG ATGCGAAACA AGCATCAAAC TGGTTAATGG GCGAAGTTTC 6480

TGCTTACTTA AACAGTGAAA AAGTTGAATT AGCCGATACT AAACTGACAC CAGAAAATCT 6540

AGCGGGTATG ATTACTTTGA TTAACGATGG CACAATCAGC TCGAAAATTG CGAAGAAAGT 6600

GTTCAAAGAG TTAATTGAAA ACGGTGGAGA TGCTAAAGAA GTTGTCGAAG CAAAAGGCCT 6660

TGTTCAATTA TCTGATCCAG CGCAATTATT ACCAATGATT AACGAAGTTT TAGACAACAA 6720

CCAACAATCA ATTGATGACT TTAAAAACGG AAAAGACCGT GCAGTTGGTT TCCTTGTTGG 6780

ACaAATTATG AAAGCAACTC GTGGACAAGC AAACCCAGGT GTGGTTAACA AGTTATTGCA 6840

AGAAGAGCTT TCAAAACGAT AAGAAAGTAG CAGGCAAACC ATGAAAAAAG CTAGAGTGAT 6900

TTATAATCCA ACGTCAGGAA AAGAGTTAAT CAAAAAGAAC TTAGCCGATA TTTTATCTAT 6960

TTTAGAAGAA TGTGGTTATG AAGCCAGTGC ATTTGCGACC ACACCAGAAG AAAATTCAGC 7020

ACGCAATGaA GCACATCGTG cTGCGCGGGC AGGATTTGAT TTACTAGTAG CTGCAGGTGG 7080

AGATGGGACC ATTAATGAAG TCGTGAATGG GATTGCTCCG TTGAAGCGGC GCCCTAAAAT 7140
GGCTATTATT CCTGCTGGAA CGACGAATGA CTATGCACGG GCCTTGAAGA TTCCTCGTGA 7200
TAATATCGTT AAGGCAGCAG AAGTGATTAA AAAAAATCAA ACTGTCAAAA TGGATATTGG 7260
CCAAGCCGGC AAAAATTACT TTATCAATAT TGCGGCGGGT GGTCATTTAA CGGAACTGAC 7320
TTATGAAGTT CCGTCAGAGT TGAAAAGTAT TTTTGGTTAC TTAGCGTACT TAGCCmAAGG 7380
AGCCGAAATG TTGCCGCGAG TGAAGCCGAT TAAAATGCGC ATGACGTATG ATGAAGGTGT 7440
GTACGAAGGC AATGCATCAA TGTTTTTCCT AGGGCTGACT AACTCGGTGG GTGGTTTTGA 7500
GCAAATCGTA CCAGATGCTA AATTAGACGA TGGTAAGTTT TCATTAAksA TTGTAAAAAC 7560
AGCCAATATT TTTGAGATTC TTCATTTAGT TGCGTTAATG TTAAATGGTG GAAAGCATGT 7620
TGAAGATCAT CGACTGATCT ATACAAAGAC CAGCTATTTA CATGCAGAAA CGTTAGAAAA 7680
GAACAATAAA ATGATGATTA ATTTAGATGG TGAATATGGA GGCGATGCTC CAATGACCTT 7740
TAAAAATATG CATCAACACA TTGAAATTTT TGCAAATGGT GATGCACTGC CGTCCAATGC 7800
AATTATGGGT TCTGTCTTAA CTGGTAGCGA TGAGATTGTC GTAGAATCAG AAGACGAAGA 7860
GGAAGAAGCA TATAmCGAAG CCAGCAAAGA ATTTGTCAAA GAAGTTGAAC GACTAACAGA 7920
CGAAGATATT GATGGCGATG GAAAGATTGC GGAAAAAGAA AAGCACTAGG TAGTCCGTGA 7980
ATAGCTAATG AAAGATTTAG TTGATAAGAT CTTTCATTGG CTATTTTTTG ACTAAGGGAA 8040
AGGGTCTCAT TCTTTTATTT ACCCAAAAAG CTCAAACGGA ATTGCTAAAT CGCCATGrAA 8100
ATTCGTTGTT GCTTTTCTTA GGTATTTTCA TTAGAATGAG TGAAGTGTAT TAAAAAGAAG 8160
GAGCGGACTT ATGAAAAATT ATCCCAGTAA AGAAAAATGA CGTCATTGAA GTAGAAATCA 8220
TTGACTTGAC GCATGAAGGT TTAGGCGTGG CAAAAGTGGA CCATTATCCA TTATTTATTG 8280
AAAATGCCTT ACCAGGTGAA AAACTTGAAA TCAAAGTCTT AAAAACAGGC AAAAGTTTTG 8340
GTTATGGGAA AGTATTGACT GTCCTAAAAA GCAGTGAACA GCGTGTGCCC GTGAAAGATG 8400
AAAACTTTAC AAAAGTCGGG ATTAGTCCGT TGCAACATTt AGCGTACGGs GcCCAmCTAT 8460
CATTTAAAAC mCAmCAAGTA GAAAATGTTA TGCAACGTGT CGCTAAATTG CAAGAAGTCC 8520
CTGTTTTACC AACAATTGGG ATGAACGATC CGTGGCATTA CCGGAACAAA GCACAAATTC 8580
CCGTGCGTAA AATTGACAAC CAACTACAAA CAGGTTTTTT CCGGAAAAAT AGCCATGATT 8640
TAATTCCGAT GGAACATTTC TATATTCAAG ACCCAGAAAT TGATGCAGCA ATTGTTAAAA 8700
TCCGTGACAT TATGCGTAAA TATAGCGTAA AACCTTATAA CGAATCTGAT AATACAGGGA 8760
ACTTACGCCA TATTGTCGTT CGTCGTGGCT ACCATACTGG GGAGATGATG GTTGTTCTAA 8820
TTACGAGAAC ACCGAAATTA TTCCCAATTA GTAAAATTGT GCCAGATATT TTGGAAGCTA 8880
TTCCGGAAGT AGTCAGCATT GTACAAAATG TAAATCCAAA ACGAACCAAT GTCATTTTTG 8940
GCGACGAAAC CATCTTATTA CATGGGTCTG AAAAAATTAC AGATACTATT TTTGATTTGA 9000
AATTTGAAAT TTCCGCGCGT TCATTTTATC aAGTAAATCC ACaACaAACA GAAGTGATGT 9060

ATCmAAAAGT TAAAGAGTAT GCGGCTTTAA GAGGCAATGA AATTGTGGgT GGaTGCCTAT 9120

TGCgGGaTTG GaACGAnTGG cTTAACTTTA GCCCAAGATG CCAAGCCAGT TTACGGTATT 9180

GAAGTAAnTG GAGGAAGCGG TCCAAGATGC CG 9212

(2) INFORMATION FOR SEQ ID NO: 335:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1230 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 335:

CCCTGGGAAA TTGGnCCATT AAAATGGAAA TnAAGAATTC TACAGCATTA CTGTGTAAAA 60

GGACATATTT AGCAACATCG CAATGATGAT TCCGATTAAG GCACTATAAG CGGCATGACC 120

AACTGCTGAT AGGTCTTTCT TCGTAAAGAC ACCTACTAAT GACATCGCGC CAAAGGTAGC 180

TGCTGCACTT AAAAATGCGC CGGCCACTGT TCCACCAGTA TACATTAATA ACGTCACGGA 240

TAAAGTGACA CCATTTAATA ATGCATACGC AACAAATCCT GCAATCGCAA TGGTTGGATT 300

TTTCTGTGCT TTTGCGCCTA GATATAAAmC TAAGCCaATT TCAACAAGCC AAATCATCCA 360

AAAAGCTAAG CTTGATTGGA ATAAAGTAGC TAGAACAGGT AAAAAGACAG TTGTCATTGC 420

ATAAGCCATG ATTGCGCTAA TACCAATTCC TAACGCTAAA AATCCGTAAA TTTTGGCATA 480

AAATTTATTC AAACCTGTGC CAGTTTGGAT ATACTCTTGA TTATTCATAA GTnAAAGTCC 540

CTTCTTTCTC TGTTAGTTCA TCAATCGTCA CTTGGGGTGG TTGAACCATC ACCACTGCAT 600

CTAACACGCG ATTTTCTTGC TCATTAACAG CCTCAACTGA AATGGTAATC ACATCTTTCA 660

TTTTGTCGAC TTTGATTACT TCTAATTGGA AAGTCAATGT TTCGTAATGA AAAACGGGTT 720

CCACAAAGTT CaCAGAGAAA TTCaCAACAT GTGaACCTGG TCCTGGTAAA TGTTTAGAAA 780

TCGCACTAGT GATAATTCCC ATCAGCATAA TTGATGGTAC GATTGGTTTT TCATATTCTG 840

TTTTTTGGGC ATAATCATGT TGGATATATA AAGGATTCgC ATCATTGGTC AAGCCCAAAT 900

AAAGTAATAA GTCCTTGTCT TCAATTGACT CTGTCAAGGA CAACGAATCT CCCTCTTCGA 960

TATCTTCGAT GGTTTTGCCT AATTTTCTCG GTTTTCCGAT ATTCAAATTA GTACACCTCC 1020

GTTATTGATA AGATAATTGT ACCAAAATGT ACTTAAAAGG CAAGCTAAAA AGTAGTCTTT 1080

AAAGGATGTT TAAACAAAAC CAACTTATCT AcCATGCtAG CATTCTAAAA AAAACAACGG 1140

aACAGGaGCA GCTTCTTATT TtTTGAAGCC AGATACCCCT GGnCCCGATT TTGGTTATGT 1200

CCCCTATAAA ACGGTCnAGT nGGGTACCTA 1230

(2) INFORMATION FOR SEQ ID NO: 336:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 920 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 336:

GTCACAAAAG AnGTTTCAGT CAATACTGAA ACTGGTTTAA CACAACAAGA GGCCCAACAA 60

CGGCTGAAAG AAAATGGCCG CAATCAATTT GAAGAGGCTA AGAAGGATTC TGTGTTAAAG 120

AAATTTATTC ACAGTTTATC TGATTTTACA ACGATTATTT TATTAGTTGC TGCGGCAATT 180

TCGTTTTATA CAGCAATCGT AACAGAACAT GGCGAATATT TTGAAGGAAT CTTAATTATT 240

GCAATTGTGA TTATCAATGC TGTTTTAGCC ATTGTTCAAG AAGGAAATGC TGAAAAATCT 300

CTGGCGGCTT TACAAGACAT GAACAAGCAA AGTAGTGCAG TCTTACGAGA TGGCAAGGTA 360

ATCGAAGTCG ATGCTGAAGA ACTAGTAGTA GGAGATGTTC TAGTTCTTGA AGCTGGTTCT 420

ATGATTACAG CCGATGCACG ACTTATTCAA GCTTCTCAAA TGCGCGTAGA AGAGTCmGCA 480

CTAACaGGTG AAAGTGAACC TGTTGaAAAA GATCCTACAT ACGTGGGGcA TGACGATGAT 540

GGCTTAGGaG ATCAAATAAA TATGATTTTC AAAgGCTGTA CCGTTGTGAA TGGCCGCGGA 600

CGAGCAGTAG TAACTGCAAC AGGAATGnAT ACGGAATGGG TAaAtTGCTG GGTTACTGAA 660

TAACAGCGAT CAGCAAAAAc aCCATTACAa AAACGCTTGa ATCmATTrGG CmAACGTATT 720

aGTTTgTTAG CCtTAGGGGC TGCAGCAATT GTTTTCATTA TTGGTGAACT GCAAGGAGAA 780

CCGCTTCTTG AAATGTTTAT GACAGCCGTt TCCTTAGCGG TGGCAGCTGT ACCAGAAACA 840

CTAACCGTGA TTGTAACGTT AACCTTAGCA TATGGTGTTC AAAAAATGGC nAAGAAACAT 900

GCGATTATTT CGTTCGATTG 920

(2) INFORMATION FOR SEQ ID NO: 337:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2381 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 337:

AATACTATAA AACCTTTTTA TACATTTGTA CACAAAAAAC CTCACTTTCT GAGGnAGATT 60

TTATTATTTA GCATTGCCAA AAAGTCAAAA GCAAGAGCCC CCAAAACTTC TTTGTTAGGG 120

TTGcCTTGAA AATAAAGAAA AACAAAAAGC CGAAAACGCT AAAGACGTTT TCAGCTTTTT 180

GTTTTTTaAT GGrATAAATG GcTCTTTTGT GCTAAAGCCG CAAATTTTTT CGATCTGTGT 240

TCCAGATGTG GATAAAGAAT TGCACCGCCA ATACTAAAGA CGACTGAAAT CCCCACCATA 300

ATCCAAATGG CAAACCAAGC ATTTGGCCAG TAAATGCCTC CTGCTGATTC CCTTAGCAAG 360

TTCACGGCAT GCGTAAACGG CAAGAACGGA TTAATCATTT GGAAGAATTT ACCTGAGACT 420

TGAATTGGGT AGTTCCCACC CCCACCAGAG ATGGACAGTA CCAAGATTAT TATCGCAATT 480

CCTTTCCCTA CATTGCCAAA GAGAGCAACA AGCACATAAA CCATTATCAT AAACGTAATG 540

GCAATCAATA CCGCAAACCA GACGCTATAC GCTGGATTAC GAACATCTAC GCCTAACGCA 600

AAATAATTAC CTAAGGTTAC AATCAGTGCT TGACCAATCC CCATGACAAT GAAGGTAAAC 660

ATCCGTGCTG AAAATTGTTC CCGTTTAGAG AAGCGTTTTT TATCTTTTCC TTCAAGATAA 720

ACATCCGTTG TCGCAACACT TGAGAATAAT ACAGCACCAA CCCATAAACA CAAAGCTGTA 780

TAGAATGGCG TACTTGCCGA ACCATTATTA GCAATCGGAT AAACAGCATG TTCTTTGACT 840

TCAACTGGTT GTGTAAAGAA GTCACTTTCT TTATTGGCAT CTAATTTTAG TAATTTAAGA 900

ATTTCGCCTA AATCAACTTC TTTTTCACCT TTACGGATTG CATTCGCTGC TTTGTGTAAT 960

CCTGTTTTGA TATTTGGCCA ATCATTAATA ATTAGCTCGT TGGCTTTGTC CAAGGCCGCT 1020

TCAACATCTG GCATTTTCTC ATTGACTGTT TTCAAGGTAT TCGTTAAATC TTTTCGGATA 1080

CCTGGATAGT CATTTTGCAT AAAGGCTGCC GCTTTGCTGA CTTTATCTTG AATGACTGGT 1140

AAATCGTTTT TATACAAATC AGCACCACGA TTAATGCCAT TGACAATGGT TTCCATATTG 1200

CCATTCAACA TCGTATTCGC ATCATGAATT TCTTGTTTAA TTGCCGGCAT TTCGGCTTGA 1260

TATTTTTCTA ATAAAGAAAT CGCATTAGTG ACCGTTTGAC TTGTTGAGCT CAATAAACCT 1320

TCAAAGTCAA TTTGTTGCGC TTTGTTTAGT TGACCTTGAG CGkTTTGGgA TGGTAGCAAT 1380

TAATTTGTCT AAAATCGTTT TTACTGTGTT TTCAATTGCA TCAGGGGTAA TACTATTGAC 1440

GGTTGCTCTA CTTGTTGAGC TcTGCTTGAA TAGCATTTAA AtCTGCATCA GTTACCGCAT 1500

TCACATCAAT GTTATCAACA TCAGTTTTTA ATCTAACTAC AATACCATCA ATCGTATTCA 1560

ATCTATTTAT TGTGTCTTGT AAATCAGTGT TACCTGTGGC TTCTTGTATT TGCGTAAACA 1620

GTTGAATCAG CTGTTGGATA TTTTCATGTT GTTTTCCTAG GCTTTCAGAC AGTTGTTGCT 1680

TAATTGCTTC TCTTTCTTCA ACTGGTAAAC GATGATTTTT TAAATCTTCT AGTGCTGTAT 1740

TTACACCAGC GCTAATTGTT CCAATTGCTT TTAAAATCGT CCCCACACTT TGAGTAATTC 1800

CAGGTAACGC CTTTTCCAAT TGAGTGGCGC CATCTAATGT GACATTCCCT AAATCATTGG 1860

CTTGATCACC TAGTTTACGA ATATCTGGTA ACGCCGTTTG AACTTGTTGA ATGATTTCTA 1920

ATCCTTGTTT TGCTTCTTGG ATTCCTTCGC TCATTGTTTG TTCCACAGAG GCAAAGTCTT 1980

CATCAATCAT AGCAATTTGT CGGCCAGCGT TTTGAATTTC TGGAATTTTT TCTTGTAACG 2040

TCAAAATAAC ACTAGCTTGT TCTTTGATGC TTGGCATTTT ACCATTTAGC TCAACAATCT 2100

TCTCTCCTAA AGCATCAACT TCAGGTAAAT ATTTCATCGC ATCGTTGGCT TTCGCTAATT 2160

TTTCTTTTAG CTCAGGCATC TTACCATGTA AGTCTGTGAC TTGTTTTGCG TACGTATCAA 2220

TGGTTCCAAT GTTGGCATCT GTATCTAAAA TCATGctCTT AACTTTTtGA ATGCTAACCA 2280

nGTTTTTATC AATAtCATAG cCAATAayCA TkGAAgGTTT gaTaACGtAc TACTTGcCGG 2340

tTTwAAATGG AAATCCTTCC CGAAAATTTT GGTGGAACCG G 2381

(2) INFORMATION FOR SEQ ID NO: 338:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13154 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 338:

GCCAGATTAT CGTTCCTTAA AAGGTGTCTA TCAAGGAATC CAGCAACCGG aATACTTATT 60

GAGCCGGACT TGCTTAAAAA CAGAGCCAGA AAAGTTTTAT CAATTCGTTA AAACGCTCTA 120

TCATCCAGAC GCACAACCTA ATATCATTCA CCAGAAAATG GCCCAATTAG AGCAAATGAA 180

ACGGGGAAAA ATTGTTTCAC AAAATATTGA TGGGCTTCAT CGAAAAGCAG GTAGCCAAGA 240

AGTGGTGGAC TTTCATGGTA ATTTATATGA ATGTTATTGC CAAACTTGTG GTGCAACGGt 300

TCCTTGGCAA GACTATTTGC TTAGTGATCG GCACGCTGAT TGTCATGGAC AAATCCGTCC 360

AGCTATTACT TTGTATGAAG AAGGACTATC CGAAGAGGCA ATTGAAAAAG CCATTCAGGC 420

CGTTGCTTCT GCCGACTTAA TCGTGATTGT CGGCACGAGT TTTcAAGTGC ATCCATTTTG 480

CGATTTGATT CATTATAAAC AACCAACGGC AACTATTTTA GCCATTAATC AAACGCCGCT 540

ATTTTTACAA CAACCTTATT ATTTTTTAGA AGCGAAAGCC GAAACGATAT TTGCAGAATT 600

AACGATAAAG GAGTAACGGA ATGGAAGAAT TTAAATCAGA AAAAGAAAAA ATGATCGCAG 660

GTGCACTATA TTTTGCCTCA GATCCTGAAT TAGTGGCTGA TCGCAAAAAA GCCCGTGAAC 720

AAATGGCGTT AATTAACCAA CAACCAGACA CTTATATTCG TCGTCAATTA ATAGAAGAAA 780

CGTTTGGAAA AGTGGGTGTC GGGACGTACA TCGAACCGAT GATTCAATTT GATTATGGCT 840

ACAACATTTC TGTGGGAAAA AATTTCTATG CCAATTTTAA TAACGTTTTT TTAGATGTTT 900

GTCCAATCGA AATTGGCGAT AACTGTATGT TTGGACCCAA TGTGCAACTT TATACAGCCG 960

AGCATCCATT ACAAGCGGCC AAGCGTAATA GCGGAATGGA ATCAGGGAAA CGAATTATCA 1020

TTGGCAACAA TGTTTGGATT GGCGGTGGAG CGATTGTCCT TCCTGGCGTA ACATTAGGCG 1080

ATAATGTAGT GGTGGCAGCA GGTGCTGTTG TGACCAAATC GTTTCCAGAA AACTGTGTGA 1140

TCGCAGGTAA TCCAGCACGG ATAATTAAAG AATTGACTGA AGATGACGCC CCAACAACTT 1200

CTTTAGAACA ACAACGAGCA AAGATTAACC AAATCGATAA GGAACTGGTC CGTTTATTAG 1260

AACAGCGGAT GGACGTCGTC GCAGAGATTG CTGCAGTGAA AAAGAAAGCA GGTCATGCGG 1320

TTTTTGATTC TGAAAGAGAA CAACAAGTAT TAGAAACTAT TCTGAATCAT GTTGAAAATG 1380

CCGAGTATGA AGAAACATTA TCTGAGACTT TCCAGGGAAT TATGGACGCC TCTAAACGTT 1440

TCCAAGAAAA ACATTTAGGA GAGTGAGGCG AAGGTCGTGG ATGGAGAAAT CAGAAGAAGA 1500

GAGATTGTTA AAAGATTAAG AGAGACCGAA AAGCCAATTA GTGCCACTCG TTTTGCAAAA 1560

GCTTTTGATG TCAGCAGACA AATTATTGTA GGAGATGTCG CTCTATTACG GGCGACAGGG 1620

GTCGAAATTG TCGCAACAGC TCGAGGGTAT ATGTTGGAAC AACCTTTAGA AGGAATTGAA 1680

CGGAAAATTG CTTGTCAACA TACACCTGAA CAAACGAGAG AAGAATTGTC AGTCATTGTT 1740

GCCAAAGGTG GAGAAGTGGT AGATGTGAGT ATTGCACATC CGCTTTATGG GGAATTAATC 1800

```
GGCAGTTTGC GCATTCAATC AGAGAAAGAT ATTGATAAAT TTATGGACAA GTATCAAAAA    1860
GAGCATGCTA CATTACTATC AGTTTTAACA GGAGGCGTCC ATCTGCATAC TATTCGCTGT    1920
GCAGACGAAG AAACTTTCCA AGAAATCAAA GAGGCACTTC GAAAAAAAAG TATTTTGTTT    1980
GAAGGATAAC GAGAAAAAGA GAGGACGTTG ATGTTTGTTC AAGCCCTTTC TTTTTTTTCT    2040
TAGAAAATTA CCAAAAAAAG ATAGAAGAAT TTGTGAATTT TCTTCTATTT TTTTAACTTT    2100
TTGCTAGAAA TTTTTTATGA AACAGTATAC TATTAAAAAA GGAGCCATTT AAACAATAAC    2160
TATGTCAATT ATATATGTTG GTACATAAAA aTAAATGGAG AAAAATAAAA AAGGAGAATG    2220
AGTATGAAAA AACAATTTGC CCTGTTCTCT ACCTTATTAC TTTTAGGTTC ATCAATGCCA    2280
GTAGGTGCAT TTGCAGAAGG TAATAACTTG GTAGCTGAAA ATAGTTCTGT CGCAGAAACA    2340
ACCGCAGAAG CAACAACTGC TGAAGCCACT GAAACGACAG CGACTAGTGA AACAACAGAA    2400
GCGACAGAAG AATCAACAAC TGAAACAGAA AGTTCAACCG AATCAAGTGA ATCTGCAACC    2460
ACGGAGTCCA CGGAAACGTC AGGAACCGAA ACGACAGATT CCACAACGGA TTCTACATCC    2520
ACAAGTACGA CAGAATCGAC TACTGACTCC ACCTCGACGA GTACAACAGA GTCAACCACA    2580
GATTCTACGT CAACAAGTAC AACTGAATCA AGTACCACGC CAACGACCAC ACCAAGTTCT    2640
TCTAAAGAAC AGCCAAAACC TGGCACGAGT ACTTCTGAaT CAAAACAGCC GGCGAAACCA    2700
GTGACACCAA CAGCGCCAGC TGAAAAGCCA GTAGAACAAC CAGCAGCGTC AACACCGCAG    2760
CCTGAAATAG TCCCACCAGT TACAAATGAA ACGGTTGGGT TAGTGGAAGA CGATGAGACG    2820
TTTACGGTCA GCAAAACGAA GAAGACAGAA GAGTTTATTC AAGAAATCGG TGAATCTGCT    2880
CGGAAAGTGG CCAAGGACAA AAACTTATAT GCTAGTGTGA TGATTGCTCA AGCAATTTTA    2940
GAAAGTGGCA GCGGCAATAG TAAATTAAGC CAGAAACCAA ACTACAATTT GTTTGGTATT    3000
AAAGGCGATT ACAAAGGTCA ATCTGTTTCT TTTATTACCT ATGAAGACAA TGGTTTCGGT    3060
AACCTATACA CAGTAGAAGC AAAATTCCGT CAATATCCAA CGTATAAAGA ATCAATGGAA    3120
GATTATGCGA AATTATTGAA AAATGGTTTA GATTCCAATA AAGATTTCTA TCATGGTGTC    3180
TGGAAAACGG AAGCCAAATC ATACAAAGAA GCGACACGTT TCCTAACAGG AAAATACGCA    3240
ACGGACAAAG ATTACCATAA GAAATTAAAT GCGTTAATTA AAACATACGA TTTAACTTAT    3300
TATGATAAAG AAAAAGCAAC CGTTGAACCA ATGGAATCAA ATTTCCCAGC CTATAATGGC    3360
AAAAACTATG ACACCTTTAA TAGTTATGCT TGGGGCAACT GTACACAATA TGTCTACAAC    3420
CGTATTACGC AATTAGGTAA ACGTGTGGAC TTAACAATGG GGAATGGTCA AGATTGGGGC    3480
GAAACAGGTC GTGCACGCGG GTACAAAGTA AGTCGTACAC CAAAAGCAGg TGCAGCGGTT    3540
TCCTTCCCAG CTGGTGTTTT AGGCGCAGAC AATACTTATG GTCATGTCGC ATTTGTGGaA    3600
AAAGTCTTTA AAGATGGTAG CATCTTAATT TCTGaAATGA ACGTGAAGGG ATTAAACGTT    3660
GTTTCGACAC GTACAATCTC AGCCGATGAA ACACATTTGA TGAACTATAT CGTTCCAAAA    3720
GATAAATAAA AAAATCAGCT CTTTGTCAAT AAGAACGAAT GAGTTGGAAA AwTCAAAGCT    3780
```

AGCACTGAAG AAGCAATTTT AATTCTTCAG TGCTAGCTTT TTTTTGTGGC AAAAGTTCTC 3840

TCCAATAAAG TTTAATACAT TCAGATATTG CCAGGAGGCT TCAGGTCGGG ATGAGTTAGC 3900

AACAGTCGCT GAGAGCTATA TTTTTTTTAG GTAGAATCAT TTACTTTTAG GTGGAAGATG 3960

TCTTATAATT GTAAGCGGAT ACAAAAATAA AAAAAGAGGT GGGTAAATGA GTAAAAAAGA 4020

AATAAATCAA GTAGTTGCCA GTAGCTATCA ATTGTATATT AATGGAGAGT GGACAACAGG 4080

TAGTGGTAAC AAAATGATTG CTAGTTACAA TCCTAGTAAT GGCGAAAAAT TAGCAGAATT 4140

TGTAGATGCC ACAAATGCAG ATGTGGATCG AGCTGTAGAA GCAGCCCAAG AAGCGTTTCA 4200

GACATGGAAA GATGTTGACG TTGTAACAAG AAGCAATCTT TTGTTGAAAA TTGCTGATTT 4260

GATTGAAGAA AATCAAGAAC ATTTGGCTAT GGTGGAGACT TTAGATAATG GAAAACCGCT 4320

TCGGGAAACG CAATCGATTG ATGTCCCTGC CAGTGcAGAC CATTTTCGGT ATTTTGCTAG 4380

TGTGATTCGT GGAGAAGAAG GATCTGTCAA AGAATTTGAT AAGGATACGT TATCCATTGT 4440

TGTGAAAGAA CCCATCGGTG TTGTGGGTCA AATTATTCCG TGGAATTTCC CCTTATTAAT 4500

GGGTGCTTGG AAATTAGCAC CAGCGTTGGC AGCAGGTAAC ACGGTTGTGA TTCATCCGTC 4560

CTCGAGCACA TCATTAAGTC TGTTGGAATT GTTTAAAATT TTTGATCAAG TCTTGCCGAA 4620

AGGAGTAGTG AATTTAATCA CCGGTCGTGG TTCTGATTCA GGAAATTATA TGTTGGCACA 4680

TCCAGGGTTT GATAAGCTAG CTTTTACAGG CTCAACAGAG GTGGGGTACA CTGTCGCTAA 4740

AGCGGCGGCC GACCGCCTAA TTCCAGCCAC TTTAGAACTT GGCGGAAAAT CAGCCAACAT 4800

TATTTTTGAA GATGCCAATT GGGAACGTGC ATTAGAAGGC GTGCAGTTAG GGATTTTATT 4860

CAATCAAGGG CAAGTTTGTT GTGCTGGGTC TCGTGTGTTC GTTCAGTCAG GTATTTATGA 4920

TCAATTTGTA GAAGCTTTAA AGGAAAAGTT TGAACAAGTG AACGTTGGTT TCCCGTGGGA 4980

AAAAGATGTT GAAATGGGCG CTCAGATCAA TGAGCATCAA TgGAAGAAAT TTTAAAATAT 5040

GTCGAAATTG GTGTGAAGGA AGGAGcTACG CTGATTACTG GTGGGCAACG TTTAACAGAA 5100

AATGGGCTAG ACAAGGGGGC GTTTTTAGCA CCTACGTTAT TAGCGAATGG TACGAATGCA 5160

ATGTGTGTGG CCCAAGAAGA AATCTTTGGT CCTGTTGCAA CAGTGATTAA ATTTGAAACG 5220

GAAGAAGAAG TCATTCGTTT AGCTAATGAT TCTGAATATG GTCTAGGTGG TGCCGTCTTT 5280

TCTCAAGATA TCAATGTGGC ATTACGGGTT GCTCGTGGTG TACGGACAGG TCGAATGTGG 5340

GTCAACACAT ACAATCAATT GCCTGCGGGC GCGCCATTTG GCGGTTATAA AAAATCAGGA 5400

ATTGGTCGAG AAACGCATAA ATCAATGCTA GATGCTTATA CGCAAATGAA AAATATTTAC 5460

ATTGTGACAA AAGAAGAAGC AGATGGACTG TATTAAGTAG TGTAGTGAAT TCTAGATAAT 5520

AAACTTATTT TTGGTTATTG TTATTTTTCG AGCTAAACAA AACAATAATT TTTTTAAAAA 5580

GTTGAGTGAT GAGTAAAGTA AAACAGATTA AAAGTAAAGC CTGGTACAAA AATCACTTTG 5640

GATTTTTGTA CCAGGCTTTA AAGCGGAGTA ACAATAGTGT TCACTCTTGG TTGTTTTTAA 5700

TTTTTGACTA TTGTAGAAGG AGTTATTTTG CTTCTTCTGC TTCAAATTTA GCAAAAGCTG 5760

TTTCAATAAC TTCTTTTAAT GCAGAAGCAG AAGCTTCCAT TTTTTCTTGT TCAGCaTCTG 5820

TTAATGGAAT TTCAATGACT TGTTTAACTC CtTGGCGGTT GATGATCGCT GGTGcACCGA 5880

TATAAATATC GTTTTGACCA TATTCACCTT CTAAATAAAC AGATAATGGT AACACAGAGT 5940

TTTCATCGTT TAGGATAGCT TTAGTGATAC GCGCTAGTGC AACCGCGATT CCATAGAAAG 6000

TAGCTCCTTT TTTCTCGATG ATTGTGTAAG CAGCGTCGCG TACGTTGAAG AATAAATTAA 6060

CCATTGCTTC TTCATCGACG TCAGGATTAT TTTTCACCCA TTCGTAAATT TGTAAGCCAG 6120

CGACATTCGC ATGTGACCAA ACTGGGAATT CTGTATCTCC GTGTTCCCCT AAGATGTAGG 6180

CATGGACATT TCGTGCATCA ACGTCAACTA ATTCGGCAAT TGCTTGACGG AAACGAGCAG 6240

AATCTAGTGA AGTTCCTGAA CCGATTACTC GTTCTTTCGG GAAGCCAGAG AATTTCCAAG 6300

TTGAATAAGT CAAAATATCA ACTGGGTTTG CGGCAACTAA GAAGATACCG TTGAAACCAG 6360

AATCAACAAT TGTTGTAACG ATTTCTTTAT TAATTTTCAA GTTTTTATGA ACTAAGTCTA 6420

AACGAGTTTC GCCTGGTTTT TGAGGCGCAC CAGCTGTTAA GACAACTAAG TCTGCATCAT 6480

GGCAATCGTC ATAAGTAGCA GCATAGATTT TTTTAGGAGA AGTAAATGCT AATGCGTGAG 6540

ATAAGTCCAA CGCGTCTCCT TCAGTTTTTG GTACATTAAT ATCAATAATC CCAACTTCTT 6600

GAGCAATATT CTGAGTTACT AAAGCAAAGG CATAGCTAGA ACcTACGGCA CCGTCCCCGA 6660

CTAAAATTAC TTTTTGGTGA TCTTTATTCC CTGCGGCTGC AGTCATGTGT ACCATTCCTT 6720

CCTCTACATT CTTTTTCGTG AGCAACGAGT TTTTATTTCC TGTTGCTAAA GTGATAGTAA 6780

CATGGTTTGT AAAAAATGTC ACGAACTTTG ATTGTAAAAA TGATACAGTT TTCAGACCTG 6840

TTTCAAAATG GTAGCTCTTA CATGTGCGTA ATTTCACTAT CTTACGCATA TTTAAATTTC 6900

TAAAAATTTT TTTACATCTT TCTTAATCTT CAAAGACTGC GAGTTTTTTT ATTTCTATGG 6960

TATAATGAAT AAAACTGGGT AGCGCATACG CTTCCCAGCC ATTTTGTGTT GGCAAGCTTA 7020

TAATCCAAGG AGAATGAAAG ATGAAAGTAA TCGTAGGACT GGGAAATCCC GGTAGTAAAT 7080

ACAAAGAAAC AAAACACAAC ATTGGCTTTA- TTACGCTAGA TGAAATTGCG TATCGGCAAA 7140

ATGTTTCGTT TAATAATAGT AATTTTGAAG CAGATATTGC AGAGTTTTTT ATAGGGACTG 7200

AGAAAGTATT GTTAGTGAAA CCATTAACAT TTATGAATGA ATCAGGTCGC TCTGTTGGTC 7260

CGTTGTTGAC CTACTTTGGG GTAGATGAGG AAGATTTGAT TGTTATTTAT GATGACTTAG 7320

ACTTAGAGAT CGGTAAAATT CGTTTACGGC AAAAAGGTAG TGCTGGCGGC CACAATGGGA 7380

TAAAAAGTCT GATTGCCCAT TTAGGAACAA ATGTCTTTCC TAGAATTAAA ATTGGCATTG 7440

GtCGGCCCTC aAAAAATGAT ACAGTTATCC ACCATGTACT AAGTACTTTT CCCAAAGAAA 7500

CACACGAAGA GATGTTACTT GCTGTAAAAA AAGCAGCCGA TGCCGCTTTA TATGCATGCG 7560

AAGGACATAC GTTTGTTGAA ACAATGAATC AATTTAATGG AAAATAGCAG AAAGGAGGGC 7620

CCTTGTTTGA ATATTATCGA ACGAATTAGC CAAACAGAAA TTGCTAAATC ATGGCGCAGT 7680

CAACTAATGA CAACAGGAAC GCGCCAACTA ATTACTGGGT TATCAGGCTC TGCTAAAACA 7740

```
TTACTGATGG CAGGCGCTTT ACAAAAGACG CATCAAAAAA TTGTTATTGC GGTCCCAAAC     7800
CTATACTATG CCAATCAATT AGTAGATGAT TTTTTAAATA TTTTACCTTC AGAACAAGTT     7860
CATCTATTTC CAGTAGATGA AGTGTTATCC GCCGAAATGG CCTTTTCTTC ACCAGAAGCG     7920
CGCGCAGATC GAGTGAGTGC GTTAAATTTC CTGATGACAG CGAATGCAGG AATCGTGGTC     7980
GTACCAGTAG CTGGACTACG AAAATATTTG CCAACGAAAG AAACGTGGCA AAATGCGCAA     8040
CTTTATTGGG AAATCGGGAC AGAAGTAGAA CCTGAGATCC TCGCGCAACG CTTAGTGTTA     8100
ATGGGCTATG AACGCCAAGC AATGGTCGGA AAGCCTGGCG AATTTAGTAT TCGCGGAAGC     8160
ATTGTTGATG TCTATCCGCT AAATGCCGAA TACCCTGTTC GGGTGGAACT GTTTGATGTT     8220
GAAATTGACT CAATCCGTTA TTTTGAAGCA GATACCCAAC GTTCATTAGA AAATCTTGAA     8280
GAAGTAACGA TTTCACCAAT GACTGATTTA GTATTCTCAA AGGCAGATAT GAATCATGGC     8340
ATGACGAGGT TACAAGATGC TCTAGAAAAA CGTTTAGCGA CAgCTAAAGA TCAAACGGAG     8400
ACCGATTTTT TAGAAGAATA TTTTGGTCAG TTGCTGTCTT CTTGGGAACA AGGGATTCCT     8460
ACAGAAAATG CTCACTATTA TACAGATTTT CTGTACCAAC AAAAAACTAC GCTGTTAGAT     8520
TATTTACCAG AAAATAGTTT GCTTTTTGTG GATGATTATT CTCGTATGAT GGAAACGGAG     8580
CGAGAAATCG CTAGAGAGGA AGCGGAATGG CAAACCCTGA AAATTGAAGA AATGCGCGTG     8640
TTTTCTGAAC AAACTTTTGG CGTAGACTTT CATGATCAAT TGCGAAAAAC TGCACGAAAT     8700
ACGACCTTCT TTTCACTTTT TCAAAAAGGA ATGGGGAATC TTCGTTTCCA AGAAGTCCAT     8760
AATTTCCAAT ATCGTTCAAT GCAACAATTC TTCGGTCAAA TGCCGTTGCT AAAAGCGGAA     8820
GTGGACCGTT GGCAAAAGCA AAACCAAACG GTCGTTGTTT TTGTTCCTAC GAAAGAACGG     8880
ATTCAAAAAG TCGAAGAGCT ATTTCAGGAC TTTGACATAC CTAGCGTAGT AAGCAATTGG     8940
GATCAGTTGT TAGATGGGCA TGTCCAAATT GTCCAAGGGG CATTACAAAC AGGGTTTGAA     9000
TTGCCTAAAA ACAAGCTCGT GGCGATTACG GAAAAAGAAA TTTTTCATAC AACGACCAAA     9060
AAACGGGCAC GTCGTCAAAC AATCTCTAAT GCGGAACGTT TGAAAAGTTA TAGCGACTTA     9120
AAAACGGGCG ATTATGTCGT TCATGCGAAT CACGGGATTG GTAAATATAT TGGTATGGAA     9180
ACCTTAGAAG TTGATGGGGT CCATCAAGAT TACATGACTA TCTTGTACCA AAATGATGAC     9240
AAGTTATTTA TTCCAGTAAC ACAATTGAAC TTGATTCAAA AGTTCGTTGC TTCCGAGTCT     9300
AAAACGCCGA AAATTAACAA GCTAGGCGGC AGTGAATGGA CGAAAACTAA ACGAAAAGTT     9360
GCTTCGAAAA TTGAAGATAT CGCGGATGAT TTAATCTTAC TTTATGCGAC AAGAGAGTCG     9420
GAAAAAGGCT ATGCTTTTCC GCCAGATGAT GCGTATCAAA AAGAATTTGA AGAGGCTTTT     9480
CCGTATTCTG AAACAGATGA TCAATTACGT AGTGCCGCCG AAATTAAACA CGACATGGAA     9540
AAATCTCGAC CAATGGATCG TTTGTTAGTT GGCGATGTAG GGTTTGGTAA AACAGAAGTG     9600
GCTTTACGAG CGGCTTTCAA AGCTGTTAGC AACAATAAAC AGGTTGCCTT TTTAGTGCCA     9660
ACAACTATTT TAGCGCAACA ACATTATGAA ACGATTCAAG AGCGGTTTGA GGGCTTTCCT     9720
```

```
GTAGAAATTG GTTTATTAAG TCGTTTCCGC ACGAAAAAAC AGCAAAATGA AACAATTGAA      9780
AAAATTAAAC ATGGCCAAGT AGATATTGTG ATTGGTACCC ATCGTTTATT GTCGCAAGAT      9840
ATTAATTTTA GTGATTTAGG CTTGTTGATT ATTGATGAGG AACAACGGTT TGGAGTGAAA      9900
CATAAAGAAC GACTAAAACA ATTGCGTTCA CAAGTGGATG TCTTGACGTT AACAGCAACG      9960
CCAATCCCAA GGACCCTTCA TATGTCGATG TTAGGTGTCC GTGATTTATC CGTCATCGAA     10020
ACCCCACCAG AGAATCGTTA TCCGATCCAA ACGTATGTCA TGGAAAATAA TCCTGGGGCG     10080
ATTCGGGAAG CGATTGAGCG AGAAATGGCC CGAGATGGTC AAGTCTTTTA TTTATACAAT     10140
CGAGTAGATA CGATTGAACG AAAAGTTGAA GAGTTACAAG CACTAGTTCC AGATGCACGA     10200
ATTGCGTATG CACATGGGCA AATGACAGAA GTACAACTGG AGAATACTTT GTTTGATTTT     10260
ATTGAAGGAC AATATGATAT ATTAGTCACT ACAACAATTA TTGAAACAGG AGTAGATATT     10320
CCTAATGCTA ATACCTTGTT TGTAGAAAAT GCGGACTACA TGGGCTTATC TACGTTGTAT     10380
CAATTACGTG GTCGTGTGGG GCGTAgcAAT CGAGTGGCCT ATGCTTATTT TATGTACGAA     10440
CAGCAAAAAA TCCTAAACGA AtTAGTGAGA AACGTTTAGA AGCAATTAAA GACTTTACGG     10500
AATTAGGTTC TGGGTTCAAA ATTGCGATGC GTGACCTTTC CATTCGAGGG GCCGGGAACC     10560
TTCTTGGTGC CCAACAACAC GGcTTCATTG ATTCGGTTGG GTTTGATATG TACTCGCAGA     10620
TGTTATCTGA GGCAGTTGCT CGTAAACAAG GAAAAATATT CCAAGACCAA AAAACGTCTG     10680
TTGAAATTGA TTTGGGCATT GATGCTTATA TTCCAGGAAC GTACATTACA GACGAACGAC     10740
AAAAAATTGA AATTTATAAA CGGATTCGTC AATTAGAAAA TATGGATATG TACGAAGAAT     10800
TAGAAGCGGA CTTGTTAGAC CGCTTTGGAG AGTATCCTGA CGAAGTGGCG CATTTATTAA     10860
CCACTGGCCG TATCAAAATG GaTGGCGACC GTGCCTTGTT AGAAAGTATT CGTAAACGCG     10920
ATCAAAAAGT CAAGTTTGTG TTAAGTAAAA TTGGCACAAA AACATACTCT GTGGAGCAGT     10980
TGTTTGAAGC GTTATCGGCA ACCAGCTTAA AAGCGGATCT AGCAGTTGAG AAAGAACAAA     11040
TGACTATTTC CTTGAAGTTG CCGAAAGATT GTAAAGAAGC TGTTTGGATT CAAGAAATTG     11100
CTGCTTTTAC AACCGCATTG CGACAAGAAA AATACAAACA TTCGCAAGGA ACGGATTCCC     11160
TTTAAAAGGA GGCCTAGAAG ATGGCACAAA AAGAAATGCA ACGAATGATG CAAGGTGCCG     11220
TTGTGTTAAC CATTGCTTCC TTCATTGCAA AAGTGTTGAG TGCGTTTTAT CGCGTTCCTT     11280
TTCAGAACTT TGTAGGGGAT GAAGGATTTT ACGTGTATCA GCAAGTTTAT CCAATCTATG     11340
GGATTGCGAT GACATTAGCA TTGTCAGGCT TACCGCAATT TATTTCCAAA ATTGTTGCTG     11400
AACAACCAGA TATTCGGAGT CAAAAACAAG TGTTGCGCCA ACTATATCCC TATGTGTTAT     11460
GGTTAGCGAT TGCTTGTTGG GCCTTCTTTT TCTTTGGTAG TCAAGAAATT GCTATTAGTA     11520
TGGGGGATGC TGCACTCCAA CCATTAATGG AAGTTGTTTC CTTTACCTTT TTACTCGTTC     11580
CAATTCTTTC TTTCTATCGT GGGAATTTTC aAGGTCATTT GTTGATGGTG CCTAGTGGTA     11640
TTTCACAGGT GATGGAACAG TTTGTGCGTG TCGGGGTTAT TCTAGTAGCT GCGTTGTCCT     11700
```

ATCATTACTT TGGTGGCTCC ATTTACCAAA CGGGGACGGT GGCCATGAGT GGTGCGTTGG 11760

CTGGCGGTAT CTTAGCGGTA CTTGTCTTAT GGTACTACAA TCGAAAAATT CTTAGTGGAA 11820

GTACGGAATA CTTGCATCAA TGGAAGATTA TGCCTCAAAC CACAGGACTA TTTAAACGTT 11880

TAATGATTGA GGGTGGTTTG GTCTCTTTAT ATAGTGCTTT TCTGATTCTA TTTCAATTAA 11940

TTGATTCCTT TAAAGTAAAA AATGCCCTCA TGCTTTTCGG GCTCTCTGAT TTAGCTGCGA 12000

AAGTGGACAA AGGGGTCTAT GATAGAGGAC AACCACTAGT TCAATTAGGA TTGGTAATTG 12060

CTTTAGCTTT AAGCTCGACC TTTTTACCAG GTTTGACAAA ATATTTTATG AAAAAAGACC 12120

GTCAACAATT TTTACAAGTT GCAAAAATCT TTTTGCGCCT TACCACAACC TTGGCATCGG 12180

CGGCTTCGAT TGGTCTAATG ATGTTATTGC CGTACATGAA CTTTACGCTT TTTAAAGATT 12240

ATAAAGGCAA CGATGTTTTA GGGGTGTATG TTTTATCCAT CGCTTTTATG GCCATTATTC 12300

AGGCCTACCA AAGTATTGAA CAAAGTCGCA ATCGTTTTAA AGGTCCTTTA GTGGCAGCTG 12360

GTGTGGGCTT ATTAGTGAAG TTACTGACAA CGGGCTTTTT CACCATTCGT TGGGGAACAT 12420

TAGGTGCTAG TTGGTCTACT ATTCTTGGGT TACTAGCGAC GTTATTCGTT TTAGTCCGAC 12480

AATCGGATGC AGCGATTAAT TGTTTTGTTC GAGAACGAAA CTTTTTGAAA AAACTGTTGC 12540

TTAGTTTAGC AATCATGATG CTGTCATTAC TGGTCTATCA AGGAATTATC TCCATCTTGT 12600

TCCAAGGTGT GCATCACCGA AGTCAAGCGT TTTTCGTTAC AGTTTTAGGT GTTGCAGTGG 12660

GTGGTACAGT TTTTATTAGC ACCATCATTA AGTTAGAATT ATTTACGATA CGTGAATGGT 12720

TAAGTTTACC GTATGGCGCA AAAATTTTAC GGATGAGACA AAAAAAATAA AAGAGGTGAA 12780

TGTGATGCGG TTAGATAAAT TTTtAAAAGT TTCTCGCATT ATTAAAAGAA GaACAGTTGC 12840

CAAAGAAGTC GCTGATAAAG GCCGGATTCA AATCAACGGA GTCTTAGCCA AATCTTCTAG 12900

TACTGTGAAG ATTGGTGACC TTGTCAAAAT TCAATTTGGG AACAGAATTT TAGAAGTAGA 12960

AGTCCTGCAA CTGAATGACT CTACGAAAAA gAAGATGCCA CAAAAATGTA TGAAATAAAA 13020

TCAGAAACAC GTGTCTCTGA AGAATAAACA GGTGAGGATG GAGCGCTTGG CAAGTTCATT 13080

GAAAAGATGA ATATTCCAAG CGCTTCATCC TTTCAGCGTT TTAATGAAAC AGACGCGCCG 13140

CTTTTGAAAA AAAT 13154

(2) INFORMATION FOR SEQ ID NO: 339:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 528 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 339:

TCGTCGTTGC TGAATGCGAC AGCAGTTGGG CCTGTAAAAG ACTCGTCTAA GCCGTCTAAA 60

CCAACTTTTT TAGCTGCACG AGAAAGGATT GAGTTTTTGA TGACTTTCAT TTCTACGCCT 120

GCATCACGTA ATTGTTTACG TAAGTTAGTG ACTTCTTCAA CTGTTAAACC ACGGTAGTCC 180

ACGATTACAA CTGATGCTGC TGATTCAAAT TTCTCAGCGG CTGCTTGAAC TAGGGTTTCT 240

TTTTTAGCGA TTGCTGCTTC ACTCATTTAT TTCACCTCCT GATTTTGATG GTGGAGTTGG 300

ACTTGGTTTA CGTTAkGTAA AGTAGAAGAA TCaAATGAAA CTTcGCTnAC TTCGTTCGCC 360

AAGTATTGTT TTACATCGGC TTTTTTTCAA AAAGTCGTGT AAAACAACAA AACTCCATGC 420

CACCGTrGAC ATAGAGGGaC CGAATTGTAT ATAACTCaAC AATTGTCCTC GGTAGGaAAT 480

TmAAGACTTG CGTCACCTAC TGTCTTCGGG ACAGTTAATT ATTAACCC 528

(2) INFORMATION FOR SEQ ID NO: 340:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4126 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 340:

GGGGAAGCCA AGCAAGAACA GAAGCGACGG GTTATGGAAC GGTTTATTTT GTTAAACATT 60

TATTAGcAGA TAAAAACGAT ACGTTTGAAG GCAAAAAAGT CGCTGTTTCT GGTAGTGGGA 120

ATGTCGCAAT TTATGCCATG GAAAAAGCAA CTGAGCTAGG CGCCACAGTA ATCACATGTT 180

CCGATTCGTC TGGTTTTGTC TATGATCCAG AAGGAATTGA TGTGGCGTTA GTAAAAGAAC 240

TAAAAGAAAA AAATCGTGAG CGCATTTCAA AATATGTTGA AACACGTAAA GGCGCAACTT 300

ATTATGATAA AGAATCTGTT TGGAATTTCG AAACAGCGTA TGACATTGCG TTACCATGTG 360

CAaCACAAAA TGAAATCmAC gAAAAGCAAG CCGCTATTTT AGTGAAAAAC GGAGTGAAAG 420

TCGTGGCAGA AGGCGCAAAT ATGCCTTGTA CACTAGAAGC AGTCGCCGTT TTTGCTAAAT 480

CAGCTGTAAT TTATTGTCCA GGCAAAGCCG CTAATGCAGG TGGTGTTGCG GTTTCTGCAT 540

TAGAGATGAG CCAAAATGCT GAACGTTTAG CTTGGTCCTT TGAAAAAGTT GACGGTATGC 600

TAGATCAAAT TATGCAAAAC ATTTATGAGA CTTGTCGTGA TACAGCTAAC GAATATCAGG 660

CACGAGATAA CTTTGTTTTA GGTGCGAATA TTGCTGGCTT TGAAAAAGTA GCGGCAGCGA 720

TGCTAAGTCA AGGACTCGTT TAATAACAAA AAATGCTAGG ATTTCCTAGC ATTTTTTATT 780

GATTTGTTTG ATTAAGAGAA TAAATATCTC TAATTAAATA GTTAAAATTA AACAATTGGT 840

CTATCCAAAA GAATTTGCTT GAAAAAATGT CTAAAAACAC GTAAAGTATA CATGTAGTAA 900

GTTTACTTTC GAGGaGGAAA AGGAATGTCA CACmTTCmAT TaGATTATTC CAAATTaGCA 960

CCATTTGTTG CTGACCATGA ATTGGAATAC aTGcAAcACm AGTCACTGCA GTGGgACAAa 1020

GCATTACGTG AAGGtACAGG AGCTGGtAAT GATTTTACTG GcTGGaTtGA TTTACCgaAA 1080

ACTATGACAA AGAAGAATTC GCGCGCmTTw AAAAAGCAGC AGCAAAAaTT CAATCTGACT 1140

CTGAGGTATT AGTAGTCATT GGTATCGGTG GTTCATATTT AGGTGCCCGT GCAGCAATTG 1200

AATTTTTGAC ACATTCATTT AATAATCTAT TATCCAAAGA AGAACGGAAA GCGCCACAAA 1260

TCTTCTTTGC TGGAAATAGC ATCAGCTCAA CTTATTTAGC TGATTTAATC AATGTGATTG 1320

GTGATCGTGA TTTTTCTGTG AACGTTATTT CAAAATCTGG AACTACGACA GAACCAGCGA 1380

TTGCTTTCCG CGTCTTTAAA GAGTTATTAA TTAATAAATA TGGCAAAGAA GAAGCCAATA 1440

AACGTATTTA TGCGACAACT GACCGTGCTA AAGGAGCAGT AAAAGTTGAA GCTGATGCGG 1500

AAGGTTGGGA AACATTTGTG ATTCCTGATG ATGTGGGCGG CCGTTTCACT GTTTTAACGC 1560

CAGTAGGTCT TTTACCAATC GCTGTTAGCG GTGCGGATAT TGATCGTTTA ATGGAAGGTG 1620

CAAATGATGC CCGTAAAGAA TACGGCGCAA CGAGCGACTT AAAAGAAAAC CAAGCCTATC 1680

AATATGCTGC GTTACGTAAC ATTTTATATC GCAAAGGCAA AACAACTGAA ATGTTAATTA 1740

ACTATGAGCC TGGCATGCAC TATTTCTCTG AATGGTGGAA ACAATTATAC GGTGAATCAG 1800

AAGGAAAAGA TGGCAAAGGT ATTTTCCCAG CAGCCGCAGA CTTTTCAACT GACTTACATT 1860

CCATGGGTCA ATATGTTCAA GAAGGTATGC GTAACTTATT TGAAACAGTG GTTAAAATTG 1920

AAAATCCACG CCATAGCATT TCAATTCCAG AACAAAATGA AGACTTAGAT GGCCTAGGCT 1980

ATTTACAAGG AAAAGAAATT GAcTTCGTCA ATACAAAAGC GTTCGAAGGT ACGTTGCTTG 2040

CTCACACAGA TGGTGGTGTG CCAAACATGA TCGTTAAAGT GCCAACAATG GATGCTTATT 2100

CATTAGGTTA TGTTATGTAC TTCTTCGAAA TTGCAGTGGG CATCTCAGGT TACTTAAATG 2160

GGGTTAATCC ATTTGACCAA CCAGGTGTTG AAGCTTACAA ACGCAACATG TTCGCTTTAT 2220

TAGGTAAACC AGGATTTGAA GAATTAGCAA AAGATTTAAA TGCACGCCTA TAGAATAGTA 2280

TTCTAAGTTA ATTTTTAGCt TTGTGGTGAT ACAAAGATGT TTGCTTTGAA ATATTTTTGG 2340

TGATACATTT GGTGATACAT TTATAAAAAA ATCAGGTAAC GTTTTAGTTA CykGATTTTT 2400

kGTGTgCATT kGATATAATt CGAGTTTATT TACAGCTCGT TCTTTCACTT CAGGAGAGAC 2460

ATGTGCATAT ATCTCAGTCG TTTTAACATT CTTATGACCA gmATGTCTTT AATGTCCTCA 2520

AGTGGAACTC CTGCTTGTCG TAAACGAGCA GCATTAGTAT GGCGACCGTC ATGAATTCTG 2580

ATTTTTGGTA AGTTTGCTTT TTGAATGATA CGGTGCCAAG CTGTGTCAAC GGATCGTTGA 2640

CGAATAGGCA AGCCAACGTT TTTTCCTCTA TGATTGATAA ATAAAAAATC TTGGTTTGaG 2700

TTAGTCATGG GTAATACATT ACTTGTAAGA TCATAATTAA TACTATAGAG TTCTAAGACG 2760

GAGTTTCTCA CTCTGTTTGT CATTTTTAAA CTTCTCTTCC CAGCATCAGT TTTAACATCA 2820

TCTAAAATTA CTTGACCAGG TTGCTCTTTT CTATACAAAC GCTGCTtGTC AAaTGTGAGA 2880

aTATyATCTC TAAAATCAAA GTTAAACCAT CGTAAGGCCA TTACTTCGCC TTTCCGCAAG 2940

CCTAAGTCGA AAATGATTAA AAAAAATACA TACCACATAA AATCGTCAAA ATTTCTAGCG 3000

GTGTCAAGAA ATAGATTTTC CTGCTCTAAA GTGTAGAATT CTAGTTTCTT TTCTTTATTT 3060

GTTTTTTCCT TAACCTTATC TAAAGGAAAT TCAACGAAGT AGGTAGGATT GCTTTTTATA 3120

TACCCTAATT TTTGTGCCTT TTTGAATGCG TTGGAAAGAG CTGCATTTAC AATTTTTACT 3180

GTATTGTAAC TAAGCCCATT TTTACTTTTT CCTAGTTCTT TATTCGTAAA CAAGCTATTA 3240

ATAAACTTTT GGTGATCCGC AGgTTTATAT TTATTTAATT GATAATTACC TATATAGGGA 3300

TCAATATACA TACGGATAGC ATCTCGATAG ACAATTCGTG ATCCTTCCTT AACGTTATTT 3360

TTGTATTCGT TTATCCAAAT ATTCATAAAT TTTGAAATAG TCATTTTTTG TGGATTTTCT 3420

TTAACGAATT CTTCAGATGC AATTTGCCCT TCAATTATTT TTGCAGCTTT CAAAGCTTCC 3480

TTATGACTTC TAAAACCTCT CTTGTGTAtT TTtTGTTGCC TGCCAGTACC ATTGCTACGT 3540

CCATTTGAAA CGAaGTATTC CCAACGTTCT TGACCATTTT TCAATTTATA AGATTTTATA 3600

GAAGCCATTA ATCTTCACTC CTGAACATAG ATTTTGGAAG GGTATAAAGT TTAGGTTTAT 3660

ACCAGCCTTT TTCATAGAAC TTATCCATTT TTAATTCTTT TTCTTTCCAT GAAAATAAGG 3720

GGAGTGTGAC CATATTATCA GTTTCAATGT CTTTTTTTAG TTTTAATTGA TCCTTCAGCA 3780

CATCTTGATA TTCTTCTTCA CTATAATATG GGTTtGTTAA TATTGGATAA ATTGTTACTT 3840

CTGGAACATC TCTTTTCAAT CTTTCTCTCA ATTCATATGG TGTCATCATT ATAGTCAATC 3900

CTTTCGTAGT TTTACTGTTT AAATAAAATT AGTAACATAA AATAGAAGAT TTATATTAAA 3960

AATCATTTAT TTTTTCAGTA TTCAGCTCAA TGGATAAGAA CGTATGTnCT TTTTGGTTTT 4020

AAAAGAAAAG CCCATAGGCT AATCTTACTT AGCGGATTTA AACGTACATC CGCAATTATT 4080

ACACnCCATG TGTTTTTnTT ACCCnTTTTT CCAGCGAAAC CAGCTA 4126

(2) INFORMATION FOR SEQ ID NO: 341:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13884 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 341:

GACGATTTAG GCGTTAAATT AAAAATCGAA GAACTAGGCT TTGATGCTCT GTTAGGGGCC 60

TTAAAAACTG GCAAGATTGA CATGGTCATC TCTGGCATGT CGCCAACGCC TGAACGATTA 120

AAAGAAGTAG CTTTTTCTCA ACCATATATG AACGTTCAAC AAAGAGTCAT TGTTCGGAAA 180

GAAGACAAAG ATAAATTTCA TTCCACTAAA GATTTCGAAG GCGTTCGTGC AGGTGCCCAA 240

AAGCAAACGA CACAAGAAGA ATTAATTAAA ACCGAATTAA CCGGTGCTGT CCCTACTTCT 300

CTTCAAAAAA TTCCAGATTT AATTATGAAT CTAAAAAATA ATAAATTAGA TGCAGTCGTT 360

ATGGAAGAAC CGGTTGGCGA ACTATGgTTT CTCAAAATGA GGATTTAGCT TTTGCTGATG 420

TTACCTTTGA AAAAGGCTCA AAAGATACTG CCGTTGCTTT TCCTAAAGAT GCGCCCGTTT 480

TGACGGAGAA AGTAAATGCC TCTATTAAAC AAATTAATGA TCAAAAATTA ATGGACGGTT 540

ATAAAAAAGA AGCCAATCAA TTAATGTTCC AAAAAGATCA AAGCTTTTTA CAAAAGTATG 600

GAAAGTTCTA CCTTAACGGT GCAGGCTATA CCATTTTCCT AGCTTTCATT GGTGTTTTGT 660

TTGGTGCCAT TTTAGGGGCA TTACTTGCCT TAATGAAGTT AGCCAATTCC AAAATTTTAC 720

GTGGCATTGC CATTGCTTAT ATTGAATATG TTCGTGGCAC TCCTTTGCTT GTTCAAATTT 780

```
TTATTGTCTA TTTCGGGACA GGCGTTCTTG GCTTAGACTT ATCAAAAGTA GCAGCTGGTT     840
GTATTGCCCT TTCTTTAAAC AGTGGTGCAT ATGTAGCAGA AATTATcGTG CTGGTATTCA     900
GGCAGTTAAT AAAGGACAAT TGGAAGCTGC TCGTTCTTTG GGAATGAATC AAAACCAAGC     960
AATGCGATAC ATCATTTTCC CACAAGCTAT TAAAAATATT TTACCAGCGT TAGGCAATGA    1020
ATTTGTTACA GTAATTAAAG AATCTTCCGT TGTTTCAGTC ATTGGTGTTT CAGAACTAAT    1080
TTTCCAAGCT GGCAACGTTC AAGGAGCAAG TTTTAAACCC TTCTTACCTT ATTTAATTGT    1140
ATCGTTAATT TATTTCGTCC TAACCTTTAC TATCTCACGT CTATTAGGTG TTGCTGAAAG    1200
GAGAATGAGT ACTAGTGATT AACATTAAAA ACTTACAGAA AACATTCGGA AAAAATGAAG    1260
TACTTAAAGG AATTGACTTA GAAATTGAAG CTGGTGAAGT TGTCGTAATT ATTGGCCCTT    1320
CTGGTAGTGG TAAAAGTACT TTTTTACGTT GCTTGAACTT ATTAGAAACA CCAACTGGTG    1380
GCAGCGTTGA ATTTGAAGGA AAAAATTTAT TAGATAAAAA TACCAACATT GATCAACTTC    1440
GTCAAAAAAT GGGAATGGTA TTCCAAAATT TCAATCTTTT CCCGCATAAA AATGTCTTGG    1500
AAAACTTAAC GATTAGTCCT GTTAAAGTAA AAAAAGAAGC CGCTTCTGAA GCCACAGAAC    1560
ATGCTCTTTC TTTACTGGAA CAAGTTGGTT TATCAGATAA AAAAGAAAAT TATCCCTCTC    1620
AACTTTCTGG TGGGCAACAA CAACGGGTAG CGATTGCTCG TGCTTTAGCC ATGCGCCCTG    1680
ACGTGATGCT GTTCGATGAA CCAACTTCTG CCTTAGACCC TGAAATGGTG GGCGAAGTGC    1740
TATCCGTTAT GAAAAATTTA GCTATTGAAG GGATGACAAT GGTAGTGGTT ACCCACGAAA    1800
TGGGCTTCGC AAAAGAAGTC GCGGACCGTG TCATTTTTAT GGATGCAGGA ATCATTCAAG    1860
AAGAAGGCAC ACCTGAAGAA ATTTTTGATG CGCCACAAAA CCCAAGAACA CAAGATTTTC    1920
TAAGAAAAGT TTTATAAAAA GGAGCACCTA GTATGGATCG CAAGAATATC GCAACACGTT    1980
ACCAACAACC TACTGAAAAT TTACTCATGG ATATTGCAAC ATTGGCGAAA AAGACCCCAA    2040
ATTTAATCGA TTTATCGATT GGTGACCCTG ACTTAATTAC CGATGAGCGG ATTATCGAAC    2100
AAGCAGCTAA CGATGCCAAA AATGGGCATA CAAAATATAC CGCTTCTGAT GGTAGCGAGG    2160
CATTCATTGA GGCGGTCATT CAGTTTTATC AGTCACATTA TCAATTATCT TTTCAACCGA    2220
ATCAAGTCCG GGCCACAGTT GGCGCTCTAC ACGGGATGTA CCTTGCCTTA CAAGTGATTC    2280
TAAATCCTGG AGATGAAGTG ATTATTCATG AACCTTATTT CTCTCCTTAT AAAGACCAAG    2340
TCCTTTTAGC AGATGGCGTT CCGGTCTTTT TACCAACCTA TGAAGAAGAT GGTTTCCAAA    2400
TTGATGTAGC TCTTTTGAAA GAAAAAATCA CGCCTAAAAC CAAAGCGATT ATTCTTAATT    2460
CACCGAATAA CCCGACAGGC GCTGTTTTCT CTGAAGAAAC TTTCAGAGAA ATTGCACAGG    2520
TAGCGATTGA ACACAACTTA TACATTCTTT CAGATGAAGT TTATGAAGCC TTTTGTTTCC    2580
AAGAAACGTT TACTCCCATG GCTACGTTTG CTCCAGAAAA TACTATCACT TTTGGCAGCT    2640
TTTCAAAAGC CTTTGCAATG ACTGGTTGGC GTATTGGTTA TATGATTGCC CCAGACTATA    2700
TTAACGAAGT AGCCAAATTA ATTAATGAAG GCGTTACGTA TTCTGCGCCT ACGTTGTCAC    2760
```

AGCAAGCAGG AATCTATGCG TTACAACATT TTGACGAATT TGTAGACCCA ATCGTTGAAG 2820

TGTTCCAAAC ACGTTTAGAG TATGTGGCAC AACGAGTAGC AAAGATTCCT TTCCTTTCGC 2880

TTCATCCAGT CAAAGGAAGT ATATACGCCT TCATTAACAT TCAAAAAACC GGACTAACTT 2940

CCGTTCCTTT TGTGGAAAAG CTTTTAAAAG AGACACAAGT ATTAGTGATT CCTGGAAAAG 3000

CATTTGGTGA AACAACAGGC GATGAGTACA TTCGTTTAGC CGCAACACAA AATTTAGATT 3060

TATTAAAGGA AGCGTTCGAT CGAATTGAAA GAATGACCTT TGAATAAAAA AAGCTTCTGC 3120

ATTAGCTAAT AATCATAGCT AATGCAGAAG CTTTTTAGAT TTCTTCTATC AATTCAATCG 3180

CTTGCCCGCC AACATGCGGT GGAATTGACA TAGTCATCGG CGGCAAGCCC ACGACTGTGA 3240

AACGGACTTT GGTCCCTTTT TTCAAGTGAG ACAGTGGGAT ACTTTCCTCT AACTGTTCAA 3300

CAAGTATGTT TAAAATTACC CCATGAGAAA ATAAGGGAAC AATGGATTCT GGATCACTTT 3360

CTGCTTGGAT ATTTTCTAAA TGTAATCGCA TTGACAATGG CTGCGTCTCA TTTAACACAC 3420

CAGCCTCTGC TAATTTCCCT GTAAAAACAG CAAGTTTTTC AGTCATTTGT TTGCCTCCTT 3480

TTTAAACATC ATATTGAAGA GAATTAATTC CATCCCCTAA AAATTCTGAA GCGTTGTTAA 3540

ATTCGGCAGG ATCATATTTG GTATAGGAAA CTTGTAATGT CAAAGTAATA TTTCTTAAGG 3600

TTTCAGAATG GGTAAATTTC TCTTGAGCTT GTTGTTTGAA TTCTTTTCTA GTAGTGTATC 3660

CGTCATCCCC ATCGCGGAAA ATTAATAACG CCCAAGTCGG CGGTGTGCCG TCGATAAAGT 3720

ATGGTAAGAA GCGATCATCC GTTGTCTCTT TAATCGCCGC TGTCACGATT TGTAGCAATT 3780

CATTTTTTTG TTCAGGACTT AATAAGTTAC GAATACTTTC CCAATAGGCC ACTTGGATCA 3840

CGACTAAATA TAACGGCGCT TCATAACGTG TAGAAATATC AGAAAATAAA TCGAATCCTT 3900

CCTCGTACAT TTTTGCTGTT CTTAAATTAG TTACTGCATC TAACGCATTT AACCGTGAAG 3960

TTTCTTTACG TAAGCGAATA TTTTCTTCTT CAATCGTACG AATTTGATAA GTCACTGCAT 4020

AGATTGTTAA ACAGTAGAGC GGCGGCATGA TTAGCCAAAA TAATGAGCCA TAGCTAAACA 4080

AATCATGATA GACATACTGA TATACAACAT AGGTCAATTG TAGAAAAATA AATATAAGCG 4140

AAAATGTTAA CCCCGCTGTA ATTCCTACAA AATAGGTAAC GATAATTAAG ATAACTGAAA 4200

TTAATAGACA ACTTAAGTCC ACAAAGAAAT CATTCGTTCC TAAGGCCATA AAGATGGTTG 4260

TGCCGACAAC CAAAACAATT CCTACCAACA AAGTTAAATC CAAAAGTAAG GTCACTCGTT 4320

TTTGATTCAC TCTATTCCCC ACCTTTCAAA TTCTTCTTCA CGATGAAGAA CGCACTTAAT 4380

CCAATAATGG TCATACCAAT TAAGAAGACA AATAGATAAA TCATTAATTT ATGATTATTG 4440

ACAATTCGTT TACCTAATGA TTCGTTGCTA ACGTTGCTTA CACGTTTTTT GAAACGATAA 4500

TCATAGCGAC GATAGTTCGG ATCGACCACA ATGGCATCTC CTTTGTACAT AGAAGTATTC 4560

TTCTCTGTAT TCACTTGTGT AGAGGCTAAA AAGACACCTT GTGATTTAGC TCCTGTTAAG 4620

ATTAACGCCG CGGCTTTCGC ATTATATGGA GAAAACATTA ACTGCGCCGT ACCGATTTGT 4680

TTACCATAAT CTTTTTCAAT GCTCAGTTTC TCGTTTGAAA CAAAACGTGT GAAGTCCTTA 4740

TCATAATGGA AATATAATTG ATCGTTTAAC TTACGAATCA TCGGATTATC CTTAGGCGTA 4800

CCAAAAATAA TCAAGTTATG ATTTTCTAAT GCCGCATTTT TAGGCGCTTT TTTATAATAC 4860

GTAATCTCAC CGACATTACT TTCGGCATAA TTACCAATTA AGTTAAATAG ATTGGTCAAC 4920

ACTTTAAAAT AGTTATCGTC CATTTTTTCT GGTAATAAAA TCGCTAAATC AGCAAAAGAA 4980

CGGCTTCTGA TAAAAATGTT CGGATAGTTG TTAAATAATA TGTCATTCAG TTCTTCTGTT 5040

TGAATAAACG CATTCGAATT ATTTTCAATA AAGGCCCATG GAGTCTGACC ATTTCGTAAA 5100

ACTTCTGGTG ATTTAACATT CAAATCAAAA GCAACTTTTA AAACAAAACT ATCAGCAATT 5160

TCTAAGTTCT TTGGAAATTC GAGATTTAAT TCATCACCGT TCGCACGAGC AGCGGTTAAT 5220

TTCTTGCTGC CAATTGGTTT GTCATTGGCA TAGACCGTCA CTAAAGAAGA ATCAAAATCT 5280

AAGTTTTCAG CGTATTTAAA ATGTAAATGC ACGCGACTTC CATTGGCATT GTTTCGATCC 5340

ACTGGTAAAT TGACAAAGTA CGTTTGTTCT TGGTGATAAG CCCCTTCTAG TTTGTCCCCC 5400

GTTGAGGTTA ATGGATAATT TCCATCAAAT TCTAAAGTTG AACTAAACGT ATCGGTATTC 5460

TCATCAACTG TTGTCTCTTC TTTATCTGTT TGGGTCATTA ATTCATAGTT TGCTAAATAA 5520

CGTCCCGCAC GAACAAGTAA GTCCTCATCT TTAGAAGTAG CTACTAAAAC ATGCTTATCT 5580

GGTTGATTGA AAAATTTTAA GACAGCCTGG TCCTCCACGC GTTTGCTATC AATTTGACTT 5640

TTATACTGAT CTGGTAATTT ATCATAGCTG GCAATAATCA ACTGATAGGG TGCTGACTGT 5700

TCTTTGTTCA AAGAAGCCAT TGGCAACAAT TCTTCTGAGG TAGTAATCAA ACGAGCAGCA 5760

CCAGcTAACG CATGGGTCGC TGCAGCTAAT TCTTTTTCCG ATGCATTTTC TGGTGTCAAA 5820

ATAACACTGT GTTTATTCGC AATCGTGTCT GCTCCCACAA AATGATTATA AAAAGAATGA 5880

ATCGTATTCT CTGGTAATTG TAAATCGTAT TGGAAGTTTA CGTTAGATCC TTCATACATC 5940

GTTAACCAAT TTGCTGGAGT TTCAATTAAA TTATACATGT CGTTTCCTGC CTGATTAATA 6000

ATTTGGCCTT CAACAGTCAA TGTATTTGTC TCTTTAATCA GTTCCAAAGG GATTTCAATC 6060

GTCTTTTGTT GAATACCAGT CGTATTTTCA GGACGCCAAG AATAAAACTT CACACCATTA 6120

ACTGCCACAG TTAAATCTGA GGTCTGATTA TTTTTTAATT GTGTGATTTG GtAGGTCATA 6180

TTGAATGATG CTtTTTTTAC GTCCCAATAA TCAATTTTTG TAAAATACAT GGTGCTACGG 6240

ACAGAAGTAC CTGTTaAAGA GATTGTTTTA TTTTGAAAAG GTTGTGTAAA GGTATGGtkG 6300

tCTkGTtCtT tGTTTnCGGC ATATACGGTC GGACTtGCGA CACCAATCAG AATTGCCAGA 6360

AAGAACAACC CTATTAAAGT GACTGTTTTT TGTTTCATTG CACGTTTCCT CCTTATTTAA 6420

ATCGTTCAGT TTTGTACCAT TTTGTTTGTG TTTGTTTATG AAAAACATTT TCTTTAATAT 6480

ACATCACCAT GCCATAAGCG GCGACTACTA GCCACATCTG ACTGTATGTA ATATACATCA 6540

AGGCAATAAT TAAGATATTC TCTAAAGTCA TTTCGCCTTT TTCGGTTGTA ATACTGACAA 6600

AAGTACTGAA AATAAATAAC AAAATCGCTA ACAACCACAA GGAATTGCTA AATCCTTGCA 6660

GGGTCGTTGT TAAGTAACCA GACATTGATA AAACTAACAT GATGTCTGAC AAAACGAGCG 6720

ATGTCATTAA GAGAAAATAA ATGGACAAGA AATATAAAAT ATCAAAACGG ATCCGACTCG 6780

CTTTAGGATT AAACAATAAT TTAGCATTTT TAACAATGAC ATAAATATTG CCTTTGACCC 6840

ATCGCGTCCG TTGTCGGAAC CAAACTGACA AAGTTTGTGG CTCTTGCTCC CACGTCACTG 6900

CTTTTGGTTG AAACTTAATT TTATAGCCCA TCATATAGAT GCGGAAACTG ATTTCTGTAT 6960

CTTCGGCTAA CGCTTTGTCA TCCCAGCCGC CAATGGCTTC AATAATACTA CGACGAACAA 7020

TGAAATTTGT GCCAGGAATC GTGCATAACT TAAATAATTG AAAACGACCG GCTTGTGCCA 7080

TCCATTGAAA GGCAAGTGTT TCGATATTAA TAAAGCGTGT CAACAGGCTT GCATTGCGAT 7140

TTCTCGTTCT AAATTTACCA ATAACCGCGC CGTATTCCTC ACTAGCAATT AACTCTCCAA 7200

CTAAATAACG GAGTGCATTC GGTTCAGGTG TATTGTCCGC ATCATAAATG GCAATCACGT 7260

CCCCCTTGGC TTGCTTGAAA CCAATGTTTA AGGCATTAGA CTTACCTTTC CCACCTGTTA 7320

TATTGTCTGT ATTAATAATC TGCAAATTTC GATTAACCGC CTTTTTTTGG ACGTTAGCTA 7380

AAAGCTCAGC ACTATTATCA CTTGAGTTGT CATTAATCAC AATAATTTCA TAGCGATCTT 7440

GCGGATAATT AAAATTCAAT AACGCTTCAA CTGTTCGCAC AATGACAATT CCTTCATTAT 7500

GGGCCGGCAC CATAATTGAA ACAAACGGCG CGTCTTCGTC CTCTACAGGT GGTGGTGCGA 7560

CAGGATCTGT TTTTAAATAA TAAACATAGC CAAAAATAAC TAAGAGAATA TTTACAATTA 7620

GTAGACTCCA AATAATGATA GTTGCTGTTA ACATTACAAT GTCAGCAATG GTCATTTGTC 7680

TTCCTCCTTA TTAATATATT TCGACCGATA ATTCTTTCGT CCAATCATAA ACAAAATAAT 7740

CAGGATGATT AAAGTCAGCG TGACAACGCC AATCAAAATA TTGTTTGTAA TACTGAAAAA 7800

GAATGAAAAT AACCCTGTAA ATCGTTGTTT TTCAACTGTT GAAGCCCCTG TATCAGCCAA 7860

ACCATTGACA ATTTGATTGT TTAAGAAGAA CCGATTATTC CGAAATTCGA TTTTCTGCGT 7920

TTGTGTTTGA ACAGTGAAAT GTTGATCCGC TTCTGAAACA CTAAACACTA TTGGTGCTTC 7980

CTTGACTTCT TGAATTAGAT GATCCACTTC TTTTTTAGAA TTTGGAAATG AAAAACTAAT 8040

GGTGGTCGGC ATTGGAAATC GATAGTCATT CGGATTATCT TTATCCGTCC ATTCAATCCC 8100

CGATAAATAA TCATACGGTA AATCAAACAA AGCATTCTTA TAAGTGATTG ATTCTCCCGT 8160

TTGGTTTCGA TAAAAAACGC GTTCAATCGG CGGATTTTCT CTTAGAATCA CTGTATCACT 8220

AATTGCTAGA CCATCCTCTT GGTATTGAAG GTCTTGATTC CAGTATCCTG GAGCAGAAAT 8280

GCCTACTGGA TAGACATGTC TGTCGACCAA TAAACTAATC TGCTGCTCCA TCAATTGTTT 8340

CAAGGCACGT TGGTCATTTA AATCAACATT GTTAACGACC GGTGTTTCTA AGAAAATGAC 8400

CCCCGTCGAA GCAAACGCAC GTAACGCATT CGTAAATAGT TTATAAGGTA AAATCGTATT 8460

GTTTTGTGAA ACGCTCGTAC TGGATAAAAT ATAAGGATAC GTGGTTTTAA CTAATTCTTC 8520

TTGCAAATAG GTTGCCACAG ACATATCTCT TAACGGGTTA AAATCAGAAA TCGTCAAAAA 8580

CGGCCGTGAC ATTTTTTCAC TTTCCAGCCA TTGACTAATT ACGTCTAAGC TTTCTAAAAA 8640

AACAGCGCCT TTGCGTTCAA ACATTGGTAG AAAAGCATGC CCATTCTCAA TTACACCAAA 8700

TGGATATTCT TCTGGTGCTA ATTCTTGTGT TTTTAATCGG CCAACCGTCT GACCTTGCGT 8760

ATTTTCCAGC ACAACTGATT GATCTTGAAA AGGCAATACT TGAGAAAAAC GGTTTTTTTC 8820

ATCTTCTAAC CGATATTGAC GATGACTTAA TTCTTTCCAC GTACCGCTAA AATATTGTTT 8880

TTCATCGTCA CGCATGTTTC GACCAATGTG TAATTGTTTG CCTTTAAAAT GAACTCGATC 8940

TGCTAGAAAA TCAGAGGCAA TAACGCCGGC TTGTTCAGGC CAATTCACCA TAGAAATCAA 9000

ACTGTCATAG CCATCATTTA GCAATTCGCC AGGTACATAA TCTGATACCG CCACACTCTG 9060

AACTTCTACT CCAAAACTCG TCAACACTCG TTGTAAAGCA TCCACATCAT TTTCTTTTTT 9120

CCCAGCAATA TTCAAGCTGT CATAGACAAG TAAGACATGT TTACTTGGTG CAGCCAGCGC 9180

CGGTTTGGCA AAGCCACAAA AAACGAAAAA GAGTAAGAAA CACACCCAGC CAATAAGATA 9240

TTGTTTATTT TTCAATTCCC TCTACCTCCC CAATGACTTG TGGTGTTGTA TCTTGCCACA 9300

TTCGTTTACC AATAACATCA GAAACAATTT GTAACAAGAG AAAATCTAAA AATACTGAGA 9360

CTAAAAACAA ATGACGTACT AAGTCACTAT CCCCATCAAT AATAATTGTT GAAATGAAAA 9420

CAACTAATAA ATTTAGATAA ATTCCTGTCA TTAAAAAGAA GCGAAAAACT AATCTGGGTT 9480

CGTTyCTCTT GAGTCCTCTG TATAAACCAA CGCCATATAA aCCAAACAGA ATCACGCCTA 9540

ATAATAGATA GAAAGCAAAG TTTTTAGGAA AAAAACTGGC TTTAATACGG TTATACACCA 9600

TAAAAAATTT AGATTGtTCC AGCGGTGCAC GCCCCGTCGT TTTCTCAAAA TTCCCCACAT 9660

TATTAGGCTT TACTAAGTAA ACATTTGGGA CCGCTGAATT CAGTCCTTGC CATAATTCAG 9720

ATGGATAATT TAAATAATTC AATAACACCC ATCCAAAACT CGTTTTATCA AAAAAATCTT 9780

TTTTCATTTG TTCAGAAGTT GGTGACACAG GACTGTATTC TTCAAAGTAG GAACGTCCCT 9840

TTTGCAACCC ATATTGCGGG TTGATTCCGC CCTCTGACAT CCGTTTATCT GGAACCTTAC 9900

TATCCTTCAT TACACCGCGC GTCATACTGT GGTAGACATC ACGTTCATAG ATATCGTTAG 9960

GAACGATAAA AGCCGCTAAC GTTGTAAAAC CAATCAACGC AATTAAAAAG GCCGTCATTG 10020

ACAATCGTTG ATAACGTTGT GACAAGTAGA CAAAAAATCC TAATCCAGCA ATAATTACGC 10080

CAATAATCAG CAAATATTCT TGGCGTGTTG TTCCTATTAA AAGTACTACT GCTTGTGCCA 10140

AGATAACTAA AACTATTCTT CTGAGCCAAG TTCGTGGTTT CGTCACCCGA TAAAAATAGA 10200

CAGAAAATGC TAAGACACTC ACAGAAAGTA AAAAACTCGT CGCTTCTGTA TAAAAGGAAT 10260

TAAAATAAAT CGTATAGGAG GTATCCCCTA AGATGAAAAT AACCAACAAG GAGATAAGAT 10320

AACTTTTTTT CAAAGATACG TCTTCGGTTA AGCCTCGCAC CAGAATATAA ATTGACACTA 10380

AAAAACAACT AAGATAAACG AAACCTAAGA AACGAATATC GAAAAGTGTC TGACTATAAA 10440

AAAGTTTATT AAGCCAAATG GCAACGGTAA TGAAAAATTG TTGAGTGGAA ACAAAGTGCG 10500

TCGTTGTTTC ATTAAAATAA CGCATAATTG AAAAGTCCTT AACAAAATAG ACCACATTGT 10560

CGTAGCTATC ACCATTGACA CGATAAAGAC TATTTGAATT TAACACACGA AAATACTCGC 10620

CATTATCTCC CAAGCCATTA ATCGGATAAA CGAATAACAA AATGCTTGTC ACTAACCCAA 10680

```
TCCCTAAAAC GGCCAAAAGA CTTGGAGAAA TATAACGACT CAACCATTCA AAAGCCGAGT    10740
TCATGTATTT TTTGATCTTT TCCATGCTTT CACCTCATCT AGAAATCATA GGCCAAAAGA    10800
GCCATTAAGT TATTATAGGA ATAGACTTGT TTCGTAACTT TATCACCAAA TCCTCCATAA    10860
ATCGGACTGC TTGAATCCAT AATTTGAAAA TTATTAAGTA CTGTAATTGC CGCACGATAC    10920
AACTCTTTGT CATTTTCGGC AACACCAATT AAAGCCGCTA TCGCATAACT AGCTGCAGAT    10980
TGATTTTTAT CCACGGGCGA TCCATTTAAA TCATACGAGT TAAAAAGCGT TCCTTGCTGA    11040
ACTTGTTTTC GCACAAAATC AATACTAGTC TGCTTCGCCA TACCCACTTC ACTTAAATGG    11100
ACAATCGTTA ATAAAGATTC AATAATATTA AGTGTTCCGT TGTTTTCGTA TTTATTTGTT    11160
TGATGATTAA ACTTCGTTTG ATAAAGAGGT AAATCATCAG AAATATAGCC ATCATCTATT    11220
AATTGGTAAT GATACTGCAA GTAGTCTGCT GAAATCCCAA ATTCTTTGTA TAGATAACCC    11280
AGCGTAATTA AGTCAATATA AAACAAAGAA GTTTCTGAAG ACTGTTGTTT TGATTTACTA    11340
TCGTAAAAAT CAATTAGAAT ATTATCTTTC ATAGAGGTTT TCATAAAACT CTTGCCTAAT    11400
TTTTTAATTT CTTGATCATA CTGATCTGTT TTAAAATGAC CCCCTGCTTC AATTAATGAA    11460
CGAATGATCC GTAAATCATC AATCGAGGCA TTGACATTGT ATTTATTGCC TTGCTCGTCA    11520
ATTCGATAAC TAAATTGGAC GCCGTCGTAA AAGGTGTCTT TCGTCTGCTT GTAAAAATTA    11580
TCAAAACGTC CCTGCGTATC AAAAGCTAAA TTACGCAAAT ATAGACCCGA AGACTCACTT    11640
AGCAATTCAT GTCCTGTGGC GACATCTTTG CGTTCAGCAT CTGGTAAATA ATTGGTTAAA    11700
AAGCCGCCGT CTGTTTTGAG CCGGGTTTCA ATAAAAAACT TCAACAAATC TTTGCGTTCA    11760
GCTGTTTCTT TkGTGTGAAA TTCATGTTTT GTTAAAAAAT GTTGTTTGtC TGTCGACATT    11820
AAGCGTTGAA TTTTTGGATT CACTTTTTCC ACCTCTACTT TATAGGATTT TTCTCTAAAC    11880
AAAAGAAAAT ATCCTCCCAG AAGCACCACG AATACCCCAA TAAAAATCCA AAGTTTTCTC    11940
TTCATTGTTA CGCCTCTTTT CCTCACTTCT TTATTGTCAG TGTACACAAA AGTACTTGTT    12000
TTTGCGAAAC TTACGCTACA TCTTCAAGCG CAAGAAAAAA AGCAATCCAT AAAAATAAAA    12060
GGATTGCTTT AACCGTTTCG TCGATCATTT CCTGTCATCG TAATCTCTTT TGACAGATAA    12120
CGCACACGTT CCATAATTCG TCTGGCTTTT AGCGGCTCTT GTTCTCCTCG TTGCGTCACG    12180
GTTAAATGCT CTTCTAAAGC TTTTAAATCT AGATTGGAAG AGAAAAAAGT AACTAATTGT    12240
TCTTGCATTC GGTATTGCAA GATTACGCTT AAAACGTCGT CACGAATCCA ACTAGTCATG    12300
GATTCCGCAC CAATGTCATC AATCATTAAT ATAGGGGATT TTTTTACTGC ATCGAGCTTT    12360
TCACCAACTT GATCGCGACC AATGGCTTGC TTCATTTCTA CTGTAAATGT TGGAAAATGA    12420
ACAATTGTCG TAGTAAATCC ACGTTCTGCT AAAGCATTAG CCATCGCACC TAATAAGAAA    12480
GATTTTCCTA CACCGAAGGG gCCTTGTAAA TAGAGCCCTT TATGAAATTC TTTTGGTGTC    12540
GCAGGATATT CTCTTAAAAA TTGCATTGCT TCTGcTAACG CTTTGGCTCG ACCTTGAGAA    12600
GATGGATCGA AATCTCGTAA ATTTGCTTCT CGGACATCTT TTGGCATGTC CATTGCCCTT    12660
```

ACTCGATTAC GAATTTCTTC TTCCTTCTGA CGGGCGATTA AATCAGCCGT TGGGACATAC 12720

GTTACAtCAA TGTAATGAAA ATtCAACATT AATCTtGGCT CATAGCCAGG GGCAATCATt 12780

GCCGGATCGT TtACTTGAAA TTTCyTTTTT TCyTGAACAA ATTCATACaG CTTAGCATAA 12840

CTTTTACGAA TATCTTCATC GGTTAAACGC TCTCGGTTTG CTTGTATAAA GGCTTGAACA 12900

TCTTGATCTT TCAATACTtC GTTAACCAGC TCTTCATATC GTTCATTGAT GTCACGCTTC 12960

TGGATAATTT TTGACATTTC TTTTCCTACG TCTTCCATCA TTGGTCACCT CCTTGGTTTA 13020

AAAATTCTTG AATTTGTCGA TCTAACTCGG CTTGTTCTTC TGGGGAAAGT TTTGTTTCAT 13080

CGTTTGGCTG ATCTGCCCAT TCTGGCAATC GCTCTTGTCG AACAGTTTGG CGTTTCCCTG 13140

AATAATTTTG TTGTCGCTGT TTTtGTTTGG TTTGGCTTTG TTTTGCCAAT TCTCTGACAT 13200

GTTCAATTGC TTTTwCTGGT GAATGAATGC CaCTnTGTCC CCaTyCATyC GCAATyCGAT 13260

TCaCATACTC CGCTTTTAAA GTCGGTTGTT GTTGAATATT GTAAACATAG TTAATCAAAA 13320

TATTAATCAC ACTAGTCGGT AAGCCTGATT TGGCAACCAA ATATTTCACT AATGAACGTT 13380

CTTGTGGTGT CGTATAGCCG CCTCGACTAT TTTTTAAAGC TTCTAAATAT TGAATAGGCG 13440

GATTCTTTTC ACTATCCATA ATCATTTGAA TATCCAGTTC AGaAAAACCA TTCATTTTTA 13500

ATGCGTTATA TCGATACGTT nGTGGCTcTy CCAcCGTCaA TTGTGCTTGC TTGTTTgkCm 13560

CTACGTTTGT TGaTTGTGGT TTTTTATCAG GATCaAACGC TTGATGAATC GTTCGCACTA 13620

ATTCCTTAGG CAAAATTCGA CTTGTATAAT AGTCAAATGA TTTCGAAGCA AAGTCTACCA 13680

TATCTAACTC ATTGATGCCA TATAATTGAT GAAAAACATA GACTTCTTCT TGAAAACTAG 13740

CCTCATCTTC AGGCAGTTGT AAACCTAAAC GGTGCAAACC TTCCACAAAA TAGTGCCAAT 13800

CAAAAGAATC CGTTGATACA GAAATTTTCT TCGATTTTTC TTTTTGAGAA AAAGTTTCTT 13860

GAATCGTTGT CAATCGGTTC GCTT 13884

(2) INFORMATION FOR SEQ ID NO: 342:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 784 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 342:

ATTAATACCT GTATTAATGT CTTTTACAAC TTGAAATTTG ATTTTTTCCA TGGCGACATG 60

CTTCTTATCC CAGTATTTTG GATTTTTAGC AATCGTCCAT TTATCTGAAA TACCAACACC 120

GTTTAAATCA GTTAATGTAA ATGCTCCATT CGTTACAATG TTTTCTTTTG TTGTTCCATA 180

TTCTTTACCG TACTCTTTTA CTGCGTTTTG TTGGACTGGA AAATAAGCAG TCAATGCTAG 240

CAAATCTGTA AAATAAGGTG TAGGTTCTTC TAATTCGATT TCTATTGTTT TGTTTCCCTT 300

AGATTTTACA CCAAGTGTTT CTTTTTGTTG CTTTCCGATA CTAATCTCTT TTGCATTTTT 360

TATAGCAGCA AACAGTTCTA CGTTTGGCGA GGCAGTTTTA GGATCAAGAA CTCTTTGCCA 420

TGCCGTTACA AAATCATCAG CAGTAATATC TGTwCCATCG GCCCATTTTG CACCGTCTCG 480

AATCACAATT GTATATGTTT TACCAyyATT GGATATTTTT GGTTGCCCAG CTGCAATTGC 540

TGGTTGTGGT TGATTTTTTT CATCTAATCG GTACAACCCT TCCATTACGT TGGCCAACAA 600

TGTAAyACTT GTTTGATCCA CTGCTGTTGC TGGrTCTAAA CTTTGAACTG GTGTACCTGC 660

CGATAAAGTT GCTATTTGTT CTACCTTAGT TTCATTTGTT TTTTTTGAAT CTTCTGTTGA 720

TTTATTTACT CCACAACTTG CAAATACnAA CAACACnAAT ACTAAAACAC CAACTnTTTT 780

AAGC 784

(2) INFORMATION FOR SEQ ID NO: 343:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3907 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 343:

CCCTTnATTA TTCCTATTTT TTGGnGnTAA AAATCAGAAA TTACTTGTAG GATAAGCTGT 60

CCAATAGGTA nnATTAACTA ATAGGTAAAG TAATTCGTTA TTCAATGGTT TCCCATCTGT 120

CAATATATCG ATAATTCTTT TTACAAACGG AGGGGAATTT GAACAGAAAC GATTGCGGCA 180

ACCATACTTA TTAGTAAGGG TATTAAAATG CATAACATTG TTTTCTTGTC AATAAATAGA 240

TTAAATAAGT TATATTTTTC CTTATGCTAC AGGGCGCTCC CAAGGATAAC GACCGTATTG 300

TTGTTTATAC CAATTAATTT CGTAAGTCGT TAGCTTAGCT GCTAAAGTAT AAGGGGTAGG 360

GAATTTCTTA ATCATCGCTT TAATTTTTGC TAACATTAAT TTTCCTCCTT TCATTGTTAT 420

GCCTATAAGA TTATAATAGA TATAGACAAT TACATTTTAG CATACTTTAA ATAGATTtTT 480

AcTATTTTTG TTAAAATATT GATAATTAAT TTTGAAAGTG GGTGAACACA AATGTTTAAA 540

TTTTATATTT TTATATTTGT AGTATTGATC GTAATATTCT TGTGGTATTT TTATTCATTT 600

AAACCTCTTT TGAACAATAA AAAAGAGGTT ACCCTACTAT CTTTAATTGC AACAATATTA 660

GAATCTCTTT TTTTCTGGCT GTAATCCTTT TAAGGCAACA AAATTATTTA TAACAAATGT 720

TGGTTGTGTT GCAAAGTTTT AAAATAAAGT AAGCTACTTC AATTTTTACT GTTTAGCTTA 780

AGCTGATTTG AATAGAGTAT CAATTTCTTC CCACGGGGAA AAGTCGAAGA GACCAATTTC 840

TCTTCGACTT TCTTTGTATA AAGCAGAAAC AACTTCGATT CCTTTAATCG TTACTGCAGC 900

TGTATAGGTG GACTGAAAgT TGTTGCCATA AGGAAGCTTC CCCTTCAACT GTCGGTGATC 960

CTGTTCAAGA ATATTATTGA GGTATTTCGA CTTCCAATGC TTCACTTTTT GGTATAGAAT 1020

TCCTTCTTCT TTCAGCTCTT TGATCGCTTT TAATGAAGGA GCATATTTAT CTGTTACAAT 1080

GGAACGTGGT TGACCGTAGA CTCTGATTAG ACGCTTGAAA AAGGCTTTAG TAGAGACAGT 1140

ATCACGATGG TTTCGTAGCC ACATATCTAA CGTATTGCCC TTCGAATCAA TGGCTCGATA 1200

AAGATACCGA TCTTGACCTT TCACTTTGAC ATAGGTCTCA TCAATCCGCC AGGATTTTGA 1260

TAAGGATCTA CGTTTATTCC ATAGATGATA AAGGATTTTT CCATACTCTT TGTTCCAGCG 1320

GTAAATTGTG GTATGATCCA CGCCAACACT AAATTTACGC ATGATTTTGG CACAATCACG 1380

ATAGCTCAAA GAAAAGTGAT AGTATAACCC GATGGCTTCT ATAATAATTC GATAGTCATA 1440

CTGTTTTCCT TtaAACTGAT TTTTCATGAC TGTTCCCTCT CGTTCTTCAT AGATTTACTT 1500

GAAATTATAC CGTAAAATAG GGcTAAATAA AACTTTGCAA CAAAACCCAG CTCTTTTCTC 1560

TTGACTTTTC TTTATGTAAA AAGCCTTGTA TTGCAAGCAA TACAAGGCTT TTTATGCTAT 1620

AATAGAATTG ATAGTAAAGT TTAATGACGG TGGCTATCCA ATCGAAAGTG AGGTGGTGCG 1680

TATGAACGTA CTTCCAAAAT CtATAGAAAG GAGAAGCCTT TTGTCCGTTT CGGAAAGTTT 1740

GCAACTAATG CyTGCGTTTG GTGGGTTTAC ATTGACCCTA tCcACAACCA TTGTTGCCAy 1800

cTGAcTATAA AGACAAAAAG AAATAACCGT CACTtTAGCA GGAGACCGGT TATTTCTAAA 1860

TAACTTATAA AAACTCTGCC ACCGTCTTTA GCGGTGCTAT CAATGAGGGC GTATTACCAG 1920

TACGTCCTCT TTACATAATT ATTATACCAC ATTTTGTGCG AAAAGGTAGG CGAATGCCTA 1980

CCTTTTTtCA TGCGCGCCTA AACGACGGTG GCGGTGGGAG TGTGGCGCGC ATGAnAAAAC 2040

AGATCTATTG ATTGTAGCCA TATTCAGACA ACCCGTTCAT GGACGGTTGT TGGTAGACAT 2100

TTtGTTGTGT TGGTGCTTGA TnCATTGCTG GTTGtTCATT AACCGTTGGG GTTGATTGTC 2160

CTTGTGGGAC TGCGTGTTGT TTTTCTTCTC GTTTTTGTAC GGTTTCTTTG GACTCTAAAA 2220

ACTCAAAATC TTCTACAACT AATTCTGTCA CATAATGCGT AACGCCTTGT TGGTCATCAT 2280

ACGAACGAGT TTGTAGTTCC CCATTTACAC CAACTAATGC GCCTTTTTGC GTGTATTTTG 2340

CAAAGACTTG GGCGGGTTTG CCCCATAAAA CACATTGAAT AAATGTGGTC GGCACATTTC 2400

CGTTAGTGTC TGGTCTTTTT CGCTTACAAG CCAATGTAAC GTTTGCGACA CTCTTTCCGT 2460

TAGTGGTGGT ACGTAGTTCG GCTCCTTTTG TTAAACGCCC GACAGCTGAT AATTGAATCA 2520

TGATTGTTTC CTTCTTTCTG ATTTTTTATT TATATCATGC CTTTTTTAGT GGCAATAAAT 2580

ATCTTGTTAA GTGGAAGTTT TTTTGAAATG AGAGTGATTT GCCTGTCTAG GCAAATCACC 2640

TCCTTCTATC TGCCATGCCC TTATGAGGGC ATGGCTTTTT CAATCTCTAT TTGCCATTGT 2700

TCAATGGCTT CTTCTACTTC TTCTGGTACG TGTCCCCGAT TGTATTTCCC TTGCACTTGT 2760

ATCACTCTTT TTTCACGGTA TTCTAAGGTA AAATATGGTT TATCTGGTTC TTCTTTTCTT 2820

CGGATAAAAA CAATCGTTGT TTTTCCTTTT GTATGTTGTT CAAGGTAATG TTGGCTACCA 2880

ACACAATGGT GCAATACCTG CCCTTCCTCC ACAATTTCTT GCAACGAATG TGGGGCGAGA 2940

AAAAGTAAGT CGTCAATTTC GGTTTCTAAT GCTTTCAGTT GTTTTTTTCG TTCTTGGTAT 3000

TGCTGTTCTT CAAGCTCTCT TTCAAGTAAA TTCAACGTAT TTGTGGCATT GTCATGGGCT 3060

AGTTGCAAGT TTTTAGGGAA AAGCACATCA TCAGTAAATG GTGTATTCAG TTCCTCTAGC 3120

ATGTGTAAAT AGTCCATATA GTAGCTAAAA CGCTCGCCTT TTCGCATAAA CCAATGTTGA 3180

AACTTGGTTA AGTTCACTTC TTTAGGTAAT TGCTTAATTT GTGAATAATG TAAATAATTT 3240

TCAATACCTT CTATTCTCGC AACTTTTCGC TCCTTCAAAG TTTTTTCTAG CATAAGCATA 3300

TGAAACGTTG GGTTGGTGGT TTTTAGTCGC TGTTTATTGG CTTTTAGCCA GTTCTTCGTA 3360

ATAATTCGCA TATCTACTTT TTTATGCCAT GTGCCAAAAA CATTTCGTTC GTCATTAAAA 3420

AGTATTTCGT TTGCTAGAGT AGTTGCGCCT ATCTTTTGCA AAAACTCAAT TTCATTGCGG 3480

TATTTATAAA TATGCGCCAG TTCTTGGCGA CTAATTACAG GTAACTTGAG GTATTTCAAT 3540

TCCGAATTGG TCGTTAGATT TTCTTGCCAG TTATTTGGAT AAAAAACAGT GTTCGTGTAA 3600

GCACTACTCA TGTAACCATA ATTTGCTAAC AAGCCCTCAT GCCAATGTCC GTTCTCATAA 3660

TGGGCTTTAA CATGGGTATC GTTTGAAAAC CGTTCaAGGT TTGACGGCTC ATAGATAAGG 3720

GTTTCTGTTC CATCTGTAAT CGTCTGCCAA TAACTATGGG TTTGAATTTC AATTCGtTTT 3780

GCACGAACGA GAATGATTCC AAAGGAGTAA AATTCGTTGG AAACGATAAT CGTGAGTATT 3840

TCGTTAATCG TGTTTTTATG GTGGGGGAAT TCTTGCGAGA ACTCGCAAGA ATTGTCTCCT 3900

TTnnGTn 3907

(2) INFORMATION FOR SEQ ID NO: 344:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 482 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 344:

TAAAGAGTCA AACAGAACGC TTTTTTAGCT GCTTCTTTAT TGCTATACTA CAGAGGGAAA 60

AAACGAAAAG GAGTATGAAA AATGTTTTTT AAAGAAATTC ATCATGTGGC AATCAATGCC 120

TCAAATTATC AAGCAACTAA AAATTTTTAT GTCGAAAAGT TGGGCTTTGA AGTTCTTCGC 180

GAAAATCATC GTCCCGAAAA AAACGATATT AAATTAGATT TAAAATTAGG CAGCCAAGAA 240

TTGGAAATTT TTATTTCTGA TCAATTTCCC GCTCGCCCCT CTTATCCTGA AGCTTTAGGA 300

TTACGTCATT TAGCTTTTAA AGTTGAACAT ATTGAAGAAG TGATTGCyTT TTTAAATGaG 360

CaAGGAATTG gAAACAGAAC CGCTACGAGT AGaTGaTTTT aCGGGTAAAA AAtGAcCTTC 420

tTTTTGAtCC AGATGGACTA CCyTTAgAaC yTCCwgGAAT AATAAATAAA GGGAGAGnAA 480

nC 482

(2) INFORMATION FOR SEQ ID NO: 345:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22960 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 345:

```
ACAGTTCGgT CCCTaTCCGT CGCGGGCGTT GGAAATTTGA GAGGAGCTGT CCTTAGTACG     60
AGAGGACCGG GATGGACTTA CCGCTGGTGT ACCAGTTGTT CTGCCAAGGG CATTGCTGGG    120
TAGCTATGTA GGGAAGGGAT AAACGCTGAA AGCATCTAAG TGTGAAGCCC ACCTCAAGAT    180
GAGATTTCCC ATTTCTTTAA GAAAGTAAGA CCCCTGAGAG ATGATTAGGT AGATAGGTTG    240
GAAGTGGAAG GCTAGTGATA GTTGGAGCGG ACCAATACTA ATCGGTCGAG GACTTAACCA    300
AAGAATGGAT AAGTAAAAGC AACTTGGTTA TTTTGATTCA AACTTCAATC CAGTTTTGAG    360
TGAATGAAGA TTCAACTCAA TAAGAAACAA CACCCAGTGT GGTGGCGATA GCGAGAAGGA    420
TACACCTGTT CCCATGCCGA ACACAGAAGT TAAGCTTCTT AGCGCCGATT GTAGTGAAGG    480
GTTTCCCTTT GTGAGAGTAG GACGTCGCCA CGCGAATTGT TGTTTTTGGA GGTTTAGCTC    540
AGCTGGGAGA GCACCTGCCT TACAAGCAGG GGGTCAGCGG TTCGATCCCG TTAACCTCCA    600
TTTTTTGAGC CGTTAGCTCA GTTGGTAGAG CATCTGACTT TTAATCAGAG GGTCACAGGT    660
TCGAGCCCTG TACGGCTCAT ATTCATATTT CATTTGCGGG TGTGGCGGAA TTGGCAGACG    720
CACTAGATTT AGGATCTAGC GCCGCAAGGC GTGGGGGTTC AAGTCCCTTC ACCCGCATAT    780
TTAGCCGGCT TAGCTCAGTT GGTAGAGCAT CTGATTTGTA ATCAGAGGGT CGAGGGTTCA    840
AGTCCTTTAG CCGGCACCAT TTTGCGGAAA TAGCTCAGTG GTAGAGCACC ACCTTGCCAA    900
GGTGGGGGTC GCGGGTTCGA ACCCCGTTTT CCGCTTAGCC AATTCATTAC GCCGGGGTGG    960
CGGAACTGGC AGACGCACAG GACTTAAAAT CCTGCGGTGA GTGATCACcG TACCGGTTCG   1020
ATTCCGGTCC TCGGCATTGT TTTATATTTA GCACCCATAG CTCAACTGGA TAGAGTGTTT   1080
GACTACGAAT CAAAAGGTTA GGGGTTCGAC TCCCTTTGGG TGcATAGTTG AGTGATTGAC   1140
TAAGTTTTAA AAAGGTGCAC GTGAGTTTCT TTTTAATTCG ATAGTCTTTT TTTATTACTT   1200
TAACTTCATA GCCCGGGAAG TAGCTCAGCT TGGTAGAGCA CTTGGTTTGG GACCAAGGGG   1260
TCGCAGGTTC GAATCCTGTC TTCCCGATTG TGAAAAACAG TACTGATCAT CAGTACTGTT   1320
TTTTTTTTTT GTTTAAAATT TAAGTATCGA AGTAGGTGTA TTTTTTTGTG ATGTGTAATT   1380
CTTATCGAAA TTGCGACTAA AAATACAGGA ATAAAGGACA TTGAGAATAT ATTCGATGGC   1440
TAGTTGAGAG GAAAAGGTTG AGACTTTGAA ATCTGCATGT TCAACGTTAG GAATGGGCAA   1500
AATCACTGTG GCGTATTTAG CTAATTGGCT GTTCAAGTCT GAAGTAATGG CAATAATCGG   1560
GGTGCCTCCT TCTGCAAAAA TCCGTGCTTC ATTTACGATT TCTGCAGTGT TACCACTGTA   1620
AGAGACGAGG ATGGCTACGT CATTTTTGGT GGCATTGATG GCTAAATAAA ATTGTTCAGC   1680
TTGATAAGGG ATTAAATGGA TAAATAGCTG AATGCGTAAT AACTTTGTTT GAAAATCCGT   1740
TAAGCGAATG TAGGAGTTAC TGACACCCAC TGCTAAAATA TTTTCAGCGG TTAAAAGTAA   1800
ATCAATGCTT TTATTTAACA TGGCCTCGGT TAATAACTGT TGGGTGGATT GAACAGTTTC   1860
AGTCGTTAAC TGCGCAATTT TCTTGGCAAT TAGTAAATCA GAATCTTGCG CTAAAAAAGG   1920
CATATTCGCA TCAACATTCC GGATTGCTTG GTAGTGAATT TCTAAGTCGC GGGACAAGCG   1980
```

GATTTTAAAT TCTTTGAAAC CGCTAGTCTC TGTTTTTTTA CACAAACGAA CAATCGTGGA 2040

AGTTGAAGAG TTGGTGGCTT CTGCTAGTTC ATAAATCGAC ATATGCAGGA CTCGTTCAGT 2100

ATGAGATAAA ATAAAATCAA TAATTAGCCC TTCGTTATAT GTATAGCGTT CTTTTTGTTC 2160

TAAACATTCG ATTAGATTCA ATGGGTTCAC CTCTCTGTGG TACTTTGTTT CGCTTTCAGT 2220

ATACCTTAAG CGTATGAATT ATGAAACATA GTTACAGAAT TAGTTAGTCG TATTTACATA 2280

AGAAAGCGGT TTCTTTATGA TGAGGATATC ACAAAGGATT AAGGAGAGAA GTGAAATGGA 2340

TAAGTTAACA AACTGGGTGG AGCAAGCCGT AGTGCCTAAA GTTTCAAGGA TTACAAGTTT 2400

ACGTTATTTT CAAGCCTTAA GAAATGGTTT TTTTGCGATT ATGCCGTTAA CAATTATTGG 2460

ATCGATTTTT ATGTTGATTA CAGATTTTCC AGTAGCAGGG TATGGCGATT TTATGGCTGG 2520

AATTTTCGGT GCAGGTTGGG CGGATATGAT TTCACCCGCT TATCGAGCAA CTTTCAACAT 2580

GATGGGAATT ATTTTTGCCG GGACGATGTC CTATAAATTG GCGGAAAGTT ACGAAATGGA 2640

TCGTTTAACT TCATTAATTT TAGGAATTGT TGCTTACGTA GTCGTTTTAC CGAAAACGGT 2700

GACCACTGAA TCAGGCGAgT TGTGACGAAA GTTTTATCGT TTGATTGGTT AGGAACGCAA 2760

GGGGTGATTA CGGCCATTAT TATGTCAATT CTTTCGGTTG AGTTGACTCG TTTCTGTATT 2820

AAGAAAAAAT TAGTCATCAA AATGCCTGAC AGCGTGCCTT CTATGGTGAG CCAAGCTTTT 2880

AGTGCATTGA TTCCTGGGAT TTTTGTGGTA GCGGTGGCTT TATTGATTAA TGGGATTGGG 2940

CTTTCGTTTG CGGATTCCTT CCCACAATTA ATTTATGCGG TGATTCAAGC GCCACTGCAA 3000

GGTTTAATCG GAACCCCTTT TGCAATTATT ATTGTGGCTG GCTTAAATGG ACTGTTCTGG 3060

TGGTTCGGGA TTCATCCAAC TGTCATCAAT TCCATGCTAT ATCCTATTTT ATATGCGAAT 3120

GCAGACAAAA ACCAAAGCTT AGCGGAATTA GGACAATTAA CTGCTCAAAG TGGCAACTTT 3180

GGCACGGTCC AAATGCTCGA TCAATTTGCT ACAATTGGAG GCGCTGGTTG CACGATTGGC 3240

TTGGCTATTG CAATGGCCAT CGTCGGACAT TCATCTAGAA TGAAAGCCAT GTCGAAAATT 3300

TCGTTTGTTC CAGCATTTTT CAACATTAAT GAACCCTTGA TTTTTGGACT CCCAGTTATT 3360

TTTAACCCTT TATTATTGAT TCCGATTGCG GTCGCGCCGA TTGTATCGGT CTTGATTGCC 3420

TTTCTGTCAA TGAAAATTGG GTTCATGCCA ATGTTTACAA ACATTCAAGC ACCATGGcTA 3480

CGCCCTTTCT GTTTTCGGGC TTTTTAGTGG GAAAATGGCA AGGTGCAGTG ACGCAAGTTT 3540

TAGCTGTTGC CGCTAGTGTC GTTATTTATT ATCCATTTGT AAAAGCCTTA GATCGACAAT 3600

ATCAAAAAGA AGAAGCCCAT GAAGTACCAG AAACATTGTA GAGAAAGAGG AGTGAAGGAA 3660

ATGCATGTTG AGAATGATGA AAATATTTGC TGTGAAGTGG AAAAAGGAAT GGAAAAAATG 3720

TATCATCCGA TGGTGCCCGA GTTGGTTCAA ACCTCTTCCT TTTATTTTCC AACTTACGAA 3780

GAATTTATGG CAGCGAGTAT AGATGAAAAA AATAATTATG TTTACACACG TGGCACCAAC 3840

CCCACAACAG AAATTTTAGA GAAAAAAATT GCTCGTTTGG AACGTGGAGA AAAATGTAAA 3900

GTTTTTGCCT CAGGAATGGG CGCTATTTCG GCAACGCTGT TTACGTTGCT CCAACAAGGT 3960

GATCATGTTT TAATGGTCAA TACAATTTAT GGAGAATCTG TTTCGTTTGT GCAATATTTA 4020

GAGAAATTTG GCGTATCCTT GACCAAAGTA GACGTTGCAG AAACAGAAGA ACTTTTTCAA 4080

TTTGTTCAGG AAAACACCAA AGTCATTTAC TTTGAAAGTC CGTCCTCACA AAAGTTTGAA 4140

CTGCTAGATT TAGAAAAAAT AACTGCGTTA gcACAAAAAA TTGGCGCATA CACTGTCATT 4200

GATAACACGT GGGCGTCACC GTTATTTCAG CATCCTTTAT GCCATAATGT GGACCTGGTG 4260

ATTCATTCTT GTTCCAAATA TATTGGTGGT CATTCGGATA TTGTAGCGGG GGCTGTCATT 4320

GGCAAAGCAG CGCTTGTAGA TCAAATTTTT GAACATGGTC ATCAAGCATT AGGCGCAGTC 4380

AATAGTCCCT TCAATTCTTG GTTAGCGTTA CGAGGTATTC GTACCATGCC CGTTCGGTTG 4440

GCGCATCAGT CTCAAGCCAT CCAAACGGTG CTGTTGGCTT TGCAAAAGGA TCCGCGAATT 4500

AGTCGCTTAT ATCATCCTTT TGTAGGAAAT CAACAACAAC AAGCGTTAGC AGAAAAATAT 4560

TTGACAGGTT ATGGCTCATT GTTTGCGATT GATTTAAAAG ATACCGATTT TGAACGTTTG 4620

AAAACATTTG TGAACGCATT GACGCTGGTC ACAATTGGGG TCAGTTGGGG TGGCTTTGAA 4680

AGTTTGGCCT TGCCTGTATT CAAAGGCAAT AACTTGGCGG CTGTTCAGAA AAGAGGCTTG 4740

TCGCCAGCGC ACATTCGCAT GTACGTTGGT TTGGAAGAGC CAACATCTAT TATCGAAGAT 4800

ATTCAACAAG CCTTGGATCA AGCCTATGGC GAAGAATCGC TTTTGTGAGA CATTAAGATA 4860

ATCAGCTGAA CAAAAGCAAC TCcTTCGGAA ATAAGCCGAA ATTCTCCAAA AATTAAAGAA 4920

CAATTTTCGG AAATTTCTTC TTATTTCTTG GAGCTAAACA CTTTTGTCTC AGCCTCTGTG 4980

ATTCAATAAC ATATTTTCAT TTTTAGGAGG GGAACTTAAG TCAGTTTGAC TGAGGTTGCT 5040

CCTCCTGTTT TGCATTTTTT TATTAAAGTC AAAGCGTTCC TCTTTATTTA ATAGGCaAAG 5100

TTCGATATAA TAATGGTCAA TATTAGTCAA GAGAGAAGAA GGATGGAAGA CTATGAGTAA 5160

TCaAAATACG TCAGATTTAA TCGAGGCTTA TTTAAAAAAG ATTTTAGAAG AAAGTAATAA 5220

AATTGAAATC CGTCGGGCGG AAATGGCTAA TTTATTTAAT TGCGTTCCTT CACAAATTAA 5280

CTATGTAATT AATACACGTT TTACCATCCA ACGAGGcTAC GCAGTTGAAA GTAAACGTGG 5340

TGGAGGCGGC TATATTCGCA TTGTCAAAGT ACAAATTTCA GATAACGATC AATTACTAAA 5400

ACAAATGGAT CAATTGATTG GCGCAACCTT AACAGAAAAA GATGCGCTAA CATTTATTCA 5460

AACATTGTAT GAAGAAGAAG TGATTACGAA AAAGGAAGGC AATTTAATGC TAGCTGCTTT 5520

AAGTAAAAGC ACATTGAATG GTCTGGGGAA TCACGAAGAT TTTCTAAGAG CTCAAATAAT 5580

GCGTTCATTT TTAGAGCGCT TGAGCTATGA GGAGGAATAA CTATGGATGA ACTATTTACA 5640

GAAAGTGCGA AAGCCGTACT GGCGATTGCC CAAGAAGAAG CGAAGTATTT TCGTCACCAA 5700

TCCGTCGGTT CTGAACATTT ATTATTAGCA CTCGTTTTAG AACCAAATGG GATTGCCGGA 5760

AAAACATTAC GACAATTGAA TACTGACACA GAAGACATCC GTGAAGAAAT TGAGCATTTA 5820

AGTGGCTATG GCACCATGCA AAGTCCAATG GGAAATAACA ACTTATACTT ACCTTATTCA 5880

CCAAGAGCCA AACAAATCTT TGCCTATGCT GGAGATGAAG CGAAgcsCTT AGGGGCTCAA 5940

AAGATTGGTA CCGAGCATTT ATTATTAGGC TTGTTAAGAG ACGAAGAAAT TTTAGCGTCT 6000
AGAATTTTAG TGAATTTAGG ATTAAGCTTA TCCAAAATGC GTCAATTATT ATTGAAAAAA 6060
ATGGGCGTTT CCGAACCCAA TGGCGCACAA CGTCGTCGGA ATGGTCAAAA TAAAAATGCG 6120
CCACAAGGGA CGCCAACATT GGATTCTTTA GCAAGAGATT TAACCAAGTT AGCACGAGAA 6180
CAAAGCTTGG ATCCAGTAGT TGGTCGCGGT ACCGAAGTCC GTCGTTTGAT TCAAATTTTG 6240
AGTCGCCGTA CCAAAAACAA TCCTGTCTTA GTCGGTGAAC CAGGTGTTGG GAAAACAGCG 6300
ATAGCAGAAG GGTTAGCCCA GAAAATTGTC AACCGAGAAG TCCCTGAAGA TATGCAAGGC 6360
AAACGTTTAA TGATGCTAGA CATGGGTGCC TTGGTTGCTG GAACCAAATA TCGAGGTGAG 6420
TTCGAAGATC GCTTGAAAAA AGTTGTCGAT GAAATCTATC AAGATGGACA AGTCATTTTA 6480
TTTATTGACG AATTGCATAC CTTGATCGGT GCTGGCGGTG CAGAAGGGGC GATTGATGCT 6540
TCAAATATTT TAAAACCTGC TTTAGCACGT GGCGAGTTGC AAACAATTGG TGCGACAACT 6600
CTGGATGAAT ACCAAAAATA TATTGAAAAA GATGCCGCTT TGGAACGTCG TTTTGCCCGT 6660
ATTCAAGTAG CAGAaCCAAC CCCAGAAGAA GCAGAAGAAA TTCTAAAAGG CTTGCGTTCC 6720
CGCTATGAAA AACATCATGG CGTAGAAATT ACAGATGAAG CTCTACATGC AGCGGTTCAA 6780
TTATCGATTC GCTATTTAAA TGACCGCCAA TTACCAGATA AAGCAATTGA TTTAATGGAC 6840
GAATCTGCAG CAAAAGTTCG TTTAGATAAG GCAGATCAAC CTTCAGAAAT TAATGAATTG 6900
CGTACCGAAA TTTCTCAATT AATTACGGAA AAAGAAGAGG CCATTCAAAA CCAATCATTT 6960
GAAAGTGCTG CTCGGATTCG TCAAAAAGAA AAACAAGTGA TGGAAAAACT AGAAGAATTG 7020
ATTGCAGTGA AAGAAAAAAG TCTGTCTGGT TACAGTACAC AAGTAACAGA AGAAGACGTG 7080
GCTGGGGTTG TTTCACAATG GACAGGCGTT CCTTTGCAGC AGCTAGAGAA GAAAGAAAGC 7140
GAACGATTGA TGGAATTAGA AACCATTTTA CATCAACGGG TCGTTGGCCA AAATGAAGCA 7200
GTTGAAGCTG TTTCACGCGC TATCCGTCGT GCGCGTAGTG GCTTGAAAGA CCCAGCAAGA 7260
CCGATTGGTT CCTTCATGTT CTTAGGACCA ACAGGGGTTG GGAAAACAGA ACTTGCCAAA 7320
GCCTTAGCCG AAGCGATGTT TGGCAGCGAA GAAGCTTTGA TTCGCGTGGA CATGTCTGAG 7380
TTTATGGAAA AATACAGTAC TAGCCGTTTA ATCGGTTCTC CTCCAGGTTA TGTCGGCTAC 7440
GAAGAAGGCG GTCAACTAAC TGAAAAAATT CGTCAACGTC CTTATTCCGT TATTTTGCTA 7500
GATGAAGTAG AAAAAGCCCA TCCAGATGTC TTTAATATTT TATTACAAGT CTTAGATGAT 7560
GGCCATTTAA CGGATGCCAA AGGACGAAAA GTCGATTTTC GTAACACGAT TTTAATCATG 7620
ACCTCCAATT TAGGGGCAAC AGCGATTCGT GAAGAAAAAC ATGTTGGCTT CAATGTGAAA 7680
GACATTTCTA AAAATCATGA ATTAATGCAA AAACGAATCA TGGAAGAATT GAAAAAAGCC 7740
TTCCGTCCaG AATTCTTAAA CCGAATTGAT GAAACGGTTG TATTCCATTC CTTGAAACAA 7800
GAGGAAATTC ATGAAATCGT AAAAATTATG AGTCAATCTG TCGTGAAACG GATGGCTGAA 7860
CAAGAGGTTA AAGTGAAAAT CACACCAGCT GCAATTGAAG TTATTGGTAA AGTCGGGTTT 7920

GATCCAGAAT ATGGCGCACG TCCAATTCGT CGGGCCTTGC AAAAAGAAGT GGAAGATCGT 7980

TTAAGTGAAG CCCTTCTTTC TGGTCAAATT CAGTTAGGGG ATAAAGTAAC TTTAGGCGCT 8040

AGCAAAGGCA AAATTACGTT AAATGTGCGG GCACCAAAAG CACCAAAGAC AGAAGCGAAA 8100

GAATTACAAA CTGTATAGAT GAGATGCAAA GGAGGTCGGG ACAGAAGTGT TTAACTCCGA 8160

GAAATAAGAA GGAATTTCCG AAAATTGTTC TTTAATTTTT GGAGAATTTC GGCTTATTTC 8220

CGAAGGGgTT GCTTCTGTTC TCGCCGTGTA TCAGTTTTTG AGCGTGGAGC AAAAATCCAA 8280

AGTGATTTTT GTCCCACGCT CCTTTTTCTA TTTTCCAGTG GCGATTTCTT TTTGAGAATT 8340

GCCAGATTGA TTTTTCTGTG CTACGTTAAT AAGGAGGGGT GATGGAATGG AACAATCTTT 8400

TTTTCGAGAA TTACTACATT ATCAACGTCC AAAGGATAAT CCTTTTCGAC TAGTGTTTGC 8460

TGGCTTGGTT ATGTTCAGCA TTCTTTTTTT GGGCTATATC AATCATCAAT TATTGATTTC 8520

AAGCTTTGGC TCGTTAGGTA TTTTTACTTT TTTATATTAT CAGCCGCTAC CATTAAAACA 8580

GTTAATGACT CGTTTATCTG TGGTCGGTAG TTACCTTTTT TTAGGAAATT TGTTAGGCAT 8640

GTTATCGACC CATATTGCAT GGCTTATTCC CATTGTTGTT GCACTGGTTG GTTTTGGTGG 8700

CCGCTTTTTC TTTGAAGTCT ACGATATTTC AAAACCAGGG GCCTTCTTTG GTGTGATGGT 8760

GACAGCGATG GGTGCCAGTA CGAGCATTCC GCTAGCAAAA ATTCCGTTCA TGAGTGGCTG 8820

TATGTTATTA GGCATTGGCT TTTCCATTTT GTTTGCACTA GTGTTGCATT TTACAGAGAA 8880

GGAAATCAGT GTCCCGCTAG CTAAGCGATC ATTACGGGAA CGAATTTATC AGCATCCAGA 8940

AGCGCCATTA GATAGCATAT ATTATAGCTT TGTTTTATTT TTTGCTGTAT ATATCAGCGA 9000

AAGTTTACAT TTACAAAATC CGTATTGGTT AGTCGTTTCC TGTGCATCCA TTCTACAAGG 9060

AGACAATTTA CGTGCGATTA AGCAGCGAAA TATTCAGCGA ATTTTTGGCA CAACGATTGG 9120

ATTGGTTATT TCCGCTTTTC TGCTTAATTT AGCTTTGACA ACTTTTGAAT CGATTGTCGT 9180

GATTACAATC TTATTTGTGA CCGTTGAGTA TTTTATTCGT CGTAATTATG GATTGGCGCA 9240

ATTTTTCACA ACGCCTATGG CTCTGATGCT TTCATTGCTC GTTCGTCAGC AGTACGTGAT 9300

AACATTGATT CAATTCCGCT TTTTGGGGAT TGTTTTAGGC AGCTTGTTAG GGcTTAGCAG 9360

CAGCGTGGCT ATTCACCGTT GTAGTAACGT TTTATGAACG AAAATATCAC TTGAACGAGT 9420

GAnGAAGAAC AGATGACCAG TAATTAAAAA AAGAGCATTT GGTAGGTAAG CGGATGGCGC 9480

TTTTCTACCA AAATTTTTTC GGAAACATTG TGAAAAATGT ACGTTTAATA CAGATAGGCT 9540

CTTTTAAAGT TTGTGAACGT ATGATACTAT TAAACGGTCA TAAAGAAGGA GGCTATATTT 9600

ATGATTGAAC TGATTGCAAC TGCAGAATCT gTTGAGCAGG CTGTAGAATT ACTCGCAACA 9660

GGTGTGGATA CATTATATAT TGGAGAAGAA ACATTCGGGT TACGTTTGCC CGTATCTTTT 9720

ACACGAGATG AACAACGCGA GATTGTGAAA CGTGCACATG AAATGGGCAA ACAAGTGCTA 9780

GTTGCAGTAA ATGGCATTAT GCACCCTGAA AAAATGAAAT TAGTCCCAGA GTATCTGACA 9840

TTTTTAAAAG AAATCAACGT CGATAAAATT ACTTTGGGTG ATCCGGGTAT TGTGTTTATT 9900

ATGCAACGAG ATGGATTAGA AATTCCTTAC GTTTATGATG GCGAAACCTT GGTAACGAGC 9960

TCACGTCAAA TCAATTTTTG GAGTAAACGA GGAGCCATCG GAGCTGTCCT AGCTCGGGAA 10020

GTACCTTTTG AAGAAATGGT CGCAATGGAA GAAAACTTAG CAGTTCCAGC TGAAATTCTT 10080

GTCTATGGGG CGACCTGTAT TCATCAATCG AAACGCCCAT TAATTCAAAA TTACTACAAT 10140

TACACAAAGA ACGACAAAGG GGTTACGAAA GACGAAGGGC TGTTTATTTC TGAACCCAAA 10200

AAACCAGAAA CCCACTATTC CATTTATGAA GATAGTCATG GTACGCATAT TTTTGCCAAT 10260

AACGATGTGA ACTTGATGAA TGAGTTAACA AACTTGTATG ACCATCACTA TCGTACATGG 10320

AAGCTCGATG GCCTTTACAC GCCAGGTGAA AATTTTGTCG CGATTGCTCG CTTATTTGCT 10380

GAAGCGAAAG CAGCACTTGA AGCGGGAACT TGGTCAGAAG ATCAAGCAGC TCATGCAACA 10440

GCTGAAATTG AAAAATTACA TCCAACAGGT CGTGGCTTAG ACAATGGTTT CTTTGATTTA 10500

AGTGCCGATG CAATAAAATA AATTGATTCA TTTAAGGGGA AGAAGAACAT GAAAGTAATG 10560

ACAGAACGAA CATTGAAGCG TCCCGAGGTG TTGGCTCCAG CCGGGACGCT AGAAAAACTA 10620

AAAGTCGCAA TTCGTTACGG TGCCGATGCC GTTTATATCG GAGGGAATGC CTACGGTCTG 10680

CGTACCGCGC AGGAAATTTC ACGCCAGAAG AAATGGCGGA AGGCGTAcTT TGCCCGCGAA 10740

CATAATGCGA AAGTTTATGT CGCTGCCAAT ATGGTCACTC ATGAAGGAAA CCAGGAAGGG 10800

GCAGGAGCGT TTTTCAGAGA ATTGCGAGAC ATCGGAATTT CAGCTGTGAT TGTTTCCGAT 10860

CCAGCTTTAA TCGAAATTTG TGCGACAGAA GCACCAGGTT TACCCATCCA TTTATCGACA 10920

CAAGCTTCAG CAACCAACTA TGAAACATTA GAATTTTGGA AAAATGAAGG CTTAGAACGT 10980

GTGGTTTTAG CGCGCGAAGT TTCAATGGAG GAAGTTGCAG CGATTCGTGA AAATACGGAC 11040

ATTGAAATTG AAGCCTTTAT TCATGGAGCA ATGTGTATTT CATATTCAGG ACGTTGTACT 11100

TTATCCAACC ATATGTCGAT GCGGGATGCC AATCGTGGCG GTTGCTCTCA ATCTTGTCGT 11160

TGGAAATACG AGTTATTTGA TATGCCTTTT GGGACAGAAC GTCGGAGCAA AACCAGCGAA 11220

GGGGAAGTTG AAGAAGAGTT TTCAATGAGC GCCGTAGATA TGTCTATGAT TGAACATATT 11280

CCAGAATTAA TTGAAAATGG TGTCGATAGT TTTAAAATTG AAGGCCGCAT GAAATCCATT 11340

CACTATGTTT CGACAGTCGC TAATGTGTAT AAAAAAGCCG TCGATTCTTA CATGGAAGAC 11400

CCTGAAAACT ATGTCTGCCA ACAAGAGTGG ATTGATGAAT TATGGAAAGT AGCACAACGT 11460

GAATTAGCAA CTGGTTTTTA TTACAATACA CCAAGTGAAA ATGAACAATT ATTCGGCGAA 11520

CGTCGTAAAA TTCCGCAATA TAAATTTGTT GGTGAAGTGA TTGCGTACAA TGAAAAGACG 11580

CAAGTCGCAA CGATTCGTCA ACGGAACCTT TTCAGTGTCG GCGATGAAAT TGAATTTTAT 11640

GGTCCTGGGT TTACCCATTT CCACCAAACC GTTAAAGAGA TGTATAATGA AGACGGTGAA 11700

GCGATCGATC GGGCACCTAA TCCCATGATG TTATTAACAA TGGAAGTGAC GCAACCTGTT 11760

GCAGTCGGCG ATATGATTCG TAAGAAAAAA TAAACCAGCC AAAAGGAGGA GTAATTATGA 11820

GTGTCCGCGT GAATAATTTA TTAAAATTAC CGTCTTTGCG AGAAGCAACG GTACTTTCTG 11880

GTAAAAACCA ATTGGACACA TCTGTTGCTT CTCTTTCTTT TTTAGAAGTC TCTGATATGT 11940

CTTTGTTTTC GGAAAAACTT CACAAAACAA ATGAATACCA TGCTGGCGAA ATTTTGATTG 12000

GTTCATTTTG TGCCATTCGT CATGATGTGG TGAAGCAATG CGAAACGATT CGGCATCTCC 12060

ATGAATTAGG CGAAGTAGGC CTAATCTTGT ACTACGTTGG GATTATCGTA CCGAAAGTTG 12120

CAGAGGAAGT GATTCAATTA GCGGATTCTT TAGGTTTCAT TTTAATCCAA ATGCCTAAAA 12180

ATGATCCCAC ATTACGCTAC AACGAAGTCA TTTATGAAGT CATGAAATTG TTAGTCAATA 12240

AAAATAAAAT GGCTTCTTTT GCAACGGACC TCTTGGAAAA AGTTTCCCTT TTGCCTGAAC 12300

AAAACCGCAG TGTGGAAATG ACCTTGAAAA TCCTATCCGA TTTTTTAAAA GTCAATATTG 12360

CGTTAACCAC ATATACCAAT GAAATCATCA GCACCATCAA TTGGCCGCGA AgCACGAAAT 12420

TGCCCTTAAC CAAGTTACTT TATCAATCAC AAGAGAAAGT AGCGCTGGGA GAAACGACCT 12480

ACTTCATTAC CTCGCAAAAA CTTCCTCAGA AAGATGGACA AACATTGAAG CTTTACTTCA 12540

TTCGAGAAGA AGAGCAGTTA ACCTCTTTTG AGATGAATCA AGCGGTGGAA GTTGTTCAGA 12600

TGGCTCTGAA TTTATGGGGG AAAAACTATG ACGAAGTCAG TGaGTATGCC TTAGTCCAAG 12660

CAATCATTAA CGATGAAAGT GATAAGATGT ACCAATTAGC CAAAAAATTA ATGATTGATA 12720

TTTCTTCTGT CGAAACGATG TGGCTTTTGC CCATTTGGCA ACAAGGACAA GAAATCCACG 12780

CTATTCAGCA AGATTTGAAA GATTTCCTAC GTTACTATTA CCAAACAACG ATTGTCCAGC 12840

AAACAGAAAA TTATCTGATT GTTCTCTTAG GTAATTATTT ACACAAACAG CCTGAAGCCA 12900

CGATTAGTAA AGAATACTTG CAAGCCACAG ACTATCAGGA CTATTTAGGT GAAATCATCC 12960

TTTGTCCGAA AGTTCGTAAC ACGACTGAAG TGCGTGTCAG CTACCAAACC GCTAATCAAG 13020

TCGCGCCGTT TTTAAAGCGT GTCTATCCAC AAAGAAAAGT GTTATCAATC GCGGAAATTC 13080

GTACAGTTCA ACAAATGATG ACGTGTCTCC AAGCAGGAGA AGAAGCAATC CAACAACGCT 13140

TAGCAGTTAT CCAACCTCTC TTAAAAGAAT CAGAGCAGTT AATGACATTA GGAACGTTTT 13200

TACTAGATGC AAGTAGTGAT TTTAATGAAT GCGGCGAACG CCTATTTGTT CATAAAAATA 13260

CGGTCAAATA TCGTATCAGT AAAATTAATG AAGCGCTCGG CTATGATATG ACTAACGCCG 13320

CCGAAGCCTA TGAAGTCTAT TTGGCGATGA TCTTATATCG CTTGGTCACG AATGAATGAA 13380

AAAAGCAGAC AGAGGTCGGG ACAGAAGTGT TTAACTCCGA GAACCACGTA GGTACTGTTC 13440

AACACTCTCA AAGGCTCAAA TACATCGAGT AATCGCAGAT GTCATCTATT TGCTTTGCAA 13500

ATAGTGTTTC TCAGCAATAA GAAGGAATTT CCGAAAATTG TCATAAAACA CAACGAGGTA 13560

CGAGTTGATG TTGCATGATA CCTACCCGGT GTTCTTTAAT TTTTGGAGAC CAAGTAGGTA 13620

TTGTTCATCA TCGAAtCACT ACGTTCTTCG TGATTCTCAA CAATTTCGGC TTATTGTTTT 13680

CTCGCCAAAG GCTCGTCACA GTTTTATCCT TGTAAGTTCC GAAGGACCTT CACCAAGTAG 13740

GTACGATGAT TTCATCTGTT CGTGCTTCAC AGTGAAACAT CGCAACTCTT AGAGCTGTAG 13800

CTCTTTTGAA GGCTGTACTG TGACGAGACG CATGTCGAGA AAACATCTTA GAAGTTGTTG 13860

AGATCGAAGC AACGCGTAGT GTTGCCTCTG TTTCCGCCGT TTATCAATTT TTAAGsCTGG 13920

GACAAAAATC CAAAGTGATT TTTGTCCCAG GTTCTTGTCT ATTTtTAtTT TAAGCCAAAA 13980

CTTCCTCGAT ATTTTTTTTA ATTTGAAAAA GACCTATGAC TGATTTTTCG TTATAATTGA 14040

AAAAATAATT AAGGAGGAAT GAGAATGAAA TTAAAAAACG AGCAACAACA ATCGTGGCTT 14100

AGCTTAGCAT TTGTTTGGGC GGGCGCAATG ATTAGCGTGC CAGGATTAAT TATTGGAGGA 14160

ACGTTAGTTG CAGGCATGCC TTTATGGGAA GCTTTGTTAA CAGGATTTAT TGGTTACGGT 14220

ATTATTGTTA TTTTAATGAT TTTACAAGGA ATTCAAAGTT CCGACTTACA AGAACCTTCT 14280

GTCAAAGTCG CTTCTCAAGT GTTTGGCATT CAAGGTTCGC AAAAGATCAT TTCTATTATT 14340

CTAGCGATTG CTTGCTTAGG TTGGTTTGGT TTGCAAGCCA ATGTCAGCGG CGGCGCCTTC 14400

ACAAACTTTT TAAAAATTTA CGGCATTGAT TTACCTGTGT CGCTTTCTTC CTTAATCTGG 14460

GGAATTATCA TGCTTATTTC GGCTCTTTAT GGCATAAAAA TATTGAAAAT TCTGAATTAT 14520

TTTGCGGTTC CTGTGTTAGT TCTGGTTTGT CTTTATGGCT TGGTGGCTTC GTTAAGAAAC 14580

AACGGTTGGG CGGCTGTTAG TCAATATACC CCTCAAACAG CAGGAAGCTT TATGTCAGGT 14640

CTTTCAATGA CTGTCGGTTC TTTTGCATTA GGCGCAGTAA TTGCTGGAGA TTACTCTCAA 14700

TATGTTTCTT CTCGTAAAGA TGTGGTCAAA GCGGCCACCT TAGGGATTTT ACCGACAGGC 14760

TTGTTGATGA TTGGCGTTGG TGCGGTCTTG ACCATTGCAT CGAACACCGC AGATATCACG 14820

GAAGTTTTTA TGAATCTCGG TTTTCCAGTT TTAGGAATTA TTGCGCTAAT TTTAGCAACT 14880

TGGACCACGA ATGCCGTGAA TGCTTTTTCT GGCGGACTGG CCTTGATTAA TGTCTTTGAT 14940

ATTCCAAAAG AAAAAGAAAA AGTCGCTGTC GGTGCGGCCG GCGCGATTGG AACCTTATTA 15000

GCAGTTGTGG GGATTTTAAA TTACTTTACG CCGATTATGT CCGTCCTTTC TGCGATGGTG 15060

CCACCAGTGG CTGGTGTAAT GATTGCTGCC TATTGGTTGA TTAACAAAGG CGATCGAACC 15120

AAATGGCAGC CAACGCCTGG CGTGAATAAA TTAGGTGTTT TTTCATGGCT AATTGGTGCT 15180

GCAGTGGGGG GCATTCCGGT GATTATGAGT TTCTTTCCCA ATGCCCCTCA ATTGCCTAAC 15240

CAACCGCTAA TTGGCATTGT TTTATCGTTT GTTATTTATT ATTTCGGTGC AAAACGTCAA 15300

CAGACGAACA CCGAATCATT AGAAATGGAA TGAAGAAGAG GAGTGTCGGA GAATGCGTTA 15360

TTTAGATGCA GAAGCAATTG AGAATATAGC CACAGGAGCA GCTTTTTTAG GGACAGGAGG 15420

TGGCGGTGAT CCCTACATTG GTAAAATGAT GGCTTTGTCC GCCATCGAAG AAAACGGACC 15480

TGTCAAACTG GTTTCTCCAG AAGAAATTGC CGCGGAGGAT TTTTTCCTAC CCGCCGCAAT 15540

GATGGGTGCC CCATCTGTCG CAATTGAAAA ATTTCCCAAA GGCGACGAAT TCGTCCGTGT 15600

CTTTGAGAAG TTAGGAAAAT ATTTAGACCA AGAAACGATT GCGGGAACCT TTCCAATGGA 15660

AGCTGGGTGC GTCAATTCAA TGATTCCAAT TGTTGTTGCA GCGAAgcTAG gCATTCCCTT 15720

GGTGGATTGC GATGGTATGG GTCGGGCCTT TCCAGAATTG CCTATGGTAA CGTTCCATTT 15780

GAATGGGATG TCAGCGACCC CTATGGCAAT TACCGATGAA AAAGGAAATA TTGGCATTAT 15840

GGAAACGATT GATAATACTT GGACAGAACG TCTTGCTCGT GTTCAAACGG TTGAAATGGG 15900

CGCCAGTGCT TTAGTGAGCA TTTATCCCGC GACAGGCAAA CAATTACAAG ACTATGGGAT 15960

TCACAACATC GTGACATTAT CAGAAGAAAT TGGCAAAGTG ATTCGAGGTA CCTATGCAGA 16020

TGAACAAGAA AAACGCCAAG CATTAGTAGA AGTTACGGAT GGCTTTGAAT TGTTCCAAGG 16080

AAAAATTCTA GATGTGGAAC GAGAAGTAAA AGGTGGCTTC AATTTGGGAC GTGTCAAATT 16140

GAGTGGCTTA AACAGTGACG CTGGTTCAGA AGCAGTCGTC CATTTTCAAA ATGAAAATTT 16200

AATTGCCGAG AAAGATGGTC AGGTGATTGC GATGACGCCT GATTTGATTT GTATGGTAGA 16260

TTTAGAAACT TTAACGCCTG TGACAACAGA AAGCTTGAAA TACGGCAAAC GTGTCCAAGT 16320

AATGGGCTTG AAAGCGAATG CCGCTTGGCG AACGAAAAAA GGTATCGAAA CAGTGGGTCC 16380

TCGGTATTTC GGCTACGAAA TGGATTATCA ACCACTAGAA AACTTAGTAG CAAAGGAGGA 16440

CAAATAAATG TATAAAATCG GTATCGATGT CGGCGGCACA AATACAGATG CTGTTATTTT 16500

AGATCATCAA TTAAACTTGA TTCATTCAGT GAAAGTTCCA ACAACCGATG ATATTCAAAC 16560

AGGGATTGCC GGTGCCTTAA ATAAAGTCTT GGCAGAATCA GCAGTGGACC CAACCAAAGT 16620

CACACATGCA ATGTTGGGCA CGACCCAATG TACGAATGCC ATTGTTGAAC GAAAAAAATT 16680

AGCAAAAGTT GGGGTCCTTC GTTTGGGCTA TCCAGCAACC GCTTCGGTCT TGCCTTACAC 16740

AGCTTGGCCC AAGGATTTAG TGGCCACTCT TTCAGAAACA TATGCTTTAG CACATGGCGG 16800

CTACGAATAT GATGGACAAC CATTGACTGC ACTGGACGAA GAAGAATTAC GTGGGATTTT 16860

GGCCTCTTGG CGAGGAGAAG TAGAGGCAAT CGCAGTAATT GGCGTTTTTT CTTCTTTAAA 16920

AAATGACCAA GAATTGTTTG TTCAAGCACT TGCGAAGGAA GTGTTAGGGG CAGATGTTCC 16980

TGTATCTTGC TCCTCAATGA TTGGGTCTGT CGGTTTAATT GAACGGGAAA ATGCGACAAT 17040

TTTGAATGCT GCGTTGCACA AAGTAATTAA AGTGACCAGC GAAGGCTTTG AACAAGCCTT 17100

AGAACAAGAA AAAATTCATC ACGCCCAAGT GTACTTATGT CAAAATGATG GTACGTTAAT 17160

GTCGTTAACG TATGCCAAAC AGTTTCCTAT TTTAACCATT GCTTGTGGAC CAACCAATAG 17220

TATTCGTGGC GCTTCATACT TAGCTGGCTT GAAAGATGCG GTTGTTTTAG ATGTTGGTGG 17280

CACAACGTCC GACATTGGCG TGTTAGTGGA TGGCTTTCCA AGAGAATCTT CACTGGCCGT 17340

TGATGTTGGC GGGGTACGGA CCAATTTTAG AATGCCAGAT ATCGTGTCAA TCGGTGTCGG 17400

CGGCGGCAGT CTAGTGCGTG AACAACCAGA TGGTTCTGTG ACAGTTGGCC CCGATAGTGT 17460

CGGCTACCGT ATTACACAAG AGGCCCTTGT TTTTGGCGGC ACACAATTAA CGACCACTGA 17520

TATCGCGGTT CGCTTAGGCC ATGCCCAAGT CGGTGATCCT AGCAAAGTGG CTCATTTAGA 17580

CCAAGCGTTT GCGGAAAAAG TGTACCAAAA AATTGGCGAA TTGGTCAGTG AAGCGATTGA 17640

TCGCATGAAG ACCAGTTCCG CAGACGTGAC GGTCGTTTTA GTTGGTGGTG GGAGTATTAT 17700

CATCCCCGAA GAATTGACAG GCGTTAAAGC GCTAATTCGC AACAAAAACG GAGCAGTGGC 17760

GAATGCCATC GGTGCCTCTA TCGCTCAGAT TAGCGGACAA TACGAACAAA TTTATGTTTA 17820

TTCAAAAATC CAACGAGACG CGGCGTTGAC TGATGCCCAA GAAAAAGCCG TCCAACAAGC 17880

TGAACTTGCG GGTGCAGTCC CTGGCACAAT CGAGTTAGTC GAGGTGGAAG AAACACCACT 17940

CGCCTATCAT CCAGAAAATG CAACCCGCTT ACGAGTAAAA GTGGTTGGGA ATATGTATTG 18000

AGGACCAGTG CAAAGCAAAA GTTATAAGAA GAGTTATTTG AAATATGAGG TCGGGCTAGG 18060

GCTGTCGATT GAAGTTTTAT CCTGAAGAGT ACAGAAGAAG CTATATCTGT CTCGCCGTTT 18120

ATTTTAAAAT TCAGAgtTAg agTaAGTCyA AATGACTTAC CTTCCTAAgT TCATTTTTTT 18180

GTGTTGACAT ATTTTTTTAG AAAAGGnCAA AATGTTAACA TGGTATGTTG TTGGGGTGTT 18240

ATGAAAAAAG GAGAGAAAAT AATGAGTATT TTAAAAGAAC ATTTAAAAAG AAATTGTGAT 18300

GCACGAATGG TTGGTGTTTT ATCGTTAAGC ATAAGAGAAG CATATCGCTA CTTAATGGAG 18360

CTAACAGAAA GAGAATCTAT TTTTCAACGA GCAGAAATGA AAAAAGTTTA GGGACACGTT 18420

CGGCATGGAT TGGTTGATGT TGGTTTAAAA CAAGTGTTAA CTTCATCAGC AATCCCACAT 18480

GAAATAGCTG ACAAAGCTAG TAGTCGATAT GTCAATGGTC ATACGTACTT AATGATTGAA 18540

ACCAAAGGCG CAATTTTAAC GCCAGCGAAA GTTTTATCAG AAGCTTCAGT TCCTAAAAAG 18600

GCATTATTTA GAAATCGTGG GAGCTTATTA AATAAGCAGT ATAATTTATT TGACAAACCA 18660

GAAGATTTAA ATGAGTATTA TGATGCCAAT CAACCGCCAT TCTTATTACT AACTTATGGC 18720

GGATCGAATC ATCAATTACG TTTTGTTCGA TTAGGGTTAC CAGATATTGG AGTTGGTCGT 18780

TGGATTGACC AAATTGATAT TACGCAGGCA CCAGTTCTTT TAAAAAATCC AGAGGAAGTT 18840

CGTAAGGACC TACATTTGAC TTTTACGTCA ACCGCTGACG AACTGATAAG GAGAGGACTA 18900

GAAAATGAAC GAGAGGTTGA TTTTTAATCC AAAGCAATTA ACTTCGGCTC GCATTGCACG 18960

AGGATTAACA ATGAAAGAAT TAGCCGAAAA AGCAGAGCTA TCTAGGCAAA TGATTTCGAA 19020

TTACGAATCG GGAAAAACAA TTCCTAAAGC AGATAGCATA TTAAAACTAA TTTCTGTTTT 19080

ACAATTTCCC CGTAGTTTCT TTTCAGCAGA TACATTTGAA TTGTATTCAG GCGCAACTTT 19140

TTTTAGAAGC CGAAGTGCGG CGACGAAGAA AGTAAGAGAT ATGCAAAAAG AGCGGTTAAA 19200

ATATGTGCAA GAAGTGTATA AAGTATTAGC AACATATGTG AACTTTCCCA AAGTTTGTTT 19260

ACCAGAATTA ATTGAAAAGA GTATTTATGA AATCACAGAA GAAGAAATTG TGAAAAAGGC 19320

TGAGGAATTA CGGAAAATTT GGGGCTTAGA CTTGATTTCG CCAGTCCCTA ATCTCATTCA 19380

GTTGGCTGAA CAAAATGGTG TGATTATTTC AGAAGCCAAC ATGTCCAATC CAACACTAGA 19440

TGCGGTCTCG CGTTGGATGA TTGGTCGGCC GTTTATTATG CTCACAGATA ATCATGAATC 19500

TGCTGTCCGT CGTCGTTTTA ACGTGGCGCA TGAACTAGGA CATATTTTAT TACACAACGG 19560

TGTTGAAAGT ATTCATGAGT ATTCGCAAAC AGAATTAAAG AACATTATTG AGTATCAAGC 19620

AAATCTATTT GCTGCGCATT TTCTCCTTCC TTCAGCGGCT TTTTCTGATT CGCTCTTGTC 19680

TATTTCGTTA GAATACTATA TTGATTTGAA AAAGTATTGG AAAGTTTCTC TTCAATCAAT 19740

GATTCAAAAA ATATACACAT TAGCGTTAAT TAATGATGAT CAACGTCTCT ATTTAAATAA 19800

AAGAATTGCT AAAAATAAAT GGAAAACGAA AGAGCCTTAC GATGAACAAC TGACTATCGA 19860

ACAGCCAGAA GTTTTCCAAA AAGTTTACCA CATGATTTTA GAGCACCAAG TCATTCCTGC 19920

AAGTCAACTA AATCAAGCTT TTAAATTACC TAAAGAAGAA ATAGAAAAGA TATTAGGAAC 19980

AGCTATTCAA ACGGAGCAGG ACAAAACAGA AATTGTTCTT GAATTATTAC GTTCTAACAA 20040

GTAATCAAAA AGAAGGAGGA ATAACTATGG CTAAATTATT CGCCTACCAA ATTGGTCAAA 20100

ACCCTCGTAT TCAAACGGAC CTTTTGGTAG ATCCTCAGCT GTTTGAAGAT GAGCATGGCT 20160

GTATGGGTGC AGTGGGCTTC GGGTTAGCAG ATTGTGTTCA AACCGGAATG TTTACAGATA 20220

TTGAAGTGAT AAAACGGTAT CTTCACGAAG CAACCTACGT TTTTATTAAT GGTGACTTTG 20280

ACCGACTAAG TTATTTAGAA ATTGGGATTG CACTTTCCTT AGGCAAAACA CTTTACGTCA 20340

TTACAATGAA CCCAAACGTA ACGAAAGAAG ATTTAGGAAT TCCCTTCGAC AATGCTACGA 20400

TTGAATTCCT TTCCCCCAGT GCTTTTACGG AACGAATCCA TGAAACAGAA GCAGCAGAAA 20460

ACTAAACAAA AGTAAGTAAT ACGACTGTCT TTTTTTGCTA TTTTTTGCGA TAATAAGGGA 20520

GAAAGTTCAA ATTGAAAGGG AGCCAGAATA AATGAAAACC TATGATTATA TTGTTATTGG 20580

TGGCGGTAGT GGCGGGATTG CTTCTGCTAA CCGCGCAGGA ATGCACGGAG CTAACGTCCT 20640

TTTAATTGAG GGCAACGAAA TTGGTGGGAC CTGTGTCAAC GTTGGTTGTG TGCCTAAAAA 20700

AGTCATGTGG CAAGCCAGCA GCATGATGGA AATGATGGAA AGAGATACAG CAGGCTACGG 20760

GTTTGATGTT GAAATCAAAA ATTTTAGCTT TAAACAATTA GTTGAAAATC GTGAAAAATA 20820

TATTGACTTC TTACATGGTG CGTATAATCG TGGTTTAGAC AGTAACAACA TTGAACGTAT 20880

CCATGGTTAC GCTACGTTTA CTGGGGAGCA AACAATTGAA GTCAATGGGA CAGAATATAC 20940

GGCGCCGCAC ATTTTAATCG CGACAGGTGG TCGTCCGAAA AAACTAGGTA TTCCAGGAGA 21000

AGAATATGCC TTAGATTCAA ACGGCTTCTT TGCGTTAGAA GAAATGCCAA AACGTGTGGT 21060

CTTTGTCGGT GCGGGTTACA TCGCAGCCGA ATTAGCAGGA ACCTTACATG GCTTAGGCGC 21120

AGAAACACAT TGGGCCTTCC GTCATGAACG ACCATTACGA AGCTTTGATG ACATGCTTTC 21180

TGAAAAAGTC GTAGAACGCT ATCAAGAAAT GGGCATGCAA ATTCATCCGA ATGCAACACC 21240

AGCTAAAATT GAAAAAACGG CGCAGAATGA GTACGTTATT ACATTTGAAA ATGGTGAATC 21300

AATCACGACA GATGCCGTGA TTTTCGGGAC AGGCCGTCAG CCGAACACGG ATCAATTAGG 21360

CTTAGAAAAT ACGAAAGTAG CCCTTGATGA AAAAGGCTAC GTCAAAGTAG ATAAATTCCA 21420

AAATACCACG CAAAATGGCA TCTACGCAGT CGGCGATGTC ATTGGTAAAA TCGACTTAAC 21480

GCCAGTCGCG ATTGCCGCTG GTCGGCGCCT GTCTGAACGC TTATTTAACG GACAAACCGA 21540

TTTATATCTG GATTATAACT TGGTGCCAAC GGTGGTCTTT ACCCATCCAC CAGTTGCGAC 21600

AATTGGCTTG ACGGAAAAAG CAGCGCTTGA AGAATATGGT GAAGACCAAG TGAAAATCTA 21660

TCGCTCTAGC TTTACACCAA TGTATTTTGC TTTAGGAGAA TACCGTCAAA AATGCGACAT 21720

GAAATTAATC TGTGTCGGCA AAGAAGAAAA AATTGTCGGT TTGCATGGCA TCGGCATCGG 21780

TGTTGACGAA ATGCTACAAG GCTTTGCAGT CGCTATTAAA ATGGGCGCGA CCAAAGCTGA 21840

TTTTGATAAT ACTGTTGCGA TTCATCCAAC AGGCTCAGAA GAATTTGTAA CAATGCGATA 21900

AAAGGCGAAT AAACCCCAAA AAAGAGCGTT GCTGATGCAA CGCTCTTTTT TGAATACACG 21960

AGTTATACTA TTTGATTCGT AAAAAGTAAT TTTTTATATC ATACCAATAA CCACTTTTCC 22020

ATCTTTTAAC TACTGACTTT TCCAAGAAAT ATGCTATGaT TAAAAAAATC AGCAAAAACAG 22080

AGGGTAGGGA GTTAGAGGAT GGATAAAATT AGAATCAACA ACTTGAAATT TTTCACAAAA 22140

AATGGTGTTT TGGCAGAAGA AAAACGTCTA GGACAGCAAG TAGAAATCGA CCTTGAAATG 22200

CAGTTGTCGT TAGCTGAGGC AGGGCGAACA GATGATGTCA CTCAAACTGT CAATTATGCC 22260

GAAGTGAATG AATTGATTGC CCAACGAGTA AACAATCACT CGTACGATTT AATCGAAGGT 22320

TTAGCGAGTG CTATTTTAGA TGATATTTCC GCAGATTATC AAGAACAATT GAACAAAATT 22380

ATCATTAAAA TTAGAAAATA TAGTGTGCCG ATGCCCGGGC TTTTTGACAA TATTGAAATT 22440

GAAATGGAGC GAGAAGTATA GATGATTGGC TATGTAGCAT TAGGTAGTAA CGTCGGCGAT 22500

CCATACGCAA ATTTAGTGAA AGCCCGCAAT TATTTAAACG AAGACCCGCA GATTCAGGTC 22560

GTGGCTTCTT CAAAACTTTA TGAAACGGAT CCTTACGGCT ATACAGACCA AGATGATTTT 22620

TTGAACGCCG TATTGAAGAT TGATACTAGC TATACCGCTG AAGAGTTATT GGCTGTTACC 22680

CAAGGGATTG AGCAAGCAAT GAAACGTAAA AAATTAATTC ACTGGGGTCC GCGGATTATT 22740

GATATTGATA TTTTATTATT AGGGGGGCAG TCAGTCACGT TGCCCCACTT ACAAGTTCCT 22800

CATCAAGAAT TAACAAAACG GTCATTTGTT TTGATTCCCT TAAGTGATGT TTACGCAGAG 22860

GAACAACTAT TGGGCGCGCC AATAGCTACG TGGATTCAGC GCAGTGGCAA TCAAGAAGAA 22920

GTGCGAGTAA GCGAAAAGGA ATGGTAGAGA TGGAGCAGGA 22960

(2) INFORMATION FOR SEQ ID NO: 346:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2966 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 346:

ATAAAGGAAG AtGCCTTCAC CGATTCCACC AGCTAAAACA GGAGTGACGA TATwgAAAAG 60

GCTATGTTTC CAACCGrGAC CtAACAgTGT TCCaACAAGG GTCCCAACGC CCATTGCTAA 120

AATCATTCCT AATGCCATCG GGACAATCAT GCGCATTAAA CCTTGGACTA AAATTTTGCG 180

GTTCATTCCa AGGATACTTC CACAGACCAA ACATGCAATA TAGAAATAGA GAAAGTTTGC 240

TTGTTTCATT AAAATATCTG TTGAGTCTAA GACATTTTGG TTGAGTAAAT TGAAAAAGAC 300

CATAATTGAT GGTACTAATA AAGATAAAAT TGCTGGACCG CCAAAATGTT TCAAAATAGG 360

AATATTTCCA CCAATTGTTC CAAGTAACCA ACCCATTGTT AAAATTACGG CAAAACCACC 420

```
TAACATATTG ACTGGTAGTT GTTCTAATAG AGCCGTAACA AGAATAATGC TAGCTGTTAC     480
AAGATAAACA GGTAACGTAA CCGAACCAAT CCTCGTTTTG GTTAATTTGT TGCGCCAATC     540
TGTTTCATTA GCAGCTGTTG CGGGcAATTT CTTTTCCATG CTATTTCCTC CTATAACTTT     600
AATCTATATT CATCATAATG TAAACGTTTT CTGTAGAGTA GATTATTTAA TTAATTTAAC     660
TAAAGAAAAG TTTTTTAACC AAATCTTCCT ATGCTATACT TTATTTGAAG AGAAAAGAGG     720
GATGGAATGA AGCGTGGCGG GAAATTATGG TGTTTCTTAT CAGTGATTTT TATTGCAACA     780
TTAATTATTA TGATTACTTT CTTTTACGGA GTAACGACGG TTCAAACAAT TAAAGAAGTT     840
CGTAAAAACC AAGAGCAAGC ACTTTTAGCC GTCGGGGAAC AGTTGGCTAT TGAACCGAAT     900
GTTATTGAAG CCTTAAAGAA TGATCACTAT TCAGATGAAT TGGAAGCGTA TACGGTTCGT     960
TTAGGTGAAA TCCATCAGTT GGATTTTATC GTGATTATGA ATATGCAGGG GATTCGTTTA    1020
ACACATCCTG ATCGCCAAAA AATAGGTAAG CATTTTGAAG GgGCGATGAA GTGCGAGCAT    1080
TAAAAGGGGA GGAACATTTG TCTGTTAGTC AAGGGAGTTT AGGAGAATCA TTACGAgGTT    1140
TTGTCCCAGT CTATGATCAA GGAAAACAAA TTGGGGTTGT gGCAATGGGC ATaAAATGAC    1200
CTCGCTTTCA CAATTGATTG AACGAACmAA AAATGATTAC ACAGTGAGTG TCCTTTTAAG    1260
TGTTGGTTTT GGCTTTATAT aGCCATCGTG GtTTCnTATT ATTTAAAAAA caACTTCACG    1320
ATTTAGAGCC AAGAGAAATT GCTCGTCTAT TAGAAGAGCG AaACGCCATG CTTGAGGAAA    1380
CGAAAGACGC AATCCTTGTG ATTGATACCG ATCAAAATAT TTTATTAGCA AATATTGAGG    1440
CGACTAAAAT GTATCATAAT ATAACAAATA GTGAAGAAAA TTTACTAGGA AAGAAACTTT    1500
CAGCATTGGT TTTATCTCCT GAAAAATTAG TCGTACATTC AAAAACAGAG CAATTCTATC    1560
GTCAAAATGG CCAGGATTAT TTCGTCTCAA TCGCCCCGAT TAATGTCCGT AAAAAAACAA    1620
TTGGGCATGT TATCTTTCTG AAAAATGCGA CAGAAACGTT TATTGTCGCA GAACAGCTTG    1680
TCAGCACaAC AACaTATGCT TCGGCATTAC AAAGTCAATC ACACGAATTT ATGAATAAGA    1740
TGCATGTCAT TTATGGCTTA GTCGATTTAG AAGATTATGA GGCGCTAAAA CATTATCTAG    1800
CCGATTTATT GAAACCAGAA AAGGAATTTG CACAGCGCTT AGCTATTTTG GTCAGAAATC    1860
CTATTCTAGC AGGTTTTTTA AGTGGGGAAA GAATCAAGTT TGCAGAAATT AAAACACAAT    1920
TAGCTATTGA AATTTATCCA GAAATTCCAC CAAATAAACG AGATGAAGAT ACGCAAAATT    1980
TGATTGCTAT CTACCGTTAT ATTCATCGTT TCCTGATGGA GCAAACGTTG CCAGAAGAGA    2040
TCATTGAAAC CATTGATTAT CAGCCAGGAA GTTTGACGAC GACTTATTCG TTTGCTTATC    2100
CAAAAGAACA ACTGGAACGC TTTGAGCAAG AGTTTTTCAC TTCTTATTTG GCAAGATTAT    2160
TAGAGAATGC GGAAGCGACA TTGACTTGGG AAAACCAACA AAACAATTGG CTTGTCCTTC    2220
GCATAAATGT TCATTACGAA GGAGCAGAGG AAAATGAACC TATTGATTAT TGAAGACGAT    2280
CCTATGGTTG CATATATACA TCAAAAATAT CTTGAAAAAT TAATTCACCA ACCGACTATT    2340
TTTACGGTTG CAACAATTGC TGAAGGTCTA CTACTAACGA AAGAAAAACA ACCAGCACTT    2400
```

GTTTTATTAG ATGTTCATTT AAAAGATGGC AATGGCTTGA CCTATCTGGC AACAATTAGA 2460

GATGAAAAAA TTGATACAGA AGTTATCCTG ATTACGGCTG CCAACGAGTT GGAAAATGTT 2520

AAACGAAGCT TACATTTAGG CGTTTTAGAC TATTTAGTTA AACCCTTTTC TTTTGAGCGA 2580

TTTCAACAAA GTATCGAAAA CTATCAGAAA AAAACTGCTC AATTCACGCT GGAAACGAAG 2640

GAGCTTTCGC AAACAAAAGT GGACCAATTA TTTCATTATT CACAGACAAA CGCACGTAAA 2700

AATGAACAAG CTCTTCAAAA CATGTCCTTA GAAAAAGGAC TAACACmAGC AmCATTACAA 2760

TTATTGTTAA AAAAAATAGA CGAATTTACC GACTATTTTA CGATTCAAGA ATTATCAGAA 2820

GCGAGTCAGC TTTCACATGT TTCTGTTCGC AAATATGTGT TGTTTCTAGG AAAAAAATAA 2880

CCTTTTAGAA AGTAAAAATA GCTATTTAAA AGTGGGGCGT CCCTATCAAT CGTTCnGAGA 2940

ATTTAGTATT TTTTAATTTT TAGGAA 2966

(2) INFORMATION FOR SEQ ID NO: 347:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1580 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 347:

AAATTATTTA AAATTCAGAC TCATTAGTAG ATCAGATTAA TATCTAACTT TAATGCTCAA 60

AAGTGCTGAT GATGAATAAG CCATCATGAC AGAATTTAGC GGCATCATAA AAATTATTCA 120

GATCAAACGT CAACTATGAT TATAGCAGAA AAACAATGGG TTCAAAATTG AGATAAGAAA 180

CCAGCTAGAC TGTCrACGTA AAATGATAGG CTAAAAAAAG TGATGTGTGC TTGmTACrAT 240

TTTTATACAG TATTTCrTGA TGTrATAAAy AAATTkTCTA TTAAaaTaTG TtnACTAaaA 300

AGACAAACTT TTTGGGGGCA CTACATATaC CTTGTTGGTC TTTTnTATAT TCCAATTGTT 360

ATTAGGTATG TAGCAGGACA TTTAAGGTAT ATTTTGATTT GTGTTCTGGG TCAGCGCACC 420

AAAAACAAAA ATGAAACGTG TGTCAGAGGA AGTATGTGTG GTTTTCCAAT AAAAACAGCT 480

GATGGTCCAA TTTTAAAGAA TAATATAGCA CAGGAGAATA TGAATGAAAA AAGGAAAGTT 540

ACCATTAATG ATATAGGGAG AAATwTAGTG TATTTArGGA ATTTGAGTAA CGGTAAaTAG 600

TTGATGAATA TATACTTTTT TAGTrGAAAA AAAACACTGA TAGGAGGTAT AATATTGAAA 660

AAGAGAATAT GTATCACTAC TGGAGTTGCA ATGTTATTAG GATTATCACC GATATTAACA 720

ACAGGGGCTA CAGTATACGC AAATGAAGAG AGTAACTACG TATAAGTAAA TTCTACAACA 780

AAAGAGCTAG AGCAACGGAT GGAGGCGCTA GCCAATCAAT ACAAAGGAGT TTCTCTATTT 840

GAAACCGAAT TCAATAAAGA AAATATGACT CCTGTTCAAT TAGAATTATA TAACAAAGTT 900

TTAGATCTAC AATATGATAT TGCTTTACAA AATGGAAGTG CAAAAGGGAT GTcTAGAGAG 960

CAATATAAAC AATATCTATC AAACGCGATA CAAGGTTTTA TAATACAACC TAGAAATACT 1020

GGACATGGTT TGATTTCTGT TGACGTCTTA GAAACAGTTT TAGATACTGC GCTTAATGTT 1080

GTATTAATTA TCGGAAGTTT TGGGTCAATG TCTGCAATGG TTAAAGAATT GGGTCAAGAA 1140

GGCGCAAAAA ACTAATTACA AGATAAAGCA GTTGGATTAA TTATTAGTGT GATTCTAAGT 1200

TCTGGAAAGC AAGCTGCAAA ATTATTAGAT TCTATTGACA AAATACCCAA TAATGGCTAC 1260

ATTGAATTAA CTTAATAGAG TTTACGAAAG GATTGAAAAA GATGTACCTT GTAATAGGGT 1320

CAATTCTAAT AATCTTTATA GTCTCAGTTT TTATCTCTAC TATAGGGGCT TTGAATGAAA 1380

AAaTAAAAAa TAAATTCTAC TTATTTGTTA AATTTTCGTT TAATATTTCT ATTGCTTGTA 1440

TTTTTTTAGG TGGTATATCT TTTCTTTATG CTTACTTTGT TTTTGATGCG TTTAAAAATT 1500

TTAAAATTAT CTTGATATCC ATATTTTTAT TTGTAATCTC TTACAAAGGT TATATGGATA 1560

AAAAATATTT CCCAnGAAGA 1580

(2) INFORMATION FOR SEQ ID NO: 348:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 632 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 348:

AAGATATGGT AAAAATTAAG GCTAAATCTT CGCATCTTTT TGCATATCAG CTGTAGAAAC 60

AAATTGTAGA TTTGTTTTTT TACCATCATT AAGCTTTATT TCTTTTTCyT CaATtATCAA 120

rGAATTACTA TCaCTTTCtA AATTAACATT TATATTATCT TTGATTTGTA TTTCATTCTT 180

ATTATTTTTt TCTTTTATGA ATACTTTTAT TTCACTACTC TTGGAATAAA ACTCCAAaGT 240

TTGTTCTATA TATTTTATTT CTTTTCCATT CATATTACTA GAAATTTCAT TTGCCATATT 300

ATGAATCTTT TCTTTTCTCG TCTCCAAATA TGTTTTAGGA AAGATAAAAA AAACTAATGA 360

ATGAACTAAT ATAATTATTA TTCCAAGAAC AGAAAATATT TGTATGAACA TTTTTGGAAA 420

TATCTTTAAT TTTTTCATTT TCTCTCCAAT TTATATCCTA CATTTCTTAT AGTAGTTATA 480

CAGTCTAATT GTAATTTTTT TCTAAGTTCT TTTAtATATA CATCTATTAC TCGATCATAA 540

GGAGkTTCyT CGCTATCTTT CCATACATAA TCaATAATTT GCATTCtTGT TAACACTTGT 600

CCAATAATAA CCAATAAACA TTTAATACTC CA 632

(2) INFORMATION FOR SEQ ID NO: 349:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1868 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 349:

AGTGCCTTAT GCCAGAGTTT CaTCCmyTGG ccAAAaCTkG GAACGGCAAA TTCAAGAGTT 60

AAAAAAATTA GGAGCGAAAA AAATATTTGT AGAGAAAAAA TCTGGCGCAA GTATTGaACA 120

```
ACGACTAATT TTTACAGAAG CTATCTATTT TGTGAGAGAA TCCGATATTT TTATGGTAGA     180
AGCCATTGAC CGATTAGGCA GAAATTACGA TGAAATTATT CAGACGGTTA ATTTATTGAA     240
AAATAAAAAT GTTCGACTCA TAATTACAAG CCTTCCTATT ATGGCTGAAG TGGTTGGTAA     300
TCCATTATTT GATCGTTTTA TTAAAGATTT GATAGTCCAG ATTTTGGCAA TGATTGCGGA     360
ACAAGAAAGA ACAGAATCTA AGCGACGACT AGAACAAGGA ATTAAAATTG CAAAAATGAA     420
CGGTGTTTAT CAAGGACGTC CTGAATTATA TAGTCCAACA GCAAAAGATC TCCAAAAAAG     480
AGCTGTCTAT AGAAATATTA TAGAAGAGTT ACAAAAAGGC ACTGCTATTT CAAAGATTGC     540
AAAAAAATAT GGAATTAATC GACAAACAGT TTATAGAATA AAAAAAGACT ATGAGTCAAA     600
TCAAGCAGAT TCTTAATGAA TCATAAAAGA AATAAGAGGG CTTATTTGGT GAAAAmCAGA     660
AAGGAGTAGT TATTCTGCAA ATTATTAAAG ATACAAAAAA CACGTAGAAA AAGAAAATAA     720
GCCAATGGTT ATTTTTGATA TGATGTCAGT AAAAGTGGAA AAATTATTAT TAGAGTTGTC     780
AGAAGAAGTA TTTCGAGTGT GTCCTAAAAC AAAGAAAGAA TTTAGAAAAA TGGTTTCAAG     840
ATATAGCTTT AGAAATGGAG AGAATAAGAT GGTTGTTCTA AAAAAATATG ATTACTGGGA     900
AATATAAATT ACTACGTGGA ATAAGGCAAA TATGGTTGCT TATGTTGATT GTGACCAATT     960
GGCTAGAAGA GAGTTAGGAA AATATAATCA TCTTATGTAT AGTAATTGTA AAAAATATAA    1020
ATTAATTAAG AGCGTTTGCG TAGAAATAAG CGCAGAAAAA AATAAAAAAA GGTGAAACCT    1080
TGATAAAATA GTAATAGGCA ACAACTAACT CCTCAATATT AATATAATAG TGTCAATTAA    1140
AATGCATAAA TTAATCCCTA AATTGTTAGT TATTACCTTC ATAATAAATA GAATTGTGAA    1200
ATTTATTGTC TAAGGTGAAT AGTTGTGAGA AGAGAGTACG ATTATCAGAA AATTGAAAAA    1260
CAAGTGTTTA CCAATGGCTG TGAACTAGAA TCCAAAATAC GTACTACGTG AACAATGTGT    1320
AGTATGTATT TTGGTGTTTT CTTAATAACT AATTTTGAAT AATCATTCAA TTCTTTTATC    1380
AATAATTGAA CGTAAGACTT TTTCTCTTAA ATCCAATGAA TTAATCTTGT TATTTTGTTC    1440
ATATTCATGT ACTAACAAAT CACAAATATA AAGTTGTGAA ATTTTTCCAG CTAAAGAACC    1500
ACCATTTAAA AATTCTTCGA TTGCTGTTTG TAAAACTAAA TCTGCAGATT TTCCTATCGG    1560
TGAATGAATA TAATTCGTTA TAGCAAGTAT TTTTGCTCCA TTATTTTTTG CAATTTTTAA    1620
TGAGTCATAA GTATCTTTTG TTTTTCCAGA TAAAGAAAAG ATAATAACTA AATCTCTATC    1680
TGTTAACAAA GAAGCAACTT GAGCTTGATA ATGAGGaTCT AATACAGCTT TAGCTTGTAT    1740
TCCTACCCTC AAAAACATAC TTTCCAAATC AAGACTAGTA TTcCACTAGA TCCTACACCA    1800
AAAATATATA nGcTGCtTGC tGatTAATAA GTTGGGATGG CTTTGGTnAA TTGCCTCCAT    1860
TAATTAAA                                                             1868
```

(2) INFORMATION FOR SEQ ID NO: 350:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3274 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 350:

TTCTAATCTT GAAAAAAACA GACATTTTTC CTGATGTCTG AATTATTAAC CAAAATACTT 60

TGAAAGTCAA ACTATTTTTC TTATTTTTTG AGAAAAAGTG CTGGAAACCG CATGAAATAA 120

TAGTTTGATA ATCAAAATAA ACTCGGAGGT ACAGTATGAA AAGAGTGGTC ATAACCGGCA 180

TGGGTGCCGT AACACCATTA GGCAATACCG TAAAAGAATT TTGGCATAAT CTAGTGGATG 240

GCAAGCTAGG AATTGGCAAA ATTACTAAAT TTGATtCTGA AGATACTGGG GTGGCTTTGG 300

CTGGTGAAGT GAAAGAGTTT GACCCAAGTG CTGTTTTAGA ACGCAAAGAG CAAAAAAGAA 360

TGGACTTATT TTCTCAATAT GGCTTGGTCG CTGCTTTAGA AGCGTGGGAA ATGAGTGGTT 420

TAACAGAAGC AACGATTGAC CCGACTCGTT TAGGTGTGAT TGTCGGTAGT GGGATTGGCG 480

GGATGACTAC CTTACAAGAC CAAGTACGGG TAATGGATAA AAAAGGCGCA AAACGCGTAA 540

CACCATTTTT CGTGCCCATG GTAATTGCTA ATATGGCAGC GGGAAATATT TCGATTCGTC 600

TAGGTGCAAA AGGACCTTCT CAAACTATCG TGACTGCTTG TGCTTCTGCG ACGAATGCGA 660

TTGGTGAAGC TTTTCGCACG ATTAAATATG GTTTAGCGGA CATGATGGTA ACTGGGGGGA 720

CAGAAGCCAC AGTATGTGAA ATTGGGATTG CAGGCTTTGC AGCATTGAAT GCTTTAAATA 780

CAACAGAAGA TGCGACTCGG GcTTCGATTC CTTTTGATAA AGAACGAAAA GGTTTTGTGA 840

TGGGTGAAGG CGCAGGAATG CTGGTTTTAG AAGAACTCGA GCATGCACAG AAACGTGGCG 900

CAACTATCTA CGGCGAAATT GTTGGATATG GCAGCAACTG TGATGCTAGC CACATGACCG 960

CTCCTTTAAA AGACGGTAGT GGCGCGGCAG CTGCGATGGA AATGGCGATT GCAGAAGCAG 1020

GGATTACACC GGAACAAATT GGGTATATCA ATGCTCATGG TACGTCAACC CCAGCGAATG 1080

ACGCTGCGGA AACAACGGCC ATCAAACGTG TCTTTGGTGA GCGAGCATCG CAAATTCCAA 1140

TTTCAAGTAC AAAGAGTATG ACTGGACATT TATTGGGCGC GGCTGGCGGT ATCGAAGCGA 1200

TCGCTTGTGT GCAAACATTG CAAGAAGGAA AAGCTCATCC GACAGTAGGC TACCAAGTTG 1260

CTGATCCAGA TTGTGATTTA GATTATGTCA CAGAAGGTGC ACGTGACATT ACAGCGGACT 1320

ATACAATTAG TAATTCGTTT GGCTTTGGTG GCCATAATGG TGTAATTTGC TTGAAGAAAT 1380

GGGAGGAAAA CTAGTATGAA AAAAGTAATG ACTGCAACAG AAATTATGGA AATGATTCCT 1440

AATCGCTATC CGATTTGTTA TATTGATTAT GTGGATGAAA TTATTCCAAA TGAAAAGATT 1500

ATTGCAACAA AAAATGTGAC AATTAACGAA GAATTTTTCC AAGGACATTT CCCTGGAAAT 1560

CCAACAATGC CAGGCGTTTT GATTATTGAA GCATTGGCAC AAGTAGGTTC GATTTTAATC 1620

TTAAAAATGG ATCAATTTGA AGGTGAAACA GCCTATATTG GCGGTATCAA CAAAGCCAAA 1680

TTCCGTCAAA AAGTGGTCCC TGGTGATGTC TTGAAATTAC ATTTTGAAAT CGTCAAATTA 1740

CGTGACTTTG TCGGCATCGG CAAAGCGACT GCTTACGTGG AAGATAAAAA GGTCTGCGAA 1800

TGTGAATTGA CGTTTATTGT GGGACGATAA GAATCAATCA ATTTTGGATA CGTTAGTTTC 1860

CCTAAGAAAA GGAATAAGTC ACAATGGCTT ATTCCTTTTT AAATTTATAT AAATAACACT 1920

AACGCAAAAA AACATGCTAT AATCATAAAA AGCTTATGTG AAGGTGGTAA GCTTATGAGT 1980

AAAACAATTT ACAGGAGGGT TTTTCTTTTG GATATTTCAG TAGAGTTTTC GGGACATAAA 2040

TTGGCGAATG TGTTAATGAA CGCATCAGGC ATTCATTGTA TGACAATTAA GGAAATGGAC 2100

GAGTTAGCGG CTTCTCAAGC AGGTGCTTTT GTCGCTAAGA CGGCAACACC TAATCCTCGT 2160

CAGGGCAATG AGGAGCCTCG GTATTTTGAT ACACCATTAG GAAGTATTAA CTCAATGGGC 2220

TTGCCTAATT TAGGGATTGA TTATTATTTA GATTATCAGA TTGCACGTCA AAAGGAATTT 2280

CCTGAGGAGT TACGCTTTTT ATCTGTATCT GGTATGAATT ATGAGGAAAA TATCGCTATT 2340

TTAAAAAAAG TGCAAGAAAG TGAATACACA GGAGTGACAG AATTTAATTT GTCTTGTCCC 2400

AATTTGCCTG GTAAACCGCA AATTGCTTAT GATTTTGAAT TAACAGAAAA GCTTTTAACA 2460

GAAGTTTTCC AATTTTTTAC TAAACCATTA GGTGTAAAAT TACCCCCATT TTTTGATATT 2520

GCACATTTTG ACGCAATGGC TGAAATTTTA AATAAATTCC CTTTAGTATA CGTGAATAGT 2580

ATTAATAGCA TCGGTAATGG TyTaTATATT GACAGTGACA AGGAmGAAGT GGTcATTAAA 2640

CCAAAAGGAG GCTTCGGTGG ACTGGGCGGC GAATATGTCA AACCAACAGC GTTAGCCAAT 2700

GTTCGTGCGT TTGCGCAACG TTTGAAACCA GAAtCAAAtT ATTGGAACGG GCGGTATTAC 2760

ATGTGGAAAa GATGTTTTTG AGCATCTTTT ATGTGGTGCG ACATTAGTAC AAGTCGGCAC 2820

ACaATTGCAT CAAGAAGGTC CACAAGTTTT TGAGCGTCTA GCCaAAGAAT TACAAGAAAT 2880

CATGGCAGCA AAAGGTTATG AAAGTATTGA AGAATTTCGT GGGAAATTGA AAGAGATGTA 2940

ACAAAAAAAC CGTCAAGAAA TTTTCTTGAC GGTTTTTTGT TATTTCACAG CCTTAACATT 3000

TTCCATAAAG GCTTCCAAAT ATTCAGGTTG TGTTAATTGT TGGTTACAGG TGTCGCTATC 3060

ATGGCAAATG TAATTGCCTT TTTTCGTGTA GGTTCCGTCG CCGCCTGATT TGGTGGTAGC 3120

TAGGAACATA GACACATTCG TGATTGTTTG ACAGATGGAG CAAACACCTT TAACAATGGT 3180

GGGTGAAAGG GTGCCGTGGA CGCCGATTAA TTTATTTTCT CGATAAGCTA AGATAAATAT 3240

TTTCTGAGTC CCAGCGTCGT TCCAACCGAC ATAG 3274

(2) INFORMATION FOR SEQ ID NO: 351:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4760 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 351:

TnGGAAACCC AGACTTATTA ACTTTTAnGA TTTTCCACTT nGCATTCGTA GAACCTTTAC 60

TCAACCAATC TAATCGAGTC TtTCAATAAG CAAATTAAAA GATACAGCCG TAGAAAAGAG 120

CAGTTTCAAA ATGAAGAATC ACTAGAACGC TTTCTAGTCA GCATTTTTGA TACATACAAT 180

CAAAAATTTC TAAACAGAAG CCATAAAGGT TTTCAACAGG TAACCGATAC ATTAGTTTCA 240

ATGTTTACTG AGTAACTAAT TATTTTGCAG GAGGACAATT TATTTACACA AAATTATTGA 300

CGCTCCCGTT AACTTTTGAA TCCCTAATTG ACTAGCTGCA TAAAATAGTA AATATTTTAT 360

CGGATTTTGC GCTAAATAAA TAATAATACT AGATTTTCTT AAAAAGTCAC TTCCAGGAAT 420

ACGTATATTT TCTAAATTAA AACTTAAAAT AACCACTAAC ACAGCAACTA TTGGTAACGT 480

AAAATAATTT TGTGCATCAT AAAAAACACC TAATTGTTGC GCAAATTTCA CTTCCTTTAG 540

TCCTACATAA GAAAAAATAA CAATACCTAA AACTAAGAaT AAATTATTTA AAGGCTTTTC 600

AAATACATAG TTTGCTAAAA AATAGCCTAT AATCACATAA CAAATCGTAG ArATCATCGA 660

AAAgATATAG AACCATAAAA AGGCACTTGT TTAATCCATT CAATGTCAAA GTACCACTTA 720

CCATAAGTAG AAGTAAATAC TCCTATCAGA AAGCTAACAA TTCCAATTGT TAATAATACT 780

TTCTCACCAA AATAACTTCG GATGTAATAA ATAATAAATA CACTCATTAC TAATGCGATT 840

AAATACCAAG AAACACCTAT TCTGGGGACA TAGCCTAATA ATATGCTCAG TATAAATGAT 900

ATTTTTCCTC TTAAATCCAT TTCCCCTTTA ACAAAATTcA TAATAGATAC GATAAAGAAG 960

GGAATGTAAA aGATCAACCA CGAAAGATAT ACTTTTAAGA TTCTTTTTAT ATAATTGAAT 1020

AGTATTTTTT TATCTTTTTC TTTTTCTCTA AAAGAAATTT TTTTAAACAA AAAATATCCA 1080

GATACTATGA AAAAAAACGG AACAGAAAAG CGGGCTAATA TTTTCAAATC ACTACCCAAA 1140

ATAGATTTAG AGGGAAAGGT ATGAATTCCA ACTACACTTA ATGCAGCAAT AAATTTGAAG 1200

GCATCCAATA CATCATTGTT TTTAGTCATT TTTCCCCTCT CTATTCTCCC ATTTTTCTTC 1260

AGATTTCATG ATAGATAATT CTTGCAATAA TGTTTTAATG CGATTTTCGT GTTTTGAAAT 1320

CAACACAGAA TTTTTAATTT CCATGATTAA CAGGACAATC ACCGCAACTG ATAATAAGAA 1380

GTTTGATGTT GTTTCAAACC CAAAGGcCAT TGCTAACCAT TCTGCCACAT GCGGGnAAAT 1440

aGCaAAAnTA ATTAaGACmA TTGAGATTAA TAmCCAAATC AAAGCATTTT TTAGGAGAAA 1500

AATATTTTTA TTAATCCCtC GAATAATGTA TAAGAAAAAG CCAATTGCAA ATAAAAATAA 1560

AGCTAGCCAT AACACGGTTG GTAGCATTAG TCTCCCTCCT TCATAAAGGC AGCAATCAGA 1620

ATTGCTGAGC CGACTTCCAG CATATACTTC ACAGAAGCCA AGGCACGAAT CGAAGAAACA 1680

CCACCAAGCC GTTCCATCAT ATTGACAGGT CTTTCAACAA TCACAAAGCG TTTTTTAATT 1740

AAGTGAACAA TCGATTCTGG CTCGGGATAA TTTGTCGGAT AACGCTTTGC AAAAAAAGCA 1800

ATCACTTTCC GATTTCCCGC TCGATAACCA GAAGTCACAT CATAAATTGT TTTACCAGAA 1860

GCCATTCGAA TGCAAAAAGA TAACAAACGA ATACCAAAAC GGCGCATTTT CGTACTTTGA 1920

AAAGACGCTT CGTTTCCAGG AATGAAACGA GAGCCAATCG AAAAATCACA TTTACCCTCT 1980

GCTAATGGTT CCAGTAAAAT AGGTAAACTA TTGATATCAT GTTGACCATC ACCATCAAAC 2040

TGAACCGCAA CATCATACTC ATTTTCTAAG GCATATTTAT AGCCGGTCTG TACCGCCCCG 2100

CCaATACCTA AATTCaACAC TAAATGAATG GCATTAATTT GGTTGGCTTC TAAAATTTGT 2160

TTGGTTCCAT CTGTCGAGCC ATCATTAATT ACCACATAAT CCAATTCATG TTGAAAATGC 2220

GTAACTTCCT GTTTAAATGT TTCAATGGAA GCAATCGTCC GCAAAATGTT TTCTTCCTCA 2280

TTGTAAGCAG GAATGATTAA AAGTACCTTC ATAAAGAGAA ACACCTTTCC ACTTTTTCTC 2340

TAACAAAACC CCTCACTTAC ACAAAACAAT TTATTGTTGA ACATTTTCTA ACATTTGTGC 2400

TAAGGCTTCT TGCCAAGTAG GAATTTTAAA GCCTAACGCT TCTGTTTTAC TTAAATCCAT 2460

GACAGAATAT TGTGGTCTTT GCGCCTTTTG TGGAAATTGA GTTGAATCCA CTGGCAATAC 2520

TTCCACTTCA GTATCTTTTA AAATTTCTTT AGCAAACTGA TnACcAGCTA CAGCTGTTTT 2580

CATTAGACAA ATGATAAACA CCAAATGGCG CTTTTTCTGC AATAACAAAT GCCATAAATT 2640

CGGCTAACGT TCTTGTCCAA GTTGGGCGGC CAAACTGATC GTCCACTACC TTTAGTTGGT 2700

CTCTTGTTTC CGCTAATTTT TGCATGGTAA AGACAAAGTT ATGACCGTAT TGACCAAATA 2760

CCCATGATGT TCGAATAATA TAATAGTCGT CCAGGATTTC TTGGACCGCT TGTTCACCTA 2820

ATAACTTCGT CCGACCATAC TCATTTAAAG GATTCGGTTG ATCATCAATC GCATAGACGC 2880

CTTCTTTTTT AGTTCCATCA AAAATATAAT CTGTACTTAC GTAAACAAGT GTCGCCCCCA 2940

CTGCTTTAGc AcTTCTGCAA CATGACGTGT GCCATCTACA TTCACTTTTT CATCTAATTC 3000

CTTGCCTTCG TCTTCTGCTT TATCAACAGC TGTATAAGCA GCACAATGAT AAATGACGCT 3060

TGGTTGAATT TCTTTCACTT TTGCCAACGT ACTATCCGCA TCAGTAATAT CTAATTCTGT 3120

GGAATCCGTT GAAACATACT CAATTCCTTG TTCGTCTAAT AGGTGACGTA ATTCTGTGCC 3180

CAATTGGCCA TTTCCACCAG TAATTAAAAT CATTTTTTTC GCTCCTTCTT GAAAAAAAGG 3240

CAGGGCAAAG TAGCTTTTGT ACTTTGCCAT CGCCTTTCGA CTGTTATCTC ACATGTCAGA 3300

AAAAATTACT GACCATTTTG TGCATATTTT GCTTCAACAG CTTCTTTATC TGCACGCCAC 3360

CAGTCTTCAT TTTCTGTATA CCACTTGATT GTTTCTGCCA AGCCTTCACG GAAATTCGTG 3420

AATTCTGGTT CCCAACCTAA TTCTTCGCGT AATCTGGTTG AGTCAATCGC ATAACGTAAG 3480

TCATGTCCTG CACGATCGTT GACATGCTCA TAGGCATCAA CAGGTTGTCC CATTAATTCT 3540

AAAATTAATT CCATCACTGT TTTATTGTCT TCTTCACCAT CGGCTCCAAT TAAATATGTT 3600

TCGCCAATTT GCCCTTTTGT TAAAATTGCC CAAACAGCCG AAGAATGGtC GTTTGTATGA 3660

ATCCAGTCAC GAACATTTTT ACCAGCACCA TAAAGTTTTG GTGTAATGCC ACTTAAGACA 3720

TTGGTAATTT GACGAGGAAT AAATTTTTCA ATATGTTGaT ATGGcCCATA GTTATTTGAA 3780

CAATTTGAAA TTGTTGCTTG TAAGTTGAAA GAACGAACCC ACGCTTTGAC TAACAAGTCA 3840

GATCCTGCTT TTGTTGAAGA ATAAGGACTT GAAGGATTGT AAGGCGTTTC GGCAGTGAAT 3900

TTTTCCCcTT CGCCTTCTCC ATGTCCTGGT AAATCTTCTC TTAAAGGTAA ATCGCCATAA 3960

ACTTCATCAG TCGATACATG GTGGTAACGA ACATTGTTTT TACGGCAAGC TTCGATCAAT 4020

GTGTACGTCC CAATCAGATT GGTTTGTACA AATGGGAATG GGTCGTTTAA TGAGTTGTCG 4080

TTATGCGATT CAGCAGCATA GTGAACAACC GCATCTGTTT CAGCGACTAA ACGATTAACT 4140

AATTCAGCAT CAGCGATATC TCCTACGACT AATTCCACGC GATCACTTGG TAGACCTTCA 4200

AGATTTTTTT CATTTCCTGC ATACGTTAAT TTATCTAATA CAGTTACATG AACTTCTGGA 4260

TGATTCTTTA CAACATAATG AACAAAATTT GAGCCGATAA AACCAGCTCC ACCTGTTACG 4320

ATGATTTTTT TCATAATTGA AACTCTCCTA CTTTTTTCTT AAATTTCGCC ATAGATAAAC 4380

GGATTTTCCG CTTCAAATTC TTTTAAGGTT GGATGTTTTT GATCCTTTTC TGACGTAATC 4440

GCTTTTTCTG GCGCAATTGG CCAATCAATA GctAACGCTG GATCGTTAAA AGCAATACCG 4500

CCGTCTGCTG CCGCATTGTA GTAATTGTCA CACTTGTACA TAAAGTTCAC GTTCGGTGTC 4560

AAAGTGACAA AACCATGGGC AAAACCTTTT GGTACTAAAA GCTGACGGTG ATtATGTTCT 4620

GATAAAATGT AGCCTTCCCA TTGACCATAG GTTGGACTCC CTTTACGAAT ATCAACAATC 4680

ACATCTAATA CCGCACCTGT TACAACGCGA ATTAGTTTTG TTTGCGCmTT CGCCTTTTTG 4740

AAAATGCAAT CCACGCAAGA 4760

(2) INFORMATION FOR SEQ ID NO: 352:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1028 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 352:

AACCGGGAAA GnTCTAAAGA TTCTTTCAAC TTACAGGAAT TTTTTTAGGC TTTTCTTGGC 60

AGTTTCnATG AAAATGCCAA GTAAACCTAG AAGGGATGGT AACTGATTAA TTGAAAAGGT 120

AGGGAAAGAG CCTGGATGTT GAATAAAGTC GACCTGAAAA AAGGGTTAAG TACAGAAGAA 180

GTTGCCAAGC AGAAAGAGTT AGGCCTTCAA AATAATTATG AGGAAAATGT AGCAAAATCT 240

ACAAAAGATA TTATTTTTGA TAACGTCaTG ACGTTATTCA ACTTTTTAAA CTTTGCAATC 300

GCTGTTTGTT TACTATTTGT CGGAGCGTAT TCCAATCTGG CTTTCTTAGC GATTATTATT 360

GTGAACATGT CAATCGGGAT TTTCCAAGAA ATTCATGCCC GTAATTTAGT TCAAAAATTA 420

TCTATTGTGG CAAAAGAAAA TGTTCATGTT GTTCGAAATG GCGTACAACA AGAAATTGAT 480

ACAAAAGAAC TAGTTATGGA AGATATTGTT ATTATCTCTG CAGGGGAACA AGTTCCTTCT 540

GATATGGAAG TTATTGACGG TAAAGTCGAA GCGAATGAAG CGTTGTTAAC AGGTGAATCA 600

GATTTAATTG AAAAGGAAAT CGGCGATACC TTGCTTTCTG GAAGTTTCAT TGTTAGCGGA 660

CAAGCCTATG CCCGTGTTAT CCACGTTGGT GCCGAAAACT ATGCTGTGAA AATCACCCAA 720

GAAGCTAAAG TGCATAAACC AATTCAATCA GAGTTAGTCA ATTCGATTCG TAAAGTCTCT 780

AAATTTACAA GTTGGGTGAT TATTCCTCTA GGGATTATTT aTTTGTTGAA GCTTTTTGGC 840

TAAGAGaTGC TGGCATTAAA ACATCTGTTG TTGCTTCTTC AGCTGCTTTG TTAGGGATGT 900

TACCAAAAGG ATTAGTACTA TTAATTAGTA TTGCGCTGAC AACAGGTGTC ATTAAGTTAG 960

CCAAAAAACG TATTCTAGTA CAAGATATGT ACTCAATTGA AACATTGGCC CATGTGGACA 1020

CTTTGnTT 1028

(2) INFORMATION FOR SEQ ID NO: 353:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2749 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 353:

GCGgTGGCTT AGGACTCGCT ATGCTACTAG AGAGACCTCA GCAAAAAAAA AACAGCCGAT 60

TTTATCAAAT GAGTCCTGAG GAACGCCTGG CTTCTCTTCT TAATGAAGGC CAGATTTCTG 120

CTGATACAAA AAAAGAATTT GAAAATACGG CTTTATCTTC GCAGATTGCC AATCATATGA 180

TTGAAAATCA AATCAGTGAA ACAGAAGTGC CGATGGGCGT TGGCTTACAT TTAACAGTGG 240

ACGAAACTGA TTATTTGGTA CCAATGGCGA CAGAAGAGCC CTCAGTTATT GCGGCTTTGA 300

GTAATGGTGC AAAAATAGCA CAAGGATTTA AAACAGTGAA TCAACAACGC TTAATGCGTG 360

GACAAATCGT TTTTTACGAT GTTGCAGATC CCGAGTCATT GATTGATAAA CTACAAGTAA 420

GAGAAGCGGA AGTTTTTCAA CAAGCAGAGT TAAGTTATCC ATCTATCGTT AAACGGGGCG 480

GCGGCTTAAG AGATTTGCAA TATCGTACTT TTGATGAATC ATTTGTATCT GTCGACTTTT 540

TAGTAGATGT TAAGGATGCA ATGGGGGCAA ATATCGTTAA CGCTATGTTG GAAGGTGTGG 600

CCGAGTTGTT CCGTGAATGG TTTGCGGAgc AAAAGATTTT ATTCAGTATT TTAAGTAATT 660

ATGCCACGGA GTCGGTTGTT ACGATGAAAA CGGCTATTCC AGTTTCACGT TTAAGTAAGG 720

GGAGCAaTGG CCGGGAAATT GCTGAAAAAA TTGTTTTAGC TTCACGCTAT GCTTCATTAG 780

ATCCTTATCG GGCAGTCACG CATAACAAAG GAATCATGAA TGGCATTGAA GCTGTAGTTT 840

TAGCTACAGG AAATGATACA CGCGCTGTTA GCGCTTCTTG TCATGCTTTT GCGGTGAAGG 900

AAGGTCGCTA CCAAGGCTTG ACTAGTTGGA CGCTGGATGG CGAACAACTA ATTGGTGAAA 960

TTTCAGTTCC GCTTGCTTTA GCCACGGTTG GCGGTGCCAC AAAAGTCTTA CCTAAATCTC 1020

AAGCAGCTGC TGATTTGTTA GCAGTGACGG ATGCAAAAGA ACTAAGTCGA GTAGTAGCGG 1080

CTGTTGGTTT GGCACAAAAt TTAgCgGCGT TACGGGCCTT AGTCTCTGAA GGAATTCAAA 1140

AAGGACACAT GGCTCTACAA GCACGTTCTT TAGCGATGAC GGTCGGAGCT ACTGGTAAAG 1200

AAGTTGAGGC AGTCGCTCAA CAATTAAAAC GTCAAAAAAC GATGAACCAA GACCGAGCCA 1260

TGGCTATTTT AAaTGATTTA AGAAaCAwTA AAAAAcAGTT CAGCAGAAAT TATTCTGCTG 1320

AaCTAtTTTT TTTCACATTA gGTAGCCGTT TCAGGcCACg AATTGGTTTT ACTTTTAAGA 1380

CATCTAAGAA GAAAGTGAAA ACAGGGATTC CTACAATAAG TCCCCATACA CCAAAAAGaC 1440

GCTCACTAAC TAATAAAATG ACAAACGTAT AAAAGATTGG TAACTCTGTT TTACTAGACA 1500

TGAATTTAGG ATTTAGCACG TAAGATTCAA ATAAATGGAC GATTGTAATT AAAGCTAAGA 1560

TATAAATCAC ATCGTTCAAG CCACCTTGGG AATAAGCGAT GAAACTTAGC GGAATGCAGG 1620

AAATAATCAC ACCAGCAACT GGGACTAAGC TCAAAATGAA GATCATAATG GCTAAACTAG 1680

GTAGCTGAGT AAAGCCGATT ATCGCCAAGG CCAAGGTAGT AATGACTGTA TTCACCACAG 1740

CAATGAAAAA CTGAGCTTCC aTTACGaCAC CaAATGTGTT GACGAACTTG TTGGCAAAAT 1800

aATaGATATC cTGmnaAAAC CmATCmAAAT CACtTTTTAA AAwTAACTTr GAAAAATCAG 1860

CCaTTTGCTT CTTCTCAATC ATAAAGAAGA AACTTAAAAT AAAGGACATG ACAAATGATA 1920

AGCCGACAGC GCCAATATCT TGAATGTAGC GTAAAATCAT TGAAGCACCA TTCTGGAGTT 1980

GAGAGAGTAA GTTAGATTTT TCCAAGTATT GATCAATAAA TTGCAATACT TGATTTGTAT 2040

CTGTATGTGG ACTTTGATAA AAATCAACTA CAGAATTAAT CATTTGAGTC GTTTGGTTAA 2100

CTAGAACGGG AACATATTTA GTAATCGCTA AATAGATTAG AAAAACAACC AATGCATAGG 2160

TTAAGAGAAC GATGACAATC GTCGGGATTT TCATATAGCG TTGAACAAAA TGGACCAAAC 2220

GAACAGCTAA ATAGGTAAAG ATAAATGTTA AAAGAATTGT TGTAATCATA CTTCTGGATA 2280

AATAAAGCAC AAAAATAwTC AAAGCAAGGA CGCAAAAACG TCGCAAACGT TCATTTTGGA 2340

TAAATTTTGG TTACAAACTC ATAGGGGAAA CTCCTTCTTT ACTGGATAnT CAATCCTGTT 2400

AAGAATAACA TAACCATACC GATGGCAGCT AGGCCGCAAA CAATAATATG GCGGCGTTTG 2460

TACATCACTA CAATGGCGAT GAACTCTCGA ATAAACGCGT TAGGTAAAAA ATAAAAAGCG 2520

GTTTTTGCAC CAATGCCATC AGCTTTTAAG CCAGCCATTC GAGCGAAAAG ACCCGCCCGG 2580

AAAAGGTGAT AATTGTTTGT AGTAAAAATG GCTTGAAATT AGCATTTCCA AAATCCTCAA 2640

TCATTGTTTC TTTGGAAAAA CGCATATTTT CTAGTGTGTT TTTTGAATGA GCTTCTACTA 2700

AAATATCATC ATCAGGAATA CCTTGTGTCA AAGCATACAT TTTCATCGC 2749

(2) INFORMATION FOR SEQ-ID NO: 354:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 793 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 354:

GGACGCGATT GATTTTTACG AAGGAATGTA TAAAAACnCG GCAAcGaATT yCAcTCAAaT 60

TGACTGGCAT AACTATCaGC GGGCAAATTG GCGCGAGGTA GAAACAAAAG GCGACACGTT 120

GACATTGCAT TTTGCCAACG CGGGCGGTGC TGGGGGCTCA TACACACAAT TCACGAAAGT 180

TGGTACGAAT ACTGTTGTTG TATCATTTGA TGGCAATGCT GCTTATCCCG ATAACCCAAG 240

TAGCGTTTTA CTCGTGCAAA ATAGTGATTA TAAAGTTTTA AGGACTTTGA ATCAGTAATA 300

GGGGAAATAT ATCGTTTACT TAAACATAGG CAAAGGCGGG TTGCTTACCT TTTTATAAGT 360

AACCTACCTT TGCCTATGTG CTATTTTTTG GGATAGAGGT GCTCAGAGTA GGGGATGTTT 420

TTATGGGAAA TTTTTATGCT AGATCTCTGT TTGACTAGGC TCAAATAGAT CTGAATCAAT 480

CGTTGAAGTG aTATCTGTTT ATATAAAGAT AGGCTAAATC GTCAATTTTT TACAAATCGT 540

CGTtCATGTT AAAATAAATG GaGAAATGTT TTAGTAwTCA TAGTAAGTAT TTtATCGTAG 600

AATAAAAGAT ATACTAACAA nnGGAGCGAT GAAAATGCTA AAAGAAGAAA GATTAAAGGC 660

GATAGTCTCT CTAGTAGATC AAAAAGGGGC AATCAAAGTT ACTGAAATTA TGGAACGATT 720

AAATGTTTCG GATATTGACT GTTCGTAGAG ATTnTGACAG AATTTGGAAG CCAGCGGGGC 780

GTTTTnAAAC GAG 793

(2) INFORMATION FOR SEQ ID NO: 355:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20860 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 355:

CAGTTGGCAG TGTTTCGCTA CATGACAAAG ATATTCCGTT ACAGACAAAA ATTTATTATG 60

AAGTGAAATC TTCCGAACGT CCAGCTAACT ATGGCGGAAT TACCAAAGAA TGGGGCATGA 120

ATGATGTCTT GGACACGACC CATGATCGTT TCACAGGGAA ATGGCACGCT ATTACGAATT 180

ATGACCTTAA AGTAGGGGAT AAAACGTTAA AAGCAGGAAC AGATATTTCT GCCTACATTC 240

TTTTAGAAAA CAAAGACAAT AAAGACTTAA CGTTTACGAT GAATCAAGCA CTATTGGCCG 300

CCTTAAATGA AGGAAGCAAT AAAGTAGGCA AACAAGCTTG GTCTGTGTAT CTGGAAGTCG 360

AACGGATTAA AACAGGTGAC GTAGAAAACA CGCAAACAGA AAACTACAAC AAAGAGCTGA 420

TGCGTTCTAA TACsGTGGTG ACrCATACGC CTGATGATCC AAAACCAACC AAAGCCGTTC 480

ATAACAArAA AGGGGAAGAy ATTAATCATG GAAAAGTGGC TCGTGGTGAT GTTCTTTCTT 540

ATGAAATGAC TTGGGACTTA AAAGGGTACG ATAAGGACTT TGCCTTTGAT ACAGTCGATC 600

TTGCGACAGG CGTTTCTTTC TTCGATGATT ACGATGAAAC GAAGGTGACA CCAATCAAAG 660

ACTTACTTCG TGTCAAAGAT TCTAAAGGGG AAGACATTAC GAACCAGTTC ACGATCTCTT 720

GGGATGATGC CAAAGGCACG GTGACGATTT CTGCCAAAGA CCCACAAGCC TTTATTTTGG 780

CGCATGGTGG GCAAGAATTA CGTGTAACTT TACCAACAAA AGTTAAAGCC AATGTTTCTG 840

GTGATGTGTA TAATTTAGCG GAACAAAATA CATTTGGTCA ACGAATTAAA ACCAATACCG 900

TTGTCAACCA TATTCCAAAA GTGAACCCTA AAAAAGACGT GGTTATTAAA GTCGGTGATA 960

AACAAAGTCA AAATGGTGCC ACAATCAAAT TAGGGGAGAA ATTCTTCTAT GAATTTACAA 1020

GTAGTGACAT TCCTGCAGAA TACGCTGGTA TTGTGGAAGA ATGGTCGATT AGCGATAAAC 1080

TAGACGTCAA ACATGACAAT TTAGTGGCCA nGGTCTGTGT TTGCCAATTC TAmkTTTGTy 1140

TTAGCAGACG GAACCAAAGT GAATAAAGGG GACGACATTT CGAAACTATT CACGATGACC 1200

TTTGAACAAG GGGTAGTGAA AATCACGGcC AGTCAAGCCT TTTTGGATGC GATGAATCTA 1260

```
AAAGAAAACA AAAACGTTGC GCACTCATGG AAAGCGTTCA TTGGTGTAGA ACGAATTGCG    1320
GCAGGAGACG TTTACAACAC GATTGAAGAA TCTTTCAATA ATGAAACGAT CAAGACTAAT    1380
ACGGTAGTGA CGCATACGCC AGAAAAACCA CAAACACCAC CAGAAAAAAC AGTGATCGTG    1440
CCACCAACAC CAAAAACACC ACAAGCACCA GTAGaGCCAT TAGTGGTAGA AAAGGCAAGT    1500
GTAGAGCCAG AATTGCCACA AACAgGCGAA AAmCAAAtGT CTTATTAACG GTAGCTGGAg    1560
TTTAGCTGCn ATGCTTGGCT TAGCAGcTTA GgCTTtAAaC GTAGAAAAGA AACAAAATAA    1620
TTGAAGACGA mArGAaGGAC GAaCTGATTG TCCTTCTTTT TTTAGTTATG GAGGGGAAaC    1680
ATTGGAAGCA GTAGTAGTnG AAgAGAGGCA AAGGGGATGA AAGAAATTGC CATCCAAGAA    1740
AAAGATTTGA CCTTACAGTG GAGAGGAAAC ACAGGTAAGT TAGTTAAAGT TCGATTrAAr    1800
AATACACGTG CAATGGAAAT GTGGTACAAC AAACAAATTA CCGAAGAAAA CATTCAAGAG    1860
ATCACCACGT TGAATATyAT TAAAAATGGA AAATCTTTGG CATTAGAAGT ATATCCAGAA    1920
AAAAGTATCT ATGTGAAACC AAATTTAGGC AGAATCAATG TGCCTGTCTT TTTTATCAAA    1980
ACACCTATTA ACAGAGGAGT ATTTGAAGAG ATTTTCGGCG AAACGTTAAA AGCATAAGTA    2040
AAGGAGAAAG AAAAGCCAAT GTATGTATCT TTTGTTCTAA GTTGTTTGCT TGGATTTTCT    2100
GCCTACAGTC TATTAAATAG GCTAAATTCG TTGGAATTTG TGGATGTTTG GTTAGACAAA    2160
GAAACACAAA AATCACACT AAAACGCTGT TTTTATGATA CGTCTTTCAA GAAACAAACA    2220
CTAAAAGAGT TAGAACGAGT ATATTTCCAA TTAAAAGAAA TAATCAACGT GCAAATAAAC    2280
AAGCGGTCTT TAAATACGAA TGACATACGT AATGTACGAG AACTAGAGGA AAAACAACAA    2340
GAAATAAAAC GATTCATGTT AGACGTTTTA GAAGATGCTT ATTGGAAAGA ATTAGCAAAT    2400
ATGCCAGAAG ACCAACGACA CTTAGACGAT TGGGATTTCT TTTGAAATTG GTTACTGAAA    2460
GGAAGCGATG GATTACAATT TGTTTGATTT TTAACCAGCC ATGCTAAGTT TTTCCGTTTC    2520
GGTTCCTTTG GTTTTTTTGT TAGCGATAAT AAAAAAATGA TGCCTGACAC AAAAATATAG    2580
GAGGGAAAAA ATGATTATTT TATCAGGACT TTTTATTTTA GGGATAGGAA TAGTTGGAGG    2640
TTATCAGCTA GCCACGTTAC CTAAATTAAT TGAAATGAAA CAGCACAAAG CAATACAGAA    2700
CCATTTCAAT GTTAAAGGAA ATGAGTACAC CTATTATCAA GAAGACAGTG AAAATTACAT    2760
TTTATCCTTG GAAGATACAG AATATCGAAT TAAATTTTCC AAGAACACCC CGTTAAAAGT    2820
GGTATTCACT GAAATTTTAG AACCTATGTA AGGGTGAGTT TATGTATATC TCTTTTGATT    2880
TAGAAGCAAC ACAAGCAGAA AAAATAGCTy TACCwGTCCA TCAACGATTT TCGCTyGTTC    2940
AAGGTCTTTT GyTACTTGAA CAAGCCAATC AAGAAGTTTT GGCCACAAAA CAAGGAGGCA    3000
CGATAGCCTA TTTTTTATAT ACAGGAAAAG AACAAGAACT ACCTATTTTA GAAGCGGAAC    3060
TTTTATTACC GGTCAAACAA AAGGATATGT TAGACGTATT rCTTGAACCA GTCACAAATG    3120
ATCCAACGCT TGATCCAATA GAAGTAAAAA ACTTCTTACA GTTAGTTAAT CAATCGTTGC    3180
CAAAGAAGCA ACGAGCTAAA AGTAACAAAA AArGTACAGA AAAGAAGGmA ACTAyTGATA    3240
```

CAAAACCAGT GAAGAAAAAg TTTTTTTgTC CGTTAGTCTT ATCACACTTr TTGGATTGAT 3300

TCTACTTATG AGTGGTTTTT TATTAGGGAA AACCAATTCT ACACCTTAyC AAGAAAAAGA 3360

ATTAGAGGTr AAAAAAGAAC TCAATACCTT AACAGAACAA GTCAATGAkC AAAATAAAGT 3420

AGAAACATTT ACTCGATTCT TTCTAACTAA TTACTATTCA GGAAAAACAG ATGCCAAAAC 3480

CAGACAAGCA GTTTTAAAAA AGTTTGTAAA AAAAGAAAyC TTATCTGATT TTTTATCGGA 3540

GAAAAGCCGA GCGAGAAGTA TGTTTCCCTG GGAAGTAAAG AAGAAAGACA AACAATGGAC 3600

GATAGyGTTC ATTGTTGTAT TAATCAATGA AAAAGmCGAG CAATCGACAA AGAAAATTAC 3660

GTTTACAGCT GAAGAAACAA AAGAACAAyT GGTCGTAATG GAACGACCGA CAGAAGAAGA 3720

TTTTGAACTT ACAAATTAAA ATTAAAATGG AGGAAATAAT GATGAAAAAA ATTCTTTTTG 3780

CTAGTTTATT TAGTGCCACA CTACTATTTG GGGGAAGTGA AATTTCTGCT TTTGCACAAG 3840

AAATTATCCC TGATGATACT ACGACACCGC CCATTGAAGT ACCAACAGAA CCAAGTACAC 3900

CAGAAAAGCC AACAGATCCA ACACCGCCAA TTGAGCCACC TGTAGACCCT GTAGAGCCAC 3960

CTATTACACC AACGGAGCCA ACAGAACCGA CAGAGCCGAC AACACCAACA GAACCTACAA 4020

CTCCTACAGA GCCAAGTGAA CCAGAACAAC CAACGGAGCC AAGTAAACCA GTAGAACCTG 4080

AAAAACCAGT TACACCAAGC AAACCAGCAG AACCCGAAAA AACTGTGACA CCAACTAAAC 4140

CAACAGAATC TGAAAAACCA GTACAACCAG CAGAACCAAG CAAGCCAATC GACGTTGTTG 4200

TAACGCCAAC AGGGGAATTA AATCACGCTG GAAATGGTAC ACAACAGCCA ACAGTCCCTA 4260

TTGAAACAAG TAATTTGGCA GAAATCACGC ACGTGCCTAG TGTAACAACA CCTATTACAA 4320

CTACAGACGG AGAAAACATT GTAGCTGTAG AAAAAGGTGT TCCACTTACA CAAACAGCAG 4380

AAGGGTTAAA ACCTATTCAA TCrAGTTACA AAGTATTGCC TAGCGGAAAT GTAGAAGTAA 4440

AAGGTAAGGA CGGTAAAATG AAGGTTTTAC CATACACAGG TGAAGAAATG AATATCTTTT 4500

TATCTGCCGT AGcgGtATCT TGTCTGTAGT ATCTGGGTTT GTCATCTTTA AAAAACGCAA 4560

AGCTAAAGTA TAAGCTAGCG TATAGATAGG AGCGTACCAT GAATAArAAA AGAATTrTCT 4620

TTGGGCTTCT CTCTTTTTTC TTGCCTATAT TTTTAGTCTT TGGAGGCTTA CTTTTTTTTC 4680

TTTTATTATT AACGAGTACG TCAGATACTT CAAAAAATGA TTGTATTCAG CCAAGTATAA 4740

ATAATCCAAC TGATGCGACA GATACACCTA AATCGATCGA GCAGTTTGTA AAAAGCCATA 4800

AAGATGCTTA CCTTTTATCA TGGAAAGCAG GTGGCTTTTT ACCGTCTGCT AGTATTTCTC 4860

AAACGATGGT AGAAAATGGG TTTAATTTTA CTAATCCATC GGGGACGTCA TTTTGGCAGG 4920

CrCACAATAT GGGCGGTGTT AAAACGTCAA AAAAAGAAGA TTTTCCTGTA ACTTTAGCAA 4980

CATTCGGCCA AGATTCTGTT GATATTTCTG GTACAAAGCC AGGGTCAAAC GTCGGTGATG 5040

GCACTGGTGG GGCATATACC TGGTTTAAAG ACTACAATGC TGGAATTGTT GGAAAAGCAG 5100

AATTTATGGC ACACCAGACA CTGTATACAG GTGCTATCAA TAATACTGAC GGATTAAGTA 5160

CTTTATCAGC TATTTATTCA GGAGGATGGG CTACAGACCC TACTTACCTC ATGAAGTTAC 5220

AGGCCACATA TAATAGCTTA GGCAAGCAGT TTCAATGGTT GGACCAAGAA GCAATACAGA 5280

AATATGGTAA CGCGCCGTTT AAAAAGAGCG AACTTGTCCT AAATATTCCC GGAAAATCCC 5340

CAATCACAAA CGAAAAATAT GGTAAAAATT CTGATTGCGT TGTTACTTCA GATACATCCG 5400

ATCAAGTTAC AGGACAAAAT ACTGCCCCAT CATTAGAAGT ACCAAGTGCG TATAAAGGAA 5460

AGTTAACACT TCCGCCTATC GATTCGAATG ATTACGCTGG AAATAACTAT CCTTTTGGAC 5520

AATGTACGTG GTACGCCTAT AATCGTATGG CGCAAATTGG CAAACCAATT GAATGGTTTT 5580

CAGGTGATGG CGGAAATGGC GCAGGTTGGG CTAATTCTGC TCGTGCGAAA GGGTACACAG 5640

TAGTAAAAGT AAAACCAAGT GTTGGTTGGG CAGCTAGTAT GCAAGGTGGC ATAGGAGGTT 5700

CTGCACCGCC CTATGGGCAT GTGGCAGTTG TAGAGTATGT CAATTCTGAC GGCAGTATTC 5760

TTGTGAGTGA AGCAAATGTT ATTAATCAAG GCTCTGGTAC GCGTTCATGG CGAGTATTAG 5820

ACAGAGCAAC GGTTGAACAA ATTGATTTTA TTCAAGGAAA GGGAGCGTAG CAATGAGGGA 5880

GCGTTTAAAG CAAGCACAAC CACTAGTGAT CGTTCTTTTA CTTGTTTTAC TGACCACTTT 5940

AGGAGGCGGT TATTTCAATG AGAAAAAAGA AAATGCACAA ATAAAAAATC AAAAAGAGAA 6000

AGAGCAGCAA GTCCTAAAGA AAACTTTAGA AGAAAAAAAC GGACAACTGA ATGATTTATC 6060

GAGAAAAATT GAACAATTAT TGGATGAAAA AGAAGATCCT GTTAATGAGC CATTAACAAA 6120

AGCAGTGAAT GCCTTATTTT CACTTkTGTT TGACTATGAT TCTAGCAATA AAGAACAAAC 6180

AGTCGCTAAA AGAAAAGAAA AAGCAACTGT TTATGCGACA GAGCAAGGGC TATCCGGGGT 6240

ATTCCCAAGG GATkCAGATA AAGTTGTTCC TAGTGTTGTG ACCGTTTCTC AATTAGATAA 6300

GGCACCTGAA ATTTATCGAA CCACACAAAA rGATACAGGA AAAGAAAAAA CAGCATTGGT 6360

CCGGCTAGAA TATTCAGTTC GAATCGCAGG AAGCCAArCG CAACGTGGAA ATTTTATTTA 6420

CAAGGTAACT TTCAATGAAG AAAGCAAGAA ATTTACAACC ATTGAGAATT TAGGTGAGGA 6480

AAATCAGTAA AAAAGTACGA ATTACAACrC TTATTAGTTT AATGCTTGTT TTAGGGGGGA 6540

TAAGCCTCTT TTTATTTATG GTAAAACAAA ATGAAAAGGC AATGCTCAAT CAGGAAAAGG 6600

AGCGAGCGGC CGTAGTTTTA GCAGACAATG AAACACGTGT CTTATTTTTT TACAAAGATA 6660

CGTGTGCTGA TTGTCAAAAG GTCTATCCGA yGGTTTACGA AGAGGACCAG GGCGAAAACA 6720

ACATTGTATT TATCAATTTA AATCAACCkA AAAACCGTCA CTATATAAAG ACCTATACGT 6780

TGGAAAAAGT ACCAACGTTT GTTACACCGA AAAGAAGATA CGTAGGAACA GATACGAAAA 6840

AAATTCAGCA ATTAATTGCT GAAAATAACG ACGAAAGAAG TGAAGATTAA TGCACGAAAA 6900

TCATGTGAAT GAAAAGGAGA CAGCCGTAGA AAATACGGAA CGTATTGCTA AAAATTATGC 6960

GTATGAACGC CCAGCGATTC AAACGGCGCT ATTTATCTTA TGGCGGGTGC ATAATAAGCA 7020

GTATCAAACm GGCGCTCGGA TATTTTACGA CGAATTAGAA AAGGCCACCA AGACGTCTAA 7080

GACCGCTTAT AAAGAAGCAT TAGCTTTTTT AGAAGGTGCA GGCATGGTGG TAAACGAGGT 7140

TGTCGTAGAA AGTAAAGTCC CCCAATCGTT AATTCAACGT TACGGGATTT TAAAAGATGA 7200

GTAACAAAAA GGCTGTATTC AATCCrCAAG CAGGGAAGCA AGCCyAGAAA GAAAAAGTGA 7260

GGAAAGAGAC ACGTGTTTCT TCATTGGTTC AAATAATCCG AGCAATCTTC AGGAAGAAAG 7320

AGAAAGAACC AGAAAATTCT CGAACCATTC GTTTTTCGCC ACGTTCGGCG AAAAGTATTT 7380

ACTCAATGGT GTTTTTCCTT TTATTTTTCT TGATAAGCTT ATTTTTATTA ATGAGTTTTG 7440

GACGATTAAA CACGCTTACA AGGCTTGCGA TGCAAAAGCA AGTAAATCGT GAGGAAATTA 7500

TTGAAAGCGT GAATAAAGGC GTAgCTGCAA GTGATCAATT GAAATATGAC GGAATGAAAC 7560

TTGTGGATCG ATTATTTACA GTGTCCTCAA AATTAGAAGG AAAAGAGTAC TGGGAAGCGC 7620

AAATAACCCC TTATTTAGCT GCTGGATTGC GTGCAGAAGA TTTAGGTCTT GATAAGACAA 7680

ATGATGACCG CGTTGCACGA AATGTACGCT TTATTCGATT AGAAACAGTT GATTATAAAG 7740

AATCACTTTA CAGCTTGTAC TATGATGTCC GTTTTACTGA AGGGAAAGAA TGGCGCCAAG 7800

TGCAAGTGAT TTTGCCTGTT AGTTATGCAG AAAATGAACT GAAGTTrCTT GATAGACCAA 7860

CGTTAATGAA TTTAGCGAAG ACAAGTGAGA GTAACAAAGT TGykTATGAC GAGCAACGTT 7920

TTATACCTAA AGGAAAAGAA GTAAATGAAG AAGAAACAAA AAAACTAACG GAATTTACCA 7980

ATCGTTTCTT TGAATTATAT GTAAAAAATG ATGAAAAGTT AGGACTAATT GCCAATGTAA 8040

AAGGACTAGA ACATGCGACG CTAGAAAAAG TAGATATTAC AAGTCTAAGA GAAACAGGAC 8100

AAGGGGTCTA TGATGTTCGA GGAACCTACC AATTCAGTTA TGAAGGGAAA AGTCCGCTAA 8160

CGTCTAACTT CTCATTACAA ATTGAAGCAA CAAAAGATAG TTATTTTATT AAAAAAATGA 8220

ATGGAGTGTA AGAATATGTT TGATTTATCT GGTTTAGCAA ATAATATTTC CTCACAGGTG 8280

ATTATAyTAT TGGTGATTTT AGGACTGGTA TCCTTAGTTG TAGCGGTTGC AACACAAGGG 8340

GCCGCACGAG GAATTGCGAC AGTCTGCTTA ATCTTAATTT TAATCGCGTT GGTTTTAGTG 8400

TTAAAAGATG CTGAAAAAAT TGGTACCTGG TTGAAAGACT TAATTTTCAA ACCAAATGCC 8460

GGCTTTATTT TCCCAATAAA AGGTTGGCGA TTTTGATGGA ATATAATTAC ACACGAGAAT 8520

TTAAACAACC TCACAAGATT TACAGTATCA AAGGGGTAGC ATTGCCTTTT GCGCCAAACG 8580

GCATTCGATT GGAACAAATT TTTGTAGGTA TTGGTATCCT TATTCTCTTA TTGATATTCG 8640

CGATTATTTC ATTTGTCGCT AAAATCAATT TTTTTACCAC AATTATTGCT AATTTTTGGT 8700

TGATTCTTAT AGTAGGAGTG GGCGTCCTTG TGTGGACGCT CTTTTCTTTA AAATGGGATA 8760

ATAAGAGCTT TCTTGATTAT CTAATTGGTC GAGGAAACTT CTTGTATCAA AAGAAAAAAA 8820

GATATGAACA TGGGCTACTT GTCCAATTCC ATAAAGAAAA AGTCCAATAC AAAGTTAGAA 8880

AATAACAGTA AAGTGTTTAC AAAGAGATTG TGAGGGATGA AGGTGGATAA AAGAATTGAA 8940

GCTGTGACAA AATTTTTAGA GTCTTTAGGA ACTGTAGAAG ATTACACCGA GGATGTAGCT 9000

GTTAAGTATA GAAACTTGAT TCTGAAATCG TATGAATTAT ACGAGAATAA GTATAACGAT 9060

ACTGTTGACG ACAGTTTATG TATAGAAGTT TGGTCAAACG GGACTTATGT kGTTACAAAC 9120

GAGGATTTAT CTTTTGATTG TGAGAGCGAA GAAGATTTAC AAAAGCTAAA AGAATTGTTT 9180

GTTAACACGT CATTTtATAT AACCATTAAT GAATTAAATA AAGTAGGTCA TAAGGCTACT 9240

TTATCCGTCA AAGCTAAAGC AAAAAATTTA AGGAAACTGG GGCAGTTAAT TAAAGAATAC 9300

AGAAGTTGTA ATTGCAAATA TTTAAAGGAT AAAGTAACCG AAATAATTGG TGATGATGGG 9360

CGTGTTTATT TAGATAGGAT AAGTGAACGG ATGGACTAAG AAAAACATGC TATAATGGTT 9420

AGAAACCATA AAGAAGCCTT ATTTACAAAA GTAGTArCAA AAAATACAAC TACATGAGTA 9480

GGAGGTAAAT ATCTTGATTA ATAACGTTAC ATTAGTTGGA CGATTAACCA AAGACCCAGA 9540

TTTAAGGTAT ACGCAAAGTG GAACAGCCGT AGGTCAATTT ACGTTGGCCA TTAATCGCAA 9600

CTTTACCAAT GCTAACAATG AAAGGGAAGC AGATTTTATC AACTGTGTTA TTTGGCGGAA 9660

AGCTGCAGAG TCATTAGCAA ATTATGCAAC AAAAGGGACT CTGATCGGTT TAACTGGTCG 9720

CATTCAAACA AGAAACTATG AGAATCAACA AGGCCAGCGT ATTTATGTAA CTGAGGTTGT 9780

CACAGAAAGC TTCCAACTAT TAGAATCAAG AGAAGTAAAC GAGCAACGAA AAGAACAGGC 9840

TACAGGTAAA GCTACGTTTG ATAAACAGTC AATGGATAAA CCTGATCCTC TGGATCCATT 9900

TTCGCCAGAA AATAGCATAG TGGATATTTC TGATAATGAC CTGCCGTTTT AATGATTTAT 9960

AAGTGAATAA ACAACAAArG ACGAGTGAAC AGATCACTCG TCTTTTTTTT GTACAAGTGT 10020

AAAAAAATAG TGATTGCAAA GAAAGGAGTT GATAAGGTGC CATTAAAAAG TAAAGAGCCG 10080

CGCTTAGACT TTCCAATTGA AGCCATTACG AATAATTTTC TTTTTACAAC CTCTAACGAC 10140

GTGTGGGTAG GTTATAAACT TGCACACCAA GTATTTCCTT TGAATAATTT AGACTTTTTC 10200

AAGGAATACA TAGAGGACGG AAAAGGCGTA TTCGAGCAAG ACACGTTCGA CTACCATTTT 10260

ATGAATATTC CAGAATACTT TGATCTCGAC GAACAGATTG AAGCGACAAT CGATAACCTG 10320

GTAAAGGGAA GCTTTGCGGA TTTAGGAGAA ACCTATTTCC GACAAGCGGG TGATATTCTA 10380

AAAGACGAAG TCCAAATGAA TAAATATAGT ACGTATTTGT TCATTCGTTT TACCGCACCT 10440

ATACAAGTTG CTAATCCTAT GGAGTATATA GAATTGTTCA AGGAAGTTGG ACGTAAGGCA 10500

ATCAATGCTT TAACAGGGGT TCATGTACCA GTTAGCCAAC TTTATCGCGC ATATCAAGGA 10560

TTAGAAAACA AGATTTACAA AGACTTATCA AATTTTAAAA ATGTGGAACG CATGGATCCA 10620

AGAACCATAG GTCGATTGTT TTATTATTTC TTTCACCGTG CAAATACCCG TTTACCTGAA 10680

CGAACGTTAC TTCCGGAAGA AATGACGGAG GGGATTATTG AAAATAACAC GGGATATATG 10740

ACTATTGAAC AATTGGATAA AACACATTAC CTTGCTTTTT TACCGTTGAT CAGTTTGCCA 10800

ACGAGTATGT TTGGAAGTGC GATTATACAA AACTTGCAGG ATAGTTTGAG TACCACGATT 10860

GAAACCCATG TCAAGGTTAC ATTTCGCCAC CCAGATAAAG ATAAACGCGA TGTACACAAA 10920

CGTCGTAAAA GAATCTATGA ACAGGACAAA GAGCAAACAC AAGTTGACGG TATTTTAGAC 10980

GAAGACGAAG TGATTTTATT TGGGGAAGAG CGATTGCACG AACTAAGTAA AAAGATACGT 11040

AGTAAAGAAC GACGATTGTG TCGTATGACG ACAACGTTTG TACTTTCAGC CGATAGTAAA 11100

CAAGCATTAG AAGAGAAAAT AAAAGAACTC GAATTTGTTT TGGACGGTAC AGACTACAAG 11160

```
TTGTATCGTC CGCTAGTCGA CCAATTAACA TTGTTTAATC AGTGTTTAAT TGGCGCTAAG   11220
AGTCAATTTA AAAGCTATGA GCATGTTGTT TCAACAGGTT ATGTAGCGGA CCTTGGAATG   11280
GATTTAGAAA AAGAAATTGG TAACCGCTAT GGCATGCCTC TAGGTCGTGT CATAACTGCT   11340
AAAAAATTCA AGTCCGTACA ACAGGCACTT TCTTTATCAA GCAAAATCGT ATGGTTTTTC   11400
CCAAACTTGA CTAAAAAAGC CATAGAAGGC GCACAATATA CAAATGGGAA TACGCTAATC   11460
ATTGGCCCrC CTGGGCAAGG AAAGTCAGTA CTTGTTAAGT ACATTTTCTT GTGGTTAACG   11520
TTTTTAGGAC AAAAAATACT TTATATTGAT CCTAAAAATG AAACCGTCCT TTTTTTCAAA   11580
CGAGCATTAG ATAAGTTCGG ACATATTCCA GAATTTAAAG CTTTGTATCA ACGCATTAAC   11640
TTTGTTTCGC TATCAAGTGA AGAAAGrTAT CGAGGAATGT TAGATCCATT GCTATTTCTG   11700
CCAAGAGAAG AAGCTATTCA AACCGCACGA AACGTGTTAG AAAATTTTGG TGAAGTTACG   11760
ACAGACAGTC ACACTGCAAG TGAyAAGAAG ACGCTrATTT TAGATTCTGT CAATGCGGTG   11820
ATGAAAGGTA AAGGrAAAAA ACACCTAACG AAAGTAATCG AGGTTATTCG CGAAAAAGAC   11880
CCTAAATTAG CmAATTTAAT CTCTGGCCAC AATGTTGGGT TAGGyAAAAT TCTTTTAGGG   11940
AACGATTACA GCGAACCTAT TCGCTTTGAA AATCAAATCA ATGTCTTAGG TACCCAGGGG   12000
CTTAAAATCC CTACACAAGC GGAAATAGAC AGCGGACGAT TAAACAATGA ACAAATCGCA   12060
GGCATGTCwA TTATGGAAGT TATTATGAAA ATGACAACGA TCTTCTCTAC AGATAAAACT   12120
GAAGATGCTG CGATTATTTT CGATGAAGCG AAAGGATTTG AAGACACGGC CCAAGGACGT   12180
TTTCTAATTG AAGGAAGTTT ACGyCAAGGA CGTGCGAACA TGACCGATAT ATACCTTGTC   12240
ACACAAGCGT TTATGGATTA TGATAAAGAA GATAAAAAAG AACTATTGTC TTATAAATTT   12300
GCGTTCCGGC CAAACCAAAA AGAAGCACAG AAAAAAGTTT TAGAATTTTT TGGTATGGAT   12360
ACGAACCCAG CAAATATTCA GTTAATTAAT GAACTAAAAT CAGGGACTTG TTTATTCCAA   12420
GACCATCGAG GGCGTAGCCA gCCTATTGCA ATTGATGTTT TATTCGATAG TTGGTTAATG   12480
GCCGTTTCGT CTACTAATAA AGAGGATGAA GCTACTCAAA TGGCGTTGGC AATGGAACAA   12540
GGATCATAAA AAAGGAGAAG TGAGCATGAA AAAAGTACTA CCTTTTATTG CCTTAGTCGG   12600
CTTGTTATTG TTGTCAGGTT GTGGAACAGA TATGAAAAAG ATATTGACTG CCGATGGTGG   12660
TAAATGGAAA GTGGAAGAAA CACGTGCAAC TTACACTTTT TTTGATGACG GTAAATTTTC   12720
AGCTAATGAC TCAGAGGATA GTGTTAGTGG GACATACACT TATGATGAAA AAAATAAAAA   12780
AATAACCTTT GACmTTACTA GCAGrAACtC TTTCATTaTG GAAAAAGTrG AwTmCAArGw   12840
TAmCAAGATT ACAGGGGAAA TTGGCGAAAA ACAAAGAACA CTTATAAAAC AAAAAACAGA   12900
ATAAAGAGGT GATACTTTGG ATAAAAAGTC ACTGaTAAAA GCAGCCTTAC TACTTTTAGT   12960
CGGCTGTTTT TTTGTGTTTA CACAAACTAC ATCGGTcAAT GCCGCACCTA AACCTGATGA   13020
AAATAAGGTT ATTGTTACGG AGCCAACTAC TAATGTAGAT ACAATATATG ATACTTACAA   13080
AGACAATGCT TTTGAGTTAA TGACTAAAGA AAAAGAAAAG GATTCTGCTT TTGGGGTTAA   13140
```

AGAAGCAATT ACTAACGCTT CTGCAGCAGT CAAAACATTT GTATGGGCTG GGGTAAAAGG 13200

ACTCGGAGAG TTTAATGCAG TTATGGTAAA AACACTATTT AGTATGGATA TTATAACTGC 13260

TATTAAGCAA CCGATTATGA ACCTAACGTC TAGCATAGCA ACTAATATGT TAGGAATTGC 13320

GGGAACGATT GGTATTGCCT TTGTCTTTGT TATTTTAGGT GTAAAATTTA TTGGACAACA 13380

ACGATACAAA CGGTTTTTTG GTATTTTTTT AATGACGATT CTTATTTTTA CGGGTCTTTC 13440

CGTCCTAAAA GATGCCAACA CGTCCAATTC TTTGTTTGAT ATGATGTTTA GTGTAGATAA 13500

AGAAGTTGAG ACCGCATTTG TAAATATTAA TCCGGTCTTA GGTGATGTTA GTGTACCAAT 13560

GACAGAGAAA GGGAAAGATA AGAACGGAAA TGAAGTAGAG CAAAAATTAT CTGCAGATCa 13620

ACGAGCAAAA AGTGCAGGAA ATTTAATTGC TTCACGTGTT TTTTACACCA ATGTGTATGA 13680

ACCTTACTTG CTCATGAACT ATGGAACAAG TGATGTTAAT AAAATCCGTA AGAAAACAGT 13740

rAAGTATAAA GAyAAAGAAT ATGATCGAAT CAACTTATTG TTAGACAACG ATATGAATAG 13800

CGAAGAAAAT AATAAGTTAA TGGAAGAAGT TGTAAATTAC GAAAGTAAAG ACTTAAAAAA 13860

CCGTTCAATT ATGTATTATA ACAACTGGAC CAATACTTTT TATGGCTTGT TTTATCTGGT 13920

AGTTAACTTC ATTCAAACAG TGGTGTATTT CCTGTTAAGT TTCTTACGAT TAATTATTGC 13980

kGTCATTCAA TTGTTCCTAT TGCCGTTGTT ACCGTTATTG TTGTTCGCAG GACTCTTTTT 14040

AACTGAAACs AACGTCTTTG CGAATTACTT TAAAACCTTT GGCATGACTA TTTTTATGAA 14100

GGGAATGGTT GGTTTTGCGA CTATTTTCTT TGCAAGTTTC TTATCACTAG GGTTTCAATT 14160

AAGCAATCAG ACAGAAAACG TGTGGCAGAA GATATTAACG ATACTTATTT ATCTTCTTAC 14220

ACCGCTTGGC TTGTATGTCT TCCGGAAATT CTTTGCTAAT CTTGTCACGG GACGTGTTTC 14280

ACTTTCTGAT GGTGTCGGGT TTATCGGAAA TCCTTTTGGC ACAGAAGCTA ATATGCGTAG 14340

AGCGGCAAAA GAACAAAArC AAGAGAATAA GGAGCsCAGG AAACAGGCAC AAGAAGAACG 14400

TAAAGCAGCT ATTAAAAAAC GTCAAGAAGA AGCAAGGAAA CATGGTAAAG CAGAAACTGG 14460

CTTAAAACAA CGGCCATTAA ATAAAGAACA AGAAAAACGT AGTGCrctAC GCCGTGArTT 14520

GAAGCCAAGr CCTCAACATA AAGCGCCAAA TGCTATCGAA AAAGCACAAG AAGGATTAAG 14580

ACAAGCGCAC GAAAAAGGAC GTATGCAAGA ACAGAAATCT GAAGAACAAC TCAAACGGCA 14640

GCAACATGAA TCAGCTCGyC AAAAAGAATA TGAAAAAGAT CGATTGAAAA AACAGGAATC 14700

ACTTAAAAAA ATTAATGATC AAAGCGACGA ATCAACCAAT TTAACAGCTT TAAGAAATCA 14760

AAATCGGCGT ACTGGACAGC GAAGTAAAAA ACGCGAAACA GGTCATCAAA CACTTATGAA 14820

ACGACAAGGA GAAAAAGTAC TTCAACCAAC AAATAAACCA AAAGATATGA AAAArCAAGT 14880

GGTTGCTCGT AGTCCAAGAC GAACAGGACA ACCAACGACT CGTCAAAATA TCCAACCAAA 14940

AGTAGCGGCC CAAATGGTCA AACGGAGTGG TGGAACGTTG GGAAGTAATC CTGCCTATCG 15000

ACGTCCAGAA GTACGwAAAA AAATGGATCA AGTGAAGCAA GTAACGCAAG TTGATGTACC 15060

TCAAACGCAT TCTACACCAC AACGTGTACA AyCCAAGGTA AATGAGCCAA TCACACGTAT 15120

GAATAGACAT AATACGCCAA TCATTAAGAA AGAAAAGAAA ACTAAACAAy CACCAGTTGT 15180

TCGCAGAGAA CGCAAGAAAC CTGGTACACA GAAGATAACA AACAAACCTA AAGGAGTGAG 15240

ACAACCTCGT GCAAGGAAAT AATTTGTTTA ACAATGGsTT GCTACAACCA AAGAAACCGC 15300

CAATAGGGCA AGAACAACAA AAAATGACTC GCTACCACCA ACTCCGACTA GTCTTTGGAT 15360

TAGTCGGAGT TCTTTTATAT CCTTTTGCGA TTTTAGGCAG TATTCCATTA tTAATTGGTC 15420

TAGCGGTTGA TAAAAAAGAC AAGGCTGCAA ATGTATTTGA TATGGATTAT GAGAGTTTTT 15480

TGAAACGAAA TAGTATTGTy TTTCTTAGTT TTTCAGCTGT GTTATTTGTA ATAAATGTCT 15540

TTGCATTTAT CCTCTGGATT CCCAGAGGGT ATCTGTCTGC GTATCTTCTT TTCCCCTTAA 15600

ATCTTCTTCA TACaGCACTG CGTTTTAATT GGGAGACAAT AGTCGCTTTA CTTATTGGGT 15660

CAAGTGGAAT GGGAGCCATT TTTTTGGCAT TTTCTTCATT TGTGGCAAAA CGAAAAGTGA 15720

TTTCAAAGGA AGACGAACGG AAGAAAATAA CCGAATCAAA AGCCTACAAA GACCGAGAAA 15780

AAAATAAATT TGAAGAGTCG CAACGCTTCA CAGATGAACA AGAAGAAGCG TATGAAGAAG 15840

CAGTAGAAAC GGTTGATATT GACAAGTATA AAGAATTATC CAATCAATTA TTGCTTGGTA 15900

CAAGTGAGTT CGGGTTGCCG TATATCATCA ACTTTTCAGA GTTCAATCAG CATGTTCTTG 15960

TTCCCGCAAC GACAGGTTCA GGAAAGACAA CATTACTACA ATTAATCGTT CAACACGCCG 16020

TAAAATTCAA CCTCCCTGTC ATTCTCATTG ATGGGAAAGG CGCACGCGAT ACGCTAGAAA 16080

GTATGCGAGA GATTGCACGG TATTATGATA AAGAAGTACA TGCCTTCACA GACGATGGCG 16140

ACATGCGATA TAACCCAGTA GAACACGGAA ATGACGTGTC GATTCGAGAT AAGTTGGTGT 16200

CTTTAGCAGA AACGGAAAGT GTTTTCTATT CCAGTGCCGC GAAACTATTG TTAGAAAATA 16260

CCGTCCAACT GATTGATTTA TTTTCTACAC GTGACGACGT GAATCGAAAG TTAGAAGACA 16320

TTCAAAAATA CTTATTACCT CGGAATGTTT TGCGTTTGTT TGCGGATAAA ATCGAAGGAA 16380

AAAATCCAAC GCTCTATGAG ATAGAAGTAG AAGTCAAGCA AGCAAAACCA AAAAAGAAAT 16440

CAAAGAAAAA TGAAACGAAA GTCCCGATTA TTGATGAAGA AATGGACGAA GTAAGTGaAG 16500

CAACCACACC AACAAACGAC GAGGTAACGA GTGAAGATTT AGAACTAGAA GAAGGTGACA 16560

TTCCTCCTGT AGAGACAGAA ATCATTGTCT TAAACCCAAA CACTTTACAG CTAGACGATT 16620

TTTATTATTT GTTAAAACGA AATCTTTTTT ATTTGTCAAA AAAAGaAAAA ATAATGTTTG 16680

AACGTTTGTT TACACGTTAC GAACACAAAG ACTCACCCTT TTATTTATAT GCCACTTCTG 16740

AATCATTACA AACCAATATC AACATGTTAT TAGACAGTGA GTTAGGCCAT TTATTTGATA 16800

CAACAAATGC AAAAAGTGTT CTTGATGTCC AACAAATTGT CCGTGACCGT TCACTTGTTT 16860

ATGTGTCATT CAATGGATTA ATTTACAAAG AATATATCCG TACATTGGCG CAAATGTTAG 16920

TCGGTGATGT GAATTATTTT GCGTCCGAAA TGTATCGAAA AAATGTCAAA CGTGAAGTCA 16980

TTGTGATCTT TGACGAACCT GCTTCTTATT TAAATGAACA GTTTATTGAT ATGGTAAACA 17040

AAGGACGAGG AGCGGGTGTT CATGGAATTT TCACACCGCA AACGATGGCA GATATTGCGA 17100

AGTTAGGTGA CAAATTAATG GAGCAATTAG TCGGAAATGT AAACACCGTA TTCATTGGAA 17160

AGACTAACGA AAAGGGCGAn CGGAgTATTG GAGTGAAaCA ATGGGAACGT ATCAAGATAT 17220

AGATGTAACG TCCGTCACTG AACAAGAAGA TGGCTATTCA GACGTCGGAA AATCAGATTG 17280

GTCAGGTGAT CGAGGCACAA AACGCAACGT CGATCGCTTC AAAGTGAACC CAAATGTAAT 17340

AAAAAGTTTG CGTACAGGAG AGTTTATTAT TTACAGAACC GCAGAAAATG TGAATGTCCC 17400

GCCACAGAAA GTTTATGTGC GAAATGCACT AGAATGGTTG AGAACTCACA ATGGAATTTG 17460

AAAAATATGA ACACTTAAAT TATTGAAAAA AACGGCCTAA AGGTATGCCT AATAGTATGC 17520

CTAAAGGTAT GCTTAAAAGT ATACCTAAAG GTGTGCCTAA TTAAGAAAGG TGGAACCCGT 17580

CCTCGGCGCT TGGTATCAAA GGCTTTCTCC GCCTGCAAAA TTCCTGACCA TTTTAGCGCA 17640

ACGTGGGGTG AATACCCCTC TTATAsGCTA AAATGGTCAG GAATTTTGCA ACTACGAAAA 17700

AAATACCAAG CGCCGAGGAC AAACAGAATA GAGAAAAAAA TAATGAGTAA AGTAATAAAA 17760

AATAAAGAAA AAAGTTTAAA AGATGAAGTT CAAAAGATCA ACCCCTACCC CAGTCTTGGG 17820

GTCTCATAAC TATCTCTTTC CGTTCTCTTT CAACCTCTTT TGAAAGAGAA AACTTACAAC 17880

TCCCCCTTAC GCCTGCGCCC CCTCACCAAA TCTTGATGGG GCTGTCGCCA AGACCAGAAA 17940

CCCAAAAGAG TGGTTCTGTT TGATAACTAG TGGACTATTG ATGATTGATT TAAAGTAACA 18000

ACCCCCTAAA TCCCCCAACA AAACTTGGGG GACTTAAGAT CCAAAAAAGA ACTTCTCCAC 18060

AGCATTACGA AATTCTTTTT TATTGAtAya AATAgGCGTT GCCTATTTGT TCCTTAAAAA 18120

ATTCCAGTAA ATAACTACAG AACTCAAATA GGAGGCTTCC ATGCATAAAC TAAAAGAAAT 18180

TGAAATACTA GAAGAAGAAA TACAATCACG AATACTCGAC CATCGTTATG CGGACAGTTG 18240

GGAACAAGCG ACTCAAATTT CAGCAGATAT TAAAAAAAAA GAAGCGGAAC TCGCTAGGCT 18300

AAAAAAGAAG GTTAGCAAAT GAATCCATTA CTTTATAACT ACTTTATTGA TGAAACGTCT 18360

TCTAAAAAAA ATATGATAAA AGTATTGAGT TTATTACATG AATTAAGAAT CATTTCTCGT 18420

TCACAACTGT TATTGTTACT TAACATAAAT GCGAATGTTG CTGAAAGAAC GATGAATCAA 18480

ATATTAAAAT ATCTTAAAGA CAACAACTTG ATTGACAAAA TAAAAAAAGG TAATCAAGCT 18540

TGTTATTTCC TGACAAAGGA AGGTCACCAG AGTATTGGTG GCTATTATAC TTTGCCAAAG 18600

GTACCAGAAT ATAATTTGCA ACACCATTTA CAAATAAATG ATTATTTAAT CAAAATGCTG 18660

GAACTGACAA AGGATAGAAA AAATTTGAAA TTTGTTCTGT CCGAACGAaG GCAAGTCTTT 18720

GAAACCAAaG ACCTTGCCAA CAATCGAAAT CGAAAAAAaT ATTTTGTTGC TGATTTTATT 18780

TtCCGTTTTA GAAGTAAAGA AAATAAAGAA GTCAATTGGT CATTTGAAAT TGAGTTGACG 18840

ATGAAAACTA GGCGGCGATA TCGGGAGGGC ATTTTCCCAA AGTACATTGC AGAATTAAAA 18900

CGTCTGCCGT ATGCTCGACT AATTTATGTG ACACCTAGTC CGTTAATCGA AGAAGAACTA 18960

GAACGATTTA AAGGGTATTT TATACGAAAA GAAGGTGAAG ATAATGCAGA AGTATTTGAT 19020

CGCCTGCATA TTTTTTCAGC CGAAGAATTT GAACAGGAAA TTAAACGATT ATTAGCAGAA 19080

GATCAATTTA TTAATTGGGA GTGAGTTTTA TGAGAAGAAT AGAGAAAGAA TTCAACAAAA 19140

AGCTAGCCGG TTACGAAAGA GAGTTAAAAA AATTAGGCTG TCTGGATGAT GAAACAGGAC 19200

TTATTCCCAT AAGCAAAAGA CGATGGCATG TCATTTGGTG GCGACCTGAT ACACCTGCAA 19260

AAACAATCGT GAGATCCTAT CGATTAACGC TAGATAATGA AAACTTATGT ATTTTAGGAG 19320

ATGTAGAAAT CACTATTTAT CATGATGGAA CATATGGTAT TTCTAAAGAA GGTGTCCCTA 19380

TTTTTATTAA TGACTTATTG TCTTTAAAAA AATTATTTAC CATTTTCTAT GGAACACCTT 19440

TTAACTTAAA TTTTGAAAAA ATTCGCTGTG TTAGTTTTAA CAGGTATTGT ATTACTATTC 19500

CTGAAATTTA TGTTGAAAAA TTCGAAGTAT TAATAAACTA CTCGATGATT TTAAATAGCT 19560

GCCTACATGA AATACAAAAG CACGTAGAAT ATGATTAAAA AGCTTTTCTT CTTCCAGAAG 19620

AACACCAATG AATTGGTGAA GGTCAAGGGC ATTCCGGTCA AAAATTCCTT TTCTGATGTC 19680

AGAACAGTGT CACGCAGCCT TTTTCGTTTT CAGGCCGTTC CTTTTGCCGT GTGTAACGTG 19740

CTGTTCACTG TTTGTTACAG TACCATAGCC AGTCGAAAAA ACAAGGGTAT ATAAACGAGC 19800

AAGCTCGCCC TTGTTTTTTC TTTATGTCTA CGGTCCTTTC TCGGCACAGC GAACGACACG 19860

AACACACTGG CGAAAATGAA CAACCTAATC GAAAACGAAA AAGGGGGGAG ACTACACCGT 19920

TTGACGAAAA GAAATTTTTG ACCGTCCAGC TTCACTTCCT CAAACGCCTA ACAAAAATGG 19980

TTGTTCAAAA ATCAAAACCG AAAGGAGCTG AAAACATGAA TACCCCACAT TTAACATTCA 20040

AACTGGAACA TGCCAGAAAA GAACACCAGA AATTAAGTGA AGCCATTATC ACAAATGACA 20100

CAGTGACGTT ACTACTTAAT TATGGTTGTT TGAAAAATGC GAATGACCGT TTATATCAAC 20160

TAGAATATTT TCTAAATCAT AAAGAATGGA AGGACTGAAA GTATATGAGA AACAGAATtG 20220

AAGAAcTAAA AGAACAAGCA AGAACAGAAT TAAACGAATG GGGATTGATT ATTGACGGTT 20280

GCTTTGAGGG CGACTTTGAA ACATGGATAG GCTGTTATGC ACGACCAAAA GACAAACCGA 20340

CAGCCCTTGA CCCAATCAAC GAAGAAGAAG CCAAAGAACA AGcAAAATAT GCGGTCAATG 20400

GTTTCCCTCA AGATTTTACA GAATGGTACG AGTGGGAAAT TAACAATGGA AAACTTAAAA 20460

ATTTACTATA AAAnAGaAcC TTTTGCTAAC GACAAAAgAT TTCTAAAaAG GTACTTACAC 20520

TTGAATTAAT TGTrACTCAA GTATAACTTA CCTAACGACC TTTGACAAGT GAAAATTTGC 20580

TCCCATCTTA AATTTTAAGA GGGGAAGCTT ACAAGAAATG AGGAAACGTT ATGGACTTTC 20640

AAAATATACA AACTGTAAAA GAATTAAGAT ACGCACTGAA ACAATACAAA CAAACACTCG 20700

CAACTACGAT TTTTGATGAT ACACAAGGCT ATATCAGTGT CGCCTTTAAA AAAGAAGTGA 20760

TAGAAAAAGA CGGAAAAAAC GTTCAGTTTT ATGTTTTTAT CCTAATTATG AGGTCGCTCT 20820

TTATCCAGTC GAAACGTTAC TGGCTTACCT TGAAAAGTTA 20860

(2) INFORMATION FOR SEQ ID NO: 356:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3182 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 356:

```
TTTAACTcTT TTCCAGaATA AAGAACCTGA GTCTCTCCTA ATTCTATTCC CTCAAATTCT      60
TCATAAGGAT CGCTATATGC CCAACCGAAA AAATCCAAAA ACGGCGTTCC GTTCTTGACG     120
GGAGAGGTAT CATGCTTAAA ATGTTCAATA TTATGCGTTG CCCaATCGCT ATTACCtAAA     180
GCTAAAAGAG TTAGACAGTC TTGAACAGTA TCAAAATCTA TCTCCCCTGT TTCAAAAGCG     240
TATTTAATAT TCATATAAGG GTTCAAATAA TAAGCACGGT CCAGTAATTT TTCTCCTACG     300
TTTTCGACAT AAGCCTCTAC CATTTTAATT GCTACTTCCT TGGTTAACTC CTTACGAACA     360
GCCTTTTCTT CTTTGTAGGT TTGTTGTTCT TCTGATACAT CTAAATACTC ATCTACCTTC     420
AAAGGATTCG CTTGTTGACG AACAATAAGT GTTTCTGATG AAAATTCATT TAATATCGGT     480
AGCCCTGCTA TTTGTGAATA TGTTTTACTA GCTAAATCAG CTTTCATACC AATTTTATTT     540
ACGATATTTT GCCAAAGGCA ATCGTTTTCC TCTTTTGTAT AGCTTCTGTC TGTATCTATA     600
ACTAACCGAT ACCTTAGACC AAAATCTGAA ACATTATTTT TTATCGTAGG ATAGAGAATA     660
TAACTAATTC CTTTTAGGGA TTCAGCAATC TTTTTTATAA AATTTTTTAA ATCTCTATCC     720
TCTAGCTCGT CATAATCGAT AGCGATTAAT TCTTTTCTTA ATAGATTATC GTTTGAGCGT     780
TTACCTATCA GTTCACCAGC GAGAAAATAA CTTGCTTTTC GCTTAATCGG TTCTATTGAT     840
TCTTTGGTAG CGTGTTCTCC TAGAGTTACT TTCTCTGGTA GGTACTCCAC AAGATATTGA     900
AAAGGGGAAA CGCCCTCTAC TATCTTCATT TTTTTACTAA ATCCCTTACT TACATAGATA     960
CTCaATTTTw AATCACATCC TCTATACACA TTCATTAAAC ATTTCAGCAA GTGCATTTAC    1020
TTTTTTTGAT ACTAAATATT GCAAGGTTAC CATTTGATCA ATGCATCTCG ATGTTTCATA    1080
AGCCaATATT TTATCAGAAT TTTTCTGGCC ATTTTCATAG TCTATTnTCn CnCTCGGAAA    1140
AGATGCGAAA ACCTCCkGTA AAAAGAACAG TAAAtGTTTA ATATCTCTTA ATTCGTCGGT    1200
TATTGTATCT TTATCAAATT CTGGATACAT TACTCGCACT CTCCTTTATT AAATTTATAA    1260
AAATGTTCAT CaAAATTCAT GAATAACACG GTGAAAGCAC CGATAAACAG TGCTAGTTTT    1320
AAGATTGTAG GTAGAACAGA TGCACCGCCG CCTAAAAGAA ATGCGACTAA TAGTAAATTA    1380
AtTTTCGTTT TCATTTTGAT TTTCCTCTCA TTCTTTGGTA AAATGAAAGA CAAAAGACGG    1440
CTATATTTAG CCTTTTTGGT GCCTTTAACA CGTTCACTTT GGTCGGTTAG TGTGTTAAGG    1500
CTTTTTCTTT TGCTCTCATT CATCATAATT GCTCCTTAAT TTGTCGCTAC TGGCATTTTA    1560
AGCGTTGTTT GATATAAAAC ACCTTGACCA TTCAAAATGC TCTTAATCGT GTCATAATCC    1620
ATTTTAAGAG TAAGCAATGT AGCTACTTGT TGCTCTCGTT TGTTCACTGC TTCGATCTCT    1680
TCTGACGTTA AAAAATCTAA TATCGTTGAT TTTTTACTTA CCTCTCGAGC TTCTCTTAAT    1740
TGCTTAGCAT TAAAACCTAA TGCTGTCTTA TAAACAAGGT CTGTATAATG CTTGTAGTAA    1800
```

TGCGGGGAAA GCCCTAACTC CTTGATAACA TCCGTCATTG ACTTACGCAC ACTTTTTTCC 1860

TTTTGTCGTT CCATACGTCT AGCGTAGAGT CCGTTTTTCA TGTCATAGAA CTGTCGCACC 1920

AGCGCTTTCT TAAATCGTCG AACTGGTTCG GTATTATCTA GATAAGTGAT AAGCAAGGTC 1980

GCTTGTTCCT CGTTTAATTG ATAGAATTTA GCTTTTTGCC CACTCTCCAT TGCTCGGATT 2040

TCAAATCCGA CCTTGCCAAA CTCTTCTAGG TCTTTTTGAT ACTTTTTTAT TAGTTTAGAA 2100

ACACTTTCTC TTTTAATTTG TGAATACTTT GCAATAACCT CATCGGTCGT ATAAGGTTCT 2160

TCCTCGATAT ACTGTGAATG TAAAAATACT AATTCGTTCA TCTATCACAC CTCCATTTCT 2220

TTTAAGCTTG GGAACTGACT TTCTAACCAT GGAATCAACA CAGACAAATT ATGTCGTGCT 2280

TGTTCTTCAG TAAGATTAGC TATCCTATTT TGCATAATCA TCAAATCGGC ATAAAGCTTG 2340

GCGTTTCGTT TTTCATGTTC TCTACACTCT TTTAAAATCG TTTCTGCTTC TTCTACGATC 2400

TCACTAATGG AAAAATTCTT GTACTTTTTT GGTACCCATA TTAATTTGAG CATTGTTGCC 2460

TCCTCAAGTT AAATCTGTAG TAACGGACGA TTGACTATGT GCTATCTCGG aTTCTCTATT 2520

GGCGGGTGGA TwTTTATTAT TCATTTTCGA GAAGCCAAGA TCTTAgTTGA TGCTTACTCC 2580

AGCAGGTTTT ACGATCAGAT AAACGCCTTC CtGTkGGTAG TCCtTTCgCT ATCaATTCaT 2640

CtAAAtAAgC TTCgGAACAG TTGTCtAAgA AACGTCTTAC TCCaACGCGT GTATAgATGA 2700

GGTCGCTGTC TACTTCACTG TCTCGTCTAG CTCGCTCAAT CGCTTCTAGC GTGGTTTCAT 2760

ACACTTGTTT ACTGATTGCA TTCAATTGCT TTTTATCAAG CGTATGGCAC ACCTCAACAA 2820

CTACAGGCGG TGTTTGCAGT TGCTCCATTG TCATGCTGTC ATACCTCCTT ATAAATTTTT 2880

TCATTTACCA GCCAGTCAAC TAATTTTTTG TAGACTGTTT TACGCACAGA CAAGATTTTT 2940

TCATTTTCTA TCTGTCCTAG CGTTTTAGAT GAAATACCAA TTTCATGAGA TGCCTTTTCT 3000

AAAGTaATAT TTTTTCTCGC TCGGCGCTCA CGTAGCTTGG cTATCATCTC TAAATCTAGC 3060

TCAAACATAT TCtCACCTCT tCCGATTTCC CGtTTTTTGG GaTTkGATAC CyTGaTTawA 3120

ATTCCCaAAA AAwGGGaAGT CAAGTTATTT CTCATTTTTG GGAATTAAAT TTGATTTCTG 3180

CC 3182

(2) INFORMATION FOR SEQ ID NO: 357:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 508 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 357:

AnATAnnATA AATGTAATAA TTAACCACAT TATTCCgGTr AATTGAACGG CTAAATAAAT 60

ATTATAAmmm mCAAATCCTA AAGTGAAGAG AAAAAGTAAA AmATAACAGT TACTTCTTTT 120

AAAATGTGCT TTCCATGATT CAAATGCTCG TTTAACGGTG ATTGCTTGAT AATTCATGCC 180

CTCTTCCAAA ATGAAATCAG TCATCATTTG GAGTGCTGGA CCAACACCAA AAATAATCCC 240

TCCAGCTAAA CTAAATAAAA CAAACAGTAA ATTTAGTTTG ATTACTGTCC AGGAAAGGTA 300

AAACAACCGT TGAATCCCTG TACTTTCCAT CTTCCTCTTC CTTTCTATTA TGTAAAAGGC 360

TAAGCCAAAA GCGCTCAGCT AATGCCTCAG CCTCCTGTGA TTATTTTTGA GATTGGATAT 420

ATTCGTCTAA TTGTGTTTGC ATTTCTTTTT GAACTTTATC CCaGCCAGCT GtTTTTAGGT 480

CATcCATTAA TTTTGGGAAG TGTTTCTT 508

(2) INFORMATION FOR SEQ ID NO: 358:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 902 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 358:

CAGTCACGAA TTGATTTAAA TCCACATTTG TGGAATTAAC AAATTCCACG ACCATCAGCA 60

TAAGTCGTTG GGTAGTTGGC CAGACGACTT TCAGAcCAAT GATcCTTGAy CAaTCGCAGG 120

nCCwTAGaCA GAAATTGGtT TAATCGTAGA cCaACTTGAC GATCGGCATT AAAGGCATGG 180

TTGTTTTGAT TAATATCGAA GTTtCTGCCG CCGATmAmAC CAAGGATACG ACCAGTAGCG 240

TTATCCATTA AAATATTGCC AGTTTCAATC ATTGTATTCC CATTGACATC TGCAGTGCCA 300

TCATCAAGCA AATAACCATA ATTTGCGACA GCGGTTTGCa TCGTGTTGTA AATATTCTGa 360

TCAATGGTTG ATTGGATTGT ATAGCCTTGC GACTGGATTT CGCGACGTGC TTGTTCTTCA 420

TATTGATCTA AACCAACTTG ATCTAAATCA TCACGATTAA CTTTAGCTTT TTTCATATCT 480

AAATCCATGA CAATCTCAAC TGCTTTATCT AACACCGTAT AGTATAGATA GCCTTCTGTA 540

TTGACATTGG CCTGTTCAGT TGGTAAAAAG TCTTTTTTCA AGTCGTAAGC TTTGGCTTCT 600

TCATATTCCT TTTGCGAAAT AGCTTTTTCT CGGTACATAC TGAAAAGAAC AAAATCTTTT 660

CTCTTCATTC CTAAAGACAA ATCGTCTTTC AAGGCACCAG TGTTAGTATA GGGCGTGTAA 720

ACAATGGGAC TTTGGGGAAG TCCTGCGATA AAGGCCGCTT GAGGAAGATT CAAGTCCTTG 780

GCACTTTTAC CAAAGAGTCC TTTGGCTGCT TCTTCCACGC CyGCAATATT TTCACCTTTG 840

TTATTACGAC CAAAAGGGGA GACATTTAGA TAAGTTGTCA CGATTTCGTC CTTAGAAAAG 900

TA 902

(2) INFORMATION FOR SEQ ID NO: 359:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3989 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 359:

TTTTTGTTGT TCTTTAATTT TTATTCGTTG GTTGCTTGCC ATCTGTCCAC CAAAGCTTAA 60

TTAAGGCCTG TGTGCCATCA TTTTCAAACC CTTTATCTGC AAGgCTTTCA TATAATTCTG 120

TCGCTTTTTG CGTCGCTGGC AATGGAAGGT CAAGTTTTTT TGCTTCATCT AACGCAATCT 180

TTAAATCTTT GATAAAATGT TTAACAAAAA AGCCTGGCGT ATAATCTTCT TTTAAAATAC 240

GCGGACCATA ATTACTTAAT GACCAGTTTG CAGCACTACC ACCACCGACT GTTTCCAGAA 300

CTTTTTCTAA AGTTAATCCT GTCGCATTTG CATAAACGAG CATTTCAGTC AAACCGGTCA 360

TTGTGCCTGC AATCATTAAC TGgATTCGCC ATCTTAGTAT GTTGCCCTTT TCCAGCTGaG 420

CCGTGtAGCA TAAACGTTTT ACCAAATGkT TTGGAAAATk GGtAAGACGG nGkCaTAACT 480

TTCyTGgATC GGCCACCAAc CATAATTGGT TAAAGTACCA TTTTTAGCAC CTAAGTCACC 540

ACCAGaAACT GGCGCATCCA ACGCATGAGC ACCCACTTCT GCTGCTTTTT TGGCAATCTT 600

TTCAGCCAAG GTCGGCGTAC TTGTGGTTAA ATCTACTACT ATTTTACCTG TAAGATCTGC 660

TTGAAAAATC CCTGTTTCAT TAAAATACAC GCCTTCAACA TCTGAAGGAA AGCCAACCAT 720

TGTAAAAATG ATATCACTCG CTTCCGCAAT TGCTTTTGGC GTATCATACC AAACTGCTCC 780

TTCAGCAACT AAATCATCTG TTTTACTTTT TGTTCGATTG TAAACATTGA CAGATAAATT 840

GTTTTTCATC ATGTTACGAA TAATTGATTT TCCCATGACA CCCGTTCCAA TAAAGCCGAT 900

TTTAGACATA GTAGCTTCCT CCGTTCTTAA AATAAAAAAC TTTCCTCCTT CAGTATAAAC 960

GAAGGAGGAA AGTTTTGCAG TTTATTTTCT TAAAGATGTA AATTTTACTT TAATTTTTTT 1020

TATTGAGTCC GCATTTTCTT TTTCCATTTG AAGTTGATCA AAGTAAATCA ATGTATCTTG 1080

ATAAAATTCA ATCATTTTGT TGCCAAAGTT TGTAGAAGAA ATTTCAAATA ATTTTTCATC 1140

TAAAATAGTT TGGTCCATTT TAATATTTGC TTGTATATAG TCAATCAATG TTGTGGCAAA 1200

ATCACTGTCT GTCTTAAAAG TTTTGCCTAA GCTTTCATGA TCAAATAAAT TATTTAAATA 1260

GGCGTTCCCT TCTGCAACAC ATTGAACGCC AGCGGCCATC GCTTCTGTAT AAGTTAAGCC 1320

TTGTGTTTCC GACGTTGAGG CGCTGACAAA ATAGTCAGCA GCTTTATAAT AAATGGCGAC 1380

TTCTTCATTT GGCACTTCAC CCGTAAATTG AACGTATTCA CTGACTTCTA GTTCTTCTGC 1440

TAATTCTTTT AAATCCTCTA GATAGGGACC ATTTCCAACT ATTACTAAAC GAGTTTGTGG 1500

TAATTTTTCA ATAACTTGTG GTAACCCTTG AATAATTGCT TGAATATTTT TTTCATAAGA 1560

AATTCGACTT AACGATAACA ACATAATCTG TTGCTCTTCA ATACCTAGTT GTTGACGCAT 1620

TCCAGCGATC ATTTCCTCAG TAATATCGGG ACGTAAAAAC TTATCAATTT CGATCCCTGT 1680

TGGAATAATT CTCATTGGTG CGGTTACTCC ATAATCCCGT AATTTTTCAA TTACACGTTC 1740

ACTGGGACAA ACAACACCTG TCGTATGATT TGTAAACACT CTTGAAAAGA ATTTAACATG 1800

TGATGGACGG ACAACTTTTC CTTTAGCAAT ATAATGTAAG TAATCTTCAT ACATCGTATG 1860

ATACGTATGG ATGACAGGAA TTTTCATTTT TTTCCCAACC ATTTTCCCTA AAATACCCGC 1920

ACCAAATTCA GTATGTGTAT GAATTAAATC TAATTCCAAT TCTTTGGCAA TTAAATAGGC 1980

GTACCACATG CCGCGAACCA CTACTCGGCG ATCTTTAAAA GATACGAATG GCACACTAGG 2040

```
CATTCGTATG ACATCTTCTT CAAAATCAGT CGCGTTAGGA TCGGTTGTCG TAAAAATATA   2100
CACTTCATGG CCGTGCTTTT CTAATTCATC TTTTAAGGTC TTAATAGAAG TCGCTACCCC   2160
ACTCACTTGC GGGAAATACG TATCAGTAAA AAAACCAATT TTCACTCTCA TTACCTCCAT   2220
TTCATCTAAC AATAACTGCC TTCAATATTT GTTAAGATAA GACTGTTTCG TAGACTTCTT   2280
TTAACTCGTA ACCAATTTGC TTGATACTTC TTTGCTCTGC TACTTGATAG CCTGCTTCTC   2340
TAGTGCTAGG TATTTTCCCT TCTAATAAGC CTTCAATATA TTTTTTAAAT TCTTCAATAG   2400
AATGACCCAT ATAACAATTT TCATTAGCTA CTAGCCAGCC TTGATAAACT GGAATATCAC   2460
GCACTAAAAC TTGTTGCTGA CTAGCTAATG CTTCTAAGAC AACAATTCCT TCCGTTTCTT   2520
CACGCGAAGG GAAAAAGAAG AGATTCGCTG CCGCATAGGC CCCTTCGATG ACATCTCCTT   2580
TAATATAACC TGGGAAAATA ACATTTTCTG GATGATCTTC TTTAACTAAT TGTCGAATAT   2640
TTTTTGGGAT AGAATACATT GGTGTATCAC CAAACCAAAT AAATTGATAC TCTGGTAACT   2700
GTCGCGCAAC TTCAATAAAA TCAGTGATTC CTTTACGTTC AAAAAATAAT CCCACACAAA   2760
TAATTACTTT TTTCTCTTCA TCTATTTTAA AATATTCTCG GAATTTTTGT TCCTTTTCTT   2820
CTGAAGGATA AAAGCGTGAT AAATCAATAC CATTTGAGAT CGCCGAAATA GGCACTTTAA   2880
TGCCATATCC TTCTAGCAAT GTTTTAGAAT ACGGTGTTGG TGTAATTAAA TGATCCGCTT   2940
TGGAATACAA ACTGATAAGA TATTTTTTTA CTAATGGTGC CAGTTGATTA GACCCAATAA   3000
AGGAATTACG AAAATCTTCT TCTGTCGAAT GGGCATGATA AATAACTTTT TTTCCTAGTT   3060
TGCGTGCTTT TCTGACCATA CGATGGCTAT TTACGCCATA TGTATTAATA TGCAATATAT   3120
CATAGTCACT GCAATCAGCA TCTAATGTAT ATTCTATTCC AACTTCAGAG AGCGCTCGTT   3180
TCTGGTGATC TAATGCTCGA CCAATACCTG ATTTGGCTAA GATTTTCTCA CCTTCAAAAT   3240
ATAACAATAT TTTCAAACCC GCTCCTCCTT TTCTGCTTTA CTAACTATAA TAAACACATT   3300
ATACCATAGC TAAAATTTGA TGTATTCATT CTTCGGTAGG ATCCTTTTAA TCTTATAAAT   3360
GTTTTTAAAG TCCCTTCTTC AAGCTTATCA AAAAATAGTG AATAAAACAA AAAAAGCAAA   3420
AACCAAGTTG GTTTTTGCTT TTGCTCAACT CCGGCAGTAG GACTCGAACC TACGACATCA   3480
TGATTAACAG TCATGCGCTA CTACCAACTG AGCTATGCCG GAATAATCGC GTGGCGACGT   3540
CCTACTCTCA CAAAGGGAAA CCCTTCACTA CAATCGGCGC TAAGAAGCTT AACTTCTGTG   3600
TTCGGCATGG GAACAGGTGT ATCCTTCTCG CTATCGCCAC CACACTGGGT GTTGTTTCTT   3660
ATTGAGTTGA ATCTTCATTC ACTCAAAACT GGATTGAAGT TTGAATCAAA ATAACCAAGT   3720
TGCTTTTACT TATCCATTCT TTGGTTAAGT CCTCGACCGA TTAGTATTGG TCCGCTCCAA   3780
CTATCACTAG CCTTCCACTT CCAACCTATC TACCTGaTCA TCTCTCAGGG GTCTTACTTT   3840
CTTAAAaGAA ATrGGAAATC TCATCTTGAr GTGGGCTTCA CACTTAGATG CTTTCAgCGT   3900
TTAwCCCTTC CCTACAAAGC TACCCAGCAA TGCCCTTGGC AGAACAACTG GGTACACCAG   3960
CGGnAAGTCC ATCCCnGGTC CTCTCGTAC                                    3989
```

(2) INFORMATION FOR SEQ ID NO: 360:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3287 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 360:

```
CCTnCCTAAT ATATCTTCnA TCACtGTAAC TTTACTkGTA AtATCAGATA TTGaAAATTT        60

ATTTTCyTGt AAAGTTTTCA TTTTATTAGA TAGAGAAATC ATTTTATTTA ATATATTTTC       120

CATCTCATCA AAAATAGCAG TGTCAACTTT ACCTTTCATT AAGCTTCCTA ATTTCTCAAA       180

AATATTATCT TTTCCACCAA TAACTTCAAC TGCTTTATTG ATGTCATATA TATTATTTAC       240

AAGTTTTGCA CTATTTATAT CTTCTGAAGC ACTAGCTAGA CGTTTCAATC CATTTGACAC       300

TTCTATCAAA CCATCGATAA CTTGTACATA TGTCGATGCA GTAAAAGTAC CTATTATTCC       360

TTTGAATAAA GATGAAATTT TTCCAATACT GGATTTCGCC ATATAGCTTA AAcTTCTTCT       420

AAACTCTCAA GCTTATTTTT AACACTTGTA ATATCTTCAG GAACCTTTTT ATTCAATTGG       480

TTAATTGATT CAGAAACTAG CATTAATTCT CCTGCTAGTA GTGCTACAGC TGCCAAACCA       540

GCGATTGCTG CAGGGAATGT AGCAATACTT AAACCACCTA TTACTCCAAC TAAAAGACCC       600

ATTGCACCAA TTGCTATACC GATATTAGCA ACCTTTTTAG CTATTGTTGC AATATCTGAA       660

GGTACTTTTT CATTTAATTG TGATAGTGCT TCGCTTGCAA TCATTAATTC TAGTGATAGG       720

AGTGCTACAC TAGCTAATCC TGATATAGCC ATCATTGGAT TAGCAGATGA AAAAGCGCCT       780

GCAATAACAA CTAATCCTGC CATAGCCCCG ATTGCAATTG ATAAACTCCC TAGTTTCTTA       840

GCGACAATTC CAATATTGTT TGGTACTTTC TCATTGAGTT GGTTCATCGC TTCACTTGCA       900

ATCATTAAAT CTACTGAGaT TAATGCAATA CTTGCAATCC CTTTTAAATT ATCAGTAAAA       960

TTTAATTTAC TAGCTATAAA AGCCAGACCC CCAATTGCAC TGATAGCCAG ACTCATGCTC      1020

GCTGTTTTTT tAGCTAGCCC TGTAAAATCA TTyGGTATTT TTTCaTCTAG TTGTTTCATT      1080

GCTTCTGCGC CTTCTTCAAG CACTTTTATt GCTCCGAATA ATAACGCAAG GTTACCAGCA      1140

TTTTTTAAAA AACCAGTCCC AATAGATTTT AAACTCTCTA ATGGTTTTAC GATTTCTGAT      1200

GCTCCTCCAC CTTTACCGAA TTTAGGCAAT TTTATTTTAC CGAATACTCC AAAAACTTTG      1260

GATAATATAG AAAGCCCTTT GGCACTAATA GCTATAAACT TTAAAGCTAC CCCAAATTCA      1320

AATAATCGTG GAATAAATGC ACCTAGGCCT TTCAAAGTAT CTCCCTTACC TAAAAATTCA      1380

AATATAGGTT TCAATAGATT GAATAATCCA ACAGCATCAT CTTTTACACC CTTTATTCCT      1440

TTACCTAGTC CCTCAAAAAA TGAAGCATAA TCAATAGTGC TTAAAAATTT CACTACTTkG      1500

TCAAAAAAAC GAAATATATC ATCTGAATGT TCTTCTGCAA AAGCTCCTAT TTTTTTTAAA      1560

TTATCTTCGA AAGCAGCACC AATTTTTCCA AATACACCTG CAATTCCTCC CAGGCCTTTA      1620
```

TCCTTCGCCA TCTTATCAAA TGTATCGATA ATCGAGGTCA CTCCACGTGT TATAGCTGAA 1680

TTCATAATAG ACATACCAGT TCGAACGCCA CCAACAGAAT TTCTAGCTTG TTCTTCAAAT 1740

GATTTGAATC CTTCTGCCCC GTTTTTATTC ATCTCCACAA TTCTATCCAT AAATTGATCC 1800

ATAGAGATAT TTCCATTTCT TAAAGCTTCT CCCAATTGAT CAGATGTTAA ACCAAAACTC 1860

TTCCCGATTT GATCCACTTG GGCAGGCATT GCTGTTAGTA GCGAGCGCCA TTCTATCATG 1920

TCTGGTTTAC CTTTAGAATA AGATTGAGAT ATTTGCTCAA GTGCGCTTGA TTGTATTTCT 1980

CGTGACATGC CACCCGCTAA AATTGCATTA TTTAATGCCA AGAATATATC CGCAGATTTC 2040

TGTACATCTT TATTATTTGC TGTAAACCTT TGTACTGAAG CAACAGCATC ATCTAAAGTT 2100

GTAGGCAAGC CATTTAGCCC TTTTACAAGT TTGCTTTTGG CGATATCGGC TTGGTCGGCT 2160

GAAATATTCA TATTTGACAT TACTTTTTCA AAGTTATTTA AAGTATCTAA TCTTGAAATA 2220

GAACCATCGA TAGAATTTGT AATCATACTA GTtGCCTTAT CAACAGCCTT TACAGCTAGT 2280

CCTACACCAA GCATTTGATT TAGTTTGGAT GTTAAAGGAT TGAATGCTCC TAATATTTTA 2340

TTGCTTGCAT TGTTAGCTAG CCCGCTCAAC TTATCTAGAT TTGATGAAAC ATTTATCTTG 2400

TAGGTCCGTT CACTCATACT TCGAAGACTA TTTTGTAATT TTCGACTCTC TTCTCCAGTT 2460

TCTCCTAGAC CACGACTATC AATTTGTAAC ATAGCCTTGC GATTATTCAA TCGACTAATC 2520

TCATTGTCTA GTTTACGAAG TTGTGCTTCT GAACCTTGCA ACTGAGTTGT TACTATTTGT 2580

AAATTTGCTT TCCTAGATTG TAGTTGTTTA ATTTGTkGAT TTAAGGTkGA TATTTTTtCT 2640

TTkGCAtCTT TAGTATTTAC ATCTGCCTTA ATGGTTGCTT TCTGTGCTTG CAATTTTTTT 2700

ATATCAACAT TAATATTTTT TATTTCTTCT TTTGCTTTTT TTAGTTTCGT AGTATTCCAC 2760

TCGATAGCAG GGCGTTGACC TTTTAGTTTA GCTATTCGTC GATCTAAACT ATCTAGTTCT 2820

TTTTCTATAT CTTTAATATT GTCCTTAATA CGAATGGAGA TATTTTCAAC GGCCAATTAT 2880

TCCACCTCCT CCAGCTCATC TTGCGATATA AATTGAACAA CAAATGGTTG TGGTTGCTTA 2940

GGTTTTGGTG CATTTTCTTG TGACGACTTC CAATCCATAA AGTTTTGATA ACTTTGCTCA 3000

TTAGCATAAT GGCCATATGC TACAATCAAT TCAGAAACGC CCCAACCACT CAATACATGG 3060

CTGGGACGTT GTCCTAAGAT TTTTGCTACA AAATGAGCCA TATACGAATA ATGATtAAAC 3120

TTAGCAAAAT ACGTTyCTgC GCTgTTGTAA TGAAACACCA TtGTTtAAAC TGtCACGCAC 3180

GATTATCCGA AAAAAGCATC TGCtTCTTTG AACACtTCTG GGtGATTTcT GTtAATGCAG 3240

TAAAnGTTCA ATCACTGAAA TGGAAGCaGk ATCGATTACT CTCAnCA 3287

(2) INFORMATION FOR SEQ ID NO: 361:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1104 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 361:

CAAAATTAGG CCTTATCGGA CCATGTTTTT CTACATTCAT TGGTTGAAAT AGCATTAATT 60

TTAGTCGTGA kCGAATTGTC TTATCGTTTT ATCGAACATC CACTGACACG ATTGTCTTAT 120

AAGGATGTCT GGACTCAATT CACTGAGTTC TTACGGAAAC CGTGGGACCT TCGTGAAAAA 180

GGGACAATGG CCTTTATGAC AGTGATTAGC GTGATTGCAG TATTCGGTTT AATTGTGGCT 240

CCAGCAAACG CTAAAAGTGC GCAACAAGAG CAACTAGAAA AAAATATTGC GAAAAATCAG 300

CAAAAAAATTG AAGAACGTAA AAAAGAAGTC CAAGAAAGTG GCACGAAAAA TAGTAGCACG 360

TCCACAGAAA GTAGTACAAA ACCGAGTGAA AGCCAAGACG TTCTACCCAC ACAACTCACA 420

CCAGAGCAAG TAAAAAAAGC GCAGAATTTA GAAATCACAA CTATTGGTGA TTCTGTTATT 480

CTAGATGGTG CTAGCGGGTT ACAAGATATT TTCCCTAAAA TGATTATCGA TGGGGAAGTG 540

GGCCGTCAAT TGTATAGTAG CATTTCACTA ATTGGTGAAC TAGACAAAAA GAAAATGCTT 600

AAAAATACTG TTTTAGTGAG TTTAGGTACG AATGGTCCTT TTACAGAGGC ACAATTTGAT 660

GAATTTATGA AAGCGTTaGG TAATCGAAAA GTTTATTGGA TTAATGTTCG CGTCCCAACT 720

AGAAGATGGC AAAATCAAGT GAATAGTTTA CTTAGTCAAA TGGACAAAAA ATACGATAAC 780

TTAACGGTCA TTGACTGGTT TAATTATAGT AACGCCCATG ATGATTGGTT TTATGATGAC 840

CGAGTTCATC CAAATGTGGC AGGTGGCGAG cAATACACAC ACTTTATCGC GGAGAAAATT 900

TTACAGTAGC AAGAAACTTc CAGCTCAGAT GAAAGGGGCT GGgAAGTTTT TTGTTATAGG 960

GAAAAGCAAA TATAGGTAAT TTAATGATTT TCTTGGAAAA CCTAAGCTTT TTTTTTTAGA 1020

TATCnGGGTA TACnGGTTTT TTGGTTACCA GnTCCCCGnA AGGGATTTAG CCAGGTAATT 1080

GGCCGGTACC TTACCTAAAA GATG 1104

(2) INFORMATION FOR SEQ ID NO: 362:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5273 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 362:

GATTnCGTGA AGCGGTTGAC ATATGATTGC ATCTTCATTT ATAATAGATT CATGAAAGGT 60

TATTCATGTA AACGGAATAA TCGCTATTnT ATAGGGAGTG AAGGTGTTTA AATGCAAAAG 120

AGTTATAAAC TAGAAGGACT TGATTGTGCA AGTTGTGCAG ACAAAATTGA AAAATCTGTC 180

CAAAAAATTC ACGGAGTAGA AAAAGCGCGT GTGGACTTCA TGGCTGAAAA GATGACACTG 240

GAAGTTGAAT CAGGACATGA TTTAGAAGTC GAAAATGAAG CGCGTGCAGT GATTGGTAAG 300

TTAGAATCAG ATGTCAAAGT TATTTCTTTA AAAGATGTCA AAGAAGAAGA GGGACGTAAT 360

CCAAATAAGA ATCGCTTAAT TCGTATCATT ATTGCATTTG TTTTATTTTT TAGCACTTAT 420

TATCATTAAA CCAAGTnATA ATTGGGTTGC TTTGGCATCT TATTTAGTCG TTTATGTCTT 480

```
AATTGGTGGC GACATTGTCA AACGAGCGGT AACGAATATT TTTCGTGGGG AAGTGTTTGA    540
TGAAAATTTC TTAATGTCTG TGGCAACGAT CGGGGCATTC TTTATTGGGG AATATCCCGA    600
AGCTGTGGCA GTAATGTgTT CTACCAAGTC GGcGAATGGT TCCAAArTGC TGCGGkTGAT    660
CAATCCAGAA AATCmATTGC CAAATTAATG GATATTCGTC CAGATTCAGC AAATCTTTTA    720
GTAAATGGAC AAATAAAAGC GGTCGCTCCA GATACAATTG AAATTGGGCA ACAAATATTA    780
GTGAAACCAG GTGAAAAAGT TCCATTAGAT GGTCAAATTA TTGATGGATC AAGTATGGTA    840
GATACCTCTG CGTTGACTGG TGAATCCGTT CCAAGAACGG TCAAAGTCGG CGATGAAAtC    900
TTAGgTGGTT TtATCAATAA AAATGGGgCC TtGACCATTA ATGTtACGAA AAAaTTtGGG    960
GATyCAaCAG TTaGCAAAAT TTTAGATTTA GTTGAAAATG CTAGCAGTAA AAAAGCCCCT   1020
GCAGAAAACT TTATTTCTAA ATTTGCTCGT TACTATACAC CAGTTGTCGT TGTTTwAGCy   1080
AtCCTAwTGG CAGTGaTTCC aCCatTTATT TTCCCTGATA CTTCTATTAA TGAATGGGTT   1140
TATCGAGCAT TAACGTTCTT GGTTATTTCT TGTCCATGTG CCTTAGTCAT TTCTGTTCCT   1200
TTAAGTTTCT TCGGTGGAAT TGGCGGAGCA GCAAACTAGG TGTTTTAATT AAAGGAAGTA   1260
ACTATTTAGA AATTTTAGCG AATACGGAAA cAATTGTTTT TGATAAAACT GGTACGCTAA   1320
CAAAAGGAAA TTTTGTGGTT CAAAACATAA CTAGTGTTGT CTTACCCGAA GAAGAGTTGT   1380
TGCGTTTAAC AGCAACAGCA GAACAATTAT CCACTCATCC AATTGCTATT TCTATTAAAG   1440
AAAGCTATGG CAAAGAAACT GTTCCAGCAA CAGCAATTGA AGAGGTTGCC GGACAcGGTA   1500
TCAAAGCCAC AATTGAAGGT AAAACAGTTT TAGTCGGTAA TGCGAAGTTA ATGAAACAAT   1560
TTGGTATCGA AGCACCAGAA GTGAAGGAAG CGGGAACTTT AAtTTTTGTC GCTATTGATA   1620
ATCAATTTGC AGGCTATTTA GTGATTGCGG ATCAATTGAA ACCAGACGCC ATTTCAGCAA   1680
TAAAAGAATT AAAGGCTGAA GGGGTCAAAC AAACTGTTAT GCTGACTGGA GACAATCAAC   1740
AAGTTGCAGA AGCTATTGCA AAAGAAGTTG GTGTTGACAA AGTTTATGCA GAACTTTTAC   1800
CAGATGGTAA AGTGGATCGT TTAGAAGAGT TGTTAAAAGC ATCTTCACCT AAAAAACAAG   1860
TTGCCTTTGT TGGTGATGGT ATGAATGATG CACCTGTTTT AGCCCGGGCT GATGTTGGGA   1920
TTGCCATGGG TGGTTTAGGA AGTGACGCGG CAATTGAAGC TGCTGATGTG GTGATTATGA   1980
ATGATGAACC AAGTCGTATC GCCTCTGCGA TTAAATTATC AAGAAAAACA TTGAGAATTG   2040
TTAAACAAAA CATTATCTTT GCGATTGCTG TTAAAATCAT TGTCTTGGCA CTAGGTGCAT   2100
TAGGCTTGGC TTCAATGCAA GCAGCAGTGT TTGCTGATGT TGGTGTAACT ATTATCGCTG   2160
TTTTAAATGC TATGCGTTGT TTACGTGTAG AAAAGATGAA AGATAATAAC TAAGCCTTAA   2220
AAAGTACAGC TGAGACGAAC GTTTTTAGCT GATTATCAGG TTTTGAGCGT GGGACAAAAA   2280
TCACTTTGAC TTTTTGTTCC ACGCTTTTTC TCCTACACAA TTTCTATTCG TTCTTAAATT   2340
GCCAGAGTTC TTGTAAAGAA ATCTTTAAAT TTTTCGCTTT GTGGCATATA ATAAGAAAGA   2400
ATTTGTACGA AAATTGCGCT TCGCTTTTTT TTCATGGTAC AATAAGCAAG AATGGATTGA   2460
```

TTTACAGGAG GGAACCTCAA CATGGCATCA AATGAAAAAA CACGATATAA AGCAATCATT 2520

GCAGACCATA CGTATACAAT TATTGGCCAA GAGTCAAAAC AACATATGGA TTTAGTCACA 2580

AAAATTGTCA ACGAACAACT GGCGGAAATT AAACATCTTT CTCCGCAAAC TGATACGGAA 2640

CAAGCTTCTG TTTTATTAGC AATTAATGCA ATTTCAGATC AACTAAAAAA ACAAGAATAT 2700

GCTTTAAAAT TGGAAAAACA AGTAGCAGAT TTAAAACAGA AAACCATTCG TTTGGCTGAA 2760

TTAGAGAATC GTGTGCGCCG AATGGAAATT ATCGAAGAAG AAGCACGAGA TGTTTTGAAA 2820

AAGAATGGAC AAGAAGATGT AGAAATTCAT AATCATGTGG AAGCCCAACA AATTTTGAAT 2880

GAAAATCGAA AACAGCAGAT TCAAAATAAA GGTGTTTCGG AGTAGTTCCT GAAAGGATGA 2940

CAAGATGTTA ACAATACTTA TTTTATTACT TTTAGCCTTT GGATTTTATA CAGGAGCTAA 3000

ACGAGGCTTG ATTTTACAAG TGCTCTATTC GGTAGGTTAC TTAATTTCAT ATTTTGTTGC 3060

TCGTACCTAT TACAAAGAGG TCGCTTCTCA TTTAGAATTG TATATTCCAT ATCCATCGGT 3120

TACGCCAACG TCAAAGCTTG TCTTTTTTAA TCAAGAGATT TCTTTAGACT TGGACAAAGC 3180

CTTTTATTCA GCGGTTGCGT TTCTATTATT GCTGTTTGCa GGCTGGTTGG TCGTTCGTTT 3240

TTTAGCTATC TTTTTACATG GATTAACATT TATTCCAGTT TTGAAACAGG TAAATGGACT 3300

ATTAGGTGGC GTATTAAGTG TTCTTGTGTT ATACGTGGGT CTTTTTCTAG TGTTAGCTAC 3360

GGCGTCAATG ATTCCGTCAG ATATTGTCCA AAATCAGTTT CGGTCAAGTG GCTTGGCACG 3420

AGGCATTGTA AAAAATACCC CAATTCTAAC AAAACAAGCA TATGAATTAT GGGTTGAACC 3480

AATTACGAAA TAAAGCTGAA TGGGACTGAG ATAGAGATAA GAAGTAAGTG AGCAAAGCGA 3540

AACTTCTTAT GTTATGTCTC AGTCTTTTCA ACAAATGAAA AAGTTAGAAA AGTAGGTGGA 3600

AAAATGAATC AGCGTATTCT ATCAACATTA GGGTTTGATA AAGTAAAACA ACAGTTGTTG 3660

CAATTTATTG TGACAGCTCA GGGGACAAAC GAAGTTTCGG AATTATTACC AATTGCTGAT 3720

GAAAATAAAA TTCAATCATG GCTAAATGAA ACACAGGATG GTTTAAAAGT TCAACGATTA 3780

CGTGGTGGAA TCCCCATTCC TAAATTAGAA AATATCCAAC CACACATGAA ACGGATTGAA 3840

ATTGGTGCGG ATTTAAATGG CATAGAATTA GCACAAGTAG GCCGTGTCTT GTCGACTACT 3900

TCAGAACTGA CTCGTTTTTT TGATGAGTTA AGTGAAAATG AAGTTGATTT TGAACGTTTA 3960

TATATGTGGC GAGAGCAACT AGAAGTTTTG CCTGAATTAA ATCGCCAACT CAAACAAGCC 4020

ATTGATGATG ATGGCTACGT CACAGATGAA GCTTCGCCAG CCTTAAAGGC AATTCGGCAA 4080

AATATTCGTC GTAGTGAGCA GACCATTCGA GAAGAACTAG ATAGTATTAT TCGTGGGAAA 4140

AACGCACGTT ATCTGAGTGA TGCCCTTGTT ACTATGCGAA ATGAACGCTA CGTAATTCCA 4200

GTCAAACAAG AATATAAAAA TATTTTTGGT GGCGTGGTTC ACGACCAAAG TGCTTCAGGA 4260

CAAACCTTAT TTATTGAACC TAAACAAATT TTAGAAATGA ATAATCGTTT GCGACAACAA 4320

CAAATTGCCG AAAGAAATGA GATTACTCGC ATCTTGGCAG AACTGTCGGC TGAGTTGGTT 4380

CCTTATCGCC GAGAAATTAC GCATAATGCT TATGTCATTG GTAAACTTGA TTTTATTAAT 4440

GCTAAGGCGC GACTAGGAAA AGAACTAAAA GCAGTTGTTc CAGAAATTAG TCmAGCAAAT 4500

CmTGTAGTTT TTAAACAAGC CCGGCACCCA CTATTAAaTC CAGAAAAaGC TGTAGCCaAT 4560

GACaTTGTGA TTGGAGAAGA ATATCAAGcA ATTGTGATTA CAGGGCCTAA TACCGGCGGG 4620

AAAACCATTA CTCTGAAAAC GTTAGGTCTT TTACAATTAA TGGGTCAAGC TGGATTwCCT 4680

ATTCCAGTTG AAGAAGAGAG TAAAATGGGG ATTTTTACAG AAGTATTTGC GGATATCGGG 4740

GATGAACAAT CGATTGAACA AAGTTTAAGT ACTTTCTCTT CACATATGAC GaATATCGTG 4800

TCTGTCTTAA AGAAAGTCGA TCATCAAAGT TTAGTACTAT TTGATGAATT AGGTGCTGGG 4860

ACAGATCCGC AAGAAGGGGC CGCTTTAGCG ATTGCCATTT TAGATTCATT GGGCGCTAAA 4920

GGAGCCTATG TGATGGCAAC AACCCATTAT CCTGAATTAA AAGTGTATGG CTATAATCGA 4980

GCTGGAACAA TCAATGCCAG TATGGAATTT GATGTAGATA CTTTGAGTCC AACCTATCGT 5040

TTATTAATTG GCGTGCCTGG CCGAAGTAAT GCTTTTGAAA TTTCGAAACG TCTTGGATTA 5100

GACAACAGTA TTATTGAGGC TGCAAAAACAA ATAATGGACG GTGAAAGTCA AGATTTAAAC 5160

GAAATGATTG AAGACTTGGA AAACCGTCGC AAAATGGCCG AAACTGAATA TTTAGAAGCT 5220

CGCCACTATG TCGATGAATC GGCCGCTTTG CATAAGGwAC TGAAAGAAGC CTA 5273

(2) INFORMATION FOR SEQ ID NO: 363:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13932 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 363:

GGTTTAAGGC TTCCACTTTC CCAATGTTGC TTGTAGGTTT GCTTTAAGGC TCGATTTCTG 60

CTGATGTTTG TTGCTCTTTG AGTTGTTTAA ATGTGAAGAT TTCTCGGTTG AATAGGTTGA 120

rGTCTTGATG GGTTAGGTTC ATTTTATTCC CTCTTTTTAG TATCTTCTAG TTACTCTTAA 180

TAATTTTATT TTTCAATAGC CATTGTAAAA ATAACACAAG GACTACAATT GGCCAAAGCG 240

TGATTGGGAA AAAGAGCCCT ACAAAAATGC CTGGTACAAG AATCGTCCCT AAAATCTGTG 300

TTGTTAAACT CGTAGCAAAA ATTGCGACAA CAAAAAGCGA GTCAATTGTT GGGAAAAATT 360

GCGCGACCGC AAACCCGATT AAAATAGCTG AAAAAACGAC GGGAAAAATA TCGCCACCTA 420

TCCAACCTGT CTGTAAACAA ACTTGTAAAA ATACTAGTTT CATTATCGCA GCGATTACCA 480

AGATATACCA GGCTTGTTGA CTACCAAAAC TAGGAACTAA ATGCATAAAA GATTGTCCTG 540

AAAATAGTAA TCTTGGCATT AAAATAACAA AGAGAAAAAT ACAGGCTGCA CCAAAGGTTA 600

CCAGTATGCT TTGCTTTTCT TCAAAAGGTT GGCATAGTTT AGCCAGAGCA GCTTGCAACC 660

ATTGATAGGC GTTACCAATC ACAAAAGCAG TTAATAAAGC GGGTAAAATC AGCCATAACG 720

AAGCCATTTC AATTTTTCCA GTTCCTAAcT TCGTAATAAA CGATGGGTGT CCCACCAATT 780

TCATTAAAAA GGTAAACACC ACCAACCCGT TCACAATAAA TAACCCGTTC ATCACTTGTT 840

TCAGTTTCTT TTTATCTACA GGAGCGAGCA TCGTATCGTA GCGCTGAACA AACTGTTGCG 900

GATGTAGCAA GCGCTGGATT TTTGTCCAAA ATACTTGCTG TTCTTGCTCG TCATAATGAA 960

AATAAAGGTA CCGTAACTTA TCAGATTGCC AGACTGACAA CGAAATAATC GCTCCCAGCA 1020

ATGCGGCTTC AGGACCGACA CCAGCCCCGA AAATTAAAAT TACCAAGGCG GCCAATAAAT 1080

TTCTAAACAC ACCTGAATAA TCGACCGATT GTTTCGCCTT CAATTCTGTC AGAGCCTCAT 1140

GCGCCGTTTG TGGTATTTGC CCGAAGCGTT TCTTCAGATA ACTCAATACC AACGCTCCTA 1200

GCACACACAA GACAAAATAA TAGAGAATTG GCCATTTCAT TTGCGCTGGC AACACCTCCC 1260

AGAGCAGAGT CGTTAGCTCT CCTTCAATGA TTAAAAATGC TCCTGAAATT CCCGCGATGa 1320

TACTACTTAA AAATAACCCA CTGACGATaA TTGtCCACAT TTTTCTGTTA CGCATACTTC 1380

CTCCAAATGA TAGTTGTTTC TTTtATAGTA GCAGATAGTT AGTGGnTTGT GTTTGAAGAT 1440

TTtGGGAAGG ATTaTTTTTT TGAAAAATTG AGGTTATGAA AAGATTCCAT TTTTTATTGA 1500

TCAAATTATC CTTCTTGTAT AATCGTTTCT TTTAGCTATA TATATTCCAT TAAATTAAAA 1560

ACCGCCACTT ATATTGCTTC TTAAGTTCCC ACAAACAACG CCACTTCCAG CGCCAACAAG 1620

TCCACTTCCT TATCAAAGTC CAACAGCACT AGCTGTTCAA TTTTCTCCAA TCGCTGTAAT 1680

AAAGTATTAC GATGAATAAA GAGTTTTTTG GCTGTTTGAG ACAAGTTTCG TTTTGATTGA 1740

AAATAGACAT TCAGTGTATG GACAAAGTCG GTATGTTGCT GTTGATCACT TTTCTGCAAT 1800

GGTTGAAGAT ATTTTTTTCT GAGGGTTGCT TGTCCTTCTT CTAAAAAATA ACGTTGCAAT 1860

GTCTGTCTAA AATGAAAATT CGAGTAGGAA GCAATCCTCA TTTCAGGTTT TTCTGCAACG 1920

AATTGGCGTT GAAAGGCTAA AACGTCAATC GCTTGTTGAT AAGATTCATG AATCTGGGTC 1980

AGACCTTGAA TCTTAGGTCC TATTGTTAGG AAAAGTTCCA CATCTTCTTG TACTTTCATT 2040

AACTGATTTA TGCTACTTTG TGCTAGGCGT TGCATCATTT TCAGATAATC GCCTTCTTTT 2100

TCAGGTGGAT TTTCAACTAA TAAGACAAAG GAACTGTCAA TTTCAATACA AACAATTTTT 2160

TGTTTGTCCC CTAAAATAAC TTGGTTTAAC TGGTTACCTA AAATAGACAT TCTTTCTTGC 2220

CTTTGGCGTT CATTTAAGGT TAATGAAATT AGCAAAATCG CATAGGGCTT ATCTATTGCC 2280

AAGCGATAAT ATTGTTGAAA TTGTTGGAAA ACAGCTTGGT CCCTCTTTTC ATGCGCAAAC 2340

AGCTTAAAAA ATAATTCGTT TCGTTGTTGG TATTCTTCAA ATTTCCGTGA GTATTCCAAT 2400

TGCAATTGTA AAAGTAAGTA ATGAGCTGTG ATTTTAAAAT GAGCCACATC GACTGGCCGT 2460

AATGGGGCCA AGGTGTTCCA GACGAGTAAA TAATATAAAA TAACATTTTT CTTTTTAATT 2520

GGATAAATTT CAATTGGCAC CACTTGGTTC TGGAGTTTCA ACTGTAAATG AATCGGTTCT 2580

TGATAAAAGC GTAATAATGG ATGACTTTGT TCCAGCAATT GTTTTTGGAA CTCTGAGGTT 2640

TCTTGTCTAT CTTCTAAAAT GCTAGAAATA CGAATTGGAT TTTTAGGCAA GTCAAAAAAA 2700

TTAACTCCTT CAAAGTGGTC ATTCAATAAC AAGACAGGAT TTTGAATGGT TTCTGCTAGA 2760

TCCTGCAGCA AATGAAAACT GCTACTTCCC GTTTCTAAGA TTTGGTAAAT TTTATTATGC 2820

```
AATGCGGTTT CTTGCGCATT CATGTCAAAT TGTTCTAACG TGACACGACT TTTCAACACG   2880

TCAGCAATTT GCGAAAAGGT GTAATTGTAA GGAATTGCTA AAATAGGAAC GTTCATTTCA   2940

TCTGCTAATT TGATTAGGTC TGGTGGAACT TCTTCAAAAT AACGTTGGGT TTTAAATCCC   3000

AAGGCAGCAC AATGGTGTTT GGCTAGTTCT TTCAAAAAAT TTTTCTGTAA TTTCCGATCA   3060

TCTTTAAATA TGTACCCTGT CGTTAACAGC AGCTCACCTG CtGTTAGCCa ATCTAAAATA   3120

TCAGGATTTT CTAAAATATT GACTTTTAAA ATCGGTTGGT CCTGATTTTT ATGTCCTGCC   3180

ACTAAGGTAA TGCCAGGCAA TTCTTCTAAC GTTAATAAAT CACTCACTTT TAACATCATT   3240

TATCCTCCCA ATGTACAAAT TGTCCATTCA AACTTGTAAA TCAATGTCAT TTTTAGCAAT   3300

TGTTTAGTGT GCTGGATTCT TTTATTATGA AGTCGTGAAA GCGCTATTAA AAAGGAGGTC   3360

GTTGGAATGA CAAAAATTTT AGCTTATGAC CAAACGATTT TATCGAAACT CCCTCATCAA   3420

ACATTGACTA GTTTGCAATG TTTTACGCAC GCTATTAATT ATGGTCATGT AGACACCCAA   3480

GAAATTCTCT TGACCATTTC AACGCATCCC GAACATATAG CACCTAACAC GATTATACTC   3540

GATGAAAAAA AACTTAGCCA TTTCGCTGAT TTACCTGTCC AAATTGAGCC CCTGTTTATC   3600

CGGTTAGGCC GACGACAATT AACACTAGCC GATGCCCAAA TTTGGCAACC AACGCCACTT   3660

TATTTAACAC CCAGTAAAAC AAAAATAGCT ATTTTTTTAA CGTATATTCA GCAAACTCGT   3720

TTCAATCCTT GGCAAGCGCT TTTTCCTGCT TCGTCGTCAA CCTTTCTTTT TGATTCTTAT   3780

AGTCGGACCC AGATTCAAAA AGAAATGATT CATTTTCAAC AGGCTTGGGA GCAAAAATCT   3840

TTTAAAATAG CTCTTGATTG TTTAACGAAG ATTTTGGGAT TGGGGATTGG TTTAACCCCC   3900

ACTGGCGATG ATTTTATCAC AGGATTACTT GCTAGTTTTT CTGCATTACA GCAACTTCCC   3960

TCGAATTTCC AACAGCTGGC TGTGCTAGCC AAGGAACGAA CGAATGCGGT CAGTTATGCG   4020

GAAATCCATG AAGCAGTGAA CGAGCGGTTT TCTCAGCTAG TTCAGCGTGT TTTTTTATCA   4080

ATTGAAAGTG GAAATACAAA CGAAATGGCT CAAGCGGTTT CAGCTTTAGA GGAGGTGGGC   4140

TCAACTTCAG GAAGCGACAT ACTTTGCGGC ATTATTTTTG GTTTAAAACT ATTTTGTGAG   4200

GAGAATAATC ATGACAATTC AAACAGTAAT TAAAGAGAAC GCCTATTTTG ATTCAGTTAC   4260

TTTAATGACC ATTTCAACAC GAGCCAATGA ACTGGCTGGA GTTAAAACTG CAATGATTGG   4320

GATGGGAACC GACATGAATC TTGaAGTGAT TCGGAATGTT GGTTTATATA CACCCGCCTT   4380

GGATCaTGTT ACAACAGGTG ATTTATTAAT TGTTTTAGAT CTTGATGACC AAGCGAACTC   4440

AGAAGAAATT TTACAGCAAG TCGACGAGTT ATTTACGAAG AAAAAGAAAA CCGCTTCTTC   4500

TGAAGTAACC TATAAGACGT TAGATTCAGC CCTTCATGAA GAGCCTGATG CCAACTTAGT   4560

TGTAATTTCT GTCAATGGAA AATTTGCTGC ACGAGAAGCG CATAAAGCTT TAGATCAGCA   4620

AAAACACGTC ATGCTCTTTA GTGATAACGT GACGGTTGAC GAGGAATTAG CGTTAAAACA   4680

AAAAGCACAC GAAAAAGAAC TGTTTGTGAT GGGCCCTGAT TGTGGAACAG CGATTATTAA   4740

TGGCGTCGGG TTATGTTTTG CTAACGAAGT ACGTTCAGGA GATATTGGAA TTGTCGGCGC   4800
```

```
TTCTGGTACA GGCAGCCAAG AAGTTAGTGT CCAAATCCAT AAATATGGTT ATGGGATTTC     4860
ACAATTAATT GGTACTGGTG GCCGTGATTT GTCTGCTGAA ATTGGTGGGC TGATGATGCT     4920
TGACGGTCTG GATACTTTAA TGGTTGACGA ACAAACTAAG GCAATTTTGC TTATTTCTAA     4980
ACCGCCTGCT CCAGAAGTAA CAGAGAAAAT TTTAAAAAAA TTAGCTGTTT CAACTAAACC     5040
AGTTGTTATT TATTTTATCG GCAGTGAACA AACGGAACGA AAAATAGACA ACGTCACTTT     5100
TGCTACTTCT TCTCTGGATG CAGTTCAAAA GGTGATTCAA TTTAGTCAAA CACAGGAAGC     5160
GGAAACCAGC CTCTATCAAA ATCCGACAGG CGCAACGTTA ACAGCCATCC AACAGCAACA     5220
AGCCCCTTCT CAAAAATTTG TTCGTGGGTT GTTCTGTGGT GGCACTCTTT GTGATGAACT     5280
TTTATATGCT TTAACAGAAG TCTCTGATGA TGTTTATAGT AATATTCATA AAATTTCTGA     5340
ACGGCAATTA GCAAATCCTG ACCACAGCCA AGCACATACA TTGATTGATT TCGGTGATGA     5400
TCGTTTCACA GAAGGCCGCC CACATCCAAT GATCGATCCG ACTTCTCGGA TTACTCGCAT     5460
TCTTCAAGAA GCCAAAGATC CTGAAGTGGC GGTTTTAGCG TTAGATTTTG AATTAGGTTA     5520
TGGTTCTCAC GAAAATCCCG TTGGCGTGTT AGCGGAGGCG TTAAAGGAAG CCAAAGCAAT     5580
AGCGAAAGCA GATGGTCGTG AGTTAGCCAT TATTGGCtAT GtTTTAGGGm CCGAAGAAGA     5640
TCCaCAaGAT ATTCATGAcA GCGGAAAnTT TTAGAAgATT TAGAGGTTCT CGTTGTCGAc     5700
AGCAGTCATC AACTTTGCGA AgCcACCAAA ACTTTTGTAA AAGGAGCGAA CGAACATGAG     5760
CAGAAATGAT CTTTTTAAAC AGCCTTTAGA TACGATCAAT GTAGGCATTG ACTTTATTCA     5820
TGAAGATATG AAAAAACAAG GGATTCCTTC ACATCAAGTA AATTGGGCAC CGCCAGCCAA     5880
TGGCGACCCA GAGCTATTAA AATTATTGGA TCAGTTAAAA AATCCAACAC TTTATGAAAA     5940
AATCCAACAA GCCAACGAAG AAGCGGTTAC CCGCATTATT CAATCAAAAC CTATTTTAGT     6000
AGGATTTGAT AAAGCAATTA ATGTCATGCC CGATATGACT GAAACGACCA TTTTACATGC     6060
TGGGCCACCC ATTACTTATG AAAATATGTG CGGACCGATG AAAGGTGCCG TTCAAGGAGC     6120
GCTTGTTTTT GAAGGATTAG CAAAAGATCT AGCTGATGCC GATCGGGTTG CTCGTTCTGG     6180
CGCGATTACG TTTTCTCCTT GTCATGAACA TGATGCCGTT GGTTCCATGG CTGGCGTGAC     6240
TTCTCCGAAT ATGTACGTTC ATATTATCAA AAATGAAACG TACGGCAACA CGGCTTTTAC     6300
TAATTTAAGT GAACAATTGG CGAAAGTTTT ACGATTTGGT GCAAATGATC AATCCGTCGT     6360
GGATCGCCTG ATTTGGATGC GCGATGTTTT AGGTCCATTA CTTCATGATG CCATGACTTT     6420
TTGTCCAGAA GGCATTGATT TACGCTTAAT GCTTTCGCAA GCCTTGCATA TGGGCGATGA     6480
GTGCCATAAT CGCAATGTAG CCGGAAGTAC GTTGTTAGTC CAAGCATTAA CACCTTACAT     6540
GGTTCAAACT GATTTTTCTC GTGAACAATT AAAAGAAGTA TTTGAATTTC TTGGCAGTTC     6600
TGACTATTTT TCTGGTCCGA CTTGGATGGG TGCCGCTAAA TGTGCGTTAG ATGCTGGTCA     6660
TAATGTGGAG AATAGTACAA TTGTCACCAC CATGTGTCGC AATGGGGTCG AATTTGGAAT     6720
TCGTGTCAGT GGGATTGGTG GAAATCACTG GTTTACGGGG CCTGCCCAAC GGGTGATTGG     6780
```

```
TCCGATGTTT GCTGGTTACA CACAAGAAGA TGCTGGTTTG GATATGGGCG ATAGTGCTAT   6840
TACAGAAACG TATGGTGTTG GCGGATTTGC TATGGCAGCT GCACCAGCAA TTGTCCCATT   6900
AGTTGGCGGA ACAGTGGCAG AAGCATTGAA TTATTCAAAA GAAATGCTGG AAATTACTAC   6960
GAAAGAAAAT CCCAATGTCA CGATTCCTGT TTTAGATTTT ATGGGGATTC CTACAGGCAT   7020
CGATGTCTTA AAAGTCTTAG AAACAGGGAT GCTGCCTGTC ATTAATACTG CTATTGCTCA   7080
TAAAGAACCA GGGATTGGTA TGATTGGCGC TGGATTGACT AATCCGCCAG CGAATGTCTT   7140
TAACGAGGCA TTAAAAGCTT TAGTAGCTAC AATAAACTAA TTTCAAAAAC TTAGTAAAAT   7200
GTCATTTTTT TGGAGCGATC ATATGAAAAA TAAATTGATC AATTACGAAG CCGTTTTAGC   7260
TCTTGTTGTA AAAATCTTTT TGGGAATGTT CCTGTATTGG ATCATGAATG ACTTTCGACA   7320
TGTGTCTAAC TTTTTTCCTT ACAATGACGT AGTTACCAAA GTGAATCAGc AGGGATTTTA   7380
CAAATTTATC TATTTTGTTA TGAACTTCAC GGAAGGGGAA TTTTACGGTG GACTTTTTAC   7440
AACGCTTTTC TTACTAATCG GCGGTCTGAT TGCTTGGCAA CTTTATCGGA AAAATTCAAA   7500
ATGGCAGGGA TTTGCAATTG CTGGTGGTTC TGGTGCTTGG CCTTGGGTAC TTGCTTCTCA   7560
GCTACTTTCC CTCTTTTTAA CTATTTATGT TTTTGATTTT ACACGCTTTT TTACAAAAGA   7620
AGTGTTGTGG TTACCAACGT TTATTGTGGT TGTCGGCACG CCGCCAGCTT TAACGCTCGT   7680
TTATGGTCCT GGCTGGAAAA AATTAGGAAC GATTTCTCTG TTGAGTGCGC TTTTCACCTT   7740
TCCGTTTGCC AATTGGCTGA ATGCGCAATT GATGCCCCTT TTAAATGTAC CTGGGACAGT   7800
TTCCAACGTT ACCACAATGG CTTTTGTTGG CTGGATTGTT TCAGCCATCT GTCATCAACT   7860
TCCTTGGATG GTTCCTTCTG TTGCACCAAT CGCTCAGAAG CGGCAAAAAC AACCTAAAAT   7920
TGAAAATACG CAAACCTTTC AATGGTCAAT ACGACGCACT TTTGCTGACT TCTCCGAAGC   7980
TTGGTTTTAT GGAAATGAGT TAGTTGGAGG TTTGGTAATT CTAGGCGTTT TAACAGATTG   8040
GTTTGTCAAT ATTAATCATA TTACAAATGG TTCTGGCTTG GTTCCTGACA TTTTAATGAG   8100
ACAAATGATT GCTTCGGCTG TTGGTGTTTT TCTTTATCGA AAACATTTCG CTGAAGAAGG   8160
TTGGTATCCA ACGTTTATGC CACTGGTCAG TATTGTTCCT GGTGTGATTT TAATGACGGG   8220
TGGTGGCTTT TGGCTTTCTC TGGCCATCGC CGTTTTAGCA GGAATCAGTG CGGCTCCTGT   8280
CGGTAACTAT ATTGCTAAAC GATTACCTCC GTTTGTTCCT GGTGCTGTCG GCTTTGTATC   8340
AGGAATGGCT GTTGTAACCA TTTTATTGTC AGCCGTTTTG CATACATTTG ATATTTTTGT   8400
ATAGTGAATA AAAAGAAGCG AATGACGATT CGCTTCTTTT TTTAATTGAC CAATCCCCTT   8460
TTTATCCGTA TTCTTAAGCA AAAGGAGGTT TTGCCATGTA CTCCATGTTA AAACGTGTGA   8520
TAACGGAAAA AGATTTGCTT CGGCAAATTC GTTTGTTGGA ACAGTTGTTA AATGTTCCCC   8580
AGCTGACAGC CAAGCGTTTA GCCGCGCAAA TCCAAACGAC AGAGCGAACC GTTTTTTCGG   8640
ATTTACAGTA TATTCGCAGT CAATTACCCG CCGACTGGTC CATTGAAACC GATAGCAGTG   8700
GTATTCGGTT ACGCAATCAG GGGAATGCGC AAACCAATGA GCTTTGGTCA TTATTTTTAC   8760
```

CTCAATCCAT TAGTATTCAG CTATTAAAAG AGCTACTATT TACGAAAGAG TTAGTGACAA 8820

CTTCTTTTTT GAGCACTAGT GGTGTTTCTT ATGAAACTTT GAAGCGACAT ATCAAAAAAA 8880

TGAATCAGGC CTTACGTGAT TTTCATTTAA CGATTCAACT AACCACGATG ACCATTCAGC 8940

TAATCGGGGC AGAAAGCAAT ATTCGGATTT TTTATCATCG TTTGTTGGTG CCTTTTACTC 9000

ATAATAATTA TTTTTTTGAC GATTATTCGA TTCATGAGGA GCATTATTTC CAATTTTTGA 9060

AACAAGTTTA CAGTAGTGAG TTGACGGTGG AAACGGAAGA AATTTTTGGC GCTTGTTGGT 9120

TTTTTATTAA CACGATTCGC AACAAAGCCA ATTGCCGAGT GAGTCAATTT TCTTTTGACT 9180

CAAAAGATGT CTTGTTTCAA CTTTATCAAC CGTCCTTGGC AAAGCTTTAT GCTTCTGAAG 9240

GTATCTATTT GCAAGGAGAA GAAAGTTTTT TCGCCTTCTT TTGTTTCTTA GAAAGCTGGA 9300

ATTACGATAA TGTCTATGGC GAAACGTTGG CTTCTGCGCT CCACACTCAT TACAGTCAGC 9360

TGCGAAAATC ATTACAGCAG TTTGTCACTA ATTTGTCCAC TGAAGAAGCC CGGCCTGATT 9420

TAATCCAAAC AAATTTATTG GATAATCTGT TATTGTTATT CATTAAATAT ACCGAGTCCC 9480

CAACACTTTC TGAGCAATTT CAATTGGAAT ATCAAGAATT AATGACCGAA CAAGCCGCGG 9540

AGGACCTTGC GTTATCAAAG AGCAACCAAG AACTTTTGGA GATTTTATCT CGCTACACCA 9600

CGATTGAAGA ACCGACTTAT TTTCTTAGTT TGGCTTCTTT GTTAGAAAAA CAAGCCATTT 9660

ATTCCATTCA GGCGCAAACG ATGACTGCCT ATTTCCTCTT TCAAGGGGAA CCTGCTTGGA 9720

AAGCTTTTTT ACAACAAGAA TTAGCTGCTT ATTTAGGTAC TCGTGTGAAA TTACAAGCTA 9780

TTGAATATGT GGAACTAAGC CAGTTGACGT TAAATGAAGC CGACATAATC ATTTCCAATT 9840

TTCCCTTGGA TCATCTTGAT CTTCCTGTCT TTTATCTCTC TTTAATTCCA ACAAAAAATG 9900

AACTCCGCCG ACTAGCAGAG CTCACTCTTC ATTCTTATTT TTAAAGGAAC TAGCGAGGAC 9960

ATTCGCTAGT TTTTTGGCTT TCTTTCGTCA TGAATCGTTT GGTTATATAT TTTTCCCCTT 10020

CAACAAGGAA AAACAGACTC ACAACATGGA CGAAAATCAT TGCTAATTGA AGGCCATTTA 10080

AAGAAGTTGT GCCAATTAAT TgctGCGCTA CTGGCCAGTA AACTACTAGC GCTTGTAGTA 10140

AAAAGAGGAT TCCTAATGAA GCAAACAACA CTTTATTTTG GAAAAAGCCT GTGCTTAACG 10200

ACGGATCAAC CAACTTACGA CAGTTGATCA TATACGCTGc tTGCGCTAAG ACGATACTTT 10260

GTAAGAGGAA CGTTTGCTGT AAGGCTTGGC CCTCAAATTG TAACGCCATG AGGTAACCTG 10320

GTACCATAAT GAGTACCGAC ACATAGACGA TTCTAAAAAT ACTGTACTTT GTTAAAATGC 10380

CTTCATTGAC ATCTCTTGGT GGTCGTTTCA TGGTATCAGC ACTCGCTTTT TCAAAACCTA 10440

ATGCATAAGA CAAAGTAATC GTAGTTACCA TGTTTACCCA TAATATCTGG ACAGGTGTCA 10500

GCGGTAGAGG ATGATTCATT AAAAGCGCCC AGACAACAAT TAAACCTTGT GCTAAAGAGG 10560

TTGGCAAGAA GAAGGTAATC GTCTTTTTCA AGTTATCAAA AATGCGGCGT CCTTCTTTCA 10620

CGGCTTTAGC AATCGTATGG AAATTATCAT CTGCCAAAAC CATATCTGCG GCTTGTTTGG 10680

TTACTTCACT TCCTTTAATT CCCATGGCAA TTCCGACATC AGCTTTCTTT AGTGcTGGCG 10740

```
CATCGTTCAC GCCATCGCCG GTCATCCCGA CAATTTCACC ATTGTTTTGT AAAGCCGTGA    10800
CAATTCGTAA TTTATGTTCT GGCGTCGTTC GTGCAAATAC ATCGACTTTT TGAACGTGCT    10860
GAGCTAATTC TTCATCTGAC ATAGCATCAA TTTCTAATCC TTCAAGGACT TTTTTCGTAT    10920
GCTTTAAACC AACTTGCTCG CCAATGGCTT GCGCTGTATC TTTATGGTCT CCGGTAATCA    10980
TTTTGACGGA AATACCTGCT TCTTGAGATT CTTTGACGGC TTGAATGGCA CTTTCTTTTG    11040
GCGGATCGAT GATGCCTGCC AGTCCAGCAA ACGTCAATCC TGACAATGTT TCATGTGTCA    11100
GTTCCTGTTG TGACGTAACT GTCTTATAAG CGAAACCTAA TACCCGTTGT CCTTTTTGCG    11160
CTAATTGCGC AgCTTGTGCT TGCCAAGCAC CTTTTTGATT ATCAGATAAC GTAGAAAGTT    11220
GTAACAATAC TTCTGGCGCA CCTTTGACAT AAATAATTGA ACCTTCTGCT TGCGGATGGC    11280
GAGTGGCCAT GTATTTATAA CTTGAACTAA ACGGAATTTT CTTTTCTACA GGGCGTAAAG    11340
ACAGTTGATC CTGATCAACA TATTGTAGCA AAGATAACTC GGTAGGATTC CCCTGCAAGT    11400
CGGCGATTTT TTGTTGGTCT TGTAATTTCA ATTCTTGACA GTTGGCCATA ATCTCTTTGA    11460
CAATGGTCAT TTCTTCAGTC ACTACATCCA TTACCGTCAT TTCATTTTTA GTTAGCGTTC    11520
CTGTTTTATC TGAACAGATA ACTGTCATTG AGCCTAATGT TTCAACGGAA GGCATGCCTT    11580
TAATAATCGC ATTTTCTCTA GCCATCTCAT GAACACCCAT CGATAAAATC ATCGTTAACA    11640
CAGCAGGTAA GCCTTCTGGA ATCATTGCTA CAATTAATGC AATCATCGCA GAAAATAAGA    11700
GGCTCCATTC CATGCCATGA CGAAAAGTGG TAAAGAAAAT CAGAAACACG ATCAAGACCA    11760
TAATCCCTTG GAAAATTTGC TTGTTTAATT GGTGCATTTT CCGCACGAGC GGCGTTGTTT    11820
TTTGATCTAC TGACTGCAAC GCGTGATTAA TTTTTCCAAC TTCAGTGGAA TCTCCTGTTT    11880
CTACTACAAC GCCCAAGGCT GAGCCCGCTT GGACAAGAGT TCCTGAAAAA GCCAAATTCT    11940
TTTGATCGCC TGTTGGCAAT TCCTCATTCA AGGAACCGGT AATTTTCTCC ACGGCTTCTG    12000
ATTCGCCTGT CAAAATTGAT TCTTCAATCA TCAGATTATG AACATCAAAT AAGCGCAAAT    12060
CTGCTGGAAC TACATCGCCA GCTTGTAACG TGACTACATC TCCCAGAACC AATGTTTCAG    12120
AAGAAACGGT TGTTTTTTGT CCATCAATTA GAACGACTGC CTCTTGGCCC ATCATCTGTT    12180
TTAATCCGTC TAATGATTCT TCTGCTTTAC GCTCTTGCCA ATAACCGACA AAGCCATTGA    12240
CAAGCACCAC TAGAAAAATG ATACTTCCTT CGACCACTTC GCCTGTGGCA AATTTTAGAA    12300
TTGCGGCAGC GAGTAAGACC ACCATTAGTA AATCTGTAAA ATGTTTGGCT AGCTTTTGCC    12360
ACGTTTTCAG TTGTTGTTTT TCTTCAATTT TATTTGGACC ATTCGTTTGT AAACGTTGCT    12420
GACGTTCTTC GCTAGAAAGA CCCGTTTGCT TCGTCTTAGT CTGCTCGATT ACTTTTGTTG    12480
AGGAAAGTTG ATACCATTTC ATTTCGTGTA CCTCCTATTT ACTTGATACC TTTATTATAG    12540
AGAGGATTTT CTTCACAAAA TGGTCGATAA TTTCAGTGGC CCCTGAAAAA AATGGGGCCA    12600
AGATTATTAA AGAAAGGAAC AACCCGCATT TTAGCGGGTT GTTCCTTTCT TTTAAATATT    12660
TTCTTGTTCG ACCACCAATT GACTTTTGCC AATTAACGCT TTAACATCAA TAATTTCACC    12720
```

ATTTTTACGT CGGTCTGCTA ACTCAACTGA GGTGGTAAAC AACAAGTCAC TAGAAGAGTT 12780

CAACGCTGTT TCAATGGAAT CTTGGACTAC GCCGACGATA AAGCCAACGC CGACGACTTG 12840

CATGGCAATG TCATTGGAAA TGCCAAACAA ACTACATGCA AGCGGAATTA ACAACAAGGA 12900

ACCACCAGCA ATTCCTGAAG CCCCACAAGC AGAAACAGCA GCGATGATGC TCAACAATAA 12960

TGCTAAATAG ATTGGTACGG ACATGCCTAA TGTGTGGACT GCCGCCAATG TCATAATAGT 13020

AATCGTAnAT GCTGCGCCGC CCATATTAAT CGTTGCGCCT AGCGGAATAG AGACCGCATA 13080

AGATTCTTCA TTTAAATCCA TTGCTTTCGC CAATTCCATA TTAATGGGAA TATTGGCCGC 13140

AGAACTGCGT GTAAAGAATG CGGGAATCGC ACTTTCCTTT AATACAAAAA AGACTAACGG 13200

ATACGGATTT TGCCGAATAT TCCAAAAAAC AATCGCTGGA TAAACGACTA AAGCAACCAC 13260

GGCCATCGTA CCAATTAACA ATAATAGTAA CTGACCATAC TTGGCTAACC CTGCAATGCC 13320

TGTTGTTGCC ACCGAATGAA ACACCAAACC TAGAATTCCA ATTGGCGCAA TGCCGATAAT 13380

CATTTGAACG ACCGTGGTAA TCCCTGTTGA AAGATTGGCA ATCACATCCT TGGTAGCGAC 13440

CGAACTTTGG CGTAACCCAA TACCGATTAA TGAGGACCAA AATAAAACAG ATAAATAATT 13500

GCCTTCAATC ATCGCTTGAA TCGGATTTTG GACAACGCTT GTTAACACAT TGCTTAAAAC 13560

ATCACTAAGT TGCGTTGGGG CTGCTTGCGC TTCCTGCGCC GCCTCCAAGA CAATCTTAAC 13620

GGGAAACAGA TAGCTTGCCG TCACGGCAAC AACAGCTGCC AAAAATGTCG CTAACAGATA 13680

CACAACTAAA ATTGATCCAA CATATGTTTT CGCCCCAGCC TTTTGTTGGG CTAGCGACGC 13740

AATAATTAAG ACAAAAACTA AAATCGGCGC AATTGCTTTT AAGGCGCCCA CAAACAATTC 13800

ACCAAGAACG CTGATGAAGG TCCATTCTGG CACCAAGAAA CCCAAGGTGG TTCCAATCAC 13860

AATCCCTATC ATGATTTTTT GAATTAACGA TAATTTCCGA ATCGCTTTGA CCATTCACTT 13920

CACGCTCCTT TA 13932

(2) INFORMATION FOR SEQ ID NO: 364:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9509 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 364:

CTGAATTGGT TATTGTTTCG ATTTTAGAAG AGAAAGAGAT TAATGTTACC AATCGTTAGA 60

TAAAAnTTAT TGGCAAGAAC AGTTGGCGAA ATTAGAGAAA CAnACGGAAA AATATCAACA 120

AGAAGCGTTA GCGAATGGCA TCGATAAAGT AAGTGTAATT GTCAATGAAG GGAACCCAGG 180

AGAATTAATT ATTAATAAGC TTATTCCTTT GAATAAACCA GATTTGTTAA TTATTGGCTC 240

CAAGTCAACT AGCAAATTAA AAAGCTTTTT TGGTAGCCAA GCAGCTTATA TGGCTCGATA 300

TGCACCAATT TCTGTAATGA TTATTCGCTA ACAAAATTGA AACaGTTAAA AAAGCCGCCA 360

CGATTTACTA TTCGTGGTGG CCTTTTTATA GTAATTTTTG TTTATATAAA TAAACAGCCA 420

AACTTAGTTG TAGCGAACTA TCAGGAAATT CTAAAGTTAA GTGAAGAATC TCTTCTATTT 480

TTTTTAAGCG GTAAATAATT GTATTACGAT GTACATACAT ATTTGTGGCA GTCAACGAAA 540

TATTTTTATT CGATTGAAAG TAGGTGTTGA TGGTTTCTAG TAGGGAACTT TCATGTTTTT 600

GATCTTCTTT TTCTAATGGT TCAATCAATC GTTGATATAT TTTTTGTTTG GCTTGTTTGG 660

TCAGGGTTTC GATCATTAAA CGATAAAAAA AGAGGTCGTC ATAATTCAAA ATAGTTCGTT 720

GTTGTAATTC TTGACTTGCT TGTGCTAAGT GGAATGTTTC TTGTGCTAAT TTATAAGAAT 780

CACACAGATT GGTGATTTTT TTGACAGTTG GCCCAATCGC TAATTTAAAT GATAATTTTG 840

GATTCAGTTT CATTAGTTGC TCTAGTAAAT TCTTGGCAAG ATGAGTGGTT TGTGCAAGGT 900

AGGTGGCTTC CTCTTGCGCT GGATTTTCTA GTAATAAAAT AAAGCCAGTT TGTGTAGGAA 960

AGGCAGTAAT TTTTTGTCGA TATCCTTCTA AAATATGATT CAATTGATGG CCAATGCTAT 1020

TCATTTGAAT AGTCGGATTA AACAGTTGTT CCAATTGTTC AATGGCTACA AGCAAGAGAC 1080

TAAATGTATC TTTTTCATTT AATTGCAGAT AATTACTAAA TTCTTGCCAA GCTTGAGGAT 1140

TGGTATTTTT TTCTTGGATA AGGTCTTGAA ACAGCGCCCC GATCTTTTGC CTCGTTGTTA 1200

AATTTTCCGA AGATtCTAAT TGAAGTTGTG GTAAAAGGTA GTGAGCACAG ACCTCCACAG 1260

ATAATAAATC AATCGTCTGT AATTTTTTGG TAGTTTGCCA AATCACTAAG TAAGATAAGC 1320

TGAGATTTTT CTTTTTGATG GGTAAAATTC GCAAAGATAA CTGATGGTTC GCTAAGGAAA 1380

ATTCAACAGT CAATGGTTCT TTATAAAATA ATAGGAGCTC ATTGTTCTGG GCTTTTAAAG 1440

CTGTCTCAAA CTGTtTTTTG AGAGTCGCGT TATTTAAAAT ACTGGTGATT CTTGTTGGCG 1500

CATCAGGTAA ATCAAAATGA TTCATAATCT GAAAACGCTC GTCTAAAAGC ATTACAGGCG 1560

CTTGAATCGC GGCACTTAAT ACTTCCAGTA ATTGGGTACT TTTTCCGCCA GCTTTGATAA 1620

TTGTATAAAA TTTATTATGC ATATGAACAG AATGTTGGCT AATCGTTTCG TCAGGTTGAT 1680

TGACTCGGCG ATTGACCAAT GAAATAATTT GAGAAAAGTT GTAATTATAA GGGATTGAGA 1740

GAATTGGTAA TCCGACGTCA TTCGCTAATT GGATCATCTC TTGTGGAACT TCATTGAAAT 1800

AACGTTGGAC TTTAAAGCCT AGAGCAGCGC AGTTCATATT TGCTAGTTCT TGAATAAGAC 1860

GTTTTTGCAA GCAAGGCTGA TCTTTGAAAA TATAGCCGGT AGTTAATAAA ATTTCTCCTG 1920

GCGTTAACCA GTCAAAAGCA TCAGGATTTT CTAAAATATT TGCTTGTAGA ATTGGCGCAT 1980

CTAAATTGGT CTCGTGCGTC AATAAGGTTA AACCAGGTAG TTCTTCTGCA TGGAGTAGAT 2040

CATTGACCGT CATCATTAGC AAGGCCTCCT TTTGTATGAA TTATCTAAAA ATTAAGTGTA 2100

TTTATTGTAA GATGTTACAA AACTATTTTA GGGAAAATAA GGTATCCTTA TACTAATAGA 2160

AAAAAGTTCC AAGTCATCAC GTGATTGGTA GTCACAAGTG ACCAAATAAG AACAATCAGC 2220

GCTTACAACC TATAGAAACG CATCGTTTAA AGATATAGCA ATAGTATACT TAATTTTTAA 2280

TGCTTTTACT ATTGTTTCAG CTACTTTAAA TTTCTCAAGT ATTTCTATAA TCTGTAAGTT 2340

TTTTGAGACT CTTATCAGAT ATGCTTCTTT TTCTATATTT TATCAGAAAA TTCTGGTAAC 2400

TTTGAAGGTT ATAGTTTGTA AAATGGAGGG ATCGGTATGA AGAAATTAAA AATGTTAGGA 2460

TGCGTCGGGT TGCTTTTAGC TTTAACGGCT TGTCAGGCGG GAACGGGAAA CTCGGCTGAT 2520

AGTAACAAAG CAGCGGAACA AAAAATTGCA ATTAGTTCTG AAGCGGCTAT TTCGACAATG 2580

GAACCACACA CAGCGGGGGA TACGACCTCG ACTTTAGTCA TGAATCAAGT TTATGAAGGA 2640

CTCTATGTTT TAGGTAAAGA AGATGAATTA GAGTTGGGGG TCGCTGCCGA AGAACCAGCG 2700

ATTTCTGAAG ATGAAACCGT TTATACATTT AAGATTAGAG AAGATGCCAA ATGGTCGAAT 2760

GATGATCCAG TAACAGCAAA CGACTTTGTT TATGCATGGC AACAAGTTGC TTCCCCTAAA 2820

TCAGGATCGA TTCATCAAGC TTTATTTTTT GATGTCATTA AAAATGCTAA GGAAATTGCT 2880

TTAGAAGGCG CAGATGTGAA TACTCTTGGG GTTAAGGCGC TAGATGATAA AACGTTAGAA 2940

ATAACTTTAG AACGGCCCAC CCCTTATTTG AAATCATTAC TTTCGTTTCC TGTTTTGTTT 3000

CCACAAAATG AAAAATATAT CAAAGAACAA GGGGATAAAT ATGCTACTGA TGCAGAACAT 3060

TTGATTTATA ATGGTCCTTT TAAATTGAAA GAATGGGATA ATGCCTCTTC TGATGACTGG 3120

ACCTACGAAA AAAATGATAC GTATTGGGAT GCTGAAAAAG TTAAATTAAC AGAAGCGAAA 3180

GTTTCAGTAA TTAAGAGCCC AACGACAGCG GTGAATTTGT TTGACTCGAA TGAATTGGAT 3240

GTAGTGAATA AGCTAAGTGG TGAATTTATT CCTGGTTATG TTGATAATCC AGCCTTTCTT 3300

TCAATTCCTC AATTCGTCAC ATACTTTTTA AAAATGAACA GCGTTCGTGA TGGAAAAGAA 3360

AATCCGGCTT TAGCGAACAA CAATATTCGT AAAGCGTTGG CACAAGCTTT TGATAAAGAA 3420

AGTTTTGTAA AAGAAGTCTT GCAAGATCAA TCAACGGCTA CAGATCAAGT AATTCCGCCG 3480

GGACAAACGA TTGCGCCAGA TGGAACAGAT TTCACAAAAC TAGCTGCTAA GAAAAATAAC 3540

TACTTAACCT ACGATACAGC GAAAGCAAAA GAATTCTGGG AAAAAGGGAA AAAAGAAATT 3600

GGGCTGGATA AAATCAAATT AGAATTTTTA ACAGATGATA CAGACAGCGC CAAAAAAGCT 3660

GCTGAGTTTT TCCAATTTCA ATTGGAAGAA AATCTAGATG GATTAGAAGT GAATGTTACT 3720

CAAGTTCCTT TTACTATTCG TGTTGATCGT GATCAAACGA GAGACTATGA TTTAGAATTA 3780

TCTGGTTGGG GAACCGATTA TCGTGATCCA TTAACAGTTA TGCGCATCTT TACTTCGGAT 3840

AGTACCTTGG GCGGCGTAAC GTTCAAGAGT GATACGTATG ATCAATTAAT TCAAGAAACT 3900

AGAACAACAC ATGCGGCTGA TCAAGAGGCT CGTTTAAATG ACTTTGCTCA AGCACAAGAT 3960

ATTTTGGTGA ATCAGGAAAC GGTTTTAGCA CCAATCTACA ATCGAAGCAT TTCTGTATTA 4020

GCTAATCAAA AAATCAAGGA TCTGTATTGG CATTCATTTG GACCCACGTA CAGTTTAAAA 4080

TGGGCTTATG TTAACTAATT GAATGAAGGT AGGAAGGGGT TTGTCATGCT TTTAGAAAAC 4140

TGTAAATTAG AAACAGGATT TTCTACTGTG AATCAGCGAA TTGTTGCGAC AGAAACAGCT 4200

AGCTATGATA TTCGAATTGA AAACGAGGTC TGTACTGCGG TCGCACCGCA GTTAAGGCCT 4260

TTAGAGAACG AAACAGTTAT CGATGTGAAT CAGCAACTCG TTTTACCTGC GCTAAGAGAA 4320

```
ATGCATATTC ATATTGATAA AACGTATTTT GGGGGACCTT GGAAGGCATG TCGTCCACTA    4380
ACGAAAGGAA TTTTAACACG TATTGAAGAA GAAACTTGGC TGTTACCTGA ACAATTACCA    4440
ACTAGTTTAG AACGTGCTTG CCAAGTTATT GAGCAGTATA TTCAACAAGG CCATTATCAT    4500
ATTCGCAGTC ATTGTAATGT AGATCCATCT ATTGGAACGA AACATATTGA AATTACTAAA    4560
GAAGCATTTA GTCGATATGA ACAATACATT ACTCATGAAA TTGTGGCTTT TCCTCAACAC    4620
GGATTACTCC GCTCACAAGT CGAACCTTTG ATGCGCGAAg CnTTACGCAT GGGGGCAACT    4680
CATGTCGGCG GCGTTGATCC AGCCCTTGTG GATCGTCATG TGGATCATTC AATCGCAAAA    4740
ATATTTGAAT TAGCTACCAC ATTTAATAAA AAAATTGATG TTCATTTACA CGCCCGTGAT    4800
ACTTTGGGAC TTTATGAATT TAATAAATTT GTTGATTATA CAGAACAAGC AAAGATGTTT    4860
GGCAAAGTCA CTTTAAGTCA TGCTTTAGCC TTAGGGAGCT TAGAAGAAGC AGCAATTAGA    4920
GACATTGCGC AAAAATTCAT CGAAACAGGT ATTGATTTGA CTTCGACTGT ACCAATTGGA    4980
ATGCCCACCA TGCCGATTCC AACCTTGGTT GAACAAGGTG TTAAGATCTC CGTGGCTCAT    5040
GACAGTTTGA CGGATCATTG GTCGCCTTTT GGCAGTGGCA ATACAATAGA AAAATTAAAT    5100
ACAGCCGCTC AAAGATTTAA AATTAGCGAT GAGTATCGCT TGAATCGTTT ATGGGGCTTA    5160
GCGAGTAATT TTGTAACGCC TCTTGATCCT AACGGGCAGC AAGTTTGGCC TAACGTCGGC    5220
GATTCTGCAG ATTTCCTACT ATTTAATGCT GAGAGCACTG CTCACGTGAT TGCACGCCAA    5280
CAGCCAATCC AACAACTGAT ACTAAAAGGA CAACTGGTTG AAGCCGTTCA AAAGGGGGAA    5340
AAAGAATGAC AGCATCCaTT GTTTATTTGA ATAATGTTCG TTTAGAAACT GGTTTTGCTA    5400
AAGACCACCA CGAATTAACG TATACTAAAA CCGCACGCTA TACTTTAGCG ATTCAAGAAG    5460
GCAAAATCCA AGCGATTATT CCGCAAAATC AAGTCACTGA AAAGCAACAA GGCACTGATT    5520
TAAACGGACA ACTGGCTATT CCCGCTTTTC AAGAGAGTCA TAATCACTTA GATAAAACCT    5580
ACCTTTCATT AGGCTGGCGT GCTAGCCAAC CTGTTAAAAA TTTAAAAGAG CGTTTAGCAG    5640
ACGAAGCCAG TGAACTTAAA TTGTTAGCTC CGTCAACAGA ACAACGAGCG ACGGCCATGA    5700
TTGAAAAATT GATTGGGTAT GGAGCCTCTT ATATTCGGAC ACATGTCAAC ATTGATCCTT    5760
ATGTAGAACT AGAAAACTTC TGGGGCGTCA AAAGAGCCTT AGAAAAATAT GCGCATGTGA    5820
TTGATTATGA TATTGTTGTC TTTCCACAGC ATGGATTATT AAAAAATCCC CAAACGGTTT    5880
TACTCATGAG AGAAGCTCTA AAAAATGGTG GAACAATGGT TGGTGGTTTA GATCCTGCAG    5940
GCATTGATTA TGCAATTGAA GAATCTTTAG AAACTATTTT CGATCTGGCA GAAGAATTTC    6000
AAGTAGGAAT CGATATCCAC TTGCATGATA CGGGGGAAGT GGGCGTTTAT ACAATTGATA    6060
AATTTTTAGA CATTTTAGAA GAACGAAAAT TTACTCAGCG AACCGCTATC AGTCACGCCT    6120
TTGCGCTGCT AGATGTACGA CCAGCAGAAA AAGAAAAACT ATACCAACGC TTGGCAACGC    6180
ATCAAACGGC GATTATGTCT ACAATTCCCT ATGATCCTCG TTATTTATTG CCACCAATTG    6240
ATGTGTTACG CCAAGCTGGT GTATCGGTTC ATTTAGGTAG TGATGGCTTT TTTGACTCTT    6300
```

GGAGTTCTAA TGTTTCAGGT GATTTATTTG AAAAATTACG TAATTTTTGT GAAATGACAG 6360

GAAAGATTAC AGAAGAGCAA CTTACTCAAG CTTACGTCCA TGGCTGTGGC AAAGTAGCAC 6420

CTTTTTCATT TGAAGAAGAA CGCCTTTGGT TTACAGAAGG CGACGAAGCA AACTTTATTT 6480

TTACAGAAGC AGCTTCTACA GCAGAAGTGA TTGCTAGAAA ACCACAGAAA CGCCAGATTA 6540

TGTTGAAAGG ACAATGGGTT TAACTTGCTT TAAAATTAGG TAGAAATAGA ATGCTAAAAA 6600

GCCAAACAGA GGAAACAAGG TTCAGCTCTG AGCAACGCAT AGTGTCAGTC TAGCGGTTAC 6660

TCGTCACAGT TTAATCCTTA TAAAGTGCTT CTGCTCACAA CGTGTATCAG CGTTTCAAGT 6720

TTGTTTGACT CACTTTAACA TAGCACTTTT GCGTATCTTC CTTTATAATA AGACAAAATG 6780

AAGGAGATGA ACCAAGCATG ACTGAACAAA CGAGAGAAAA CAAAAAAGCA GCCTACGAAT 6840

TGTGCTTACA ACAATTGAGC GGTTTGTTAG AAATAGAAAC GGATGCTATT GCAAATTTAG 6900

CAAACTCGTC CGCGTTACTT GCACAAACAT TACCAGACAC CGTTTTTGCT GGGTATTATT 6960

TATATAAAGA TGAAGAATTA GTTTTGGGGC CTTTTCAAGG GAAAGTCTCT TGTACCAGAA 7020

TTAAAATGGG CAAAGGTGTC TGTGGTGAGT CCGCGGAAAA ACAAGCTACA TTGATTGTTG 7080

ATAACGTGAA GACACACGCC AATTATATTT CTTGTGACTC AGCTGCGATG TCTGAAATCG 7140

TTGTACCGAT GGTAAAAAAT CATCAATTGG TCGGTGTTTT AGATATTGAT AGTGGTGTAA 7200

CAAATAGCTA TGATGCGGTT GACCAACTGT ATTTGGAAAA ATTTGTTACC TTGTTACTAG 7260

AAAAATCAGA TTTTTAAGCA GACTATTTTT GCTTTCTGAC AAATCATGAG AAAATAAAGT 7320

ATAAAATTTG ATCATGAAAG AAGGTGAGTC ATTTGTTAGA GCTGTATATA AATTTTAAAG 7380

GTGAAGCAAA AGAAGCAATT GCGTTCTATG AAGATGTTTT TGATACAAAA TGTCAAAATT 7440

TAATGACTTT TGGCGAAgcA CCTGAAGATC CTGAGCATCC TATAAACGAC GAGATAAAGG 7500

ATTTAGTGAT GAACGCAAGT ATCATTATTG AAGGAACCTA TGTAATGATT TCAGATGTTC 7560

CTGATATGTT TGGTTTTGAA GTTACGGAAG GAAATAACCT TTCGCTCGTG GTTAGCACAG 7620

ATGATGATGA AAAAATTGAT CGCTTATTTA AGCGATTATC TGAAGGTGGC ACTGTTACAA 7680

TGCCGTTTCT GAAACTTTTT GGTCCAAAAA GTATGGTTCC CTAAAAGATC AGTTTGGCAT 7740

CCATTGGATG TTTAATTATT ATGAAGAAGA TTAAAAAAAT CCCCAATCCG TTATTTTCAT 7800

TGGATTGGGG ATTTTTTATT CATGAAAGAA TTGCTTCTTA GCGATTTGAT ACTGCCGTTG 7860

CTTTTCTGTC CAAAATTGCT GATCGTTTTC ATCAATTTTG CGAATGACTT TTGCTGGAkT 7920

ACCCCCAACA ATCGTGTTGG CAGGAACGTC TTTTGTAACC ACAGCACCAG CAGCAACAAT 7980

AGCATTTTTT CCAATTGTAA CGCCAGGTAA AATGGTGGCA CTTCCGCCAA TCCACACGTT 8040

ATCTTCAACA ATAATTGGCG TACCGAATTC TAATTCTGCA GTTCGAATCG TTGGATCAAT 8100

TGGATGGCCG GCAGTATAAA AACTAGCACG AGGGCCAACC ATGACATTAT CACCGAAGAT 8160

GATTTTGTTT ACATCTAAAA AAATGCAATC CATATTTGCA TAAAAATTTT CTCCTACCTC 8220

AATATGGCGA CCGTAATCTA CATGGAGGGG TGGGGTGATA TATAACTCAT TGCCAGTTTT 8280

TCCTAATAAT TCTTTTTCTA GGGCCACAAT TTCTTTCACA TTATCGATAG GAAGTTGATT 8340

GATTTTTTGA GCTAACACCT TGCCTTTtGT TGATTTGTTT TCGGGTAAAA TGGTGGGGGC 8400

GTAATATAAT TCACCTTTTA ACATTTGCAT ATAATCAAAG TCATTTTGTT TCATGTAAGC 8460

GCTCCTTTAG GCTTTTTCTC TATTGTATCA AATCAATACA ATAAGTCGAG GAATATGTAA 8520

AATAAAAAAG GACAGATAGC AAAATAGCCT TGCTTATCTA TCCTTTATCT AAGGTTATCC 8580

AATTTTTTTA ACTCTCCCGC CAATTTTATC TGCTAATTTT TGTGCTTCTT TTTCAGAGTG 8640

GTAAACTAAA CAACCTGATT TATCTCGCGT ACGTTTACCA TCTGCAGTTA AATAGCCATG 8700

AAAAAATTTG ATTGCATACA TAGACATCAC CGCCTTTCTA ATTGAAAATA GTGTAAATGG 8760

AGCGATTAAA AATGTCAAAT AATAGAAGTG CTCTTAACCA TTCTTATCTT TTTTAGTTCG 8820

AAGATAAGCT TGGTTAAGAG CACCATTTCA CTTAATAAAT TAGTTTGAAA CGGGTCGTTG 8880

AGTGAAATTC CTGTTCAGGT GAGAGGGTGA TAGAACCGAA CAACGGAAAT CTTTCAGCAC 8940

CAGGACAGAC TTGGGTTTCA AATGTGAGCG CTCCGTGATT GACTAATTTT TGCCCTTGAA 9000

TATCGGGACC GTTATCAGCA AAATTTGCTG AAAAAATAAC AACTGAAGGT TGATCAGTGA 9060

ACATCTCTAA GGTGATATCT TTAGTAGGAC TTTTTAAGAT AACATCCGGG TTAATTAAAT 9120

CATGATTGGT TAAGATGAAA GGGTGATCGA AGCCGTTCAT CTTTTGGATT TGTTCATCGT 9180

TACTTGAAAA AACCTCAGCT AATTTTTTAG GTTTTGTAAA ATCAAAAGCT GTCTGGTCGA 9240

CGGAAGCTAA TTCTCCAAGC GGTAAAACTC TTTCATCAAG CGGGACAAAC TTTTGGGCAT 9300

TTATTTGTAA CTCATGCGCA TCAATGGGGA TTGTAGGATT TCCAGTTAAA TTAAAGTAGA 9360

CATGGATTGG TTGGATTAAA AAGTGTTTGT TCATCTGTAG TTGCCTGGAT ACTCAATTTG 9420

CCATTCATTT TCTTCAGTCA ATGTATACGT GACTGAAACA GTGAGATTCC AGGAAAACCA 9480

TTTTCACCAG CAGGACTCTC TAGCCAAAA 9509

(2) INFORMATION FOR SEQ ID NO: 365:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5247 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 365:

CCATTCCTAA AAAAAATTCA AAATCTTTCA GAGAAGCTGT TAAAGGAATT TTTAAGTTCA 60

CAGATATTAA CTTAGCGCGT GAGGCTAAAA ATCGATTGAT TCATGATTAT ATCGATCAAC 120

CAAAATATTC AAAAGcTTGC GCATCATTGG ATGATGGATT CGAAGACGCC TTTCAATATA 180

CCGTACAAGG AAATTCCCAC AATCGACTAA AGAGTACCAA TCTAATTGAA CGACTGAATC 240

AAGAAGTACG CAGAAGAGAA AAGATTATTC GCATCTTCCC CAATCAAACA TCAGCCAATC 300

GCTTAATTGG AGCCGTTCTT ATGGACCTAC ATGATGAATG GATTTATTCT TCAAGAAAAT 360

ACATCAATTT TGATAAGTAG AAATGGTAAA AACATTGTAT AGCATTTTAC ACAGGAGTCT 420

```
GGACTTGACT CGGAAGTCCA AACCAAAAAC AAACTAATTA TTAACCAACC TGAGCAAACT    480
GAAAAACCAA TGAAAGATGT TGCTAAATTG CTGAAAATTT TACCGCCAAC ATTTAACGAA    540
ACAGTGCTCC TTCCAGTCCA AACACAGTTA AAAGAACACT ATCCGCGAAC GGTTTCCCTC    600
GTGAAAAACT TAGTCAGTCC GCTGCCAGAA GAATACCGTT TCTTGACAGA GGTACTCTTA    660
TCTGTATTAC TCAGCGGAGA AATTTCTGAA ACGATTCCTT ACCAAGCGCT CCTCGTGGCG    720
CATGGTGAAT CGACTGCCAC AAGTATTCAA GCCGTTGCCA ACAAATTATG TGGTACCTAT    780
ATTTTCGATG CGATTAACAT GCCCTTAACC TCCTCCGTCC GAGATATCGT CGCCGAAGTT    840
AAAGAGTGGC TTTCTCAGCG GGACACATCA CAAGGCGTCA TTATGCTCGT CGATATGGGT    900
TCTCTGACAC AACTTTATAA AAGTTTGAAA CCGCAAATTC TTGGAGAACT TTTGGTGATC    960
AATAACCTGA CCACTGCCTA TGCGTTGGAA ATTGGTCATC AATTAATGAA CGAACAACTT   1020
TTCTATGGTA TTGCTAAAAC CGCTGAAAAA AAATTTAAAA CCGATGTACA GTACTTTGAA   1080
GGCTTTTCAG TAGAAAAAAA TATCATTGTC TCTAGTATTT CAGGTCTGGA TATTGCCAAG   1140
CAAATCAAAC AAATCTGTCA AAAATACCTG TATACTGACA TTAAAGTCAT TACTCTAAAA   1200
TACAAAGATT TGGTAAACAC GTTAGACATT GCCAACGCCG AAGAAAATTA TTTAAAAGAA   1260
ACCTCGCTCA TTTTGACGAC GTCTTACTTA GACAACCATA CGAATGTTGC TAGCGTGAAT   1320
CTCTTAGACA TGTTAGACGA AGATGCAGGA ACCCAGTTGA TGGAACCTTT CCAAAACTTG   1380
ATGCATCCCA ATAACATTGA TAGCATGATT AACGAATTTG TTCACTTCTT CTCGAAAGAA   1440
GGCTTGTCAG AAAAACTAGA ATTTCTAAAT CCTGACGTCA TTATTAAACA AGTCGAAAAC   1500
GTTACAAAAA ACATTGAAAA ACGCTTCGAC CTCACACTCA GCGGCAAAAT GAAATTTAAC   1560
TTAATGATGC ACAACGCGCT CATGGTAGAG CGAACCATGC TCGGCGTTGA AGACTATGAA   1620
ATCCCTGCTA ACCTCGAGGA ACTAACCATT AACCAAAAAC CATTTTTCCA AAATGCTAAA   1680
AACATTTTCT ATACTCTAGA ACAGTTTTAT CGAATAGAGA TTTCCAACTG GGAACTCTAT   1740
GTGATTTATG AAATTTTGTC TTCTCGGTGA GGGGGATGGG AAAAGCCACT TTCCAGAGGA   1800
AAGTGGCTTT TTtAGTTtGT GtTCTGCGTT AAAAATTCAT ACATATTAGG TTCTACAGGT   1860
GTTTCTAAGG TAAAGTCTAG TTTAACGACG CTTTTTTTCT TACCCTCTTT GTCTTGTTTT   1920
TCGACTTGTT GGATAGTTyC TTGAAGAGGG GCTACTtCGC CTAGATAATA GAAATCTGTT   1980
TCGTTCTCTT TTTGATTATA TTTCCGTTTC ACAAATAAAT GAAATTTATA TTCCCCAGGA   2040
TTCCGCAAAA GTTGAACTTC TGGCGAGCGT AGATTTCTAG GCGCTTTTGT ATACCAATGG   2100
AATGTTCGTT CATCAATAAA TTCATCCTCA TAAGCAATAC GTGCAGCTTT AAAATCTTTG   2160
CCTTTATCCA AAGTTACAAA AATAGCAAAA TGATTATTAC TGTTTGTATA GCCACCAATA   2220
CTTTGTTCAT TTTGATTAAA GTTCATATTT AATAATCGTA AAACATCTTT ACGACGATAT   2280
TTTTTGTATA AATTTAACGG TTGGGATAAA TCATACTCCA TAGATTTGTT AATCCCTGTA   2340
TGGATTACAT CTAACATCAA ATTGATGAAA TAAGAATTCT TTTTAAACTC TCTGAATTTT   2400
```

```
TCTGTTATTT GTAAACAATT TTCTTTCATC TTAAACATTT GTGATAACAA ATAACGTTTT   2460
CTAATTGTTG ATTCGTAAAA AGATAAATCy AAGACTTTAA CAACAGATGC TAGAGTATCT   2520
GTCGTATACG TCAGGTTAAA CGTATCAAAC AGCGCTGTGA TTTCCTCTAA AGATAACTGT   2580
TCTTTATCTG TCATTAAAAA TTCTTCAATA ACATAAAGTT CATGCGGGCG AATTCCTGAT   2640
AGCAATTCTA CAGAAACAAA TTTTAGTGCA CCGTTTTCTT CTTCACTTAA TTTAGCTTCT   2700
GTTTCCCCAA TTCTCTCCAA GAAATTATAG TATGTTTTTT CTTTACCTAT AATCAACAAT   2760
GGATCCACGC CATCTAACAC TTGAAAATCA TTCAGGTAAG GAACCCTATT CAACCGATTT   2820
TTCAACTTAA AATAACTAGC TTTTAGTTCC GACATCGCAT CTAATTTAGC TTGATCAATT   2880
GATTTAAAGA TACGTTCACG AGCAACTTTT TCAAAATTAA TTGAAGATAA ACCACTGATA   2940
TACGTCACAT CAATTGTATC GCGACGTAAA CTATTCTTAG TTTGAGaACT ATCTCCAGAT   3000
AAAGCCATAG GAATCATATA GTTATtTTTA TAATTaCCAA TAAAaTCaAT TACCGTAACA   3060
AATTCTTTTG ATGGGTCTTT TCTTAGCCCC CGACCTAaTT GCTGAATAAA AATAATACTA   3120
GATTGCGTAT TTCGCAGCAT AATTACTTGA TTTACTTTAG GGATATCAAT GCCTTCATTA   3180
AAAATATCAA CCGTGAATAT ATAATGAATC TCTCCAGTAT CTAAACGCTT AATTTCCTTT   3240
TCTCTTTTAT CAAAATTATG GTCACCAGAT AAATAAGCAC TCGGAATTCC ACGATCGTTA   3300
AACATCTCTG CTAACTGCCT TGCTTCGTCT TTACGACTAC AAAACACCAA GCCTCGAGGA   3360
CTATCTCCCG AATAACCATA ATAGTTAATT TTTTTTATTA AAAAATCAAC TCGTTCTTTC   3420
TGTAGCAAAC GTTTTAAATC GGTTGTTTCA TCAATGATTT CACCATCTTT TTCATAATCA   3480
GTTACGCCAA AATAATAAAA TGGACAAAGC AGTTCTGCCT CTAAAGCGGC TTGCAAACGT   3540
ATTTCATACG CTACATTATA ATCAAATAGC TCAAAGATAT TAAAATTATC TGTTCTTTCA   3600
GGTGTAGCCG TCATTCCCAA TAAAAACTTG GGATTGAAAA ATTCAATGGC TTTAATATAA   3660
GATTTCGCTC CAGCTTTATG GACTTCATCA ATCAAAATAT AGTCAAACGC TTCTTTTCCT   3720
AAACGCTCTT GATATTCACT TTTAGATACT GTTTGAATTG TTCCAAATAG ATATTTTGCA   3780
TCAAAATTCT TAGAATTCCC TGATAAAACA CCATAATCAT CATCTTTTCC ACCTAAAATC   3840
AATTTGAATG ATTCTAATGC TTTCATTAAA ATTAGCTCAC TATGAACAAT AAATAATACA   3900
CGTTTAGGAt GGAATtGCTG cACATCAAAk GCTGcTAGAT ACGTTTTCCC TGTACCAGTC   3960
GCCGAAACAA TCATtCCCTT AGTTGCGCCC GTTTCTCTAA GCGCTTTTAA ACTTCTAAGC   4020
GCTTCTTTCT GCATAGAGTT CGGCAAAATA TATTGcGTAT CTAAACACTC GTTTTCAGAA   4080
ACATCTTGCA CTAACATCTT TCTATTAACA ATAGGATGAT AATTTTTTTC AAATGCATCG   4140
ACCCATTCTA GGGTTAGCGA CTGTGCCTCA TGCCATTCTT GAATGAGATG CTTTTGAATA   4200
GAAGAAACCA TCTCGCCATT CTCTCTTGAA GTTAATCTAA CATTCCACTC ATAGTTAAGT   4260
TTCAAAGCAG ACATTGTTAG GTTCGAACTA CCAATAACAA AGCTTTGATA ATCCTCCATC   4320
TCAAACAAAT ATCCTTTTGA ATGAAAACCT TCTTTTTTAG AAACTCGTAC TTCTACATTT   4380
```

GGAATATTCA ACAATGCTCG AAATACGTGT GGATCATTAA AATTTAAGTA TGTTGATGTt 4440

AAGATTTTtC CAGAGATTCC ACGCAAAGCC AGATCAGCCA GTTGTGTTTT TATTGCATTT 4500

AAGCCACTAG GTGTtATAAA TGCAACGGCT ATAAAAAAAG TCTTACAGCT TTTCAAATCT 4560

TTACTCAATG TATTCAATAA ATAATCATGT TTTTTGGGAT TGTTAACAAT AAATTTAGGT 4620

GTAAGTTCGT CACCTAGAGT TTCCTTATTA ATAAATGCCT TTTGTAACGC CAATTCTAAA 4680

TTATCCATTT TGTTCAGTTC CCACACCTTT TTCTACTAAT TTwTCCACAG TAGGAATATC 4740

AGCAGGCGCC CACTCAAGTT GGTTTAATTC ATTTGGCGTT AGCaTTTAAT TTGTAGATGT 4800

TCTGTCAACT GAGGTTCACC TGATTCTAAA TAACAGATAA ATGTAGTTAA ATTAACGAAA 4860

CCAAAATCAT ACTCATATCG ACAAAAAGCA TACTCTTCCT TGACGATCGT TACTTCTATT 4920

TTTAATTCTT CTTGCAATTC ACGTTCaAGC GCTTGaACTT CCGTCTCACC ATTCTCGATT 4980

TTTCCACCAG GAAATTCCCA AAGATTTGCT AATGCACGTT CAGGTCCTCT TTGACAACAC 5040

AAAACTTTCC CCTTCTCCAC TATGATTGCT CCTACTACGC GAATTTCTTT TTTCATATAT 5100

TTTAACCCTC TAAtCTAATT GATTTTGTTA GTTTwATTAT ATTTAGGATC GCATAAATAA 5160

GCTAGGGTTG CAATATwAGt GGCGkAAATT TCATTTATTA TATTTGTTTT TTGATTTACT 5220

ATACTTTAAT TATTTAATAA CTCTTCA 5247

(2) INFORMATION FOR SEQ ID NO: 366:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1657 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 366:

GGACTTTCTC CTCCTGAGTA TTTGACTATA CTTACAGTCA ATGGTCGGAA GATCGGGTCa 60

GTACCCcTTT GGaAATmAtT TArGTACcAT TATTTAGCAC CTTtCCcTGa CTTAGAAGAA 120

GTGATTATTC GCTTAGATGA ATTTGATCAG GTTTTGTTCG CCAATCGCAC AGCTACAAAT 180

GGAAAAAAAA TGGATATGTC ATCAATTAAA AAACATGGCT TTCAACTAAT GCGAATTATC 240

TCCGATCCTT ATACAGGACA ACCACTGGGA GAATTGTACA CTGTGTTTTC AAGTCAAGAA 300

TTGCTAGGCA ACCAAGCAGA CTTATTGAAA AAATCAGGTA TTAATGCGTT TATCTATGAT 360

AGTGCTGGGA ATCAAATTTT TTCTGAAAAA GCGCAATTTA CCAAAGAAGA AGAGCGGCAA 420

CTAGATAAAA GAATGCACAC CGATAGTGAC ATTCAACAGG TTTTTCATAA TCGCTACGAT 480

ATTACGGAAA TAGAAAGCAG TGGCCGGAGT ACCATTCTCT TATTAACAAG CCGCCGAGTC 540

TTATTTCAAC AATTGTTTAT GAATTATGCG GCCATTTTAG GGATTGGGTT ACTTTTGATT 600

GTTATTTTAT TAGTGGGCTT AAACCGCCTA TTTAAACGCT ATTCTCAACA AGTCCAATTA 660

ATTTTAGAAG CGACACGAGC CATTGGCGAT GGCAATTTAA AAGAACGAAT TGACACGAAT 720

CAAGTTCAGG AAGAGTTGAA TGATATTGcT TCCGCAATTA ATTTTATGGT AGACAGTTTA 780

GATCAATACA TTCACGATAT TTATACTTTA GAAATTAAGC AACGTGATGC CCACATGAGA 840

GCGTTACAGT CGCAAATTAA TCCTCATTTT CTTTATAATA CATTGGAATA TATTCGCATG 900

TATGCCTTGA GTCGGCAACA AGAAGAACTA GCTGATGTTG TCTATGCTTT TTCAACGTTA 960

CTTAGGAATA ATATTAACCA AGAAAAAACC ACGACCTTGG CAGAGGAAAT CTCATTTTGC 1020

GAAAAGTATG TGTATTTGTA TCAAATGCGT TATCCCGATC AATTTGCCTA TAAGTTCGAA 1080

ATAGAAGAAA CAATTGCTGA TGTTGAAATT CCTAAGTTCA TCATTCAGCC GCTCGTCGAA 1140

AATTATTTTG TGCATGGTAT TGACTATCAA CGACAAGACA ATGCTATCAA GGTCCACGCC 1200

TACCGAGAAG GCGAAAAAAT TATTGTGGCT GTTGTGGATA ATgGAAAAGG AATCACCGCC 1260

AATCGGTTAG AGGAAATTCG CGAACGTTTA AATCAAACAG AAATAGACAC GGAACAATCC 1320

ATTGGGTTAA GAAATGTTCA TGAACGTTTA CAACGCTTTT TTGGGGAGAG TTATGGTTTA 1380

ACCATTGAAG GGAAAGAAGG AGAAGGGACG ACAATTCGTC TTTCGTTTGT AGCATAAGGG 1440

GTGAAAAGAT GTATCGAGTG ATGTTTGTTG ATGATGAATA TATGATTTTA GAAGGCTTAA 1500

AATGGATTAT TCCTTGGCAA GAGTTAGGCT TTGAAATTGT TAAAACAGCG CGTAGTGCGC 1560

AAGAAGCATT GGCTTTTTTa GgAaaCAGAA TCCATTGATG TTTTATTaAC AGATATTACG 1620

ATGCCAGAAA TgAGCGGGAT CgAaTTAaTg AGCAgCA 1657

(2) INFORMATION FOR SEQ ID NO: 367:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14557 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 367:

GCTAGACTCG GAGTCGCCTC CAACATCATA ACACTCTAAA AGAAACATCT ATAAAGTGAT 60

CGGTTTATTA ATTATTTAAC AAATTCGATA ACTACCATTG GTGCTGCATC GCCGCGTCTT 120

GGTTCTGTCT TCAAGATACG AGTGTAACCA CCTTGGCGTT CCGCGAAACG AGGTCCTAAA 180

TCGTTAAATA GCTTTTGTAA AGCTGATTCA ACAACGATAG ACTCATCTTC TTCACGAACA 240

CTAGCAACTT CATTACGAAC GAAAGTAGCT GCTTGACGAC GTGCATGAAG ATCTCCACGT 300

TTACCTAAAG TAATCATTTT TTCTACAGTT GAACGTACTT CTTTCGCACG AGCTTCAGTA 360

GTTACGATGC GTTCGTTGAT GATTAAATCA GTTGTAATAT CACGTAGCAT AGCTTTACGT 420

TGGCTTGATG TGCGTCCTAA TTTACGGTAA CTCACGTTGG GTTCCCTCCT TCGTTTCTGT 480

TCTCACGAGC CACTGCTGTT AAGCAGGACT GGAAACTTCG TAAAAGTCTT AATCGTCCTT 540

ACGTAATCCT AACCCTAGAT CATGCAATTT AAGTTTTACT TCTTCAAGAG ATTTACGACC 600

TAAATTACGT ACTTTGATCA TTTCTGGTTC AGATTTGTTT GTAAGTTCTT GTACAGTATT 660

AATGCCTGCA CGTTTTAAAC AGTTATATGA ACGAACTGAT AAATCAAGTT CTTCAATCGT 720

CATTTCTAGC ATTTTCTCTT TTTGAGTTTC TTCTTTTTCC ACCATGATTT CAGCGTTTTT 780

CGCTTCGTCA GTAAGGTTTA CAAAGATATC TAAATGCTCA GTCATGATTT TCGCAGCTAA 840

ACTTAGTGCT TCTTGAGGAA TGATTGAACC ATCTGTCCAT ATTTCCATTG TTAATTTGTC 900

GAAATCATCA CGACGACCAA CACGTGTATT TTCTACTTGG TAGTTAACAC GAAGAACGGG 960

TGTGTAGATG GAATCAACTG GAAGTACACC GATTGGcATA TCTTCTTTTT TGTwTTCGTC 1020

TGCTTGAACA TAGCCACGAC CAGGTTTCAC TGTTAAGCGA GCATGGAAAG TAGCTCCTTC 1080

AGCAACACTA CAGATAACTA AATCTTTATT TAGGaTCTCA ACATCACTGk CCaCGATAAt 1140

ATCGCCGGCA GTCmCGGkAG CGGGtCCTGT AAwATCGATC TCaAGGGTTT CtCyTCyTcm 1200

GCATAwAATT cmATGCTAgC cTTaATAtTC aAAATAATTT GTGTTACGTC TTCTCGTACA 1260

CCTGGAATGG TTGAGAATTC GTGTAAAACA CCATCGATTT gAATATTCGT AATTGCAGCT 1320

CCTGGTAAaG mAGACAATAA AATACGmCGT AGAGAGTTAC CTAAAGTAGT TCCGTAACCG 1380

CGTTCTAATG GTTCAACGAC GAACTTACCA TAATCTCTAT TyTCATCAAT TTTTTCGATT 1440

CTTGGTTTTT CAAATTCAAT CATTCTTATC TATACCCCTT TCAAAACGAA AAGTGTCTTG 1500

TTCAATGACG TTCCTAATTC AAACTCAGAA TGAGTGGCGC TCATTAAACA CGACGGCGTT 1560

TTGGAGGGCG GCATCCATTA TGAGGAACTG GAGTCACGTC ACGAATTGCA GTCACTTCTA 1620

AACCTGTTGC TTGTAATGAA CGAATTGCTG CTTCACGTCC AGAACCAGGT CCTTTAACTG 1680

TTACGTCTAC AGTTTTAAGT CCATGTTCCA TTGCCACTTT AGTTGCGGCT TCTGCAGCCA 1740

TTTGAGCGGC AAAAGGAGTT GATTTTTTGC TTCCTtTAAA GCCTAATGAT CCTGCTGATG 1800

ACCATGCTAA TGCATTACCA TGTGTATCAG TAATCATTAC GATTGTATTG TkGAATGTAG 1860

AATGGATATG TGCTACACCT GATTCTATAT TCTTTTTGAC ACGGCGTTTA CGACTAACTT 1920

TTTTTGCTGC CATGAAGTTC TAACCTCCTT CACTTAGGAA TTATTTTTTC TTGCCTGCTA 1980

CTGTTTTAGT AGGGCCTTTA CGAGTACGTG CATTATTTTT CGTGTTTTGT CCACGAGTTG 2040

GTAATCCACG ACGGTGACGG ATTCCACGGT ATGAACCGAT TTCCATCAAG CGTTTGATGT 2100

TTAAGTTCAC TTCACGACGA AGATCCCCTT CAACTTTTAA CTTATCAATT TCCGCACGAA 2160

TCGCATCTGT TTGTTCGTTT GTTAAATCAC GTACACGAAC ATCTTCAGAT ACGCCAACGT 2220

TTGCTAAAAC TTTTTTAGCT GTAGTGTTAC CAATACCATA AATGTAAGTA AGAGAAACTA 2280

CTACACGTTT GTCACGAGGG ATGTCTACTC CTGCAATACG AGCCATTATT CGTTACACCT 2340

CCTAATTATC CTTGACGTTG TTTATGTTTT GGATTTGCTG GGCAAATAAC CATAACACGT 2400

CCTTTACGGC GAATTACTTT ACAATGTTCA CACATTGGTT TTACTGATGG TCTTACTTTC 2460

ATGATAATAC CTCCTGTGGT TTTACGGAGT ACAATTATTT AAAGCGATAA GTAATACGGC 2520

CACGCGTTAA ATCATACGGT GATAATTCAA CTGTTACTTT GTCACCAGGT AAGATACGGA 2580

TGTAGTGCAT ACGGATTTTA CCTGAAACAG TAGCAAGAAC TTGGTGTCCA TTTTCTAGTT 2640

CGACTTTAAA CATTGCATTC GGCAAAGTTT CGACGACTGT ACCTTCGACT TCAATCATAT 2700

```
CTTCTTTTGC CACGCACAGT ACCTCCTTGT ACTTGTTTCG CGTTTTCACA CGATAAAACG    2760
GCGGAAACTC CAGCTCCCAC CTAGAAATTC TCAACTAACT TCGTTATTCA TTTTGATGTC    2820
ACAAAACTGT AGAAGTCGGA CTGATATATT CTACCATATT TAGTTATGTT TTGGCAAGAG    2880
AAACTGAATA CAACaACAGG GTTAAACTGC GGGCAATTGG AAATTAGTTT TCGTCAATGA    2940
TTTTTTTCAC ATCAGAAAAA ACAGTATCGA TTTCGCGATT GCCATCGATT GTATGCATCA    3000
AGCCTTGTTC TTTATAGAAA GCTAAGATTG GTGCGCTACT TTCAATGTTG ACTGCCAAAC    3060
GATTTTTAAC CGTTTCGGGC TTATCATCTT CTCTTTGGTA GAATTCATGT CCGCCACAGC    3120
GATCACAAGT GCCTTCAACT TTTGGCGGGT TGAATAATTT ATGATACGTT GCGCCACATG    3180
TGCGACAGAT GAATCGTCCA GCTAAACGCT CGATTAATAC ATCTTCTTCT ACATGGATGT    3240
CAATGACAGC ATCAATTTTT TTGTTTAATT CTTTAAGCAT TGTATCTAAT GCTTTTGCTT    3300
GATCCAGTGT ACGTGGGAAA CCATCTAATA GGAATCCTTT GTCTGTATCT GGTTCAGCAA    3360
GGCGTTCTTT AACAATtCCG TTTGTTACTT CGTCAGGAAC CAGTTCACCT TTATCGATAT    3420
AAGAnTTCGC TTCTAGGCCC AAGGCAGTTt CATtTTTCAT TGCTGCACGG AACATATCTC    3480
CAGTAGAAAT ATGTGGAtGC CATACGTATC AATGATTTTT TCTGCTTGGG TGCCTTTACC    3540
TGCTCCAGGT AATCCCATTA AAATGAGGTT CATATATTCT CCTCCTAATT TTAAAGACCG    3600
GCTTATGAAG CTAGCmATAT GTGAGAATGG TGCTGAGGTA AAACCTCAAC ACTCGATTCT    3660
TACTTATTGA TAAAGCCGAC ATATTGACGT TTCATCATTA ATCCTTCTAA TTGTTTCGTT    3720
GTTTCTAATG CAACCCCGAT AACGATTAGT AAACTTGTTC CACCTAAACC GATTGATTGA    3780
GGTAAGTTCC AAACCATTTG CGCAATGATT GGAAGTAAAG CAACAAGTCC TAGGAAAATT    3840
GAGCCGACAA CACTTAATCT CATTAACACG CCAGATACAT ATTCTTCTGT ACCTTTACCT    3900
GGCCGCACGC TTGGAATGTA GCTTCCTTGT TTTTGTAAGT TTTCCGCTAA TTTCTCAGGG    3960
TTTACTTGAA CAAATGCATA GAAGAACGTA AACGCAACGA TTAGGACAGT ATAGATGATT    4020
GCCCCTGGAA CTGTGTTATA ACTGAAGATT TTTGTCATAA TGTCATACCA ATTTTCACCA    4080
GCGAATTTTG ATGAGAAAGC TTGTAAAATG GCATTTGGTG TTGCAATTAA CGAGCTGGCA    4140
AAGATAACTG GAATAACCCC AGCAGCATTT ACTTTTAACG GTAAATAACT ACTTGTTGGT    4200
GCACCAGAAA CACGTTTTGT ATATTGGATT GGAATCTTAC GTTCTGCTTG TTGGAAGAAC    4260
GTTACGACTG TAACAATCAC CAAAATAGCA ATAACTAAGA TTGCAATGAA AATAACAGAT    4320
TGCCAAATAC GAGAAGACTC GATATTGACG AAGTAATCTT CATAGATTTC TTTGACTGCT    4380
CCTGGCAAAC GGGAAATAAT CCCGGCAAAG ATAATCATTG AAACACCATT TCCGATTCCT    4440
TTTTCTGTAA TTTGTTCACC CATCCAAGTC ACAAACATCG TCCCAGCAGT TAAAATAACT    4500
GCAATCATCA CAAAGGTACC TAAGGTTGGA TTGTTGACAA TCCCAGTTTG ACTTAAGCTA    4560
TTAAAACCAG CAGTGATCCC CATTGACTGA GCCACACCCA AGACAATCGT TAGATATCTT    4620
GTTGCTTGAT TTAATTTCTT ACGCCCAACT TCCCCTTGTT TTGACCATTC TACAAATCTA    4680
```

GGTACAATAT CCATTTGCAA TAGTTGAATA ATAATAGAGG CTGTAATGTA TGGCGAAACC 4740

CCCATCGAGA AGATAGAGAA GTTTTGCATG GCACTACCAC TCACCATATT CAACATATTC 4800

AAAAAGGGTA AGCTACTTAA ATCAGACAAT CCCTTTGCAT TCACCCCGGG CACAGTAATG 4860

TGCGCACCTA GGCGAAATAC AAACAAGATT AAAACTGTAA ATAAGATTTT TGATCTAATG 4920

TCTTTGACTT TAAAGGCGTT CTTTAATAGC TTGAACAATT AGATCACCTC GATTGAACCA 4980

CCAGCAGCTT CGATAGCTTC TTGTGCTGAT TTCGAGAATT TAGCTGCTTT CACAGTTAAT 5040

TTTTTCGTTA ACTCACCATC TGCAAGAACT TTGATTCCCG CTTTTTCGTT TTTAACGATT 5100

CCTGCTTCTT TTAGGACAAC AGGAGTTACT TCTGTTCCAT CTTCAAAACG GTTTAAAGTA 5160

TCTAAGTTTA CAACAGCGTA ATCTTTACGA TTAATGTTTG TAAACCCACG TTTTGGTAAA 5220

CGACGGAATA ATGGAGTTTG ACCCCCTTCA AATCCTAAAC GAACACCACC GCCTGAACGG 5280

GCTTTTTGCC CTTTTTGTCC ACGGCCAGCT GTTTTACCAT TACCAGATGA AGTACCACGT 5340

CCAACACGGT TACGTACTTG GCGTGAGCCC TCAGCAGGTT TTAATTCATG AAGTTTCATA 5400

GGTTTGGCAC CTCCTTAAGC AATCTTCTAT TACAAATCTA TAAATCTTAA ACTTCTTCTA 5460

CGTCCACTAA ATGTGAAATA GTGTTAACCA TACCTTTGAT TGCTTCATTG GCAGGTTTTG 5520

TAACAGTACT ATTCACTTTA CCTAAACCTA ACGCTTTAAC TGTAGCACGT TGGTTTTGAG 5580

GACGTCCGAT AACAcTGCGT TTTAAAGTAA TTTTTAATTC AGCCATTATT TTTGTCCTCC 5640

TTAACCGATA ATTTCTTCTA CAGATTTACC GCGAAGTGCT GCCACTTCTT CGGCACGTTT 5700

TAATTGTTTT AGACCTTCAA CAGTTGCGCG AACAACGTTA ATAGGTGTGT TTGAACCTAG 5760

TGATTTAGAT GTAATATCTG CTACACCAGC TAATTCCAAT ACGGCACGAA CTGGTCCACC 5820

AGCAGCTACC CCAGAACCTT CTACTGCAGG TTTCATAAGG ATACGGCCGC CACCGAATAC 5880

GCCAATTACT TCGTGCGGGA TAGTAGAACC AACCATAGGC ACTTCAACTA AGTTTTTCTT 5940

CGCGTCTTCG ATTGCTTTAC GGATAGCTTC AGGTACTTCT TGTGCTTTAC CAGTACCGAA 6000

ACCAACGTGT CCGTTTTTGT CACCGACAAC AACTAAAGCT GCGAAACGTA AACGACGTCC 6060

ACCTTTAACA ACTTTTGTTA CACGGTTGAT CGCTACAACG CGGTCTTCTA ATTCCAAATG 6120

TTTTGGATCA ATATAAACCA TGAATGGTGT TCCTCCTTCT CCTAAAATTC TAGTCCATTT 6180

TCACGAGCAG CTTCAGCTAA AGCTTGCACG CGGCCATGGT AAAGGTATCC ACCACGGTCG 6240

AAGACTACTT TTTTGATGCC TTTTTCAGCG GCACGTTCAG CAACTAGTTT ACCAACAGCT 6300

GCAGCTGTTT CTGTTTTTGT TCCACCTGAA ATTTCTTTAT CCAAGGCAGA GGCACTTGCT 6360

AGCGTCACAC CCGCTACGTC ATCAATGACT TGCGCGTAGA TGTTTTTGTT AGAACGGAAA 6420

ATGTTCAAGC GTGGGCACTC AGCAGTACCA GAGATTTTGT TACGTACACG ACGATGTCTC 6480

TTTTGACGTG TTTTGTTTTT ATCTGGTTTT GTAATCACAA TTTTCACCTC TTTGATTTAA 6540

TGTTTACTAC TAATATTTCA AGCAGTCAGC AGATTACGGA AAAGAGAAGC TGGCAGACAA 6600

GACAAAAGCA TTTTATTGGA AAAATCCCAT AAAATGACAG TTGCTTGTTG TTGACTTCTG 6660

TTTTCTACAC AGCTGTTATT ATTTACCAGT TTTACCTTCT TTACGGCGTA CAAATTCACC 6720

AACATAGCGA ATCCCTTTGC CTTTATAAGG TTCTGGAGGA CGAACGCCAC GGATGTTAGC 6780

AGCTAATTCA CCAACAACTT CTTTGTTGAT GCCTTTAACG ATAACTTGTG TGTTCGCAGG 6840

TACTTCAAAT GTTACGCCTT CTGGTGCTGT CATTTCGACA GGATGAGAAT AACCAACGTT 6900

TAAAGTTAAT TTGTTGCCTT GTACTTGGGC ACGGTATCCG ACCCCGATAA GTTCAAGTGC 6960

TTTTTGGAAA CCTTCAGAAA CACCAACAAC CATGTTGTTG AAGTTTGCAC GAGTTGTTCC 7020

GTGGATTGTT TTCATTTCTT TTGAATCGTT TGGACGAATG AATGTCACTT CGTTTCCTTC 7080

GATGTTCATT TTGATATCAG AAGAAAATTC GCGTGTTAAT TCACCTTTAG GTCCTTTAAC 7140

TGTAATGTTG TTTCCGTCTT GCTTGATTTC AACACCAGCA GGAAGAACAA CAACTTTATT 7200

ACCAATACGG CTCACTTAGA GACACCTCCT TGTGTATTTT TTAGAATTAC CATACGTAGG 7260

CGATTACTTC GCCGCCGATA TTTTTCGCTC TAGCTTCTTT GTCAGTAATT ACACCCTCTG 7320

ATGTAGAGAT GATTGCAATT CCTAGACCGT TTAATACTTT TGGTACTTCG TCAGCTTTTA 7380

CGTATGCACG TAAACCTGGT TTAGAAATAC GTTTTAAGTT TGTGATAACA CGTTCTTCGT 7440

TTTTACCATA TTTTAGGAAA ACACGGATTA CGCCTTGTTT GTCATCTTCG ATGTATTCTA 7500

CGTCGCGGAT AAATCCTTCA CGTTTCAGGA TTTCTGCGAT ATCACGTTTG ATTTTTGAAG 7560

CAGGTACTTC TAAAGTTTCA TGTTTAACCA TGTTCGCATT ACGAATGCGA GTTAGAAAAT 7620

CTGCAATTGG ATCTGTCATG ACCATTGAAC TATTTACCTC CTTTATGCGA GTTTACTTGT 7680

TTACCAGCTA GCTTTCTTCA CGCCGGGAAT TTGACCTTTA TAGGCAAGTT CGCGGAAGCA 7740

AATACGGCAA AGATGAAATT TACGATAAAC TGAATGTGGA CGTCCACAAC GTTCACAACG 7800

AGTATAAGCT TGTGTTGAAT GTTTAGCAGG ACGTTTGTTT TTAGCAATCA TTGATTTTTT 7860

AGCCACGTAG TTCGCCTCCT TATATTATTT TTGGAATGGC ATACCTAATT GTGCCAACAA 7920

TTCGCGAGAC TCTTCGTCAG TGTTCGCTGT TGTAACAATT ACGATGTCCA TACCGCGTAC 7980

TTTATCTACT AAATCATAAT CAACTTCTGG GAAGATTAAT TGTTCTTTAA TACCTAAAGT 8040

ATAGTTACCA CGACCATCGA AAGCTTTTTT GCTGACACCG TGGAAGTCAC GTACACGAGG 8100

TAAAGAAACT GTTACTAATT TATCTAAAAA TTCGTACATT CTTTCGCCGC GTAGnnTTAC 8160

TTTCGCACCG ATTGGCATTC CTTCACGTAA ACGGAATCCA GCGATAGATT TTTTAGCTTT 8220

TGTGATAAGT GGTTTTTGAC CAGTGATTAA TGCTAATTCT TCAAcAGCTT TGTCTAAGTT 8280

TTTAGCATTT GATACTGCAT CACCCACACC CATGTTGATA ACGATTTTAT CAACTTTAGG 8340

TGTTTGCATA ACTGAGCTGT AATTAAATTT TTCCACCAAT GATGGTGTTA CTTCTTTAAT 8400

ATATTTTTCT TTCAGGCGGT TCATTCAGTA AGCCCTCCTT CCGCAATTTT TATTTATCTA 8460

AAACTTCACC TGTTTTTTTA GAAACGCGGA CTTTTTTGCC GTCGACTTCT TTAAACGCAA 8520

CTTTAGTAGC TTCTCCGTTT GAAGGATCAA TCACCATTAC ATTAGAAACA TGAATTGGCG 8580

CTTCCACTTC AAGGATTCCA CCTTGTGGGG CTGCTTGGTT TGGTTTTTGG TGTTTTTTAA 8640

```
CGACGTTGAC ACCTTCAACG ATAACTTTAT CTTGTTTAGG AAACGCTGCT AAGACAACGC      8700

CTTCTTTGTT TTTATCTTTA CCAGTGATAA CTTTTACTTT ATCGCCTTTT TTAACAAACA      8760

TTACTGTTTC GCACCTCCTT TAACGTGGGC TCGAGATTAT AATACTTCTG GTGCTAGAGA      8820

AACGATCTTC ATGAAGTkGT TTTCACGTAA TTCACGTGCA ACTGGTCCGA AgAwACGTGT      8880

TCCGCGTGGG CTTTTATCGT CACGGATAAT TACCGCAGCA TTTTCATCAA ATTTGATGTA      8940

AGAACCGTCA GCACGACGAG CGCCAGATTT TGTACGAACG ATAACGGCTT TAACTACTTC      9000

ACCTTTTTTA ACAACCCCAC CTGGCGTTGC TTGTTTGACA GTAGCAACAA TAACGTCACC      9060

GATGTTAGCA GTTTTACGAC CAGATCCGCC TAGGACTTTG ATCGTTAAAA TTTCACGTGC      9120

GCCAGAATTG TCAGCGACTC TTAAACGACT TTCTTGTTGG ATCACGATGT GTATCCCCCT      9180

TTCAGATTTT GTGCTTCAAT CTATTAGGAA ATTCTCGTTT TATTAGATAA TAACTGCTTC      9240

TTCGACTACC TCTAGTAAAC GGAAACGTTT TGTAGCTGAT AATGGACGAG TTTCCATGAT      9300

TTTTACGATG TCTCCAACTT TCGCTGTGTT GTTTTCATCA TGCGCTTTAT ATTTTTTAGA      9360

ATATTTCATA CGTTTGCCAT AGATAGGGTG GTTTTTCTTT GTTTCAACGA CAACTGTGAT      9420

TGTTTTATCC ATTTTGTCTG ATACCACGCG ACCTTGGTAA ACTTTACGTT GATTTCTTTC      9480

TTCAGTCATA CGCGTTTGGC CTCCTTCCAC TATTAGTTAG CTTGTTCACG CAATACTGTT      9540

TTGATGCGTG CAATCGATTG ACGTACTTCT TTAATACGTG CAGTGTTTTC TAATTGACCT      9600

GTTGCTAATT GGAATCTTAA GTTGAACAAT TCTTCTTTTA ATTGTTTTTC TTTATCTAGC      9660

ATTTCGGCAG TGGTTAATTC TCTGATTTCT TTAACCTTCA TTCGATTCAC CACCCATTTC      9720

CTCACGTTTT ACGATCTTAG TTTTTACCGG CAATTTGTGA GATGCAAGAC GAAGAGCTTC      9780

ACGAGCTACT TCTTCAGGAA CGCCTGCGAT TTCAAACATG ATTTTACCAC GTTTAACTGG      9840

TGATACCCAG CCTTCTGGTG CCCCTTTACC TTTACCCATA CGTACGCCGA TAGCTTTACT      9900

TGTGTAAGAC TTGTGAGGGA AAATTTTAAT CCATACTTTC CCGCCACGTT TCATGTAACG      9960

AGTCATTGCA ATACGGGCTG CTTCGATTTG ACGGTTAGTA ATCCAGTGAG ATTCAGTTGC     10020

TTGTAAACCC CATTCACCAA ATGCTACTTC TTTTCCGCCT TTAGCTTCAC CGCGCATTTT     10080

TCCGCGGAAT TCACGACGGT GTTTTACACG TTTAGGTACT AACATGATTA TTTCCCTCCT     10140

TTCTCAGTGT TTTTTTTCGT TGGAAGAATT TCTCCACGAT AGATCCACAC TTTAACTCCT     10200

AGTTTTCCGT AAGTAGTGTC CGCTTCTTCC CATGCGTAAT CAATGTCCGC ACGTAATGTG     10260

TGAAGTGGTA CAGTTCCTTC AGAGTATCCT TCTGAACGAG CGATATCCGC ACCGTTTAGA     10320

CGACCTGATA CTTGAGTTTT GATTCCTTTA GCGCCAGCAC GCATTGCGCG TTGGATGGCT     10380

TGTTTTTGAG CACGACGGAA AGCAACACGG TTTTCTAATT GACGTGCAAT TCCTTCGCCT     10440

ACTAATTTTG CATCTAAATC TGGTTTTTTG ATTTCAACGA TGTTGATGTG AACTCGTTTG     10500

CCAGTTAATT TGTTTAATTC TTTTCTTAGG TTTTCGACTT CAGATCCGCC TTTACCGATA     10560

ACCATACCTG GTTTAGCTGT GTGAATTGAA ATGTTAACAC GGTTTGCAGC GCGTTCGATT     10620
```

TCAATTGTAG ATACAGCAGC ATCAGCAAGT TTTGTTGCGA TAAATTTACG GATTCTTAAA 10680

TCTTCGTGTA AGAATTCGGC ATACTCTTTT TCAGCATACC ATTTCGCATC CCAGTCGCGG 10740

ATGATGCCTA CACGCATTCC AATTGGATGT ACTTTTTGAC CCACTGATTG TCCCTCCTCT 10800

TAAATGTTTA AGCTCTGTTT TCATAAAGAA AACAGCCTAT TTCATGAGAC TTATTTCTCT 10860

GTTACAACTA CTGTGATGTG ACTAGTACGT TTGTTGATTG GTGAAGCTGA ACCTTTTGCA 10920

CGAGGACGGA AACGTTTCAT TGTTGGTCCT TCGTTAACAA ATGCTTCAGA TACGACTAAG 10980

CTTTCAACGT CTAAGTCAAA GTTGTTTTCT GCGTTAgCAA CTGCTGACAT TAAAACTTTT 11040

TCAATGATTC CAGCAGACTT GTTCGGTGTG AATTTCAAGA TTGAAATTGC ATCCGCAACG 11100

CTTTTCCCTC TGATAAGATC GATTACTAAA CGTGCTTTAC GAGGTGAAGT GCGAACTGTT 11160

TTTGCAGTTG CTTTAGCTGA TGTAATTTGT TCTGACATTT GCATATCCTC CCCTCAAAAT 11220

TAGCGTTTAG TTTTTTTATC GTCAGCAACG TGGCCACGAT AAGTTCTAGT TGGTGCAAAT 11280

TCACCTAATT TATGTCCTAC CATGTCTTCT TGAATGTAAA CAGGGACATG TTTCCGTCCA 11340

TCATATACAG CGATTGTGAA ACCTACGAAT GATGGAAAAA TTGTAGAGCG GCGAGACCAA 11400

GTTTTAATTA CTTTTTTCTT TTCGGCACCT TGTTGTGCTT CGACCTTTTT CATCAAATGA 11460

TCATCGACGA AAGGTCCTTT TTTTAAACTA CGACCCATGG TGAACCTCCT CTCAAAATGT 11520

ACGACACATG GTCAAAAAAA TTATTTAGTA CGACGACGAA CGATAAGTTT GTCTGATTTA 11580

GCTTTTTTAT TACGTGTTTT GTATCCAATT GCTGGTTGAC CCCATGGAGA TACTGGCGCT 11640

TTACGTCCAA TTGGTGTTTT ACCTTCACCA CCACCGTGTG GGTGATCGTT AGGGTTCATT 11700

ACGCTACCAC GAACAGTTGG GCGTTTACGC ATCCAACGAG AGCGGCCAGC TTTACCAATG 11760

TTGATTAATT CGTGTTGTTC GTTACCAACT GAACCGATTG TTGCACGGCA AGTTGCTAAG 11820

ATCATACGAA CTTCGCCAGA GTTTAAGCGG ATTAATACGT ATTTGCCTTC TTTACCAAGT 11880

ACTTGCGCGC TTGTTCCAGC TGAACGGATT AATTGTCCAC CTTTACCAGG TTTCATTTCA 11940

ATGTTGTGGA TAACTGTACC AACTGGAATG TTTTCCAATG GTAATGCGTT ACCTACTTTA 12000

ATATCTGCTT CTGGACCAGA TACTAGGCGC ATGCCTACTT CTAATCCTTT TGGTGCTAAG 12060

ATGTATGCCT TAACCCCGTC TTCGTAATGA ACTAACGCGA TGTTAGCAGA ACGGTTTGGA 12120

TCGTACTCGA TCGTTTTAAC AACCGCTGCG ACGTTATCTT TATTACGTTT GAAGTCAATC 12180

ACACGGTATT GACGTTTGTG ACCGCCACCT TGGTGACGAA CCGTAATGCG ACCGTTGTTG 12240

TTACGACCGG CATTGTTTTT TAATGGCTGT AACAATGATT TTTCTGGTGT TGAAGTTGTG 12300

ATTTCAGCGA AATCAGAACT CGTCATGTTA CGACGGCCAT TTGTGGTAGG TTTGTACTTT 12360

TTAATCGCCA CGTCTTATTT CCCTCCTACG TTGATATTGT TCAGGACTGC TTATTCAGCA 12420

GCTTCGAATA ATTGAATTTC TTTTGAATCT TCAGTTAATG TAACGATCGC TTTGCGACGT 12480

TTTTTTGTAT AACCTGCATA TTTACCCATG CGTTTGAATT TTGGACGCAC GTTTAAGATG 12540

TTTACGTTTG CTACTTTAAC ATCGAATGCT GATTCAACAG CTTGTTTTAC TAAAGTTTTG 12600

```
TTTGCGCGAG TGTCCACTTC GAAAGTATAT TTCTTTTCGT CCATAGCTAA CATGGATTTT    12660
TCAGTGATCA CTGGGCGTTT GATTACGTCT AGTAATTCCA TTATGCAAGC ACCTCCTCAA    12720
TTTGAGTAAG AGCTGTTTGT GTTGCCAAGA CTTTGTTAGC TGACACAACA TCTAACACAC    12780
TAACGTTATC AGAAGTTACT ACAGAAACGT TTGGTAAGTT ACGAGCTGAT AAAGCTGCAA    12840
AGTCGTTACC ATTTTCTAAA ACAACTAATA CTTTAGTATC GATAGACAAG TTAGCAAGAA    12900
CTTGTTTGAA TTCTTTTGTT TTTGGTGCAT CGAAGCTTAA ACCTTCTACT GCTACCAAGT    12960
TATTTTCAGC AACTTTGTCT GATAATACAG ATTTCATTGC TAAGCGACGA ACTTTTTTAG    13020
GAAATTTGTA GCTGTATGAA CGCGGTGTTG GTCCAAAGAC TACGCCACCT CCACGCCATT    13080
GTGGTGAACG GATTGAACCT TGACGAGCAC GACCAGTTCC TTTTTGACGC CATGGTTTGC    13140
GGCCACCGCC ACGAACAGCG CTACGGTTTT TAACTGCGTG TGTTCCTTGT CTTAATGAAG    13200
CACGTTGCAT GATGATTGCA TCATAGACAA CACTTTCATT AGGTTCGATT CCGAAGATTT    13260
CTTCATTTAA AGTGATTTCA CCGTTTTGAG TTCCATCTTG TTTGAATAAT GCTACATTCG    13320
GCATTCCTTA GTTCcTCCTT TCCcTACACT TATTTAGCTT TCACaGCTGA TTTGATTGTG    13380
ATTAAaGATT TTTTCGCTCC aGGAATGTTT CcTTTGaTTA AGATAACATT TCTTTCTAgT    13440
CCATTTAACA ATTCaAGGTT TTGaATtGTT TACGCGkTCG CCACCCATAc GgCCGGCTAG    13500
TCGTTTATTT TTAAATACAC GGTTAGGCGC TACTGGACCC ATTGACCCAG GACGACGGTG    13560
ATAACGAGAA CCATGAGACA TAGGTCCGCG GCTTTGGCCG TGACGTTTGA TTGCCCCTTG    13620
GAATCCTTTA CCTTTCGTAG TACCTGTAAC ATCAACAACG TCTCCTGCTT GGAAAACATC    13680
TACTTTGATT TCTTTACCTA CTTCATATTC TCCtAGCTCA ACATTCTTGA ATTCTTTAAT    13740
GAAGCGCTTA GGAGCCGTGT TTGCTTTTGC AACATGACCT yTCGCAGGTT TGTyTGATAA    13800
AACTTCACGT TTGTCTTGGT AACCAACTTG GATAGCTTCG TATCCGTCAG TTTCAACAGT    13860
TTTTaCTTGt AAAACTACGT TTGGCGTAGC TTCaACTACT GTTACTGGAA TTAATTCACC    13920
AGACTCAGTA AAGATTTGTG TCATTCCCAC TTTTTTCCCT AAGATTCCTT TGGTCATGAG    13980
TACACCTCCA TCTTAATTTT ATAGTTTGAT TTCAATATTA ACACCAGACG GTAAGTCAAG    14040
CTTCATTAGA GCGTCAACTG TTTTTGGTGT TGGGTTCACA ATGTCAATTA GACGTTTGTG    14100
AGTACGCATT TCGAATTGTT CGCGAGAATC TTTGTATTTA TGAGTCGCAC GAATAACTGT    14160
GTAAAGTGAG CGTTCTGTTG GTAATGGAAT CGGACCTGAT ACGTCAGCTC CAGTTCTTTT    14220
TGCTGTTTCC ACAATTTTAT CCGCTGATTG ATCTAAAATA CGGTGTTCAT ACGCTTTTAA    14280
ACGGATACGA ATTTTTTGTT TTGCCATCTT GTTCCCTCCT TCGCCTATTT TAAAAGTAGA    14340
CATAGCTCCA CGAAAATTTA TCCGGCATGC TCGTTCATGG CAAAGCGTCC GAGCGTGTCG    14400
CAACCTCTCG CTTCACAGCC GGCAAATCAA ATCGTTGATC TACCAAnGCT TTTTACACTC    14460
CTGTAAACAG CACCTTTTTG ATTATACTAT GAAAGGATAG TGTTAGCAAG GATTTTCTGC    14520
GTTTTTTGAA AGAATTTGnC nnGTTTAGAA AAGCATT                             14557
```

(2) INFORMATION FOR SEQ ID NO: 368:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1981 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 368:

CCCCCCGGAG TGTAAAAGTA ACTATGATGC ATACTTTGGA TAAAATTcTT CCAAAATAAA 60

mCCGTkGGCC cATkGAGACA AAGCGCTCcA AGAAACTGGC TTTATTGGCT AGATTTATCG 120

AACCATGAAC GACATAAaCT ACTTTTTAAc TGCTGTTTGT GCGGTAGAAC ATCCAGATAT 180

TGCCAGCGGC CTAATAATTC TTCTTCAGCT TCTAATCCTT GACGATCCGT TTGAATTTGT 240

TTCAAAGATT CTTGAATATT CTTTAATTCC GTCAAATAGC TTTGTAATGT TTCTTCATCA 300

AAGGTTTGTy CCAACTCTTC TAAGGTATAG ACTTTTCTTT TTATCTTTAA CCCTTTTCCT 360

GAAGAGCGTG TCAAAAATAA AAGAAGCTCT TGCAATTGTT GTAGCATCGT TTCATATTTA 420

CGTAGCTGTT CAGTTGACTG TTCTGCTTGT AAAAAAGAAT ATTGACTCTT TAAGGCTTGT 480

TCTTCTTCAG GCGAATGAAA AATCCGTTTT ACTTCGACAA TTTGCAAGCC TTGAATGGCT 540

TGTAGAATTG CTGCTTCTTT TTCGGCAGCG GCAATTATTG TCATTTTTTC CATTTTACTA 600

ACTGCCATAA ACGTCTTTCA CCCTCTCAAT AATTGAATCA ATGACTTGTT CTTTGTTCGC 660

TTCATAACGT TCGTGAAGCG CTTGTTCAGT GCGCTTTGTT TCATCTAAAA GTTGTTCTTT 720

TTCAAGTTGT AACATCTTTT CTTGCGTTGC TAGTGAATCC GTCAGCACTT TTTCAACAGC 780

TTCTTCACTC TCTTGACGTA GTTCAGCTAA ACGTTCTTGC TTCTGTTCTT CATATAAACG 840

ACAATTCATC TTAGCTTGTT GCCGATGCTC TTCCACCGCA TCTTCCGCTG CCTTGATTTG 900

CTCTAGTACT TCTTGTACCA TTGTCTCCAC CTCCTTAGTT TTTACACCTT TAGCATATAA 960

GAAAAAGGAC TAGCGAAATA GAGATACGCT CAGCCCTTTT GTTTTAGGTA TGCAAAAAGG 1020

CGCAAACTCA AAAATTGTTC TTCTCACCCC TTGAAGAACT GAATCCTTGA AGCTTTCGCC 1080

CAACCATTCT TTATTCAATA CTCTACCATC TTAGCAAAGT TTTTCGCAAA ATTCAATACT 1140

TACCCAAATG AATTAAAATA ACAAATTTTC ACTTTTATAA TAGGTAATAA AAAGGCATTA 1200

TTTTTTCGCT TTATAATTAT TGATCTTGAA TTTTTTTCAC ATAGATGTGC TATACTTACG 1260

TAAACAATTA AATTTATTTr GAGGGATCTC TGATGAATAT ACGATATATT TCAATCTCTG 1320

GTAATACACG CTCATTTGTT CAACGTTTAA CAACTTATTC TGAAGAGCAA CATCAGCACA 1380

ACGAAAAAAA TCCTACGATT ACTTTCAAAG AAATTTCGGA AAATTCTCCG TTAGAAGTAG 1440

AAACAGAACC ATTTTTTACC TTTGTGCCTA CCTATTTAGA TGGTGGCGAT GGCATCAACA 1500

ATGGCAATAC CGAAATTTTA ACAGAAGTCA TGCGAGAATA TCTTGCTTTT GAAAATAATT 1560

ATCGTTATTG CTCTGGCGTG GTGGGCAGCG GCAATAAAAA CTTTAACCAC CAATACTGCC 1620

TGACAGCTAA ACAGTATGCG GAACAATTTA ACTTCCCCTT CTTAGCTGAT TATGAATTAA 1680

GAGGTACCCA AGCGGATATT GAACGAATTT ATGCTATTTT AAAGGAAAAC CAATAAGCAA 1740

GTTTCATTAC TAAAAAGCTC ACCTATGATA GACTACATTT TGAGGTGAAG AAAAATGAGA 1800

AATACATGGT TAGCAGAACA ACTTCAGTCA ATCAGTGAAG AACCCAATTC TTTTATTATC 1860

GAGGAAACCA TCaAATATAT TGAACAATTA GAAGATGATA ATGAAAGTCT TCAAGTTGCA 1920

TTGGAAGGGA CTATTTGGGA GTCCTAAAAA TGGGACGACC ATTGGGAAAA TATTTAACnG 1980

A 1981

(2) INFORMATION FOR SEQ ID NO: 369:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1652 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 369:

TCATAGGGCC TATTTTTTGT CAAGGAATCT TCTTATCTTT TTTGTTTAAT CTATCAAATT 60

AAAATATTTT ATTTTTTAAT ATAAAATTAT TTCAATAATT ATTGAACATT TCAAATTGAA 120

TTGATTTTTT CAGACGTAAC TTTAGGcTTG AGCCAGAATT TTTTGCGCTT CATCTATTGT 180

TCGACAATCA CTGAGCCGCT GAATGTTTTC GAGATCTGAA CAGAAAAAAG CTAATTCTTG 240

AAGCATGCTT AGTTGCCGTT GACTTTTTAA AGCAACGACA AATAACAAAG TAGCTATTTT 300

TCGATCTGCT TCCGTACCAA AGTTGACCCC ATCAGGCACT TGAATAAAGC AAATCCCTTC 360

TTTTAAAACT TGTTCATCTT GTCCTTCCGC ATGAGGTAAT GCCACGAAGT TACCGATATA 420

CACACTGACT TTTTGTTGAC GTTGATACAT TGCTTTGACA TAGTCAGGCG TGACCAACTG 480

ATTTTTTAGC AGTAATTCAC CAGAAGCTGC AAGTGCTTCC TCCCAACTAG CATACGATTG 540

ATTCAACATT ATTTGCTCCG CTGTAATTTC AAACATCTCT CAACACCTAG CCTTTCATTT 600

CTTCAATTAA TAATGACGAA AGCAACTGAT AAACAATGGC TtCATTTCCT GAATGAAAAA 660

TTTCTGTATT TAAATCATTC ATAATAATCG CACCACTGAT TTTTCCTAAA ATTTTCCCGA 720

TTGTTTCATC AATCGGTATC GGCGCTAGCA TGAGTAACAT ACGTTGTAAC TGCATCGGCT 780

TTTTATCCAT ACCTTCAATC ATAAGTGGCT GGGCCAGATT AAAAATACAA AAAACAGGCT 840

GTGTCACTGC TGTACTTGAG GTATGGAATA AGGCCATTTC CGTCTGAGGA ATCCCAATAG 900

GCGCTTGCTC GTAACGTTTC ATCAATTGCT GACAGACGAT TGCACTATCG GTGATTAAAT 960

CTGCGGAAAG TTGCGCGAGA GCTTGTTGAA CGGTTTCTGC TAAATCCGCT TGATTATTTA 1020

AATCTGCTAT GAAAAATGTT TCTAATAGTT TACTAATTTC TTCCATGAAA GCGACCACCG 1080

TTTCGTAGCT TTCTTCGCCA CCTATTTTTT TCACAGGCGC TTTCCTAAGC GAGCGTTTTT 1140

CGTGACTAAT CCGCTTAAAC TCTTCTTTCA ATTGGcGGAT TTCATCGTCT AAAAGTAACG 1200

GTGAGATCAA TTTATATTTA CCATTGtATC CtGGTAATAA CGArTGgAAA TCACTAAGTC 1260

ATAGTTTTCC TCAACATTCA CCTTACCTAA ATCCGCAATC CGAAAAAAAT GAATCTGGTT 1320

GATAAAAGGA AAGTAGCGCC GTAATCGCAT TTCCAGCATA CTTGTCGAAG CCAAACCGCT 1380

AGGAGAAAAA CCAGCAATAT CAACTTCCAT AATTTTGGGA CTCCGTTCTA AAGAATTGGC 1440

AAAATGAAGC ACCATGTAGG CAATTTCCTC TTCCGATAGA TTCTTTTGCG GAAAAATTTG 1500

CGGCAACGCT TGGCGTAAAG CgCAGCTATT TCTTGATATT GATTCATGAT CCGTTCTAAA 1560

ATGGGATTCG TTAAATCTnC TTCTTGTAAA ACCGCCCGAG AAAAAACACC TGATAAATGC 1620

GTCAACAACA TTnTATACAA ATnGGTATCT TC 1652

(2) INFORMATION FOR SEQ ID NO: 370:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1742 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 370:

GTGATTCGAG nCGGTGCAAA nTGGnCCGAT GGTACCGnTA TTACTGCTGA TGATTTTGTA 60

ACGGCATGGC AAAGAGTGGT AGATCCCaAA ACTGCTTCTC CTAGCGTTGA GCTTTTCTCT 120

GCTATAAAAA ATGCTAAAGA AATTGCTTCA GGAAAACAAT CGAAAGAAAC GTTGGCAGTA 180

AAAAGTAATG GCAATAAAAC GATAGAAATT GAGTTAGAAA AACCCACACC ATATTTTACA 240

GATTTATTAG CATTAACAGC GTACTATCCT GTACAACAAA AAGCGATTAA AGAGTATGGG 300

AAAGACTATG GCACTTCACA AAAATCAATT GTAACAAATG GGGCATTTAA CTTAACAAGT 360

TTAGAGGGAG TGGGAACTTC TGATAAGTGG ACGATTTCTA AAAATAAGGA GTACTGGGAT 420

CAAAAAGAGG TTTCAATGGA GAAAATTAAC TTCCAGGTAG TTAAGGAAAT TAACACAGGG 480

ATTAATCTTT ATAATGATGG CCAGTTAGAC GATACCCCAG TTACAGGGGA ATATGCTAAA 540

CAGTACAAAG ATAACAAAGA ATTTTCCACT ACTTTAATGG CTAATACCAT GTATTTAGAG 600

TTGAATCAAC GGGAAAAAAA TTCTATTCTA CAAAATAAAA ACGTTAGAAA AGCCATTAGT 660

TATGCAATAG ACAGAAGTAA TTATGCTAAA AACATTTTAG ATAATGGTTC TATTTCTGCT 720

GTTGGTGTTG TTGCTAAAGA CGTTGCTTTT GATCCTAGTA CAAAAAAAGA TTTTGCTAAC 780

AAAATGTTGG TGCATTTTGA TACAGAAAAA GCGCAATCCT ATTGGAATAA AGCGAAAAAA 840

GAATTAAATA TTAAAGAACA AGTAACTTTA AACATTTTAA CCAATGAAGA AGAAACAACC 900

AAAAAAGCAG CTGAATACAT TCAAGGACAA TTAGAAGAAA ATCTAAAAGG TTTaAAAATT 960

ACGATAACAC CAGTTCCTGc AAaTGTACAA aTAGAGCGAG TTATGAAaCA TGATTTTACT 1020

ATTAGTcTAA GTGGCTGGCA GGCAGATTAT CCTGACCCTA TGAGTTTTTT AGGtAmCTTT 1080

GrAAGTwACA GkGTGTTGAa TTTTGGAGGT tATAGCmATA CTAAaTATGA TGAATACTTA 1140

AAAGATACCT CAGATAAGCG TTGGAAATCT CTAAAAAAAG CAGAAGAATT ACTTCTTGAG 1200

GATGCAGGAG TAGTTCCATT TTTACAAGTA GGTAATTCTA AATTATTAAA GTCAACGATG 1260

GAGAAATATA GAAACGCATT CGATAGGTGC AAAATATGAT TACmAAAAGA TGAAAAGACT 1320

AGAGTAGTTA CTCTAGTCTT TTGGTTATTT TCTTTTCGTT TAGtTTTaTA wGaTTAtTTT 1380

CyTTAGtAAG tAAAGtAwCy GGtTTACyCa TAGaACCaTT ATTTTTTTTT GAGATCATTA 1440

GATAAATGtT AwCtTTCaTt AATTTTGcTA GCTCGACcAT TTcTAAGTCy CCcTAATAGA 1500

TCGACATATT TAATgAGTTA TTTAAAGAAT CCaTGTTGTT TGTTAAAAAG TAACCAACTA 1560

AATTATAAGT TATTAAATAA TTTAGTTTAT ACTTATATGA CGGGAATTTT TGCAATAAAT 1620

CATTGTAGAT ATTCAGAACT TCTGTACAAC TTTCAAAATC TCGGTTTTGA ACGAGTTTAG 1680

TAGTAATGTT TTCAAGTAAA TAGTATGCTG CTTCACTAGT AACTGTTGGA GCGTTTGAAT 1740

CT 1742

(2) INFORMATION FOR SEQ ID NO: 371:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1862 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 371:

TAGTATTTCC TGTAACATTA TTTAGAAATA AATTTATTTC TAAATTAGGC AAGCTTTTTT 60

AATACAACAT GATTAACAGT AATAATCACT TTCTTTTTGC GATATATTAT CATTAAAAGG 120

ATATAGATAA TTTAAAAAAT ACCCCCCATT GTCGTTCAAG AATTTTGGAG TTTCCTAGCC 180

ATTAGTTTTC TTATATTCAA TTATTTTTCA TTGTCCTATA TGAAATTTTG CAACTATTCT 240

AAGGCTATGA AAAAAACTAA CCATCCTTAA AGAATTTATA TCGGCATTAG TAATTCTAAT 300

ATTTATTTCT TAGAGCAAAA CCACTAATTG TTAGTTAAAC TCTTTTTTTA TTATACAATG 360

CTTTTACTGA TAAAAGTTCT ATCCTCATTT TAATAAAAGA TTATTTAAGT TTCTATCTAA 420

TGAATTTTAC GGATTTGCTA ACGTATTATT AATTGAGTAA CCTTTTTCAA AATTGTCCCT 480

AATGCTGTGA AACCAATACT AATTTCAACA CAAATAACTA AAACTTTAAT TATCTTCtTT 540

TTCATTAGCT TTTATCTCGr AAACAAAAAT TATAAAACGT TTTTTAACTT ACCCTAACTA 600

TTCTTATTnT CTAAATGTTA CAGAATAGTT ACGGTArGTk AAATATAGTA CTTTCTTTCT 660

ATTTCATTGC CTGTACGCTC TTTTTTATCT ATATATTATC CATCAAATAC ACATTCATAC 720

TATCTATCAT AATTTTTATA TTCTATTCgT TGTACAATAG TTATAAGAAA TATATTCAAA 780

CAAAACGCCT TTTCTTACTT AACTATTTTA ATTTTCTAGA AAGCATAATC TTTTGTACTA 840

TCTATTTACC TTAAATAATG CAATTCTATT TATCATGGAG ACAAAAATTT AAAATATTAC 900

TTTATGCTTT TTGACTAACA CTATTATATT AAATTGAtAA CGTTTACTCT TGTAATTGAT 960

AACAAATCTC TTTAGTAGTA GAATAAACTT TACTATATAC TTGAAAGAAT TTATTATATA 1020

ACAACACATT TTTTTCAATT GGTTGATATA CTTTTTTATA TTTCACAAAC ATTTTAATAC 1080

AACTATCAAT GTTTGCACAC ATACCACAAC CAATAACTGC TAATATTGCA GCTCCCAAGC 1140

TAGGACCCTG TTCCGTTTCC AATGTAGTGA TAGGAAGGCC AAAAATATCA GCCTGTATCT 1200

GAAGCCATTG TTGATTTTTC GTTCCTCCAC CTACAGAAAT AATTTTTTTT ATTTTCAAAT 1260

TTTTTCTTTG TTCCATGATT ACTTTAGATT CTTTTAATGA AAAAGTTATT CCTTCAATAA 1320

CTGATCGTGT AAAATGTTTT AGACTATGTC GACTATCGAT ACCAATAAAA CTCCCCCGAA 1380

TTTGACTATC TTGATACGGA GTTCTTTCTC CTCTTATATA GGGAGTAAAA ATCAAGCCCT 1440

CACTTCCTAC TATTACTTCA TCAATATTCG CTAACAAATC AGAAAATTCT ACATCTTTAG 1500

CAAATGTATC TCTAAACCAT TTTAAAGAAT TACCTGCTGC CAACGTTACT CCCATGGAGT 1560

AwTArGAATT AGGCAGAGCT GAATTAAATA AGTGGAGATT ACCTTGATAT TCGTTCTCAG 1620

CATCGTCTTC CATTGACAAA AAGACGCCAC TAGTACCAAT TGACAGCATT GCTACTTCAG 1680

AATTAACTAt CCCAGCTCCA ATAGsCGCAC mAGCATTATC AaGCTCCACC TGGCAAAAAT 1740

TCCACTTTGG TTCCAAGnCC AAACCTATTA ATCCACGGAC CTTCTCCATA TnACCACTTT 1800

AGCTGAAGnC CCAAATAAGG TTGGCAAAAC CCCCTAGGAA ATTAACTTTC AGGAATGTCT 1860

GG 1862

(2) INFORMATION FOR SEQ ID NO: 372:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6674 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 372:

TTATTTGATT TTACCGATGG TnAnTCTTCC CACATTTACT TGGCGGTGGG AATCAAATtG 60

TTTTAGCTTT AGGCAATCAA CCGACTACGA TTTGGGCTTT AGTCGGCTTA CTGGTTTTAC 120

GTTTTGTTTT TTCAATGGTT TCTTACGGTT CGAATCTCCC TGGTGGTATT TTCTTGCCTA 180

TTCTAACTTT AGGTGCCATT ATCGGAACAT TGTATGGCAG TCTACTTGTC CAATACGTAG 240

GGATGGATCC AATTTTTGTG AAAAACTTTT TAATTTTTGC CATGGCTGGT TATTTTACTG 300

CGATTGGTAA AGCACCGTTA ACAGCGATTA TTTTAGTAAC AGAAATGGTC GGCAATTTTT 360

CTCATTTAAT GTCTTTAGGG GtTGTCGCTT TAGTTAGTTA TATTACCATC GATAGTCTTG 420

GCGGCAAGCC AATTTnATGA AAGCTTATTA GAACGTTTGG TTCCTTCTAA AGTAAGTAAT 480

ATTCGTGGCC ACAAAACGAT TaTCGAATTA CCTATCACTG CAGAAAGTTC CTTAGATGAT 540

ACGATGGTAC GTGaTTTTGT CTGGCCTAAA CAAATGTTGC ycACTTCCAT TCGGCGTGGC 600

GAATCGGAAA TTTTAACACA TGGTGATACT GTTATGCATG TTGGCGATGT GCTCATTATT 660

TTGACCGATG AAAAACTCGC TTACCAAGTC aAAAAAGAGA TTTATCAAAA AACTTTACCG 720

AATGATTTAG TGAATTAAAG AAAGCtGCTG AACCTGAGTT CAGCAGCTTT CTTTTTTTCT 780

TATGACTGTT TTTTAATTTT TTTGAATGTT TTTGACCGTT TTAGATTGAT TAGCGTTTTC 840

ATTTGTGCTa TTATATCCaT GAAATGAGGG ATGAAAATGC TTACGGAAGA ACGACATCAA 900

AAAATTTTAC AATTGTTAGA TCAAAAATCA GTCGTCAAAT CGCAAGAACT TGCTAGCTTG 960

TTCAATGCTT CAGAATCCAC AATCCGCCGC GATTTACAAG AATTAGAAGA CGCCAATTTG 1020

TTAGAACGCA TTCACGGCGG CGCTAAGCGT AtTCTTAATT TAGGCTTTGA ACAAAACATG 1080

ACGGAAAAAT CAATCAAAAA CACGCACGAA AAGCAAAAGA TTGCTTCTTT AGCCGTTAGT 1140

TGTGTACAAG ATGGCGATAT GATCTATTTA GATGCTGGTT CAACCACGTT AGAAATGATC 1200

CCTTTTTTAG TTGGGAAACA AATCCATGTA GTGACGAATT CCGTGCATCA CGCTGCAAAA 1260

CTTAGCGATA TGGAAATTCC CACTATTATG CTTGGTGGTA CTTTGAAGCT CTCCACTAAA 1320

GCAATCATTG GTGCCACTAG TATGGAACAG CTACAACATT TTCGTTTTAA CAAAGCTTTT 1380

TTAGGAACTA ACGGCGCACA TTTAGAGTTT GGTTTAACGA CGCCTGATCC CGAAGAAGCA 1440

GCGCTAAAAG CCTTGGCTAT TCATCAAGCA GAAACAGCCT ATGTTTTACT GGATCATACA 1500

AAATTTAATG AAGTAACGTT TACCAAAGTG GAAGAATTAG AGTCAGTTAC TTTACTAACT 1560

GACTACTGCC CTGCTGAATC GCTTCAAGGA TTCCAGCAAA AAACTATTTT AATGGAGGCA 1620

AACAAATGAT TTACACAGTG ACTTTAAACC CTTCTATCGA CTTTATCGTT CATGTCGATC 1680

ATCTACAAAT TGGTGATTTA AATCGGATGA CGAATGATTT TAAATTACCT GGCGGCAAAG 1740

GAATCAACGT TTCACGTATT TTAAAACGTA TGGAGACAGA ATCCACTGCC TTAGGTTTTT 1800

TAGGTGGTTT TACAGGCTCA TTCATTGCTG ATTGGTTGAA AAAAGAAGAA ATCCAAACGA 1860

ATTTCACTCC CGTGTCTGCT GATACACGAA TCAACATTAA ATTGAAATCA GATACTGAAA 1920

CAGAAATTAA CGGCTTAGGA CCTGCTCTAA CAAACGAAGA AATTCAAGAA CTAAAACAAG 1980

CAGTTAGTCG TGTTCAGGCT GGTGATATTG TTGTTCTTTC AGGAAGTACA CCTGCTTCTC 2040

TACGCAAAGG ATTTTACGAA GAATTAATCC AAATTGTGAA GGAAAAAGGC GCAGAATTTG 2100

TTATTGATAC GACAGGTGAA GATTTAATGA ACGCTTTGTC ACAAAAGCCT TTATTGGTCA 2160

AGCCAAATAA TCATGAGTTG GCAGAATTAT ATCATACAAC CTTTACAAGT ATTGAAGATA 2220

TTTTACCTTA TGGACATCGG TTACTAGAAG AAGGAGCACA ACACGTAATT ATCTCAATGG 2280

CTGGTGATGG TGCATTACTT TTCACAACAG AAGGTGTTTA TCGTTCCAAC GTTTTAAAAC 2340

GTCCACTGAA AAATTCTGTC GGCGCTGGCG ACTCAATGAT TGCTGGTTTT ATTGGCAACT 2400

TTTCAAAAAC ACAAGATCCT TTAGAGGCTT TCAAATGGGG TGTCGCTTGC GGTAGCGCGA 2460

CCGCATTTTC TGATGATTTA GCCTCAGAAG ATTTTATCCA AGAACTTATT CACGAAGTAA 2520

CAATTGAAAA AATTTCAGAA TAAATCCACT AAGGGAGGAA AATAGAGTGA ACATTAATGA 2580

TTTATTAATA AAAGATGTCA TGATTATGGA TTTACAAGCT ACTGATAAAA AAGGCGCGAT 2640

TGATGAAATG GTCCAAAAAT TATACGATGG TGGCCGCATC TCTGACATTG ACACTTACAA 2700

AGAAGGAATT TTAAAACGCG AAGCACAAAC ATCGACTGGT TTAGGCGACG GCATCGCTAT 2760

GCCTCACGCA AAAAATAGTG CCGTAAAAGA ACCGACAGTT CTTTTTGCTA AAAGTAAAAA 2820

AGGTGTTGAT TATGAAGCAC TAGATGGACA ACCTACTTAT CTATTCTTCA TGATTGCTGC 2880

TCCAGAAGGC GCCAATGATA CACATCTCCA AGCGCTTGCG GCCCTTTCTC GATTACTTAT 2940

CGATCCAGAA TTTGTTGGCA AATTGAAAGA AGCGGAAACT CCAGAAGCTG TACAAGCACT 3000

TTTCCAAGCA GCGGAAGATC AAAAAGAAGC GGAAGAAAAA GCAGAACAAG AACAAACAAC 3060

GCCAGCACCA GAAAGCAATC GTAAATATAT TATTGCTGTG ACTGCTTGTC CTACAGGAAT 3120

TGCTCATACT TACATGGCCG AAGATGCGCT GAAGAAAAAA GCAAAAGAAA TGGGCGTAGA 3180

CATTAAAGTA GAAACAAACG GTTCAGAAGG AATTAAAAAT CGTTTAACTG CTGAAGATAT 3240

TGCTCGTGCT GATGGCGTGA TTGTCGCAGC CGATAAAAAA GTAGAAATGA ATCGATTTGA 3300

TGGTAAAAAA TTAGTCAATC GTCCCGTCAG TGATGGAATT CGTAAAACCG AAGAACTAAT 3360

TAACTTAGCC ATTAGTGGTG AAGCTCCTAC TTTCCATGGC AGTGACTCGG CTGCAAGCGA 3420

CCAAGAGGAT TCAGCGGAAG GTTCAATCGG TTCTCGTATC TATAAAGATT TAATGAATGG 3480

TGTTTCTCAT ATGTTACCTT TCGTTATTGG TGGCGGAATT GCCATTGCTC TTTCCTTTAT 3540

GATTGATCAA TTTATCGGGG TTCCTCAAGA TCAATTAGCC AATTTAGGGA ATTACAATGA 3600

CGTAGCTAGC TGGTTTAACC AAATTGGTGG CGCAGCATTT GGCTTTATGT TACCTGTCTT 3660

AGCTGGTTTT ATTGCTTCTA GTATCGGGGA TCGTCCAGGC TTAGTTGCTG GTTTCGCGGC 3720

TGGTGCATTA GCAAACGCCG GAGGAGCTGG GTTCCTTGGG GCCCTAATCG GTGGTTTCTT 3780

AGCTGGTTAT GTAGTGGTTC TTTTACGTAA AGTTTTCAAA GGCTTGCCAA AATCATTAGA 3840

TGGAATTAAA ACCATTTTAT TCTATCCAGT TTTTGGTTTG ATTATTACAG GCTTATTAAT 3900

GTTAGTCATT AACATCCCAA TGAAAGCAAT TAATGATGCG TTAAATCATT TCCTACTTGG 3960

CTTAGATGGC ACAAATGCTG CTTTACTTGG TGCCTTATTA GCTGGTATGA TGGCCATCGA 4020

CTTAGGTGGA CCAGTTAACA AAGCTGCTTA TGTTTTCGGT AcAGcTACGC TAGCAAGTAC 4080

TGTTGCAGAA GGTGGCAGCG TGGTCATGGC TTCTGTAATG GCTGGCGGAA TGGTTCCTCC 4140

ATTAGCAATT TTCGTAGCAA CTCGTTTATT TAAAAATAAA TTTACAAAAG ATCAACAAGA 4200

TGCCGGCTTA ACCAATATTG TAATGGGACT TTCTTTTGTC ACTGAAGGGG CAATTCCTTT 4260

TGCAGCAGCT GaCCCATTaC GTGTGATTCC AAGTTTCGTG GTAGGTTCTG CCTTTACTGG 4320

TGCGTTGGTT GGTGCGTTTG GTATCAAATT ATTGGCACCA CATGGAGGAA TCTTCGTCGT 4380

CTTCCTATTA AGTAATCCAT TACTATATCT ATTGTTCATC TTAATTGGGG CGATTATTTC 4440

AGGAATTGTC TACGGACTTT TGAAAAAACC TGTCGAAGTA CCTGCTTAAA AAAAGCAAAA 4500

AAGCGATTCA ACCATGGTTG AATCGCTTTT TTATTATTTT CTCTTTATCG CATAATTTTC 4560

TAAATACTCA GGATTTCCTT TGCTTCTGAT ATACGTGTAC ACGCCCATCA CAAAGGCCCC 4620

AAGGGTATTG GTAAATAAAT CCCCCATCGT ATCCATTAAA GCAGCACGCC CAATAAATGG 4680

TTGGCCGTTG GACATATTAT AGCGTTGTAA GTTCATGTTC CCTAATGAAT CGCAGATAAA 4740

TTCCCAGAAT TCCCAGAAGA CGCCACAAAG CCCTGCAAAT GCAAACCCAA ATAATAAAAA 4800

TAACCACGGG GAAACATCGG CGTAcTTCGT TTTCTTCAAT AGGAAACCAG CAATACCATA 4860

ACCCACAGCA GTTAACAGCA TCGGACTAAC AAAGTGTAAA ATTTTATCCC AGAACGAAAT 4920

GATTGAGATC ATATGCAACG AGGTTCCTAG GAATACTGAA ATAAATAAGA AAAACCAGTA 4980

AAAATAAACG ATAGTCGGTG GCATAATAAT TTTTAAAAAT TTATTGACAA GATCCGGTAA 5040

AAAGATTAGA CCAATTCCAA GTAAACATTC GATAACAAAC GTCACACCTT GTTTTGTTGA 5100

ATACTTAGCC GTAAAAAATT CCCAGAGATT GTACGTCAAG CATAAGATGC CAAACACCAC 5160

TAATCCATAT TTAATTTTCG AATAATTTTT ATCCaTTTTG ACACACCTTT CAAAAGTTAT 5220

TTTATATGAA AAAAACTGTA GATGAAAGAC ACCTTCCATC TACAGTTTTA GAAAACTCTA 5280

TCGAATGAAG AGCTTATTAA AAATGTTTAT TTTTCAACAG CTGTTTCAGT AATTGTTTCA 5340

ACAGCACGTT TCATACGAAT ATCGTGTTCT AACATGTCGT TGTTTAATAC TTTACGTACT 5400

TGCGCTTCAG GCATGTTGTA TTGTTCTGCT AAATCTTTCA CTTCAGCATC GATGTCTTCT 5460

TGTGTTACTT CGATGTTTTC AGCTTTGGCA ACTGCTTCAA TCACTAAGTT TGTTTTTGTG 5520

CGTGTTTCTG ctTCGCCTTC AAATTGTTTG TGTAAATCTT CTTCTGTAGA ACCAGTTAAT 5580

TGGTAGTACA TTTCAGGAGA AATTCCTTGG CGTTGCATGT TGTTTAAGAA TTCATCCATT 5640

GAACGGTGAA CTTCGTCGTG AACCATTACG TGTGGTAATT CAACGATTTC AGCATTTTCA 5700

ACAGCTAAAC GAATTGCTTC TTCGTCTTTT GCTTCTTCAG CTGCTGCTTC TTTTGCTTCT 5760

GTCAATTCTT TACGGAATTT TTCTTTTAAT TCGTCTAATG ATTCTACTGA ATCGTCAACG 5820

TCTTTTGCAA ATTCATCATC CAATTCTGGT AATTCTTTTG CTTTTACTTC ATGAACAGTT 5880

ACTTTGAAGA CTGCTTCTTT ACCAGCTAAA TCTTCTGCTT GGTAATCTTC AGGGAAAGTA 5940

ACGTTTACTT CTACTTCTTC GCCAGCTTTT TTGCCAACTA ATTGGTCTTC AAAGCCTGGA 6000

ATAAATGAAT TAGAACCTAG TTCTAATGAG TAATTTTCGC CTTTGCCGCC TTCAAATGCT 6060

TCGTCGCCTA AGAAGCCTTC AAAGTCGATA ACTACAGTAT CGCCATTTTC AGCAGCGGCA 6120

TCTTCTTTGA TTACTAATTC AGCTTGAGAT TCTTGTTCGC GTTTCAAACG TGCTTCCACG 6180

TCTTCATCTG TTACTTCACG GTCTTGTTTA GACACTGTTA AGTTTTTGTA TTCGCCTAAT 6240

TTTACTTCAG GTTTTACAGT TACTTCAGCA GTGATAACCC AGTCTTCCCC TTTATTCATG 6300

CTTTCAACAT CAATTTTTGG TTGTGCTACT GGTTCAATTC CTGCtTCTAA TACAGCCGCT 6360

TCATATGCTT CTGGTAAAAC AGCGTTTAAC GCATCTTCGT ATAACGCTTC TTCTCCATAC 6420

ATACGGTTAA AAACAGTACG TGATACTTTT CCTTTACGGA AACCAGGTAC GTTTAAGTTT 6480

TkTTTCACTT TATTGAACGC TtGCGTTAAT CCTTTTTCGA TTACTGCTtG TTCGATTGAA 6540

AAAGTTAATA CACCATCATT AGTGCCTTTT TTTTCCCATT TCGCAGACAT TTGTTTCCCT 6600

CCGAATAACA TTTAATTTaT ACaTCAAACT TwATwakGCA TACCTTAAAA TAGACACTAT 6660

CCCTGGnAAA TGTA 6674

(2) INFORMATION FOR SEQ ID NO: 373:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1630 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 373:

TAGCATTnAT GATTGTCTTT AACTACATTC CAATTTATGG AATTATCATT GCTTTTAAAA 60

ATTATACGAT TGTTGACACG GTTTCTTCTG CACCATGGGT TGGACTTGAA AATTTTCGAA 120

TCATTATGGA AGATAGTTTT TTTTGGGAAG CAGTTCGTAA TACCTTAGCG ATTAGTTTAA 180

TGAAACTTTT TTTAGGATTC GCTATTCCGA TTATCTTAGC GGTAATGATT TTTGAAATGA 240

GAGATGGACA TTTAAAAAAA GTGATTCAGA CAATTTCTTA TATTCCGCAC TTTTTTTCTT 300

GGATCGTTTT AGGCGGGATG TTGATTTCGT GGCTTTCAAC AAATGGCTTT ATCAACCAAG 360

TCATGATGTC GTTAGGCTTA ATGAATCAGG GCGTCAATCA TTTGTTAGAT CCTGATAAGT 420

ATTGGTGGAT TGCTGTTCTT TCTGATTTGT GGAAAGAAGT AGGTTGGGGC ACGATTCTCT 480

ATCTTGCTGG TATGTCTCGC ATTGATCCAA CGTTTTATGA AGCCGCGAGA ATTGaTGGTG 540

CTTCAAAATT AACACAAATC CGAACGATTA CTTTGCCCTT ATTAGCGCCG ATTATTTCGT 600

TGAACTTAAT TCTAAATGTT AGCGGTTTAT TGGGATCGAA CTTAGATCAA ACGTTGGTTT 660

TAATGAATGC TCAAAACCAA AACAAATCAG AAGTAATTAA TTCATTTGTT TATCGTATGG 720

GGynAACGCA AGGGGACTTT TCTTATGCGA CGGCAGTTGG TTTAGGGATT TCTGTTATTT 780

CAATTGTTTT ATTAGTCATT ACAGATCGCA TTACTAGAAA AATGAACAAT GGTCGGTCGG 840

TTATTTTATA AAGGAGGCAA CAGCTGTGAT TAAAAGAAAA TCAAAAGAAA ACAAATGGTT 900

TTACTTTTTC AATACGTTAA TTCTTGTGTT GTTTTCTTTA TTAATTATCG TGCCAATTTG 960

GAATATTTTA GTTTCCTCAG TTTCGTCAAG TTCGGGTTTA GCCGGTAAAG GCTTGGTTTT 1020

ATGGCCGAAA GGATTTACGC TAGAAAATTA TCGGAAAGTT TTTTCGGATA GTAGCATTCC 1080

GCGAGCGTTT TTGATTTCTG TTTTAAAAAC AGTTATTGGT GCGTCGACGC ATACACTGTT 1140

CTGTGCGATT GTCGCCTATG GACTTAGTAA ATCTAGACTA GTCGGGCGTA ACATTTATAC 1200

AACGATGGGT GTCATTACAA TGTACTTTGG TGGGGGCATG ATTCCCACGT ATCTTTTGAT 1260

TAAGTCATTA GGACTCTTAG ATACCTTCTG GGTATATATC ATTCCAGCTT TATTTAGCTA 1320

TTATGATGTA GTTATCTTAA TGAACTTTTT CCGAGAAATT CCACCATCTT TAGAAGAATC 1380

CGCCAAAATA GACGGCGCCA GTGAGTTTCA AATTTTTTAT AAAATCTTTT TACCTTTAAC 1440

AAAACCGGCA TTAGCCACCA TCATTTTATT TAATGGGGTT GGTCAATGGA ATGACTTTAT 1500

GACCACAAAA TTATATATTA CCAAAGAaTA TCTGTATCCA TTGCAGATGA AAATTtATGA 1560

AATTATTGTt CaATCAmACT TGAGTaCCAT GACAGAGAgC GGCAGTACAG ATATGGTGGT 1620

GCAGGCAACG 1630

(2) INFORMATION FOR SEQ ID NO: 374:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 885 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 374:

TCATCCATAA TTCCTTTTTC CGCTGGTTCT TTTCCTAATG CAAAACCAGG TAACCCATCC 60

GCGACCACGT TAACAAACAA CAGTTGAACT GCAGTAAATG GCGCACCCCA ACCAAGTAAC 120

ATGGCAATTA AAACAATAAA AATCTCAGAA ATATTACAGC TTAATAAGAA GTTAATGGCT 180

TTACGAATAT TTCGGTACAC GGCACGTCCT TGCGCCACTG CATCAACAAT AGTTGCAAAG 240

TTATCGTCCG TCAAAATCAT ATCTGAAGCA CCTTGGGCCA CATCTGTTCC AGTAATCCCC 300

ATTGCGCAAC CGACATCACT AGCTTTCAAG GCTGGCGCAT CATtcACGCC ATCTCCGGTC 360

ATTGCTACAA CAGCGCcAGA ACGTTGCCAT GATTGAACAA TTCGAATTTT ATCTTCAGGG 420

GTTACACGTG CATAAACGGA TAAATCTTTA ACACGTTTAT CTAATTCTTC ATCACTCATT 480

TTTTTCAGTT CTGAACCAGA TAAGGCTTCA CTTTTATCTT TTAAAATTCC TAATTCTTTA 540

GCAATTGCAC TGGCTGTTAC CACGTGATCA CCAGTAATCA TGACTGTTTT AATTCCAGCT 600

TTTTTCGCAC GAGCAATGGC GCCCTTACTT TCTGGACGTG GTGGATCAAT CATGCCAATT 660

AAGCCTAACA AGCGAAGATT TTTTTCTAAA GCCTCTGATG TAATTTCCTT GGGtGGTTCA 720

TCATATACGG CATAACCGAC AGCAATAACT CGTAACGCTC GTTTTCCGAA ACGATCATTA 780

ACAATAGCCG CTTGATCGAC ATCTCCGAAA CCAAAACGAG GTAGTAACAC ATCAAAGGCT 840

CCTTTAGTAA TTGrGrTGTA CTTTTTCCCC CATTGATGGA CCGTA 885

(2) INFORMATION FOR SEQ ID NO: 375:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7664 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 375:

GCGTGATTAT CTATACGTTG GTTCAATGTT ATTCGGTTTG TTTTTCGGTG CTGGAAATTT 60

AATTTTCCCT GTACATATGG GACAAGAAGC GGGGGCTTCC GTCTTTTTAG CAAACTTGGG 120

CTTCTTAGCC ACTGGAATTG GCTTACCTTT TCTTGGTGTC ATTGCTATGG GTGTTTCTAA 180

AAGTTCAGGT GTTTTTGACT TAGCTTCTCG GATTAATCGT CGTTATGCTC TTATTTTTAC 240

CTTACTATTA TATTTAACGA TTGGTCCTTT CTTTGCTTTA CCAAGGTTAG CAACTACTTC 300

TTTTGAGATT GGTCTTGCAC CATTTGTTCC GAAAGACTTG CAATGGCTGG TTCTTGCTGG 360

TTTTTCCGCA TTATTTTTCT TAGCTGTTTG GCTCTTCTCT AGAAAGCCAA GTAAAATCTT 420

AGATTATGTG GGTAAATTTT TAAATCCGTT ATTTTTAGCA TTGCTAGGAA TTTTGATTGT 480

TTTAGCCTTT GTAAAACCAT TAGGTCATGT GGCCAGTGCC GCTGTTGGTG AAAATTATAT 540

TTCAAGTCCT TTCTTTAAAG GATTTACAGA AGGTTATAAT ACGCTGGATG CTTTAGCTTC 600

ATTAGCTTTT GGGATTGTTG TCGTTTCCAC TTTACGAAAT ATGGGCGTCG AAAAACCAGG 660

AGATATCGCT AAAGGAATGA TTAAATCAGG TGCTTTCAGT GTTGTATTGA TGGGGATTAT 720

CTATACCTTA CTTGCTTATA TGGGAACAAT GAGTATCGGG GCTTTTCCAA TCAGTGAAAA 780

CGGCGGGATT GCCTTAGCCC AAATTGCCAA TTACTATTTA GGCCTACCAG GGAGTATTTT 840

ACTGGCTTTT ATCGTAGTTC TTGCTTGTTT AAAAACCGGT ATCGGCTTAA GCACAGCCTT 900

TTCTGAGACT TTTGAAGAAA TGTTTCCGAA AGTTTCCTAT CAAAAATTCT TGGCAATTGT 960

TGTTATTTTA CCAGCTATTT TTGCAAaTGT CGGCTTAACC AACATCATTC AATTTGCCGT 1020

TCCAGTCTTG ATGTTCCTTT ATCCATTGGC GATTACTTTA ATGCTTTTGG TAATCATCAG 1080

TCCTATTTTC AAACATCGAC GGGAAGTCTA TCAAGCAACG ACCTATATCA CATTGATTGC 1140

TGCTGTTTTA GATGCATTAA ATAGCGCGCC TGCCTTTATC AAAGATACGG CGTTTGTAAC 1200

AACGCTTTTA GAAAAAGCCG GCCAATACTT ACCATTGTTC ACTATTGGTA TGGGCTGGGT 1260

CGTTCCAGCA ACGGTTGCTT TTCTAATTAG TTGGGTTTGG GTCATGATTT CTAAAAAAGA 1320

TCCTATTCTT GATTAATTCT TACCAAAACG ACACTAGAAA AAATGCGGAA GCGCATCTTT 1380

TCTAGTGTCG TTTTTATGTA TATTCGTTGT ACCAAACTCT TGACTATGGT AAATTGAAGT 1440

TACCAACAAG TACTGGAGGA AATTGCCCAT GTTAGTTACA ACAACCGAGA GAATTAGCGG 1500

TCAAGAATAT GAAATTATTG GAGAAGTTTT CGGGTTAACC ACACGTTCTA AAAATATGTT 1560

TAAAGATTTA GGTGCTGGAC TAAAAAGCGT AGTTGGCGGT GAAATCAAAG GTTACACTGA 1620

TATGCAACGA GAAGCACGCG ACCAAGCAAT CGAACGTTTA AAAGCAGAAG CCTCTAAATT 1680

AGGTGCAGAT GCCGTTGTGA TGATGCGTTT TGATTCTGGC ACAATTGGCA CAGACATGCA 1740

ATCTGTCGTG GCTTACGGAA CAGCTGTTAA ATACATCTAA ATTTTGCAAC TAAAAAGACG 1800

CAGACCTTTT CTTGTAAGGT CTACGTCTTT TTAGTTTTCA TATTTCACTT CTGCTAAATA 1860

CAATCCTTCT GGATGAGCAG TCGGACCTGC TAAATTTCGA TCTTTGGCTG CGATAATTGC 1920

TGGAATCGCG GTTGGTTCCA GCCGCCCATT GCCGATTTTC AATAGCGTAC CAACTAAAAT 1980

CCGAATCATT TTGTACAAGA AACCATCGCC CCGAAAAGTG AATGTTAATT CATCCCCCGT 2040

TTCATTGACA TGAACACTGG CTTCATGAAT GGTTCGGACT TTATCTTCTA CCGAACTTCC 2100

TGAGGCACAG AAAGAGGTAA AATCATGGGT GCCTAATAAA TCAGGCAAAG CTTGTTGGAT 2160

TTTTTCGAGA TCAATGGGAT ACGGAAAATA ACTAGCATAA AAACGTCGAA ATGGACTTCG 2220

CGGTTTGCCA ATATCCACTT TAAATTGATA GGTTTTTTCG ATCACATGAT AACGTGCATG 2280

AAAGTCTTCT GAAACACGTT CCCACGTGCG CACAGCAATG TCTTCGGGTG TTTGTGTATC 2340

TAACGCAAAG CGCATTTTTT CCAACGGACG TTCCTGCGGA TAATCAAAGT GAATCACTTG 2400

```
GCCCaTTGCA TGAACACCTG CATCTGTTCG TCCTGAACCA AAAACCGTCA CGGAGCGTCC      2460
ATTATTCATG CGTGCTAACG TTTTTTCTAT TTCTTCTTGA ACGGTACGCT CATTAGGTTG      2520
CTTTTGAAAG CCATGAAAAT GAGTGCCATC ATAGGCAATG ATTGCTTTAT AGCGGGGCAT      2580
AATTTTTCTT CACCTTTTCT AAGTTCTTAA TACAAATAAG CCAATGGTTA ACAAGACAAA      2640
GCCAAACATT ACTATTGTAT CTCGTAGGTG CCAATGAAGA ATGCGGTATT TTGTCCGTCC      2700
TTCACCACCT TGATAACCTC TTGCTTCCAT GGCAGTCGCT AAGTCTTCCG CACGATTAAA      2760
ACTACTTACG AAAAGCGGAA TCAAAAGTGG AACAATAGCT TTCATTTTTT GAAGTAAGTT      2820
TCCTTCACCA AAATCAACAC CACGGGCACG TTGAGCATTC ATAATCTTTT CCGTTTCATC      2880
CATTAATGTT GGGACAAAAC GTAAGGCAAT TGATAACATC AATGAAACTT CATGAACTGG      2940
AAATTTCACT ACATGAAGTG GGCGTAGCAA ATATTCAATC GCATCTGATA ACTCTAATGG      3000
TGCCGTCGTT AAAGTTAATA ACGTTGACAT GAAAATAATT AAAACAAAAC GACAGAAAAT      3060
AAATAGCCCA TTCGTCACAC CAAACTGGGT TATGGAAATA ATTCCCCACT GAAAGTACAC      3120
TTCCCCCCCT TGTGTAAACA AGGTTTGAAG AGCCaCAGTA AACAGAATTA GCCaTAATAA      3180
TGGCCGAACA CCGCGGATAA AGAAGCGGGC GCTGATTTTT GATAATGAAA TAGCAAAAGC      3240
AGTAAAGGCT GCTATTAAAA CATAACTCTG CCAATTATTT GCTAAAAAGA TAATTCCAAT      3300
AAAATAAAAA CTAGCTAATA ATTTTGCACG CGGATCTAAA CGATGGATAA AAGAATCGCC      3360
TGGAATATAA CGCCCAAAAA TCAATTTATT CATCATTTTG CCTCACCTGC TTCCTGCATT      3420
TTTTTCAGAA GCTGATCGGC TAATTGATCG GCCGTTAATG GTAATTGCTC AAAAGAAAAG      3480
CCTTTTGCCA CTAATTTTTC AGCAAATTCG GCTGCAGTTG GCACGCCTAA TTGTTTTTCT      3540
TTCAGCCATT GTGTCTCTTG AAACACTTCT TGTGGTGCTC CCGCTCGGAC GATTTGTCCT      3600
TTTTCTAAAA CAATCACGTG ATCTGCGTAA TTAGCCACGT CATCCATCAA ATGTGTCACT      3660
AAAACGATGG TCATATTATG TTCCTTATGC AGACGACTAA ACATCTCCAT CATTTCTTTA      3720
CGACCCTTAG GATCCAAACC AGCAGTCGGT TCGTCTAATA CTAAGACTTC TGGTTCCATC      3780
GCTAGCACAC CAGCAATTGC GACACGGCGC ATTTGACCAC CTGATAACTC AAACGGAGAA      3840
TGTTGTAAGT ATTTTTCATC TAAGCCGACT AAATCAAGCA TTTTCTTAGC TAATTTTTTC      3900
GCTTCTTCGT CTGAAACACC AAAGTTTTTC GGTCCGAAAG CGATGTCTCG CTCCACTGTT      3960
TCTTCAAACA ATTGTGCTTC TGGAAATTGA AAAACGATCC CAACTTTTTT ACGAATTGGT      4020
TTTAAATTCT TATTATCAGT TTCTGGCGTA ATTACACGCT CTCCAATTGT CACTTGGCCT      4080
TTAGTTGGTT TCACCAAAGC ATTTAAATGT TGTAATAGCG TTGATTTGCC ACTGCCAGTA      4140
TGTCCGACAA TCGCTGTATA AGAGCCGTCT TGAATCGTTA AATTAATATC GAATAATGCC      4200
CGTTGCTCAA AAGGAGTATT GGGTTGGTAA GTAAAATCTA CTTGTTTAAA ACGGATGTCC      4260
ATAACCAATC CACCATCCCT TCTTCTGTTA AATAATTCGT TGGCACAACA ACGCCACGCT      4320
CTTTCAAAGC TACTTTTAAT TTTTCAGGAA ACGGTAAATC CAAGCCCATT TCAACAAGCG      4380
```

CTTCTCCTGC AGAAAAAATC TTTTCTGGTG TTCCTTCGTT TGTCAATTGT CCTTGGCGCA 4440

TCACTAAAAT ACGATTCGCA TTAGCCGCTT CATCAATATC GTGGGTAATA GAAATAACTG 4500

TTAAATTACT TTCTTCTTTA ATCTTTTGAA TGGTAGCAAT TACTTCTGCA CGACCTTCTG 4560

GATCTAACAT ACTTGTCGCT TCGTCTAAAA TAATAATATC TGGTCGCAAA GCAACCACAC 4620

CAGCGATAGC TACTCGTTGT TTTTGGCCGC CAGATAGCCG AGCAGGTTCT CGCTTGGCAA 4680

AATCTAACAT ACGAACCCGT TCTAATGCAT CGTGAACACG TTCGACCATT TCATCACGCG 4740

GAATACCTTG ATTTTCTAAA CCAAAAGCCA CATCGTCTTC GACTGTTGAA CCGACAAATT 4800

GGTTATCTGG ATTCTGAAAA ACCATTCCTA CCATACGCCG GATATCCCAG ACGTTTGCTT 4860

CATTTAATTC TTTGCCGCCA ACTTTGATAG TCCCAGCAGC TGGCAGTAAC AGCCCATTGA 4920

TCGTTTTGGC TAACGTTGAT TTACCAGAAC CATTATGGCC AATAATAGCG ACCCATTCGC 4980

CTTGTTGAAT GGAAAAAGAC ACATCTTTTA AAGCGGGAGA AGCATCTTCT GGTTGATAAT 5040

TAAATTGGAT ATTATTTAGT TCAATGATTG GTTGCATCGC ATACTCCCTC AATCTATTTA 5100

TTCTTTCTAA CACTACCAAA AAAGTAGCGT TTTTTCAATC ATCTGTTCTG ATTCTTCACA 5160

CTATTTTTTA TTATACCAAA AGAACGGTCA TTATGGGAGC TCTTATCTGT CTTGACAGGT 5220

CTATTTTTCA TAAAAATAAC TTCTTGTTCT CATAAAACAA GAAGTTACTG ATTTTCTTTT 5280

AAAAATGTGT CTTCAAATAT TCAAACGCAT ATTGGGCATA TAACTCAGCA CCCATTGCCA 5340

TCGCATCTTC ATCAATATTA AAACGACCAT GGTGATGTGC CCACTCCGTA TCTTTCTCTG 5400

GATTACCACT TCCCACTAAG GCAAAACAAC CAGAAGCATG TTCAGTATAA TAACTGAAGT 5460

CTTCGCCACC AGTTGTCGGC TCTTCTTGTC TTAAGGCCGC TTCACCAAAA TTTTCCTTAA 5520

TTAGTGTTTG GGCAAACAAG GCATCCTGCT CATCATTAAT CACTGGCAAC GTCCCATATT 5580

GATAATCTAA CGAAGCCGTC CCACCGTAAA TAGCGGCTGT TTGTTCCGCA TAGCGTTGTA 5640

GCGCCTGTTC TACTCGATTA CGCGTTGCTA CACTAAAGCA ACGAACCGTT CCTTCCAAAC 5700

GTGCATTCTC AGCAATCACA TTAAAACGGG TCCCCACATC CATCCGTCCA ATCGTGACAA 5760

CTACTGGATC TAAAGGATCT GTCTCCCGAG AAACAATCGT TTGCAGATTC ATCACGAAAG 5820

AAGAAGCTAT GACTGCCGCA TCAATACAAG CGTTTGGCAT GGCACCATGG CCACCACGGC 5880

CTTTAAAATC AACAGAAAAA ATATCTGCGG ATGCGAAACT TGAGCCCACT CGACACGAAG 5940

CGGTTCCTAC AGGCATTTGT GACCAGATAT GTAAGCCAAA TACATCATCA ACACCGGTCA 6000

TTGCGCCTTG CGCAACCATT GCTTTAGCAC CTTGCGCATT TTCTTCAGAT GGTTGAAAAA 6060

TTAACCTTAC TGTTCCCTGA AGTTCTTCCT GGATTTCTTT TAATACTTTA GCCACTGTTA 6120

CTAACATCGC GGTATGGGAA TCATGCCCAC ATGCATGCAT CTTGCCAGCT TCTAAAGATT 6180

TATAAGCTAA ATCTTCATTT AATTCCTGTA CAGGTAAGGC ATCCATATCC GCCCGCAGCG 6240

CAACCACTCG GCCTGGTTTG CCGCCCACAA TTTCAGCAAT TAAACCCGTC GGTTCTGTTT 6300

TGCGATAAGT AATTCCTAAT TGATCAAGAA CTGCCGCGAC TTTTTCTGTC GTTCGAAATT 6360

CTTCAAATTG TAATTCTGGA TGTTGGTGTA AATCCCGTCG AAAAGCAATC ATTTCTTGTG 6420

CATGCTGCTT GACTAGTGCT TTGATTTTTT TCATACTTAC GCCCCGTTTC TACAAACTTA 6480

ATAAAATACT TGAAAAAATA CCCGCTAAAA TAACCGAAAC AATGGTGACT GTTACGAAAC 6540

CTGCCACTAA CATCGGTGGC ATTAAATGGC TCGTTAAAGC CGCCTGTTCT TGTTCATCTT 6600

CTGTCAATGA TTGAATCACT TCATTAGTAA TAATATAATC GGCTGGAAAG CCATATAAAG 6660

CTGTCAACGA AACAGCGAAT GCCATCTCTT TACTCACACC TAATACACGA CCAACAATTG 6720

CTGAAAAGAT GTACATCCCG ATAACACCAA TCACAATAAT CCCAATCATT GGTAAAAATA 6780

AATCCATCAA CATTTCAGGT GTGGCTTGTT TTAAGCCATC AAAGACAAAA AGCATTAACC 6840

CTAAAATCGC AAAACCAAAG CTATTTGCTT TCTGCAATGG TTGCTTTTCT AAAAAGCCTA 6900

AAGACGTTGC AATCACGCCA AAACATAAAC ACAGCACAAA AGGACTAATC GTCACAAACG 6960

GTGCCGTCCA CACAGAAACA TAATAAGCAA GTGTGGCAAC TAAGGCTAAA CGTAAAATTC 7020

GTGAATAATC TGTATGATAT CGTTTGGGTA CTTTTTCAAA TAGTTTAGGC ACATCTGGTT 7080

CTGCTCCTGT TTCTTGCCCC TCTTGTTCAT GAACCGGCTG CCATGTCCCT GCTCGATATT 7140

CTTGAAGCTT CCGCTTTCCT TCTTTTTTTA GAACAATTGA AGTTAAGGGA TAACCAGCAA 7200

ACCCTTGAAT GACATAAATT AAAATTGCCA AAACCGAAAG TGAGGCTAAA CCAGCTGATT 7260

GAGCCGCTTC TGACATTACT AATGAAGAAA CTAAACCTCC TACTAACGGC GGGATTGCAA 7320

CTAAAACTGT TTTTAAGTCA AAAACAAAAG TCCCCACAGC TAACAACAAG GCAATAATGC 7380

CCAGAATACC TGACAAAGCA ATCACAATCG TTTTCCATTG ATTAACCAAT TCTTTAACAG 7440

AGAGCAGAGT GCCCATATTA GTAATCAACA GATACATTAA TAAAGTGGCA ACCACTGTAG 7500

GAATCCCTGC CACTTCCACA ATATTTTGTG GAAAAATAGT CCAATAACCA AGAATAAATA 7560

AAACACCACA AACAAAAACG GATGGAATCC AAGCTTTTGT TTTGATAGAA ACTACATCAC 7620

CAATGTATAA AATAACCATC AGCAACGAAA AAGCAAACAT TTGC 7664

(2) INFORMATION FOR SEQ ID NO: 376:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2344 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 376:

AAAAAAATTT TCAAAAGTCA AAAATGAAGG GGATTTTCAA GATGGAACAA TCATTAATTG 60

TGAGAGGGAC ACCACAAGAG TATCTTTGCG AAATTGGTTC GTGGGAATTA TTAGAAGGAC 120

AACTATTATG CAGAAATATT CATCGTGTAT TCATTTTACA CGGGACTGAG TCATGGCAAG 180

CGGCAAAACC GTATTTTCCG ACGTTTCAAA AAGTAACAGC TGTTTTTGAA AACTATGGAG 240

GTGTTTGTAC AGATCAACGT GTGCAAGAGT TAGAAGCGCA AGTGCTGGAC AATCATTTAG 300

```
AGGCCATCGT CGCTGTTGGT GGAGGAAAAG TGGCCGACTT GGGCAAAGCC TTAGCACACA     360
AAATGGCGTT ACCGGTAATT ATTTTGCCCA CTTTAGCAGC TACTTGTGCA CCGTGCACGC     420
CATTGAGTGT AATGTATCGA GAAGATGGAG CAATGGAACG ATATGACGTT TTTCCGCAAG     480
CAAACGCTTT GGTGTTGGTA GAACCAAGAG TGTTATTACA TTCGCCTCGG TCCTTAATGG     540
TGGCAGGTAT TGGGGATACA TTAGCTAAAT GGTACGAAGC AGATGCGATT ATTTCGCAGT     600
TAGATGTCCA AGTGTTACCT ATTCAAGTTT CTCATTTTGC CGCGGAAAAA TGCCGCGATA     660
TTTTGTTAAA TGAAAGTATC AATGCTTTAA AAGCGATGGA AGAGCAACAA TTAAATCAAT     720
CGTTTATTGA TGTCATCGAA ACAAATTTAA TTATTGGTGG AATGGTGGGT GGCTTTGGCG     780
ATGATTATGG TCGGACAGCA GGTGCsCATT CCATTCATGA TGCACTAACT TTGCTGCCAG     840
CTAGCCATCG ACAATTGCAT GGAAACAAAG TGGCGTATGG TGTGTTTGTT CAATTAGCCA     900
TTGAAGAAAA ATGGCAAGAA ATCGCTGAAT TGATTCCTTT CTATCACCAA CTAGGTTTGC     960
CTATTTCTTT AAAAGAAATG GACATGGACT TAACTGAAGC CGAGTATCAA GAAGTAGCAG    1020
AmCGTGCTTG TATAGAAGGC GAAACGATTC ATTATnTGAA ACAAAAAATT cCCtGAAATT    1080
GTGAAAcAGC GtGCAaGrTT TrGAAAAaTA TaCAGCGACA AAATAaGTGC CAAaCAGCAA    1140
CTAGTTCATT TGAATTATTT GAATTAGTTG CTGTTTTTTA AGGATCCTTT TTTGAAAAAA    1200
GAAACCTTGA GAAATCCTTG CAAATTCAAT AAAAAAAAAG TATGATGATG GGAGGACAAA    1260
TTATAGTAGA AAGTAGGGAA ATGAATGACG GAACCAGCCA TTCGTTATCG TTTAATCAAA    1320
AAAGATAAAC ACACAGGTGC GCGTTTAGGT GAATTAATTA CTCCTCACGG CACATTTCCA    1380
ACACCAATGT TTATGCCGGT AGGTACACTT GCGACAGTTA AAACAATGTC ACCAGAAGAA    1440
TTAAAAGAAA TGGGTGCCGG TGTTATTTTA AGTAATACTT ACCATTTATG GTTACGCCCA    1500
GGTGAAGATT TAGTCGAAGA AGCAGGCGGT TTACACAAAT TTATGAACTG GGATCAACCT    1560
ATTTTAACTG ATTCAGGTGG CTTCCAAGTC TTCTCATTAA GTGATATGCG GAAAATCGAA    1620
GAAGAAGGTG TGCATTTTAG AAATCATTTA AATGGTTCaA AAATGTTCCT TTCTCCAGAA    1680
AAAGCGATCA ATATTCAAAA TAAATTAGGA TCAGATATTA TGATGAGCTT TGATGAATGT    1740
CCACCATTTG ATGAAAGCTA TGAGTATGTC AAACGTTCGA TTGAACGGAC TTCTCGTTGG    1800
GCTGAGCGTG GCTTAAAAGC CCATGCGAAT CCTGATAGAC AGGGATTATT CGGTATTGTT    1860
CAAGGTGCGG GATTTGAAGA TTTACGTCGT CAAAGTGCTA AAGATTTAGT GAGTATGGAT    1920
TTTCCTGGCT ATTCGATTGG TGGCTTATCT GTTGGTGAGT CTAAGGAAGA AATGAACCGA    1980
GTCTTAGACT TTACGACACC GTTAATTCCT GAAAATAAAC CACGTTACTT AATGGGTGTG    2040
GGTGCCCCTG ATTCATTAAT TGATGGAGTG ATTCGTGGGA TTGACATGTT TGATTGTGTG    2100
CTACCAACAC GTATTGCGCG CAATGGTACT TGTATGACTT CACAAGGTCG CTTGGTTGTT    2160
AAAAATGCCC AATATGCAAG AGaTTTnCGT CCGCTTGATG AAAAATGTGA TTGTwAkGTT    2220
TGTCGTAATT ATACGCGGGC CTATATTCGT CATTTGATTA AATGTGATGA AACATTCGGA    2280
```

ATCCGTTTAA CATCGTATCA TAATTTATAC TTCTTGTTAG ACTTAATGAA AAATGTCCGT 2340

CAAG 2344

(2) INFORMATION FOR SEQ ID NO: 377:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2424 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 377:

TAAGTTTTGT TTCAAATTCT TCTACATATT CTCGATACAT AAAAACACAT CTTGAAATAC 60

TCGGTCTAAG TTCCTTAACC CTCTCTTCCA ATTTTGTACA TTCTCTTTTT ACATAAGAAA 120

CAGTTAATTC GTTACTGTCT TCAAGTTCTT TTACACTCAA TTGAAGTTTA TGAATAGAAC 180

TCATCACATG AATCATTTTC TCAATAGTGG CATCCATAAG TTTTATATTT TTTTTGATTA 240

ATTCTTTATC CATTCTATCA ATCCACCTTT CTAAAACGGG TACATTTCTT CTAATTCGAA 300

CTCTCGTTCC ATATAGGACC GAACACTTTT ATTATCTCTA TTGGTAAACA ACAAGTCTTT 360

TGcATCaACa CTCaCCATTT CaATAATCTC TTCaTGCTTT TCTTTTnAAT TCTGCTAAAT 420

CAATGCGTTC CATCAATTCA TTTGTTCCTA ACAGATTATC AATCTCTTTA ATCATCTTTA 480

ATGAAGGGGC TAAATAATTT GATAGCCATC GAAGCGTCTT TTCAATAGAC GGTTTTTCAC 540

CTTGCACAGA CAATTTTAAG CCTTCCAATT TATCAACAAC ATCATGCCAA GATTTACAAT 600

AAACTTTCAT ATCATCTTCA ATGTCATAGA CAGTCAAAGA TTGATTAATA ATTTCTTTCA 660

CTAACCAATC TAGCGACTCA CCGCTTGATA GAAAATATTC TACAAACAAA AACGCCTTTT 720

CATCAGAAAG GCGAATTTCG TATCGATTTT TAATTTCATG TTTTTCTCTT GCTTCTTCAA 780

CACTCATACT TTCTTTTCTT GCTAACTCGT AATCTTTTTG GTAAAAGTTA AAATACAGTG 840

GCGATTGACG ACTACCAAAA TAAAGGGATA AACCTTTATT CACCATTTTT TTATTTTCGT 900

AAACAAATCC ACCGCTAAAA TCAAAATTTT TAAATGTGGT GTCAACAAGC CCCTGCTCCA 960

TTTTTTCTTT TAAAACATAC AAATCAAAAT TTTCTTGCCC CTTCACAATC AATTCATCAA 1020

GAGCAATATC AATTCGAGTG ATTTTTGTAT CTACTAAAAT GCCTTGACCG AAAATATCGT 1080

CATCATAAAG AGAGCGAAAA AATTCTGTCC aCGATCGGTk GTTATCTTCC TCAAAGATTT 1140

CTTCTAATTG ACGACAGCCT tCGCCAGACA AGTCCaACAT AATTCCCATA TTGACATTTT 1200

CCGGATTCCG AAAAATCCGA ATAGAGCTAT ATGAAAATGT TTCCGTGTAA TGATAAAACC 1260

CTCTAGATTC ATGAGTAAAC CAGTTCATAT CCATGTGCAA CAACTTTTCA AyAACTGTCC 1320

GAACATCTAA TGTTTTAAAT CGTATTCTTA AAAAGTCAAT TTTTGCTTCT ACCGAAGCGT 1380

GCCCACCATC TAAGTTCAAC ACATTAATAA TCTTTTGTTC TAACTCTTTA GAAATAACAC 1440

GTTTTCCAGA TTCTATTTTT GAAATTAACG AACGCTCrAT TCCTAATTTA CTTGCCAATT 1500

CTTGCTGTTT TAATTTTAGy TCTTTTCTAT ATCGCTTTAA ATCAACACCT CTCACTGCTT 1560

TAAACCTACT TCCAAAAAAt GTGACAACTT TAGTCGAAAA ATGTGACAAA ATCATAATCG 1620

CTCAATCCCT TGTAGGACAT AGCTTTTAAG CAACTTTTTT ATACGGTATA TTAAATTCTA 1680

CCCCCATGTT ATAAATAGGG GGTTTTtGAC GTAAGGTGGG CATTCGCCCA CCTTACGTCT 1740

GTTCATTTTA AAAGGATTTT CCGTTGCTAA AAGTATACGT GGGCTTTCGC TCTGCTTCGC 1800

CCACGTATAC TTATTACTAG CAACAATCCT TCAAATAAAC AGCCGTTATT CAGCTGTTCC 1860

TCGTTCTAAT GCACCAATTC TCTCTTTTAA GCTTGCATTC TCCGCAAGAA TAGATTGAGA 1920

CACCTTTAAA AGTTGCCTCA TATCTCTTTT CAAATCCTCA ATTTCTTGTT TATTAATATG 1980

TGTAGCTTTA TTCTCATCAG ACATAAATAC TTCATTTATT GCCTCATTAA GCGGAACATT 2040

ATTTTCACGC AAACGTAGCA ACTCTTGAAA ATTTTCTATA TCTTTAATAG AAAAAACCTT 2100

ATAGCTATAA CTATGACCTG CACTATTCTT TGCATATCGG CGTTCAAAAC GTGTATCAGT 2160

CATTTTTTCA ATCCTATTTG CCCAATTTTT CAGTGTTTTC AAAGCAATAG CATTGTCTAG 2220

CTCTAAAATT ACTTGATCAA ATGAATACAT TTTCCACACC TACTTATTAA ACATTTCTTT 2280

CCAAGCAGAa tCCATATGCT CCTTAAATCT TGTTTCTAAA TACTTATCAA CTTCTTGTGA 2340

TGTTGAAAAT TTCTTTTCCT CAATTGGTCT ATCTTCCTCG TAAACAGCTT CTGCATCTAA 2400

AACAGAAATT ACTAATAATT TATA 2424

(2) INFORMATION FOR SEQ ID NO: 378:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1468 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 378:

GACGCTCTTT TACCAATCAA AGAGTAAGGC ATAGCCTAAA GAAAGCAGCT CTCGACCAAA 60

ATTTTTGATA TCTTTTGGAA AAGAAGAAGC CGCAGTGATG CCTGAAACAT TTGAGATGCC 120

TGCTCGCTTG GcTAAAAGTT GCGAGCGGTA CATGTGAAAG TCACTAGTAA CAATCACTAG 180

CCGTTTAAAA GAAAGGAGTT TTTGTGAATT GACAATATTT TCTTTGGTGC GTGTGGCACG 240

ATTTTCTAAG AGAATCTGTT TTTCTGGAAC ACCTTTATTT AGCAAGTATT GTTTCATAAC 300

CGCGGCTTCC GTAGCCGGCT CATCTGCTCC TTGCCCTCCG CTAACAATAA CGGGTGTCTC 360

GGGGTGTTGT TTAATATAGG GGATAGCCAC ATCTAACCGA GCTTGCAAAA CACGGCTGGG 420

CAGTGCGGTT TCTTTAGACG TCCCTAAAAC ACGGGCCCCT AAAATTAAAA TAGCATCTGG 480

TTTATCGGTG GGTTGCGTAT TTTTGCCAGT AAAAATAAGT AGACCTAAAA TCAAAAGAAT 540

ACAAAAGCCA AAACTGACAA TGGATAAACT TGTTATTTTT AACCAATGCA TGAAACATCC 600

CTCCTTGTAA TGTTAGTGTA TACGAAGTAC CATAGAAAAA GCGATAGAAA ACAAAAGACT 660

TATAAAAAAA TCTATGTTTT AATAGAAACT TTTGAAGAGA GGGAGCGAAT GCATGCGTAC 720

AGAAGAAGAA ATGTTTCAAC TTATTATGGA TGTAGCAAAA CAAGAAGAAC ACATTCGAGC 780

TGTGGGAATG GTGGGGTCTC GCACGAATGT GAAAGCGCCA AAAGATAGTT TTCAAGATTT 840

CGACATTGTT TATATTGTAG AACCATGTGC TGAGTTTTTT GAAACAGCAA CGTGGATAGC 900

AAAATTTGGA CAACCATTAA TTATGCAACG TCCCAAGGAA ATGACCTTAT TTCCCACAGA 960

ACCGAAGACG CGGGAAACAT TTTTAATGTT ATTTGAAGAT GGCCAACGTA TCGATTTAAC 1020

GTTGTGTCCT CTCGCAGAAA AAGATAATTG GCATGAAGGC GATTCTTTAG CAATTATCTT 1080

GTTAGACAAA GATGAAAATT TGCCGCCTTT ACCTGTAGCA TCTGATAAAA ATTATACAGT 1140

AACAGTTCCG GATCAGCAGC AATTTAACGA TTGTTGCAAT GAATTTTGGT GGGTTAGTAC 1200

GTATGTAGTT AAGGGGCTTT GTCGGAATGA ACTATTTTAC GCTGTCACAC ATCTCTATGA 1260

ATATTGTCAG CAGGAACTGT TGCGCTTGCT TTCTTGGCAA GCGGCTTGGC AGGAACCGGA 1320

GCCAATTTCT GTTGGGAAAC AATTTAAGTA TCTGAAAAAT TATGTCACTC CTGACACGAT 1380

GGATCAATTA GCATCGTTGc TGGaTTTTTC ctAGTaAAGg AgCcTkGTTG GaACaGCtTA 1440

ATwAAgrCcC CAAGCCTTTT TTGGACGT 1468

(2) INFORMATION FOR SEQ ID NO: 379:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13085 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 379:

AGTTTAGGnC ATmCGTAAAC ATCTTCTCCT CCTTTATTCG TTCATTTTTT TATAGATAAC 60

GTAnGGGAmA AnGGATACGA AgCTTACTCA CTTTCGTATC AAAAAAgCAw TCTTCACTCA 120

AGTGCTACAC GCTTTTTTGG TTTTAAAaTG ACATTTAAAA GGACTGCTAC CAAAGTGGTT 180

AAGATAACTT CCGAATTACC GAAAATCGTT GTTACCCAAG TTGGAAAACC AGCTAAGCTT 240

CCAGTAGACA ATGTGACACC AATTCCAAAT GCTAGAGCTG TTCCGACAAC CCCAGTATTT 300

CGAGGAGTCA TTTCCTGCGA AGCAATCATC CGAATACCTG TCATAGAAAT AGAAGCAAAA 360

ACTGAAATCG TTGCGCCACC AATTACCGCG TAAGGAATCG TAGTTAATAA TGCAGCTACT 420

TTCGGAACAA AACCAGCAAT TAATAATATA ACTGAAGCGA AAACCAGAAC ATATTTATTA 480

ATCACTTTTG TTACCGTAAC CAGACCAACA TTTTGTCCAA AAGTTGCCAC TGGGATTGAT 540

CCAAATAAAC TGCCAATAAA GTTAGTGAAA CCACTCCCCA TAATTCCGCC AGATAATTCT 600

TCATCCGTTG CATCCCGATC CATTGCCCCA ACCGTTGTTG CTGTAAATTG ACCGATGGCT 660

TGAACAGCAT CAACAATAAA CATCACCACA AGCGTAAAGA TTGGTACCAA TTGAAAATCC 720

AGTCCAAAGT GAAACGnGGT AATTACCTGA ACGATTTTAG CATTTTGAAC TGGCTCAAAA 780

CTAACCATCC CTAAAGCTAA AGAAATCAAA TAGCCAATAA CCATTCCGTT TAAAATTGCG 840

GATAACTTCA AAAATCCTTT TGCAAAATAA TTGAAGTAAA AGACAATCGC AAAGGTTAAT 900

AATGCAACTA ACCAATTCCG GGCTGAACCA AAATCAGCAC TACCAGCTCC TCCTGCCATA 960

TACTTAATCG CAACAGGAAA CAAGGATAGT CCAATGCTTA AAATAACCGT TCCTGTTACT 1020

AAAGGCGGAA ACAACACACG GATTTTTTTA ATGaACAAAC CAACAATAAT AACAAGAATA 1080

CTGCCAACTA ATTGGGAACC AAAAATTGCT GCAATCCCAA AATCAGCTCC GATGGCCATC 1140

AAAATCGGTA CATACGCAAA ACTTGCCCCC ATCaTAACTG GTAACCGTGA ACCCACTTTT 1200

CCAAAAATAG GGAAAAGTTG AATTAACGTT GCAACTGCTG AAAAAATCAA TGATGTTTGA 1260

ATCAAAATAG TCGTATCACT AGGACTTAAT TGGCAAACTT TGGCAATCAT GATTCCCGGT 1320

GTGACAATCC CTACTACTGC TGCGACAACG TGTTGCAATC CCATCGGAAT TGCTTCAGAA 1380

ATTGATAGTT TCGCATCATA TTCAAATAAT GCTTGCTCTT TTTTTTCATT CATAGGATCG 1440

TTTCCTTCTT CTTTTTATTC TAAAAGGATT GTTCCCCTTA ATCAATAAAG AATGCAATCC 1500

TTTTTCCCTC AGAAACATCT TAAAAATTAT CCTTTGACAG CAACTGCTTC AATTTCAAAT 1560

AAACCATCTT TTGGCAACTT GGCTACTTCT ACACAAGAAC GTGCAGGCAA AACTTCTGAA 1620

AAGTAATTCC CATAGATTTC GTTAATTTTG CTAAAATCAG ACATATGTTG TAAAAACACC 1680

GTTGTTTTCA CAACATTATC ATAGGTCATC TCGACCTCTT TTAAAATCGA ACCTAAATTG 1740

ATAAAAGCTT GTTTTGCTTG TTCTTCCACT GTAGTCTTCA TTTCGCCCGA TTGCGGATCT 1800

AATCCTAGTT GTCCAGAAAT GTACAACGTA TTTCCCGCTA AGACAGAATG TGAATAAGGT 1860

CCTACAGTAG CTGGTGCCTG TGCAGAATTA ATCATTTTGT TTGTCATCGT TTTCCTCCAT 1920

TCTATGATTC TTTATTTTCA ATTAAAGTCC CTGTTTTTCC AGATAATCCT TCTTTGGCTT 1980

TTTCTAATAA CGTAATCAAT GTTTTTCGAC CTGGCTTAGA TTCAGCAAAC TTAATCGCTG 2040

CTTCAACTTT TGGTAACATT GAACCTGGAG CAAACTGACC TTCTTGCGCA TATTGTTTCA 2100

TTTTTTCTGT TGAAACATTC CCTAAGGCTT CTTGATTTTC TTTACCAAAA TTAATACAAA 2160

CTTTTTCAAC TGCTGTTAAA ATCACGAGCA GATCAGCATC CACTTGTTCA GCCAGTCGTT 2220

CACTACAAAA ATCTTTGTCG ATGACTGCAT TGACACCTTT TAGTCGGTTC CCTTCTTGAA 2280

TGACTGGGAT GCCACCACCG CCGCCAGCAA TGACTGTTTG GCCAGCATCA ACCAATGTTT 2340

GAATCGTGAG AAGTTCGACA ATATTTTTTG GTTTAGGAGA AGCGACGACT TGGCGATAAC 2400

CACGGCCAGC ATCTTCCATC ACTTGTATAT TTTTTTCTTT GACTAATTGT TCTGCTTCTT 2460

CTTTTGTCAT GAAGCGACCA ATTGGTTTCG TTGGTTTTTT GAACGCCGGA TCATTTTCAT 2520

CCACGACCAC TTGAGTGACT ACGGTGGTGA CGGCTTTACT TATATTTCGT TGACGCAACT 2580

CTTCTTGTAG CGCATTCTCT AAATCATAGC CAATGTAACT TTGACTCATG gCTACGCACA 2640

TGGATAACGG AACAACTGGA TATTGTGGAT TATTAAAGGA AAATTCTTCC ATAGCTAATT 2700

GAATCATCCC CACTTGAGGA CCATTGCCAT GAGATAAAAT AACTTCATAT CCTGCTTCAA 2760

TCAAATCTGC AATCGCTTTG GAGGTCTGTT TTACGGCAGT CATCTGCTCC GTCAAATTAT 2820

TTCCTAAAGC ATTGCCACCT AAAGCTACTA CGACGCGTTT GCCCATTGCT TCACCTCATT 2880

TTCTGTTTTA TAAAAAGGAC GATGCTTTTT CTTTTTCACA ACGAGTTTTC GTATTAACTG 2940

TCTGTTGTCG TTGCATGAAG TCAACGCACA TGACTAAGCG TTCTTGACTC GAGACAATCA 3000

CACATTCATG AATAGACTTG GGAAGGTCTT ATTTGTTCAT GTTGAAGAGG CAACTTCTAA 3060

TGATTCATTG TGTAAAGAAA AAGCATCTGT CTTCATGCCA AATTCATTAT TTTACTTCTC 3120

TTGGTGTCGC TTTATCCGCT AATTCTGTTA AGGTTTTTTG CGGATTTTTT ACTTTACTTA 3180

AGAAAATCAT AGCCGCAATA ACATACGGTT TGTAGCTAGC TTCTTTGTAA AGTTCGACCC 3240

GATAACGATC GAAAACAGAA GCTTCCACTT CGCCTTCTTC ACAACTAACG CCAGTAATAT 3300

CAGCTGGTAG ACAATGCATA TACAAAGCTT TGCCGTCTTT TGTTGTTTTC ATCAATTCTT 3360

CTGTACATTC CCAATCTTTG TGCTTTTTAT TTTGAGAAAG TAATTCTTGT TCTAATTGAT 3420

CAATTCCTGC TTGGTCACCA TTGCCATATA ATTCAGTTCG CTTTTCCATT GCCGCAAAAG 3480

GTGCCCAGCT TTTCGGATAA ACAACATCCG CATCTTTGAA TGCTTCTGCC ATACTATTCG 3540

TTTTAGTAAA GTTTCCACCA AATTCTGCCG CATTTTTCTT CGCAACTTCT TCTACTTCTG 3600

GCATAATTTC ATAACCTTCT GGATGTGCTA AGACGACATC CATTCCTAAA CGGGTCATTA 3660

GACCAACAAT TCCTTGAGGA ACAGATAATG GTTTTCCATA AGATGGTGAA TACGCCCAAG 3720

TCATGGCAAC TTTTTTCCCT TTCAGGTTTT CAATGCCACC AAATTCATGA ATTAAATGCA 3780

ACGCATCCGC CATTGCTTGC GTCGGATGGT CAATATCACA TTGTAAATTC ACTAACGTTG 3840

GGCGCTGTTC CAATACCCCA TCTTTATATC CTTCTTGGAC AGATTCTGAT ACTTCGTGCA 3900

TGTAAGCATT GCCTTTACCA ATGTACATGT CATCACGAAT ACCAATGATA TCAGCCATAA 3960

AAGAAATCAT GTTTGCTGTC TCACGAACAG TTTCTCCGTG AGAAATTTGA CTTTTTCCTT 4020

CGTCTAAATC TTGAACTTCT AAGCCCAAAA GATTACACGC TGAAGCGAAA CTAAAACGTG 4080

TTCTTGTAGA GTTGTCGCGG AATAAAGAAA TCCCCAAGCC ACTATCAAAG ATTTTAGTTG 4140

AAATATTATT TTCTCTTAAA TAGCGTAATG TATCGGCAAC TGTGAAAACA GCTTCTAATT 4200

CATCTCTTGT TTTTTCCCAA GTTAGGAAAA AGTCATTTTC GTACATTTTT TCAAATTCTA 4260

ATTTATCTAA TTTCTCAATA TATTCTTTAA ACGTTTCCAT TCCATCTGCT CCTTTTTTCT 4320

TGTCTAATAA TCCTTTACTT AACCACAATA GACCGTTGGT AATGCCGCAT AAACGGCTGC 4380

ACAACGAACC AAATCATCTT TCCAAGTTTT TTCATTTGGT GCATGTGCTT GTGCTTCGGC 4440

CCCTGGCCCG AAACCAATAC AAGGAATACC ATTACGTCCC ATAATTGAAA CGCCATTTGT 4500

TGAGAAGGTC CACTTATCTA ACAATGGGCG TTCTTCACGC ATTTCCACTG TTTCTACAGA 4560

GCCTTTCCGT GTTTCTCCAT AGAGATTTTT ATGCGTTTCC ATCAAGGCTT TTGTTACATC 4620

ATGATTTTCT GGAATTACCC ACGTTGGGAA ATAACATTCA ATTTCATAAG TTAAATCCGT 4680

ATAAGAAGGA CGATCATAGT TGTACATTGA AACCGTCACA TCATCTCCAT ATTTTTTAAC 4740

AGCTGGTAAA TTGCGAATTT CTTCTAAACA ACTTTCCCAA GTTTCTCCAG CAGTCATTCG 4800

ACGGTCTAGT GAAACGGTAC AGCCATCCGC AACAGCACAC CGACTTGGTG AAGAATGGAA 4860

AATTTGCGAA ACAGTCACCG TACCACGTCC TAAGAAACGG GCTTCTTGCC ACTCTGGATT 4920

GTACTTAGGA TCTAACATAC GAACTAAACC ACGAATCGCT GTGCTTTCTG TATCTCCATT 4980

GTTGTTTAAG GCACGAACAT CTTGTAAGAT ATCCGCCATT TTGTAAATGG CATTATCTCC 5040

ACGTTCTGGT GCTGAGCCAT GACACGAAAC GCCTTTGACA TCAACTTTAA TTTCCATTCG 5100

GCCTCGTTGA CCACGATAAA TCCCACCATC TGTTGGCTCT GTTGAAACAA CAAATTCTGG 5160

ACGGATACCG TCTTCTTTGA TAATGTACTG CCAACATAAG CCGTCACAGT CTTCTTCTTG 5220

GACAGTAACT GTTACTAAAG CAGTGTATTT TTCTGATAAT AGACCTAAAT CTTTCATGAT 5280

TTTAGCGCCA TAAATAGCCG AAACAATCCC GCCCTCTTGG TCAGACGTAC CGCGACCACC 5340

GATCTCAGTT TCTGTTTCAT AGCCATCGTA AGGATCAAAT TTCCAGTTAC TCATTTCACC 5400

AATTCCGACA GTATCCATGT GCCCATCAAA AGCAATTAAT TTTTCACCAG ACCCCATATA 5460

ACCAAGTAAA TTTCCTTGAG GATCAATGTC AATTTTATCG AAACCTAATT TTTCCATTTC 5520

GGCTTTTGCT CGCGCAATTT TGTTGCCTTC TTCTGCACTT TCTCCTGGAA TTTTGACAAG 5580

GTCTCTTAAA AATTTGACCA TATCTTCTCT GTAACCTTCT GCTGCTGCAT TTACTGCATT 5640

AAAATCCATT CCTTTTACCT CCATAAAATT TTTTTATAAT TGTCTGGATC AGTGTcCCCT 5700

TCGGTAGAAA ACATTAAAAC GTTTGACGTC TCATCCAGTT GCAATGCCTC ATGTAATTCT 5760

TGATAGTCGG GATCTTGCAT CACTGTAGCT ACTAAGCCCA TTGCTACTGC CCCAGATTCG 5820

CCTGATACAA CTTGCGGATC CCCTTTTAAT GGCGCGCCCA ACATACGCAT TCCTTTTTCA 5880

GAAACCCAAT CTGGCGCTGA AACAAAGAAG GATGTATGAT TTTTTAGAAT ATCAAAAGAA 5940

ATCGTATTAG GTTCTCCACA TGcTAAGCCA GCCATAATTG TCTGTAAATC ACCATCGACA 6000

AAGCGAATTT CTCCGTCTTT TTTAATTGCT GATTGATATA AACAGTCAGC CGCCTGTGCC 6060

TCAACTACGA CCATAATCGG TGGGTTTTCT GGAAAAGCAT TAGCAAAATA ACCAATCACT 6120

GCACCAGCTA AACTGCCGAC ACCAGCTTGA ACAAAAACAT GACTAGGTCG TTCAGAGCCA 6180

AACTTTCTTA ATTGTTGAGA GGCTTCTAAC GCCATCGTTC CGTATCCTTG CATAATCCAA 6240

GTCGGAATTT TTTCATATCC GTCCCAAGCG GTGTCTTGAA CCATGACACC GTTTTCAGTT 6300

TCTTCGGCCA TTTTATTAGC CATTCGCACA CATTCATCGT AATTAACTTC TTCGATTGTG 6360

ACGGTTGCGC CTTCTTTTTC AATATTTTCT TTTCTCGTCT GTGTTGAACC cTTCGGCATG 6420

AGTACAACTG ATTTTTGGCC TAATTTATTC GCAGCCCAAG CGACTCCTCG TCCATGGTTA 6480

CCATCAGTTG CTGTAAAAAA CGTAGCTTGT CCAAAAGCTG ATCGTAATTG ATCAGAAGTT 6540

AATTCATCAT ATGTTAGTTC CGAAACATCT TTCTTTAATT TTTCGGCAAT ATAGTTAGCC 6600

ATGGCAAACG AACCACCTAA TACTTTAAAA GCATTTAAAC CAAAGCGATA AGATTCATCT 6660

TTAACAAAGA AATCATTTAA CCCCAGATAC GCGGCCATAT GATTCAATTC TGCTAATGGT 6720

GTTTCTGAAT ACTGTGGAAA ACTTTTGTGA AAAGCCAAAG CTTTCTCAAT TTTTTCTTTA 6780

GACATCAGAG GTAAGTACTT ATCCTCCGTC TTCGGCATTT CATTTGCTGT CCATTTAATT 6840
TTTTCCAAAA ATACTTCTCC TCCTTCTATT TTTAAGTTGC TTTACGAAAA ACAAATTTTC 6900
CTTGATTTGC CAAGATGACT TGCCCAGCTT CAGCAACTTT TTGACCTCTT AAATAAACCA 6960
TTCGAGCCTG TCCTTGTGTT TCAAATCCTT CATACGGCGT ATAATCGACA TTTTGTAACt 7020
GCGTTCTCGC AGCAATCACG CCGGTTGTTC TGGGATCCCA AACAACTAAA TCGGCATCAC 7080
TGCCTTCTTG CACCACACCT TTTTGTGGAT ACATCGAGAA TTGTTTAGCA ATATTTTCAG 7140
AAAGCAGAGC AACCATTTTT TCCAAGGTGA TTCGGCCCTT CGCCACACCT TCCGTGTAAA 7200
TTAATTCTGG TCTCGTTTCA ACACCTGGCA TCCCATTGGG AATTTTACTA AAATCATCTT 7260
TGCCAACTGT TTTTTGTCCG TAAAAATTAA AACTGCAGTG ATCTGTCGAT ATTGTATTGA 7320
TTGCCCCCTC TTTAATCCCT TGCCACAAAG CCCGTTGATC CTCTCGTTTG CGTAGTGGCG 7380
GTGAGCAGAC ATATTTTGCT GCTTCAAAAT TAGGCGCATC ATACAAGTGA TCATCTAGTA 7440
AAAGATATTG AGGGCAAGTT TCTACATAGA CTGATTGTCC TCTTTGACGA GCCCGTTCCA 7500
CCGCTTCTAA TGATCGTTTT GTGCTTAAAT GCACAATATT AACCGGCAAA TCAGCCATTT 7560
CCGCTATCAT CAAATATCTT GCTACCGCTT CTGCTTCAAC AGCCGCCGGT CTTGATAACG 7620
GATGATAGTG CGGGGTTAGC TTTCCTTGAG AAACATATGA TTGGaTTAAT TCATCTACTA 7680
AATCACCATT TTCACAATGA ACGCCTAACA TTCCATTAAC TTTTTTTATT TCCTTCATCG 7740
CTTCAAAAAT TTCCGCATCT GTGGTCCGCA AATTATCATA AGCCATATAC ATTTTGAAGG 7800
AAGTAATACC TGCAGCAATC ATTTCCTGAA TTTCAGCAGC TATCGTTGGT TTCCATTCAA 7860
TCATGGACAT ATGGTAAGTA TAATCACATG AACTCTTGCC TTCAGCCAAT TGATTCCAAG 7920
TAGCTAAACA GTCTTTCAAT GAGCCGCCTT TATTAGGTGT AGCCATATCA ATAACCGTAG 7980
TCGTACCTTT GGCCACCGCC GCTTGGCTTC CAGTGGTAAA GTTATCCGCC GTACTTAACG 8040
AACCTTTGCC ATTATTTAAT TCCAGATGCG TATGCGCATC AATaAAGCCT GGTAATACAT 8100
AACACCCCGT AACATCCTCA ATCTGAGAAT TTTCTACTGG TAAATTTGCC CCCATCTCTA 8160
CAATTTGCTC GCCGTCAATT CGAATATCCA ATTGACGACG ATTATATGCA GATACGACGG 8220
TTCCACCTTT GAGCAGAATA GACATTTATT TATCCCCCAG TCGTAAATTT CCTTTTATGC 8280
TTGCATACAA TAAGCATATT CTTCAATCAT AGTTACCATT AATGCAATTA AATCTTCTGG 8340
TACATCTAAT CCTTGATCTG TTAATTGAAT TGTTGACACC GTTCCCCATA GACGCAGTAA 8400
ACAAATTTTA TCCACAGGAT TTTGAACATA GAAACCTGAA TTAGTACTGT CATAAAAATC 8460
AGCTTCACTG TTAAATAAGG TAAATTTGTC TTTGTAAGGG GCACTTGCAT AAGGGCAAAA 8520
TGTTTCACAG TTGCCACATT CGTTACACAT CCGATCAACA TGGACAATTT GTGGTTTTCC 8580
TTCGACATAC ACAACAATAT TTGCTCGGTT TGGACAGACA TCCATACAAG ATTCACAAAT 8640
TGTGGAGCAT TCTAAGCAAC GACTAGCTTG ACTACAAGAC ATCTCGTCTG TCACAAGGAT 8700
ACCCCGTTTG TTTCTAACAA ATGCGACATC ACTATTTAAA TTATCTTTTT CATAGTGATG 8760

ATTGTGGACA AGACAAATAT TATTGGCCGC TTTGGTTGCA TCAGCGATTG CTTCAACAAT 8820

TGTGGCTGGT CCTAAATTAG CGTCACCAAT GACATAGACA CCTGGTATAT TTGTTTCCAA 8880

TGTTTCTTGA TTTGAAACGA CTTTCCCATA GTTGTCTGTG TGAATGCCTA ATGCTTGGTA 8940

AAATTCTGTA TCTACTTTTT CACCAACAGC GGCAATCACT GTATCAGCAG GAATATCAAT 9000

CATTTCTCCT GTACCAATTG GTCGACGGCG TCCAGAGGCA TCTCGTTCAC CTAGGACCAT 9060

TTTTTCACAA GTTAATTGTT GATTTTCATG TTTAATTGGC GAAAGTAATT CTAAAAAGTC 9120

AACACCATCT TCTAAAGCTA AATAAAGTTC TTCTTCATCT GCTGGCATAT TGCGTTTATT 9180

CCGACGATAA ACCACTGAGA CTTTTTTAAC ACCAGGTAAT GTTGTGGCTG CACGAGCACA 9240

GTCCATGGCT GTATTTCCGC CACCAACGAC GACAATTTGT TCTCCGTAAG GATTAATCGT 9300

TGGATTTTCT CGATTTGCTT TTAAAAATTC TAACGAGTTC AATGCTCGAC CAGATTCTAA 9360

ACGAAGGACC CCATGTTTCC ATGCACCAAT GGCATAAATA ACATTCGTAT AGCCTTGGTT 9420

TTTCAATTCC GCCAATGATG GCGCCTCTTG TCCTGTACGG AATTCGGCTC CCATAAATTC 9480

CGCTAACTGA ACGTCTTTTT GAACAGATTC CATAGAAATT CGAAATTCTG GTACAATCTG 9540

GCTACAAACG CCCCCAATTG TCTCGGATTT TTCAAACACA GTAACAGGCA TGCCTTCACG 9600

TGCTAATAAA TACCCAGCAG AGATCCCTGC AGGGCCTCCA CCCACAACAG CTGTTTTAGG 9660

TGCATTTTCT TTTGGCTGTG GCTTTTCTAG TGTCGTCAAT AATTCATCAT AAGCATGTTC 9720

AGCTGCTTCT AACTTCACTT CCCGAATATG AATAGATTCT TCATAAAATT GACGTGTACA 9780

TTTAGTCATA CAAGGATGTG CACAAATGGT TCCAGTAATA AAAGGCAACG GATTCTTATC 9840

GACAATTACT TGCAAGGCCT TTAAATAATT ACCTTCACTG ACATAACGAA GGTAAGCAGG 9900

AATATCCTGA TTGATTGGAC AACCACCGTC ACTTCGACAA GGAGCAATAT AACAATCTGT 9960

TAACGGAACT GTTTTTCTTA GTTTGGTACT TTCTGGTAAT TTAATTGATT TTTGGTAACG 10020

CGCTTGTGTT TTAGCTTTCT CAACAACTTG AGCTAATTTA TCCAAGTTGA CATGGACCAT 10080

CTGAGGATAT TCTGCTGCGC TTAATACATT AGCAACTTGA TTCATTCGTT GATAGCCACC 10140

AGGTTTCAAC AGAGTCGTTG CCATAGTGAT TGGCCAAATC CCAGCATCGA ATATTTCTTT 10200

ACTATTGAAA ATATCTGCAC CGCCAGAATA AGAAATTTGT AGTTTCCCAT CAAATGCTTC 10260

TGATAATTTT TGTGCCAATG AAATACTAAG AGGAAATAGC GAACGTCCTG ACATGTACAT 10320

TTCATCACCA GGTAATTCAT TTGCTGCAAT CGTTACCGGG AAAGTGTTCG TAATTTTAAC 10380

TCCGAAGCTT AGATTTTTAC TATTAGCTAA TAATTGCAGA CGTTGTAACA TCGGTACTGC 10440

TTCTTCAAAT TGTAAATCTT CTTTAAAATG GTGATCGTCA AAAACCATAT AATCAAACCC 10500

TAGCTCATCC ATCGTTTGAC GCGCATATTC ATAACCTAAC ATGGTTGGAT TACATTTAAT 10560

AAAGGAATGG AGCCCTTTTT CACTTAATAA ATAGGCCGCA ATTCGTTCAA TTTCGTCTGA 10620

TGGACAACCG TGTAAAGTCG ACAGCGTAAT AGAATGACAG ACTTTAGGAC TAATTGCTTC 10680

GATATATAAG TCATCTACTT TTTTAAAATA AGATAGATAT TTTTTGGCAG CTGCTTGGCA 10740

TTCCGCCCAA ATCGGCGTTC CTTCTGCGTT CTGCATTTCG TTAATATAAC GATCAATTTT 10800

TGGTGATTGA ATCCCTGCCA AATCATAGCC AACACTCATA TTAAAAATGA AACCATTCGG 10860

GTCACCTAGC TCGAACTCTT TACTAAGTAA CTTCAAAACG AACCATGCTT TGACGTATTC 10920

ATCATAAGCT TGGGGAACGC GCAATTCTGT GGACCACTCA ACGTTGTAAC ATTCATCTGC 10980

TGCTGCGATA CATGGTTTAC TCACAGGTAA ATCTTCTCCA TCCAGTATCT GGACTGTCTT 11040

AACTTCAAAA AAACGAGAAC CTGTTAAATA GGCCGCAATA ATATTTTGTG CTAATTGTGT 11100

ATGAGGACCT GCTGCTGGAC CACAAGGGGT TTCCATTTTT TCTCCAAATA ACGAAATCGT 11160

TTTTGTAGGA TCCGCATGAT AAAATTTACG AATGCCAAAA ATTGTGCCAT CTTGTTGATA 11220

CTCACTAAAA ATCCAATTTA ATAAAGCCTC AATTGAGATG GGATGCATAA TATCACTCAT 11280

TAACGTCACT CCTTTTATTT ATAATGAAAG CTATTGATTT ACTCGTTTCC AGAAAGATTG 11340

TGCTTGCGTT TGCGCATCAT GCCAAACTTT CTCTTCATCA ATTCCTTGAA CTTCACGATT 11400

ACGCATCCGA ATTTCTCCAT TAATCACCGT GTCAGTGACC ATCCCACCAT TCATGCCGAA 11460

TAAAATGTGC ATGTTATAAT TCTCTTTTGT CATTGGTGTC GGGCCATGAT AATCAACAAT 11520

AATGACGTCT GCTGCTGCAT TAGGTTCTAA AACGCCTAGC TTTGTTTTAA AATAGCGATT 11580

AGCCATTTGC GGATTATTAT CAAACAGCAT TGTGGGAATT TCTGCCCAAG CCGCATTTGG 11640

ATCAGCTAAA TGATGTTTGT GGATAACATT TCCTACCTTA TAGGATTCCG TCACATCATT 11700

CGTATAACCA TCTGTGCCAA GACCCATCAA AtACCATACT CATTAAACAT GCGCATCGCA 11760

GGTGGGCAAC CGACCGCATT CCCCATATTC GACTCTGGAT TCGTGACCAC CATTGTGTCT 11820

GTATCTCTTA GTAATTCCAT CTCATGAGGT CCAATATGGA TGCAATGGCC TGCCATTGTC 11880

TGTTTTCCTA AAATCCCTAA ATCAAATAAT CGGTTCACAA TCGGCTTCCC ATATTTTTTC 11940

AAAGAATCAT GAACATCGGC TAAATCCTCT GCAATATGAA TATGATAACC AATCCCATCT 12000

GGCGTATTGG CTGCACATAA TTCTAAAGAA GCATCTGATA AAGTAAATGC TGCATGCATC 12060

CCCATCATCG CTTTTTGCAT ATCATCATTT CTTAAGGAAC TGGCTTTAAT AAAGGCGGCA 12120

TTTTCTTGAA CTGCTTCCCG CATTTTCTGC TCACCATCGC GATCAGAAAC TTCATAACAA 12180

AGGCAGGTCC GAATCCCTAA CTCGTCCGCT GCATTAGATA ATTGGGTTAA ACTCCCTGTA 12240

ATTTCTCCGT AACTAGCATG ATGATCAAAG ACAGTTGTAA CGCCGTTGCG AATACTATCT 12300

AAATAAGCAA CCTTACCACT TTGATAAGTG TCTTCTAATG TCAAGGCACG ATCAATTCGC 12360

CACCATTGAT CTACCAAAAT ATCCATGAAA TTTTTAGGAT GATAATTCGT TAAACTTAAC 12420

CCACGTGCAA AGGTACTATA AATATGATTG TGCATGTTAA TAAAGCCTGG CATAATCACT 12480

CCACCTTTAG CATCAATAAA TGTCGCTTCT GGATAAGCTT GTTTTAATTC TTCTGTTGTT 12540

CCGACTTGTT TCACAACTTG ACCATCAATC GCCACACACC CATTTTCAAA AAAATGATTC 12600

TGCGCATCCC GAGTGATTAA ACGCCCATTT CCAATTAATA ACATGTATTC TTCTCCTATT 12660

TTCCAACCTA TTCCCCACGT TTAGCAAACG CTTACAAATC AATCGCTCAA TCATTCACGA 12720

ATATAGCTTG TCTAGCTTTT CTAATCTATA TTTCCTACGA AAAGCATTTG AAGCGCTTGC 12780

ATTTTTTTAC TAACGAACTC ATTAACTTGA TTAAAGTATA TCACCTTACC TTAAAAAAAA 12840

CAATGCTACC TAAAAAATAT TTTTTAGAAT TTTGACAGCC GTTATCTTGT ATTATTATTT 12900

TTATCATGCT ATCATTATTT TTCAGAACAG GTAGTACGCG ACATTAAAGG ATTTAGCCCT 12960

GCCTATTCCA TTTAAAAATA TTATGTTAAA TTTGACACAA TCTTATAAAA ACTGTTTGTG 13020

CCAGGCACAT TAGAAAGGAG CTTATTTATG TTAACCGACG AAAAACTTAC TTTTTTAACA 13080

CAAAT 13085

(2) INFORMATION FOR SEQ ID NO: 380:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1121 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 380:

TTTTGGTATG AAACAGAAAA AGTGGTTAAT CGGACTTGTT GCACTGGGCT TGGTTTTAGC 60

aGCATGTGGA AGTGGCGGTT CGAAAACGAC CTCAAACGAA CCAGCTACAC AGAAAATTAA 120

CGTCGCATCT GGTGGTGAAC TCTCGACATT AGACAGCGCT CATTATACAG ATGTCTATAG 180

TTCCGATATG ATTGGTCAAG TAGTTGAAGG CTTGTATCGA CAAGATAAAA ACGGAGATCC 240

TGAGCTAGCT ATGGCGAAAG CAGAGCCACA AGTTAGTGAA GACGGGTTAG TCTATACATT 300

CAAGTTACGA GAAGCAAAAT GGACAAACGG GGATCCAGTT AAAGCAGGGG ATTTTGTAGT 360

TGCGTTTAGA AACGTGGTCG ATCCAGCATA CGGTTCAAGT AGCAGTAATC AAATGGATAT 420

TTTTAAAAAT GGGCGTGCGG TGCGGGAAGG ACAAGCCACG ATGGAAGAAT TTGGTGTCAA 480

AGCAATCGAT GACCAGACAC TAGAACTAAC ATTGGaAAAT CCAATTCCTT ATTTAGCCCA 540

AGTCTTGGTT GGGACACCTT TTATGCCTAA AAATGAAGCC TTTGCCAAAG AAAAaGGTAC 600

TGCCTATGGG ACTTCTGCAG ATAATTTTGT TGGCAATGGG CCGTTTGTAA TTTCAGGTTG 660

GGATGGCAAT TCCGAAACTT GGaAATTGAA GAAGAATGAT CaTTATTGGG ATAAAgAACA 720

CGTAAAATTG AATGAAATTg ATGtTCaAGT AGtGaAAgAA ATTGGCACAG GAGCCAATCT 780

TTTTGATAAT GGCGACTTAG ATTACACTGT TTTAGCAGAT ACTTATGCAC TTCAGTATAA 840

AGAGTCAAAA CAAGCGCATT TTGTACCTAA AGCCATGGTG GGTTATTTAA GCCCCAATCA 900

TCGCCGTGAA ATTACCGGCA ACGAACATGT TCGAAAAGCT TTTTTACAAG CGATTGACAA 960

AGAAACTTTT GCAAAAGAAA TTTTAGGAGA tGGCTCGACA GCTTTAAATG GkTTTGTACC 1020

AGCTAATTTT GCaAAAATCC AGATACAGGT GAAGATTTCC GCAAAGAAAA TGGTGATTTA 1080

TTGCCATATA ATATTAAAGA AGCCCAAGCT AACTGGAACA A 1121

(2) INFORMATION FOR SEQ ID NO: 381:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1278 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 381:

nTCATTGAAT GCnCGGAACA ACGGAAGTAT GGATGGCGAC CCACCAGCGG CGATGCGTTA 60

TACAGAAGCT CGTTTATCAC AGTTAAGTGG TGAGTTGTTA AAAGATATCG ACAAAAATAC 120

CGTCGATTTC GTCTGGAATT TTGACGATAC AGAAAAAGAA CCCACTGTTT TACCAGCCAA 180

ATATCCTAAT TTATTGGTAA ACGGCTCAAC GGGGATTTCA GCAGGCTATG CAACAGAAAT 240

TCCTACGCAT AATTTAGCGG AAATCATTGA TGGAACTGTG TATTTAATTG ATCATCCTAA 300

TGCAAGTTTA GAGAAATTAA TGGAATACAT TCCAGGACCA GATTTCCCAA CAGGTGGGAT 360

TCTACAAGGA AAAGCAGAAA TTAAAAAAGC CTACGAAACC GGCCGTGGTA AAGTAATTTT 420

ACGTGCGAAG ACAAAGATTG AGCCTTTAAA AGGCGGCAAA CAACAAATTG TCATTTCTGA 480

AATTCCTTAT GAAGTCAACA AAGCAACGCT AGTGAAGAAG ATGGATGAAA TCCGTTTAAA 540

TAAAAAAGTA GACGGCATTG CCGAAGTTCG TGACGAAAGT GACAGAACTG GCTTGCAAAT 600

TGTTGTTGAA CTAAAAAAAG ATGTCAATGC AGAAGGTATT TTAAATTATC TATTTAAAAA 660

CACTGAATTG CAAATCAATT ACAACTTTAA CATGGTAGCC ATCGACAACA TGACCCCGCA 720

ACAAGTGGGC CTAAAACGGA TTTTAGAAAG TTACATTACG CATCGCAAAT CAGTTATCAT 780

TAACCGTTGT CAATTTGAGT TGGACAAAGC TCGCAAACGT GAGCACATTG TTGCTGGTTT 840

AATTAAAGCA TTGTCTATCT TAGATAAAGT GATTGCAACA ATCCGTGGCA GTAAAGATAA 900

AAAAgATGCG AAGAAGAACT TAGTATCCGA CTATGCGTTT ACTGAAGAAC AAGCTGAAGC 960

AATCGTTACT TTACAGCTAT ACCGTTTAAC CAATACGGAT ATTACTGATT TACAAGAAGA 1020

AGCGAAAACT TTAGAACAAC AAATTGCTGA GTTATTGAAC ATTTTAAACA ATGAAAAAGA 1080

ACTATTCTCA GTCATGAAAA AAGAActTCG CGAAGTTAAA AAGCAATATG GCAATCCGCG 1140

CTTAACTCAA ATTGAAGAGG AAATCCAAGA AATCAAGATT GAAACAGCCG TGTTAGTTGC 1200

GCAGGAAGAC GTGGTCGTAA ACnTGACGCA CGAAGGCTAT ATCAAGCGGA GTAGTATTCG 1260

TTCTTATACA GCATCAAA 1278

(2) INFORMATION FOR SEQ ID NO: 382:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4184 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 382:

ACTTTTTGCT TAAATTTCGT GATGTTTTGC ACTGACTATT CTTCCTCTTT TTGGTAACGA 60

TAAATGATTC GTCCATTTAA CCCCGCTTCT CCAACAATTT CCGCGCGGTC AGGTAAAGAA 120

```
ATCCAAGCAT CACGAAATTG GAAAAATTCA TTTTGCTCCA CGATTAATTC AGAGATTTCT     180
CCAGCTTTTA ATGAAGCAAT TTTTTCTACA AAAGGATTTT CAGCCATTGA TTTCACCACT     240
CTTTCTCTTA AATATCAACT CTTATAGTTT ACTTGAAAAT GGCGCAATTG TCTGTAAAAA     300
AATCGAATCT TTGCCTTTTT TTAAACAAAG TCTTTTACTT GATTCTCTTA CAAAAAATGC     360
TATAATAACC TCAAAGGTCA AAAATGGTCA AGTGGTTTTA TAGACAGAAT TTATTCCTGC     420
TATCTAGTAC TTGGCTGGTT TGAGGAAAAA TGGTCTGTTT ATTAGAAATA ATAGAGCAAA     480
CAACTATTTT TCTGGCCTCA GTTATTTTGA AAGGATGTGT TTCATTCAAT GATTTGTCAA     540
AATTGTCAAC AAAATGAAGC AACAATTCAT TTATATGCAA ATGTGAATGG TCAAAGAAAA     600
CAACTGGACT ATTGCCAAAG CTGCTATCAA AAATTAAAAA ATCAAGCAAA TAATTCCCCT     660
CAAAATATGG GACAAGACCC TTTTGGCTTT GGGAGTTTAG ACGATTTATA TCGTTCTCTT     720
TCTCGCCAAA TGCAACAAGG TAATCCTTAT GAACAACAAA CACCACCCAC TCAATTTGGT     780
GATGGTGGTA ATGGAGGTCA ACCCCCTCGC GGAGGCGCTG GCGCTGGTCA AGGTTTACTA     840
GGCGAGTATG GGATTAACAT CACCGAAGCG GCACGCCAAG GAGATATCGA TCCTGTGGTA     900
GGACGTGACC AAGAAATTAA GCGTGTCATT GAAATCTTAA ATCGCCGTAC AAAAAACAAC     960
CCCGTTTTAA TCGGTGAACC AGGTGTTGGT AAAACAGCCG TTGTCGAAGG TTTAGCCCAA    1020
AAAATTGTCG ATGGCGATGT GCCCCAAAAA TTATTGGACA AAGAAGTCAT TCGTTTAGAT    1080
GTGGTTTCAC TTGTTCAAGG AACTGGTATT CGCGGACAAT TTGAAGAACG GATGCAAAAA    1140
TTGATTGAAG AAATCACTGA AGCTGAAAAT GTCATTTTAT TTATTGATGA AGTCCATGAA    1200
ATTGTGGGTG CTGGTGCTGC TGGTGATGGC AATATGGATG CTGGCAATAT TyTAAAACCA    1260
GCCTTAGCAC GTGGTGAATT ACAATTAGTC GGTGCCACAA CTTTAAATGA GTACCGTATT    1320
ATTGAAAAAG ATGCTGCCtT AGAACGTCGG ATGCAACCAG TCCAAGTAGA TGAGCCAACC    1380
GTTGCCGAAA CCATAACGAT TTTACATGGT TTGCAAAAAC GTTATGAAGA TTATCATCAT    1440
GTTAAATATA CAGATGAAGC GATTAACGCA GCAGCAAACT TATCCAATCG TTACATTCAA    1500
GATCGCTTTT TACCAGATAA AGCGATTGAC TTGTTAGATG AATCTGGTTC AAAAATGAAT    1560
CTAACTATCC AACTCGTCGA TcCAAAAACA ATTGATAAAA AATTAGCAGA AGCGGAACAA    1620
CAAAAACAAC aAGCTTCCGC AGAAGAAGAT TTTGAAAAAG CGGCTTATTA TCGTGATCAA    1680
ATCAATAAAT TACAAGCAAT GAAAGAAAAA CAAATCAGCG ATGAAGAAAC ACCAGTCATC    1740
ACTGAAAAAG ATATTGAAGC CATTGTGGAA CAAAAAACTG GCATTCCTGT CGGTGACTTA    1800
AAAGAAAAAG AACAAACCCA ATTGAAAAAT TTAGCTGTTG ACTTAAAAGC TCATGTTGTC    1860
GGACAAGATG ATGCCGTCGA TAAAGTAGCT AAAGCGATTC GTCGCAATCG TGTAGGCCTA    1920
GGCAAACAAA ACCGCCCAAT TGGTTCCTTC TTATTTGTTG GGCCAACAGG CGTTGGTAAA    1980
ACAGAATTAG CAAAACAATT AGCTTTTGAG CTATTCGGTT CAGAAGATTC GATGGTTCGT    2040
TTTGATATGA GTGAGTATAT GGAAAAACAC AGTGTTTCTA AACTGATTGG TTCTCCTCCA    2100
```

GGCTATGTTG GTTATGACGA AGCTGGACAA CTCACAGAAA AAGTCCGTCG TAATCCTTAT 2160

AGTTTAATTT TATTGGATGA AATTGAAAAA GCGCATCCTG ATGTTTTACA CATGTTTTTA 2220

CAAATTCTCG ACGATGGTCG TTTAACAGAT GCACAAGGTC GTACAGTCAG CTTTAAAGAC 2280

ACCATCATTA TCATGACAAG TAATGCTGGA ACAGGAGCTG TGGAAGCCAA CGTTGGCTTC 2340

GGTGCTGCCC GTGAAGGTGT GACTAAATCT GTTTTAGGTC AATTGAACAA CTTCTTCACC 2400

CCCGAATTCC TTAACCGTTT CGATGGTATT ATCGAATTCA AAGCATTAAG CAAAGAAAAC 2460

CTGATGAACA TTGTCAGCTT GATGTTAGAA GAAGTAAATG GCTTATTAGC TAAACAAAAA 2520

TTACACATTG AGGTTCCAAC AGAAGTCAAA GAGAAACTCG TTGATTTAGG TTATGATCCT 2580

GCAATGGGCG CTCGTCCTTT ACGTCGAACA ATTCAAGAAC AAATTGAAGA CGGCATTGCT 2640

GAATACTATC TAGATCATCC TGAAAATCAT CAATTAGTCG CAGCGTTAGA TAACGAAGGT 2700

AAAATTATCG TAACTGGCGC ACAAGAAGTG ACTAAAACAG AAACTTCAAC ATCTGACCAA 2760

GCAGAATAAT CAAAAATCCA CACGCATTAA TAACAATGCG TGCGGATTTT TTTATGCTTT 2820

TCACTTTTTA TTTTCATCAA AAAAACTATA ATAGAAGAAA GATATTCGTA TACTTGTTCC 2880

TCATAGAAAA AATTCAATCT GGAACTCCTG TACATAATTG GAAACGCTCT TTTGCAGCAC 2940

ACAACTATAA TGAGGTGATA AAATGAAATT AGTGAATGTG ACCAATAGCT ACAAACAATT 3000

AGTAAATAAA CAATTGGAAA ACACGGATGC TTACTTTGTA AAAGTCTACT CAGCAGGAAA 3060

TACGACCGCT GTCTATTCAG AAGCAGCTCA CCATGTGGAA GTTCTGATTA TGAATAAAAA 3120

TCGTCGCATT CGTCCGACAG AAGTAAAAGA AATTTTAGCA AAACTGTTAA AACGATTACC 3180

TAAAGAGGCT TATGACCCAG ATGAAATTTC TATCATTGAG TTAAACCATG TGACCGAAGT 3240

GTCTGTTCCA TTAAAAGTTT CCCTCATGGA AAACTAAAAA AAACTGCCGA AGTTTCCCCA 3300

CTTCGGCAGT TTTTTTAATT TATTGTGACG TTATTTTTTT TACAAGACTT TGATAATTTT 3360

CCGACCAGGC TTCCTCATTT GTCGGACTGA CTAATTTATT TTCCTCGATT GCTTTTTTTC 3420

TTAAGTCATC CACTTCTTTT TTGAAATTTT CCCCCGATTT TTGAGCATAA TCTTTAGAAA 3480

AATATTTTGA TTGGCGAGAA TAAGCCTTCA GTTTCGCATC TTTAATCTCT TCTTCTGTAT 3540

CTGCAAAATA TTCATCATAA CGAAGATCGG TTATTTTCCC TGAATCTTTT AACGTAATAA 3600

CAGTTAAACG TCCAGAAAGC CCCTCCTCAA AATCTTCTGT TAGACTTGTA TAATGTTCTT 3660

TTGAAGGTTC TTTGATTTCC TTGGCTAAAG CCCTAGCAGC AGGAAGATAG CCATTATTAG 3720

CGCTATTGGA TTGACCTTTG GCTGTCTGGA AGGCGCCATC TAGGCGATTT TCAGCAACCA 3780

TTTGATTTTC TAGAAAAGTC ATCGTATTAA CAACGGTCAC TAACGATTGA TCAGTTCGGG 3840

CGTTGTTCGC TTGAAAATTT GCATAGCCGG ATAAACGTTT tGTCACTCCT GCCCATTCTT 3900

CAGAATAATA GTTTTTAGGT CCACGTTCAT CAAATTCCAC TAACTGAATT TGTTGATCCT 3960

TAACAACGAC ATCAATTAAA GCCGTATAGC CGCCATCAAA CACGACTTCA TGATGATAAT 4020

AATCCCCTTG AACTTCCCCG ATTTCTGGAC GACCTGACCA ATACATCTTT TGAGCTTTTT 4080

CTTCTGGACT AATTCCTGAT TGTCCATTAT TAGCTTCTAC TGTAGCGCCA GAAACAGCAT 4140

CAACTTCAAA GGTATGTCCA TTACGCTCAT AACTCACTTT GGCA 4184

(2) INFORMATION FOR SEQ ID NO: 383:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6159 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 383:

AAAAACACGT GTGTCTGTTC CATaATAGAC ACGTGTAGGA TCTGCGCCAG GAACATATAA 60

AACGGCCATC AGCGTCACCG AAATTAAAGC GCCGATTAAA ACTAAGCTGA AAATTCGACC 120

GCGCTTACGA ACAAAGCGTT CCAGAAGAAT AAAAAGTAAT GGCCAAATTA AATAATTTTG 180

CGCTTCAACT GCCAATGACC AAATATGTAC AAATGGCGAT TGTGTCGCAA ATCCATCAAA 240

ATAAGAAAAC CCTTTAAAAA TTTGCCACCA GTTGTTGTAA TAAAAAATAC TACTAGCGAC 300

AACGCTGCGA ATATTGTTTA ATAAATCCTG CTGAAAGAGT GTAATATAAC TAACTGCACT 360

TAAAAGCATC ACAATTAATG CTGGATACAG CCGTTTCATC CTTCGCACAT AAAATCCTTT 420

TAAATCAATG GTCTGATTTT CTGCCCATTC TTGCCGCAAC AAGTCGTTAA TAAGATAACC 480

TGAAACAACA AAGAAAATCG GAACACCTAA GTAGCCACCT CGCATGACTT GGGGAAAAAG 540

GTGATACAAA ATCACGCCAA TTACAGCGAT CGTCCGAATG CCATCAAAAC CTGTAATATA 600

GCGTTTTTTC TTGCTATTTG CTGtCTGATT CATTTGtGCT TTCCAACTTT TCTATAAAAT 660

TAwTCtTTkC CyTTAtTTAA AGaAGCAAGT AACTTTaATA TACGtTTTTT CCtAGGAAAT 720

GGCAAGCAAA CTTTGCCTCC TTTAGGCTTT TTTTACAATT CAAAAAACCA GCATTTTTTA 780

AATCACAAAA AAACAACCAA AGCTTATGAA AATCACTCAT AAGCTTTGGT TGTTTACTCA 840

TATCGTTTTC GTTTaGakGA rGATGGaATC CCTAGGGCTy CACGaTwTTT TGcTACCsTT 900

cGCCGGGaTA TTTCCakGCC kGAATCTTTw AATCGCTCGA cTrATTTTGg ATCcAGATAA 960

CGGTTTACTT TTTACTTCTT GATCAATTAA TTTTTGTAGT TGCTTTTTAA CCATCGTTGT 1020

TGAGACATCC TCGCCATCTT CTTCTGATGA AAGAATACTT GTTGAGAAGA ATGAGCGCAA 1080

TTCAAAGACG CCAAACGTGG TTTCTAGATA TTTTCCATTG ACCGCACGAC TAACTGTCGA 1140

TTCATGAATA GACAAAGCAT CCGCAATTTC TTTTAAAGTC ATTGGTTTTA ATGGCCGATC 1200

TGCTTCTAAA AAGAAGGCTT CTTGACGACG CACAATTTCT TGCCCTACAC GTAAAATTGT 1260

ATCCCCCCGC TGATTCACTG CGCGCTCTAA CCATTCAAAT TCGTTTTTCT TATCCTTAAT 1320

ATATTCTTGG ACTTCTTTGT CGTCAGTCTC TTGCATTCGT TTGAAATAAT TTTCTTGAAA 1380

ATTTAATGTC GGTCGACCAG CTTTGTTGGA AATCACCACT ATTCGATCTT CCTTGATTCG 1440

GACAGTTAAA TCTGGTCGAA TATACAATTC CTCGGTCGCA TCAAAAATAT TCCCTGGCGC 1500

TGGCGATAAC GTTTGTATAT AATCAAAAAT CAATTGAATT TCTGCTAATT CAATCCCGAA 1560

TTTTTTAGCT AACGGGCCCC ATTTACGACT CACTAAGGCA TCAAATTCTT CCTCCAATAA 1620

AATATACGCT AAATTAGGCG CGGTATCGTC TCGTTCTGTT TGTAGCATCA AACATTCTTG 1680

TAAATTGCGT GCCCCCACAC CTGCTGGATC TAGCTGTTGT AACAAAGTTA ATGCATCTAG 1740

CATTTGAATG GCGCTTGCTT CGGTTTTCAC TGCCGCTTCT TCTAACGAAA TCGTTAAATA 1800

ACCATTCACG TCTATATATT CTACTAAAAA CAATACCAAT GTTCGCAAGT ATGTATCGCG 1860

GTAATTCAAA TGAATTTGAT CAATTAAAGA CTCAAATAAT GATAAATGAT AGTCTGGAAT 1920

TTGATTCATA TAATTATTTT CTTCGTCTTG ATTGGTGTAA GAACGACTAG TTATCGGAAA 1980

ATCTGTGGTG TACTGCGTGT CTACTTGTAA ATCAATCAAT GGATTCTCTA GTGCTTTCGC 2040

TTCAACAAAG GCAGCGAGTT CTTCCGTATT AAATTGTAAA ATTTGAATGG ATTGTTGAAG 2100

CTGTTGAGTC ATGGCCAACT TCTGGACTTG TTTTTGTTGT TGGGACATTC GTTGCTCAAA 2160

TTTCATTATT TTCGTCCTCC CTCTTGTAAT TTTTGTCAAT ATCAGCTAAA CTAGTATTCG 2220

TGCCGCGCCC TTAGTGTAGT GGATATCACG TAAGATTCCG GTTCTTGAGA CGGGGGTTCG 2280

ATTCCCTCAG GGCGCGTTAT TTAAATATGT TGATTTATCA ACGTTTCTGG ACACTCTTTA 2340

TTCATCTAAA GAGTGTTTTT TTCTTGTTTA GGAATGATTT TTCTATCCGT ATTTTTTTCA 2400

AATGTTGCTT GTGGCAAGCT CAATCGCACT TAACGCGACA ACATTTGTTC TGCTTCAGCT 2460

ACAAATCCAA TAAGGTTAAA ATTCCTCCGT TCCTATTGTA ACGCATAACG TTTCCTTTTT 2520

CTATCTGCTA AACAAAGTAT TACGACAAGC AAACACCTGC TTCTCTCTTA AAGTTCCCTT 2580

TCTGATTATA TAGTAAATTG CTTTTAATTT GTTAATTATT TTACCTAAAA CTTAATACCC 2640

TAATTAAATA AAGAACCTGG GGCAAAAATC ACTTTGGATT TTTGTCCCAG GCTTAAAACC 2700

TGATAATTGG CGGGAACAGA ACCGTTCAGC TCCGACCTCT AAAATACTTA ATTAGAAATA 2760

TACTACCAGC AATCACACCA ATTATGATTA CTAGCCAAAA TAATCCACTT ACAAGTAACA 2820

GCTCAGTTAA GGTCGTCTTC AGATTCTCCT CCTTTAAATG GCCTCTGCTT TGTTTTGCTA 2880

TGTAATTAAT GGCGAGATCA GACAACAGCT ATCAAAAAAA CGGACCAACA CGTGGTCCGT 2940

TTTCATTCAT TATTAAGCTT TTACGATGTT CGTTGCTTGT AAGCCGCGTG GACCTTCTTC 3000

AGTATCAAAA GTTACTGCTT GGCCTTCGTC TAAAGTTTTA AATCCATCTG ATTGAATAGC 3060

TGAAAAATGT GCAAATACAT CATTGCCATC TTCACCTGTA ATAAATCCAA AACCTTTGTC 3120

TGCGTTAAAC CATTTTACTG TACCGTTGTT CATATATATT TCCTCCTAGT GCATAAAAAT 3180

ATGCAAGTTA TTGCAAAAAA TGTTGAGTGG CAAAAGGAGA AGTATAGTTT TTAAATATAT 3240

AACTCAAATT ACTTATCAAA ATTGATTACT TAATTACTAT AACATATTTA TTAGGAATAA 3300

ACAATGCCCA CGACTCTTCC TTGGACTTTC TAAAATAATC AAAGATGCGC TTGACAAAGC 3360

CCTAGCAAAA AACACTCCTA TTCTCTCTAT TTCCAGTTAG TTTTGTTTAA CTTTAAATAA 3420

TCACTACGCG CCAAATAATC AAATAGCGGA GcTAATCCTG GTGAAAAATT TACTTCCTCA 3480

AACCGCTTAA TAAATTGTAC GCCTTTAATT TTATTTGGTT CGTTCAGGGT CACTTTTCCT 3540

TGGGCTTGTT TGGCAAAATA AAAATGTCCT TGCCAAACGA TTGAATCAAA AACAAGTGGT 3600

CCAGTAAAAA TTTCTTCAAG TTCAGTAATT TCCACTGCTA ATTCTTCTTT CATTTCACGA 3660

ATCGCACATT CCAGCGCTGA TTCACCAAGT TCTTGTTTTC CGCCAGGGAA ATTCCATGTG 3720

TCTGGCCGAT CTTTTAAAAT AAGAATTTCA TCAGATTGAT TTTTAAGAAG TACTCGCGAA 3780

AATTCTTTTT TCATTTCGTT TAACCCTCTT CTTCATTTTT TATTGAGAAG CCACACTTGG 3840

TAGCATAAAA ATATAATGCA AAATACGATC CGCATACGAC GGATTTTTTT CAAAGTCATT 3900

GACTTCTTCA TTGACTAACT CCAAATCATC GGTAATCACT CCAGCAACGT CCATAAAGAG 3960

CATTTTACTC AATACCTCTG AGTCATTAAC TGTCCAAGCA TAAACTGCTT TTTTCTGTTG 4020

GTCAGCTTTT AAAATGAAGG TATCATTCAA CGTAGTAGCT TCCATTGTAT ACAAATCAGC 4080

TTGTGTTTCT GGAAAAGTGA AATTATAGGG CAAAATAAAA CTAACTTTTA CTTTTGATTT 4140

GGCTTTTTTA AACGCTTTGA TCACATTGTA GTCGAGAGAT TGAACCTGAT GGTTATTCTT 4200

TATCAGAGGA CGTTCATATT TTTCAATAAA GTTAGAAAGC GCACCTTGAC TATCTTGTTT 4260

AGACGTTTTA ATTTCTACTA GTAATTTTTG TTTGGCTTGG TTTGCTTTCG CTAAATAATC 4320

ATCAAAACTG GCAATTGGTG CTTCTTGGCC ATTTTCTTTA ACGGTCATTT TTGTTAATTC 4380

AGCCAACGTA AATTCTTGCG GCGTGCCATC AACGCCGGCG AGTTCTTGAA GGTTTGCATC 4440

ATGCATCACA ACAAATTGAT GGTCTTTGGT TTCTTGAACA TCAATTTCAA TATAATCTGG 4500

TTTAAATTTA ATGGTTCTTT CCATAGCAGG AATGGTATTT TGTACGCCAT TACTATTGGT 4560

GACCCCTCGA TGCGAAATAA TCAAAGGTTT TGATTCTAAT AGCCCTGTTA AGTAAACAGC 4620

GTTGAAAATT ATATAGCCGC CAAAAATGGT GAATAGTCCT AGGCCCATCC CGATAGTCAC 4680

TAATTTATTA CGTTTCGTTT CTTTGTCAAA GGCAACTTTT GTCGTCGTTT CTGCTTGAAC 4740

ATTTAATTGT GTATATAAGA AATTAATCAG TAATACTGAA AGCCAAGCAT TAGAGACAAA 4800

CTGTAAAAAT TGGACAACGG TTAAATTTAA AATCCCACCT AGCAAGCTAA TGGTATCGCT 4860

CAAGGTATCT AACTTCAACT GTAGAAGATA AAGTAACACA TAAATAACAT ACGTTGAAAC 4920

AAATACCGCG ATCGTCACAA ATGCAATATT ACGAATAATG AACCATAATC GTTTGTTTGT 4980

CATTGACCAA CTAGCCTTAA CTGCCTCGCC CGCTTTTTTT CTCTCCAGAA TCATTAATGG 5040

AAGCGTGTAA ATGAAACGAA TAGCTAAATA CCAGACnnrt AACCCAAAGG CCAGTAACCC 5100

CACCAAGTAA GGAATTCGTT GGGACAAAAA TTCAACAATA AAATCTGGAA TCACAAATTT 5160

CGTCAATAGA TTACTTTGGA AAATCAGGTT ACCAAAAGGG ACAATGACAA CAAAATAACC 5220

GAAAAAGAAT AACCACGAAC GGGCACcTTG CTTTTTAAGA aCCTTAAAAG ATGCCTTGGT 5280

AACTCCTTTA AAATGAAAGC GATGGTCTTG CCGTATTTGA AAAATCCCTA ATAATAAAAA 5340

GGTAAACTGA TAAAAAACTG CACCTAAAAA AAGTAACAAC AAGAGAATTA ATAACACCAG 5400

AGCAAGCGGC TTTTCTAACA ATAAATTCAG GTTATTAAAC GAAAGATAAG GAATACCTTG 5460

AGTTTTTAGT GCGTTGGAAG TAGCCCAACT AAAAGCCCCA CCCACCAATA ATAATAAGAC 5520

ATTTAAAGTA AAAAATAACC ATAAATACTG CAAGATGTCT TTAAAAAATT GCCTTGTTCC 5580

ATTTTTAAAT GTTTGCCAAG CAGACATACA AACTCCTCCA TTCTAAAAGT ATCTGATTAA 5640

AGTTCTTCCA TACGAAGCCT AAATAAAACA ACGGCAATCC AAATTGAAAT AAATGATGGC 5700

AAAATAACAA TAATGATTTC GACTGGCTCA TCAATAGTTC CTTGTAACGT ATTTTTTACT 5760

TCTGAAGAAA CAATaAACTG ATCAAAAATC ACTTGATAAA TAGTAACATA AATCAATACC 5820

CATTGAATAT AATATTCAAT TATTTTCaTT AAAGCGATGG GTAAAACAAT TTTTTTCGTC 5880

GTTAAATTCA ATAAAAAATA TAACAAAAGA CTCGTACTGA TGATAACGAC TATCCCTAAA 5940

AGAAATAACG ACTCTAACCG CTGCCCTATT ACGTTTGCAA CGCCCATTGT TAATAATAAA 6000

TACATAGCTA CCGAAATTAA GATGCTCACA CTTAGCATAA CAGAAAACAA AAGTAATGCC 6060

TTAAGAAAAm sCCACTTATC AACTCTTTTA TTTCCAATCA ATAAACCTAA AAGAATCACC 6120

GACAAAAGTA AAATCAATTC AATCATTCGT TCCTCCTCC 6159

(2) INFORMATION FOR SEQ ID NO: 384:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1366 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 384:

GTAATAGCAG AAATGAGGAA AAGAAATGCA GAAAGAATCC ATTTATGGAC TGACGAGAGA 60

ACAGCTTGTG GACTGGTTTT TAGCCCATGG CGAGAAGAAA TTTCGCGCAA CACAAGTGTG 120

GGAATGGTTG TATACAAAAC GTGTTGCTTC ATTTAGTGAA ATGTCGAACA TTTCTAAATC 180

ATTGATGACT TTATTAGAAG AAAATTTTTC ATTAAATCCA TTAAAACAAG TCATTGTGCA 240

AGAAGCACAA GATGGTACCG TAAAATATTT ATTTGAATTA CCAGATAAAA ATATGATTGA 300

AACAGTCTTG ATGCGTCAAG AATATGGATT ATCTGTTTGT GTGACCACTC AAGTTGGCTG 360

CAATATTGGG TGTACATTTT GTGCTAGTGG GCTATTAAAA AAACAAAGAG ATTTAACTGC 420

TGGCGAGATT GTTGCCCAAA TTATGTGGGT GCAACATTAT TTCGATGAAC GTGGGTTGGA 480

TGAACGTGTT AGCCATGTCG TTGTGATGGG AATTGGTGAA CCATTTGATA ATTATGCCAA 540

TGTCATGAAC TTTTTACGAA CGATTAATGA TGATAAAGGC TTGGCTATTG GTGCCCGTCA 600

TATTACAGTT TCAACAAGTG GTTTAGTGCC AAAAATTCGT GAATTTGCAG ACAGTGGACT 660

ACAAGTAAAC TTAGCAATCT CATTACATGC TCCTAACAAC GAAtTCGTAC ATCAATTATG 720

CGGATTAATC GTAGTTTTCC AATTGAAAAA TTAATGGCAG CAATTGATGA ATACATTGAA 780

AAAACCAACC GACGTGTTAC ATTTGAATAT ATTATGTTGA GTCAAGTGAA CGACCGACCA 840

GAACATGCTC AACAATTAGC AGATTTACTA CGGAATAAGA AAAAACTTTC TTATGTTAAT 900

CTGATTCCGT ATAACCCAGT GAGTGAACAC GATCAATATA GCCGAAGCTC TAAAGAAGCG 960

GTTTTAAAAT TTTATGATGT ACTCAAAAAG AATGGCATCA ATTGTGTTAT TCGTAAAGAA 1020

CACGGAACAG ATATTGATGC GGCTTGCGGA CAATTACGCA GCAAACAAAT GAAAAAAGAA 1080

AAAGTCmAAA ACCAATAACG TAGTTAAAAT CGGATCAGAA AGGTTTCTTT TTGATTCGAT 1140

TTTTCGTTTA TTTAATAAAA AATCGTCGGT TTTTCAATTC TTGTTTaTAA ACaAAAGATG 1200

rATGCyCtAA AACtTTTATG aCAAwTTTyC TTTTtGTGGt TTTcaAnCTT GaAATyCTAA 1260

AAGGAAATAG CTATACTGGA AAGGAGATCC GTCnCCnGAT TTTTTCCTGG AAACCGCCTA 1320

TTATTTTCCG nTAnCCTTTT CCAAAAAGnT GGGAATTTTG GTACGC 1366

(2) INFORMATION FOR SEQ ID NO: 385:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1022 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 385:

CATTTCCGAT TATGCCGGCG TTTAGTTTTT TGAAAATCGA TTTTAGTGAT ATTCCGATTC 60

TACTCGGAAT GTTCTTGTAC GGACCGTTAG CAGGAGTAAT TACTGCTTTT GTTCGTTCGT 120

TGCTACACCT GTTCTTAACC GGACTAGCAC CGCAAAATAT GGTGGGAGAT TTCGCTAGCT 180

TTTTAGCAAG TAGTATCTTC ACCTTGCCAA TTTTTTATTT CTTTGGTAAA AAGAAAAATA 240

TCCGTACAAA TCGGATAGTG GGCTTAGTAA GTGGGATCTT AGCCTTGACA ATTTTCATGA 300

GTATTGCGAA TTATTTTGTC ATTACACCCA TTTACTTACA ATTATATGGT GTGACCACAC 360

AACAATTTTT AGGAACATCT TTAGCAAGCT ATGTGGCGAT TGGTATTGTG CCATTCAACC 420

TTATTAAAGG CCTCTTAGTC AGTGGTGTTT TTCTAGTACT ACATGCGAAG TTATTGCCAT 480

GGCTATCAAA AAAACAACAT ACTATTCAGA AAAAAACACC GTTAACAAAA TAAATGATAA 540

AAAACCTGCT GTTGAACAAT GTTTGACAGT GGGTTTTTAA AATTTACGCC TAAAAGAAAA 600

GAGGkTGTCA TAATCTGtCA TCATTCTGTT AAAAAATACT AAACCATCTG CATTGTAATT 660

TTAACTTTCC CTTGGTATGA TAGAAATCAT CAAAGAAAGA GGAGTTTTCC TGAGTGAAAA 720

AGTGTAAAAA TTGTCGTCAC GTGAATCGTG ATACGGAGTC CTTTTGTGAA GAATGCGGCG 780

CTCCGCTAAT GAATGAATCA ATGCATCAAG AGGAAAACCA AGCACAACCA TCAATGAATA 840

AAGGGGACGA TCTACTCCTC TCAGATCAAA AAAGAAGCTG GATCTGGGCG TTCCTTTTGT 900

GGTCCATCCG TCCTGGGGCC GGGAGCCnAT TTCCTCGGGA CCGCCATnAT TTTCCAAAGG 960

ACCACCAAAT TCCCTAATTT AATTGAAGCC GATTGAGGAA TnGGGGTGnC CCAAGGAATT 1020

AA 1022

(2) INFORMATION FOR SEQ ID NO: 386:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8145 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 386:

TGGTTCAGGT AAAGGTCGTT TTATTTATGA AATGGCTAAA GCACATCCTG AAATTAATTA 60

TATTGGTATC GATATGCAAT TAAGTATTTT ATCGATTGCA TTAGATAAAC TAGTGGCCGA 120

TCCGTTACCT AATCTACAGT TATTGCGAGT AGACGGCGAA CATTAACAGA ATATTTTGCT 180

GAAAATGAAG TAGATTTAAT CTATTTAAAT TTTTCAGATC CTTGGCCAAA GAAAAAACAT 240

GAAAAACGTC GTTTAACGTA TAAAAACTTT TTAGCGACAG ATGAAATTAT TTTAAAACCA 300

AATGGAGAAA TTCATTTTAA AACGGATAAT CAAGGGCTcT TCGAATACTC ATTwAgtAGC 360

TTTTCTAAAT ATGGCATGAT TATTGAACGT GTGTGGCTAG ATTTGCATAA TAGCGAGTTT 420

GAAGGCAACA TTATGACAGA ATACGAAGAA AAATTTTCTA GTCGTGGTCA ACGAATCTAT 480

CGAGTAGAAG CACGTTTTGT TGCAAAATAA AAACAAGCAT TCCTTTATTT TAAGTGAAGG 540

AATGCTTGTT TTTTATTATT TAAGAAAGCC TAAGTTGTCG TTCTTAGAGC ATAGGCGTCC 600

AATAAAATTC TGGCCAGAAT GGATAATGGT AACCGTGAAC GCACTTCATA ATCGGGGAAA 660

AAGGAGCCTT CTGATTTGAA GCCTACGAAC GGAATTTCAA TGTATTTCAT TAAACTAGAG 720

TGTTTGCTTG CAGTTAAAAG GATAGTCCCA ATTTCTGAGT CGTAACAGTT TTTACCAGCG 780

ACAACTAATT CTTGGGTTTC TCCGTTTAAA GAAACAAAAA GAGCCACATC CTTTTTTGAT 840

AACTTTTTAC TAATATTTTT AATAATTTCT GGGTCCGTAT GCAGTTCACA GTATTTGCCA 900

GCCAATTGAA ACTTAACCAT CATTTCCTGA GCAATTAATT CTGAAAAACC TCTAGCAAAG 960

ATAATAATCC GTCGAGAAGC CTTGATTTTT TGGATCGCAT CTTCGATATT ACCAATTTCA 1020

ATCATGTTAA TCGTGCGAGT TACTTCTTCT TCATTCTTTA AAATGGCTGT CCGAATTCCT 1080

CGGTCTACTT TTTCTAATGC TGAAAAATTT ATATTTTGGT TTTCTTTATC TTTCAAGTGA 1140

TGCTTGAAGG ATGTGAAACC TTCATAGCCT TTCTTTTTCA TGGTTCTTAC AATCGTCGCA 1200

GTTGATACAT TGGCTCGTTC GCTTAATTTT ACAATGGAAT AATCAGGGAT TTCTAAAATA 1260

TGTGTTTCGA TAAAATGCCA GAGATATTGC TCTGACTCAC TCAATTTTTG TTCCATAATG 1320

ACCAGCCCTT CTTTCTTTCA CTAAAAAAAC ATTTATAAGC CTATTTTTAA TGAAAAAAAA 1380

TAAAAAAGCT AGAGAAATTA TGAAATTGTA AATGTTTTCT ATATTTTTGC GAAAGAATGA 1440

AAACCTTTAC ATGGCATTTT GAGACATAAT AAAGAGGAGA AAGGAGGAAC TACAGTGACT 1500

ATTTATACAT GTACATTAAA TTTAGCCATT GATTTATTTA TCGAAACAGA AGAGCTAGTT 1560

CCTTTTGTCG TGAATCGTAC GAAGGAAGAT GATATTCAAG CAAACGGGAA AGGAGTCAAT 1620

GTTTCCCTAA TTTTAAAGAT GTTAGGGATT GATAATACCG CTTTGGGAGT CAAAGCTGGT 1680

TTCACTGGGA ACTATGTCGA GGACTATTTA AAAGAAAAAG AAATAACCAC TGATTTTATC 1740

GAAGTCGCTG GTACAACGAG AATTAATGTG TTTACAAAAG TTACCCAAGA TCAAAAGGAA 1800

TACAAATTAG TCAATAAGGG ACCAAAATTA TCTGAAGAAC ATGTACAACG TTTTTTGAAA 1860

AAAATTTCAG AGTTACGTAA AGGGGATTAT TTGTGCGTAT CAGGAAGTTT ACCACAAGGG 1920

GTTTCGCCAA GTATTTTGAT TGAAATTAGC CGAATTTGTT TTGAAAAACA GGTGTTCTTA 1980

ATCTTAGACA GTAGTTATGA AGAAATTTTA GATTGTTTGC CTTATCAACC ATTTTTATTG 2040

AAACCAAATG AAGAAGAACT TCAAGAGTGG TTTCATACAG AAGTGCAGAC TAAGGACGAT 2100

TACATTTTTT ACGGACAAGA ATTACTGAAG CGGGGAGCCA AAAATATTTT ACTTTCTTTA 2160

GGTAGTGAAG GGGCTTTATT TATGAATAAT GAAAAGGTTC TTTCAGGTAA TTCACCAACT 2220

GGAATGGTTG TGAATACAGC TTGTTCAGGG GATGCAATGT TAGGTACCTT TTTAGCTGGT 2280

TGGCATCAAG GACTTTCATT GGAAAAAAAT TTGAAAAGAA GTATTGCAGC AGGTAGTTCA 2340

ACAGCATTTC GCAAAGGATT GACAGATTTT TCAGATGTTC AGGAGTTAGA GCAACAAATA 2400

AAGATTCAAG AGGAGGAATA AACAGATGGC TACGTATCAA TTAATAGCTG CTACAGGTTG 2460

TCCGACGGGA ATAGCGCACA CGTATATGGC ACAGGAGGCA TTGGAACAAG CTGCGAAACG 2520

TAAAGGTATT ACGATTAAGG TTGAAACACA TGGTCAGATT GGTATAGAAA ACGAGTTAAC 2580

ACCTGCAGAA ATCCAAGGAG CGGAGGCAGT CATTATTGCC GCCGACAAAG ACGTTCAAGC 2640

AGAACGTTTT GCTGGGAAAC GAATCATTGA TGTCCCTGTC AGTGTGGGAA TTAAAGAAGC 2700

AGATCGCTTA ATTGAAGAAG CGCTCGCAGG CAAAGGATCG ATTGCGGCCG AAAACCAAGC 2760

AGTGGATGAA CTTGAACAGG AAACACAGAT TTCTTCAGGA AATGTTGGAC ATAGTATTTA 2820

TAAAAACTTA ATGAATGGCG TCTCTCACAT GTTGCCGTTT GTAGTAGCAG GCGGCGTTTT 2880

AATTGCTCTT TCTTTTGCTA TTTGGGGCAT TTATTCATTT GATCCAGAAA GTAGTCAGTA 2940

TAATGCTACC GCTGCTATGT TAAAAAGTAT TGGTGATGCC TCAATGGGCA TGATGGTGCC 3000

AATCTTATCT GCGTATATCG CAGAGGGGAT TGCCAAACGA CCTGGTTTGG TTGTGGGCTT 3060

TGTCGGTGGT TTGATTGCTA ACACTGGTGG CACTGGCTTT TTAGGAGGGA TATTATCTGG 3120

TTTTCTTGCT GGGTATTTCA TTCTTTTATT ACAAAGAGTA TTTAAAGGCT TGCCTAAATC 3180

ATTAGATGGT TTAAAAGCAA TTTTCCTTTA TCCAGTCATT GGCGTTGCAG TTATTGGTAT 3240

CACTATGTCT TTACTAGCAG ATCCGATGAA AGCTGTTAAC GAAGGAATGA TGAGCTTTTT 3300

AGCTAGTTTC CAAAACTCCA GTCCGTTGGT TTTAGGAATT ATTGTGGGTT GTATGTGTGC 3360

CTTTGATATG GGAGGTCCAG TTAATAAAGC GGCCTATGTT ACTGGGACGG CACTCTTAGC 3420

TCAAGGAAAT ACTTCTTTTA TGGCAGGAGT TTCTGCTGCC TGCATAGCGC CACCTTTAAT 3480

TACAGGCTTT GCTACATTAT TCTTTGGAAA ATACTTTGAC ACAAATGAAC GCAATGCGGG 3540

GCTAGTGAAC TTTATTTTAG GCTCCACACA TATTACAGAA GGCGCTATTC CGTTTGCGGC 3600

AAAAGACCCC TTGAAGGTTT TACCAATTAT GATGCTCGGA TCTTCAATTG CGGCAGTCCT 3660

AACGTATATG TTTGGGGTAC AAGTGCCAGC ACCTCATGGA GGATTTCTTG TTCTACCAGT 3720

TGTTACTCAC GCTGTTCAAT GGGTTTTGGC AATTTTAGCT GGTTCATTGG TAGGTGGCTT 3780

ATTGCTAGGT TTTGTACAAA AAAGTAGAAT CGCCAAACAA GAGCAATCGT TATCAAAAAA 3840

ATGAAATAAC AGGAATTAGG AGTGGATGGA ATGGCTTTTA TTAAAGAAAA TCATATTTTT 3900

TTAAATCAAC AGTTACAAAC GCAAGAGGAT GTTTTTCACT TTTTAGCAAA AAAATCAACA 3960

GAATTGGAAG TTGCTCAGGA CGCACAGGAA GTTTTCGATA AACTGAACGA GCGAGAACAA 4020

GAAGGAACAA CGGGGATGAT GAATGGATTT GCGATACCTC ATGCAAAAGC AGCAACGATT 4080

CAACAAGCGG CCATTATCAT CGTTACCTTG GACCAAGGTG TTGAGTGGCA GAGTCTTGAC 4140

AACCAATTAA CGGAGTTTGT AATTGCGCTG TTTATTCCAG ATGCCGAAGC TGGCACGACA 4200

CATTTGAAAT TACTCTCTTC TGTTGCGCGC CTCTTATTAA GAGAAGAAGT TACTTCTGGA 4260

CTGAAACAAG CAAGTTCGCC AGCTGAAATT GCTACGTTAC TAAATAATCA ATTAGGAGAG 4320

GGAACAGAAT GAAAAAAATT AGTGAACAAA AACGTAAACA TTTAGAAAAT CTGGTAGATG 4380

ATCAAGGGAT TATTGGCGCA CTAGCCATTG ATCAACGTGG GGCCTTAAAA CGCATGATGG 4440

GAAAATACAA AGAGGTGACT GCGCAAGAAA TCAGTGACTT TAAAGTATTA GTATCTCGTT 4500

GCTTAACACC AGAGACTTCA GCAATTTTGT TGGATCCAGA ATATGGCCTA GCTGCTGCTG 4560

AAAATCGAGC ACAGACTTCA GGTTTGTTAC TTGCGTATGA AAAAACGGGC TACGATGCTT 4620

CGACACCAGG ACGGTTACCC GATAGTCTTG ATGTTTGGTC GGTTAAGCGT TTAAAAGAAG 4680

CGGGCGCAGA CGCGTGTAAA TTTTTGCTTT ACTACGATGT TGACGAAAGT GAAGCAATCA 4740

ATGAGCGAAA GAAAGCTTAT ATAGAACGGA TTGGTTCAGA GTGTTTAGCC GAAGAAATTC 4800

CATTCTTTTT AGAAATTGTC TCTTATGATG CAAACAATTC GGATAGTGCT TCAAAAGAAT 4860

ACGCTAAAGT AAAACCGCAC AAAGTAATTG AAGCGATGAA AGAATTTTCA AAGGATCGCT 4920

ATAATGTTGA TGTTTTAAAA GTAGAAGTCC CTGTGAACAT GAATTTTGTA GAAGGTTTTG 4980

GCACTGAAAG TCTTTATAGC CAAGATGAAG CACAGGCCTT TTTCAACATG CAAAGTGAAG 5040

CAACACAACT TCCATTTATT TTCTTGAGTG CAGGGGTTAG TGCTACGATG TTTCAGGAAA 5100

CATTGAAATT TGCAAAAAAA GCAGGTTCCT CGTTTAATGG AGTCTTATGT GGTCGCGCAA 5160

CTTGGGCAGA CGGCGTGCTT CCATTTGTTC AACAAGGTGC TGAAGCAGCC GTTGCGTGGC 5220

TTGAAACAAC TGGTAAAACA AATGTTGAAG AATTGAATCA GGTTTTAAGA GAGAGCGCAG 5280

TATCAGTCTT TGAAAAAATA CAGTAATTAA GCTGAAAATA TATACTGAAT TTAAACTATT 5340

TGGTCATCTG AATTAAATTT TTTCAGAAAT CGTCTTCTTT AAGTGAACTC AATGTACGCA 5400

AACTCAAAAG ATATGATATA ATTATTTCGA TATCGATGTA TGAACACCCA AAAAGTAACA 5460

TTGCTTTACT TTATTTGGGG AAAGTTGGGG TATTATGGGA ATTCCTAAAG AAGGCGAGTT 5520

TGTAACCATT CAGAGTTATA AACATGACGG ACATTTACAT CGAACGTGGC GTGATACGAT 5580

GGTACTAAAA ACGAGCGAGT ATTCTTTAAT TGGTGTGAAC GATCACACGT TAGTAACGGA 5640

ATCCGACGGA CGTCGGTGGG TGACTCGTGA ACCAGCCATT GTTTATTTTC ACAAGAAATA 5700

TTGGTTCAAT ATAATAGCAA TGATTAGAGA AAAGGGGGTT TCGTATTACT GCAATCTGcT 5760

```
TCGCCATATG TTTTAGATGA CGAAGCGTTG AAGTATATTG ATTATGATTT AGATATCAAG     5820
GTTTTTCCAG ATGGTGAAAA ACGTTTACTA GACGTTGATG AATATGAATT TCATAGCAAA     5880
CTAATGGATT ATCCAGAAGA TATTGACTTC ATTTTAAAAG AAAATGTCAA GACTTTAGTG     5940
GATTGGATTA ATAACGAAAA AGGCCCCTTC TCGCCCGAGT ATGTAGATAT CTGGTATCAA     6000
CGTTATCAAC AGCTATCAAA AAAATAGCCG ATGTAAACAG CCCACAAAAG GrGCTTGGAG     6060
CATTAGCCAA AATGGCTAAT GCTCCAAGCs CTAAATTTTC CGATTTAATT GGTGTTATAG     6120
TTCTTTGTTC ATTTCGTAGT GTTCAATGCC AGCATCCATA AAAGGTGCTC CTGTTTTTTG     6180
ATAACCTAAA CGTTCATAGA AGCCTACAGC GGTGACTTGT GCTCCTAGAG TAATCTTGTG     6240
GTAGCCTTGT TCTTTGGCAA AATTTTCCGC CGCTTCCATA ATCACTCGTC CATAGTCGGC     6300
GCCACGGTAA GCTTTTTCAA CGGCCATTCG TTGTAATTTC ATAACGCCAT TTTCTAGTGG     6360
AAGCAAGCGA CaAGTTGCGA CAGCTACATG GTTCGTGTAC AAAACAAAGT GAATACAATG     6420
TGCTTCATAT TCATCGATTT CCATCTCTTC TGGTACCCCT TGTTCCGCTA CGAATACGCG     6480
TTGACGAAGC CGGACAGCAT CTAAATAAAG ATCACTCATG GTATCTTTTG TTTGAGCTAC     6540
ATGCATnGAA AaaCCCCCAg TTCAGAAcTG tkTCTATTTT AyTATACCtG AAAAAGTTCT     6600
AGCmATGAAG TATTCTaCAA TGACTGAGGG GGGGAAGGTG TGATAAAtAA AACGCAACAA     6660
TTGAATAAAn ATAAAGGAGG AGCCGCTGTG ATTAATCGAT TTAAAGAATC GAAATTATTT     6720
TTCTGGTCAG TTGAATTACT TGTCGTGGCC AtGCTATtAt TTAwTGCTTC GAAAATTAAT     6780
TTCTTATTTG CACCGATTGG GACATTCTTT TCAACGTTGT TTGCCCCAGT CTTAGTAGCA     6840
GGCTTTTkGT ATTATTTACT AAATCCTGTG GTCAATCTAC TAATGAAGAC GAAAATGAAA     6900
CGAATTTATG CCGTCTTACT GGTATTTTTA TTACTAATCA TCGCATTAGT TCTGATCTTA     6960
TTAACAATTA TCCCTAAGTT GGCAGATCAG CTAGCTAGTT TAGCATCAAG TATGCCGGAT     7020
TTCTTTAAAC AAGTGGAGAC CTGGATTTAT GAGATTGCGG AGTTACCTAT TTTTAAACAA     7080
ATTGATTTAA CAAGCTATAT TGAAAAAATG GATATTTCTT ATGCCAATAT TATCCAACAA     7140
TTTTTAAGTA GTTTATCAAG TAGCTTAGGT TCAATCGTAT CAACGGTGGC GTCAACCACT     7200
ATTGTTTTGG TCACTGCGCC ATTTATTTTA TTTTATATGT TAAAAGATGG AGATAAGCTA     7260
GTACCAGCGA TTCAACGCTT TTTACCTGAG AAAAGAAAAG ATGATATTGT GGATCTTTTA     7320
GGACAATTGA ACCAAACCTT ATCCAGTTAT ATTAGTGGAC AAGCGATTGA ATGTTTATTT     7380
GTCGGTACAT TTACGATTAT TGGGTATTCA TTGCTTGGCG TTCGTTATGC CtTTTTATTT     7440
GGTGTCATTG CCGGCTTTAC GAATTTGATT CCCTATCTAG GACCTTATCT AGGATTAGCG     7500
CCTGCTGTCT TAGTGACCAT CTTTAATGAA CCCgTCaAAG CAGCCtTATG TTGTTTAgTG     7560
GTGCTAGTGG TTCaACAATT GGACGGAAAT ATTATTTATC CAAACGTTAT CGGAAAATCT     7620
CTAAAAATCC ATCCATTAAC GATTATTTTA ATTTTATTGG TAGCAGGAAA TTTAGCAGGT     7680
TTGCTAGGTA TTTTCCTCGG TGTCCCATTT TACGCAATCT GCCGGACCAT TATCTACTAT     7740
```

GTTATTGATA TGGTCAAAGC GGGTCGTAGT GAAAAAGTTA CCAATGCCGT TCTTTTAGGG 7800

AATGAAACAA ATACGAAAGA AAATAACGGT TAATGATTGA ATCCGCTAAT CTTTCGTGCT 7860

ATGATTTTAT GGATTAGGAT AGAAAATAAC CTGTCATAAT ACTGATTGTG TCTGTTTCTA 7920

AGCCTAACCA TAAAAAGAGG AAGAACGTTA AAAGAGTACG AAAGTACAGC CTCGGAATTA 7980

TCAACAATTA AGGAGAGAGA GAACATTATG AATAATGCTG ACCCTGAGAG TCAGTCGCTA 8040

ATTGCGCAAC TTCTATTATT AGTCGTGTTG ACGTTTATCA ATGCCTTTTT GGCAGCGGCG 8100

GAAATTGCAG TCGTATCTGT TAACAAGAAT CGTGTGGAAC AAAAA 8145

(2) INFORMATION FOR SEQ ID NO: 387:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3658 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 387:

AAAAGAGGCT TTAACAGCCA AGAACATCAA ACCTAGTGAT TTAGCGAAAA AAACTGGCAT 60

CGGTAAATCT TCGATCAGCG ATTGGCTAGC TGGTCGTTAC GAAGCAAAAC AAGACAAAGT 120

TTATCGCATT GCAGATGCAT TAGACATTAA TGAGGCCTGG TTGATGGGAC AAGAAGTTCC 180

CATGGAAAAA AATGCCTCAA CTATCGACCG CATTTATAAA AAATTAGAAC CCCAACGACA 240

AGCCATCGTT TATCAATTTG CCGAACAACA ATTACACGAA CAACAAACGC AAGCAGAAAT 300

TCTCTCATTC CCTCGCCGTG ACGAAATGAC ACTGGCTGCC CACGCTGGGG ATCCAGAAAA 360

GATATTTTCA AAAGAAGAGA TCGAGAAAAT TCACGATTAC TTGGATGAAA TTGATGCCAA 420

ATATCAACAA TCGATTTCTT CTGACAAAAA AGAGGATTAA CTTCCCATCT ACTCTTTTTA 480

TTAAGTAATG GCAGGTGAAT GGTAAAGTGA ATGATTATGA AAAACTCGTC AGTTCCATTC 540

AAAAGGATGT GACGGTTTTA GAAATTGATT TATACAATCA AACAGGTTGT TATGGGTTGT 600

ATCGAAATGG CAAAATTTAT ATTGAAAAAA CATTGAGTAC CCGTCAAAAG AAGAATATTT 660

TAGCCGAGGA ATACGGGCAT TATCAAACAT CCGTTGGCAC TATTTTAAAT CAAAATTGTA 720

CAGAAAATCG GAAGCAAGAG TTGAAAGCTC GCAATGTGGC CTTAGAGCAA TTAGTTACAT 780

TAGACGATTT AATTCGTTGC TCTGAAGCTG GCTTAAGTAA CCATTATTCT TGTGCTGAAT 840

TTTTGGAAAT CGATGTTGAA ACATTAAAAA ATGTGATTAC GTACTATCGA CAAAAGTATG 900

GCGCCACTTA TCTTTATAAA GGACGGATTT TTGAATTTAG GGATTATTCT GTGATGGTAT 960

TAAATACTGG GTTAACTTAA AAaGCATGGA ACAAAATCAA TAAGGATTTT GTTCCATGCT 1020

TAAAAGTATT CTGGAAAATT CAAGAGCGTT TTTTGTACCA ATGGTTAATA TGTCCAACGC 1080

CAAGCATAAG ATTGGGGACA AGTATAAGGA CGAATATAAA TGACGGCCAA AATCCGATGA 1140

ACTGAATAAT ATGTTCCCGC ATACTATGTA GTACAAAGAA GAAAAGAAAA AAGTACAGGG 1200

```
AACTCTTGCC TGAACCCATT TCTTTTAAAT AACGATCTCC ATTTTTATAG AAAACAACAA    1260
TTCTTTTCAT CTTTGCACCT CCATGACTAT TTTATCAAGA AATCATGAAA AGAAAAAGCA    1320
CACCATTCGT CTGGTGTGCT TTTTCTTTTT AATTATTCAC CTTGATTCAA CTTACGAATC    1380
AACTCATCAA TGTGATTTCC GTATTGGACA GATTCGTCTC TTTCAAAAAT CAACTCTGGT    1440
GTTTTATACA AGGTTAACCG TTGTCCTAAT TCTTTACGAA TCAGTCCTTT GGCTTTGTCT    1500
AAGCCTTGTT GAGCTTTTTG TTGTTCGGAA GCTAAATCAG ATAACAAGCT GTAATAAATC    1560
GTTGCTTGTT GTAAATCACC GGTCACACGA ACATCTGTAA TAGTGATGCC TTGGACGCGT    1620
GGATCTCTGA TGCGTTTGTT CAAGATATCA TTCACTTCAC GCATGATTTC TTGACCTACA    1680
CGACGGTCAC GATAATTTGC CATAGtGTCT TCCTCCAATT ACTAAAATTA TTGTTTGATT    1740
TCTTCCATAA TGAAGCCTTC AATCGCATCA TCTACACGTA AATCATTGAA GTTTTCAATC    1800
ATAGCGCCAC ATTCAAAACC AAGTTTGACT TCTTTCACGT CATCTTTAAA GCGTTTCAAG    1860
CTTGCTAATT TACCTTCATA AATAACAATA CCATCACGAA TGACACGCAC GCCACTATCG    1920
CGGCGAATAA AGCCTTCTGT AACATAACAG CCAGCGATTG TGCCAACTTT AGATACTTTG    1980
TACAATTCAC GAACCGTCAT TTGACCCGTA ATTTTTTCTT CAAATTCTGG ATCTAATAAC    2040
CCTTTCATCG CTGTTTCGAT TTCTTCTAGG GCTTTATAGA TAATACGGTG TAAACGAATA    2100
TCAACTTCTT CTTGTTCTGC TTGTTGTTTC GCTTGTGGTG TTGGGCGAAC GTTAAAGCCG    2160
ATAATAATCG CATTACTTGC TGCGGCTAAA GTTACGTCGC TTTCATTGAT GGCCCCAACT    2220
GCCGCGTGAA CAATTTTCAC ACGAACGCCT TCTACATCAA TTTTTTGTAA ACTTGCTGAA    2280
ACGGCTTCGG CAGATCCTTG TACGTCTGCT TTGACAATGA TATTAACTTC TTTTAATTCG    2340
CCTTCTTTTA GACTTTCAAA TAAGTTATCC AATGTCACAC GGCTACTTGC TGAACGTTGT    2400
TCAAGCAACG CACGTTTAGC CCGTTCTTCC CCTGCTTGAC GTGCTGTTTT TTCATCTTCA    2460
AAGACCACAA AACGATCACC CGCTTGAGGA ACATCATTTA ATCCGGTAAT TTCTACAGGA    2520
GTTGCGGGTC CTGCTTCTTT ATCACGGCGG CCCATGTCAT TGGTCATCAC ACGAACACGT    2580
CCGTAAGTAT TACCAACAAC GATTGGATCA CCAACATGTA AAGTTCCTTG TTGAACGAGT    2640
AATGTCGCAA CCGGACCTTT ACCTTTATCT AAGCGCGCTT CAATCACTGT TCCAATTGCT    2700
TTTTGCGTTG GATCAGCTTT TAAGTCTTCT ACTTCAGCAA TCAATAAAAT ATTTTCCAAT    2760
AGCTCATCAA TGTTTTGATT GAATTTCGCT GAAATATTGA CGAAAATCGT ATCTCCGCCC    2820
CATTCTTCAG GAATTAATTC ATGTTCACTT AATTCTTGTT TCACATGATC AGGATTTGCA    2880
CCAGGCTTAT CAATTTTGTT AACTGCTACA ATAATCGGCA CTTTTGCCGC TTTGGCATGG    2940
TTAATCGCCT CGATTGTTTG TGGCATTACT CCATCGTCCG CTGCAACAAC TAAGATGGTA    3000
ATATCAGTGA TGCTGGCACC ACGCGCGCGC ATACTTGTAA AGGCCGCATG TCCTGGTGTA    3060
TCTAAGAACG TAATTGGTTT GCCATCAATA TCTAACTGAT AGGCACCAAT ATGCTGCGTG    3120
ATACCGCCTG CTTCACCTGA AGTCACACGA GAATGGCGCA ATGTATCTAA TAATGTTGTT    3180
```

TTCCCGTGGT CTACGTGTCC CATGATTGTG ACAACAGGGG GACGTGTAGT TAAGTTTTCT 3240

TCAACAACTG CTTCTGGTTC GAAGAATTTA TCAATGTCCG CAATATCAAC TTGAACTTTT 3300

TCTTGCGGTT CCATGCCATA GTCCACAGCT AATAATTCAA TTGTATCTTT ATCTAAGGCT 3360

TGGTTTTGGT TAACCATGAC ACCCATCATA AATAATTTyT TGATGATTTC CGCTGGTTCG 3420

CGGTGGATTT TTTTCGCAAT ATCCGCCACG TTCATACCTT CTGTATATTC TAAAACATCT 3480

GGTAATTCAC GGAATTTACG TGCTGGTAcT GCTGGgCTTC GCTGATTCTT GTTGGATGTT 3540

TGGCCTTTTT TCCCTTTTTT GGTTGGAAAC GATTGCGGTT TTGGnTTATT AAAAGTTATT 3600

CCGGGTTAAT TACGATTTGG AACCGCCACC AAAATGGTGG TTACGGGTTT GGCCGCCA 3658

(2) INFORMATION FOR SEQ ID NO: 388:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 519 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 388:

TTTAGAATAT GGGCAAATTT CGTGGGCTTT TTTAACAAGT TCTTCTGCTT TAGCTTGATC 60

TACCCCATCA ATATGAACTT CTAATTCAGC ACCAACGCTA AACCCATCTT CTCCTTGGAA 120

TAAGGATACA CGTGCAGTTA CTGTACTTTT ACCTTCGACT TTTTCTTGAT CCATCACTAA 180

TTCTAATGCA CCGTTAAAGC AAGCACTGTA AGCCGCTGCG AATAATTGTT CTGGATTTGT 240

TGTATTTTCT TTCTTGACAC CTGGCGAAnA ACCGCATAAG AAAATGATTT ATCCGGTGAG 300

TGAACTTCTC CTGCACGTCC ACCCGTATTA ATGATAGTTG TTTCATAGAT TTTTTTCATG 360

TAAATTCCTC TCCTTTGTAA AGTTTTACAC TTCTGAGAAC TTGTTACTTA CATAATTATT 420

GTAGGAAACT TTGCCAGATA CAGCAAACGT TGTGCTTGTG rATAATAAAA AATaACATTA 480

ACGTgTAwTT TTGACGtGAA AAAGGACCCC ATCTATGAT 519

(2) INFORMATION FOR SEQ ID NO: 389:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 697 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 389:

TCAAAATTAT ATCCAACTAT AACAACAATC ATTTCATACC TTATTTCTTT TTATTTTCTG 60

AGTAAAACTA TGCAACATTT ACCACTTAAT ATTGCTTACG CAAGTTGGTC AGGTTTAGGA 120

TTAGTATTAA CAACAATTGT TTCAGTTCTT ATTTTCAAAG AACAAATAAA TTTAATAAGC 180

ATTATTTCAA TTATTTTAAT AATATTTGGT GTTGTTTTAC TAAACACATT CGGATCATCA 240

CACTAATTAC TTTATTCCAA TTTCCTTTTC GGTAGAAACT TTGACCCCTC ACACTCCCCA 300

AAGATGAGCC ACCAGGCAAA AAAACGCCCT GGTGGGCCCA CTTTGTTTAG TGTATATGAG 360

GGGTCTTTTC TTTCCGAAAA GTTTTCTCGG CATAAAAGAT CATCTGATGT TTTTTATCTA 420

AAAATCATTA TTTTTCTAGA AAATCGTATC ATTTTACAAC TTTTATTTTT GACTGTATAA 480

TCTAGTTAAA TCCCGTAAGG GCGCACTTAT ACGTCGATAA ATCGACGTGT GACGAAACTG 540

ACCTGcGTcA GTTTTTTAAA ATTAAATGGA ACGGcTTGAG TTaTyCCGAG CAGTTAAAAA 600

AATtAATATG kGtGAATATC GCTACTGCAT TCtTAAAAAA GTGGgGCAGA TTTTAGCCCC 660

ATAACTTTTT GGATTACATC GACGATACGG CGCGAAA 697

(2) INFORMATION FOR SEQ ID NO: 390:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3869 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 390:

CTTCTGGTkG GCyGcTTTCt GCgGTTTcTG TTGtGCCgTT CTCtTCaGTk GTCaCAcTtG 60

GTTTTTCaGG TATGATTGGC TCaAtCGTGC CyaGTGTTTC aCTTTCaTCT GGATTTTCTT 120

CTtCtGTTGG TGAGACATGA ACTTCTGGTT GGCTGCTTTC TGCAGTTTCT GTCGTGCCAT 180

TCTCTTCAGT TGTCACACTT GGTTTTTCAG GTATGATTGG CTCAAtCGTG CCtAGTGTTt 240

CACCTTCATC TGGATTTTCT TCTTCTGTTG GTGAGACATG AACTTCTGGT TGGCTGCTTT 300

CTGCGGTTTC TGTTGTGTCT GTCTTTTCAC TTGTTTCTGT TGCCTGTTCC GAAGTAGGTG 360

TCTTTGGTTC CCATTCATCT ATTGTCTCTA AAGGAATGTC TAGACTTTCT TCTGTTGGCA 420

TTTCCACATT TTGATTGCCG CTTGCCTCGC CTTCTACAAA AACATTTTTA ACCTGTGATG 480

TATTTTTTTC ATTCGTTGCG TCTTGATTGT TTAGTTGATA ATTGATGTCA TAACTATTTT 540

CAAATGTTGC TTGATGTTGG CCTGCTTCTG TGATTGTCGA AGTGTAACGg ACGATAAAGG 600

AAGTAAAGCG TGCTTTATCC CGATAAAAAC GTAAGTTAAA GTCATTATCT GTTACGAAgT 660

CAATTTTGCC ATAACCTTGT TGCTCAAACT CGGCAAGTGA AATATATTTA GTTTCTTTGT 720

CATTCACAAT ATCAAATGTA AAACTCTCTT TATTTAATTG TTGACCACTT CCTTGTCGAT 780

CCGCAATTGA AATATCTTCT GTGACATCGG ATTTATTGAG GTTCACATTT AAAAACCAAC 840

GTACTTGATT TGACTCTCCA GCCAAATCGC CTACTTTATA AAAAAACGGA TAGTCTCGCT 900

CAATTTGGCC AGTCTCTGTG TTAGTCACTC CTTCAATCGT CAAACGTTGA GTCGCCGTTG 960

CTGTTCCGAA ATCCGTTTCT ATCACATTCG GTTGAGAACT ATTGGTGATA AGCGTTTTAA 1020

TCCCGAAAGA AAAATGCCCA TTCACATTAT GTAAAGATTC AACTTTTTCG TTAAATGTTG 1080

CTACAACATG ATCTTTATAG ATAAAAACTT CCCCTAAACC ATTTAAATTG ATTTTTCGTG 1140

GTGAACTATC GTTCTCGGTC ATTCCAACTA GTGCGTCTGG TAAAGTTAAA GTAATAGTAT 1200

CTCCAGGTTT TATTTTCTGA TTTTCTTTGT CACTAAAAGA TACTTTAATT GAGGTCATCT 1260

CTCCTTGATA TAAAGTAGTG TGATCTAATT CTACTTTGTC AACGATTGAA CTTTTGCTCA 1320

ATTCGGCGCC AAACGAAATT GGACCAATCA ATAACGCCCC CGCAATTGGT AGAATCATTA 1380

CCAACAGTAA CACTAAAAAT TTTACACTTT TTGTCATAAA AAATTATTCA CCCTTCTTTT 1440

CTCTTATTTT TTCCACTTAG TGGTCTTTTT CAGAAGTTGA CAATAGCAAA AAGATATATC 1500

TCACACAATA TTAATGCTCA TTTAAATAGC CTTTTTTTAA AATTTCTAAT TCACTTTCTA 1560

AGGAAAACGG TTTATTATGG TAGTTACATT AGATAACATT TGTTACAAAA TATCTAATGC 1620

TGCCAAATTT GAGTTGCATA AATAACATAT TTCACATCAA TAATGAATAA AAATACTTAT 1680

TTCATCCTCA TTTTGTTTCC CAGGCCATTT ATTCTTGTTA TAATAAAGTT TATGAGATAA 1740

GGAGGaAAAG AATGAAAAAT TCATATTTTG ACGGTGGTTT AGCAACGTAT ATCGGGACTT 1800

CAATTTTAGC AaCCTTGATT ACTGTCTTTA CATTAGGCAT TTGCGCACCT TGGGGAATTT 1860

GTATGATGTA CAACTGGAAA ATCAAGCATA CCGTTATTGA TGGTAAACGC CTTTATTTTG 1920

ACGGTACAGC TATGCAATTA TTCGGTCATT GGATTAAATG GCTTTTATTA ACTATAATTA 1980

CTTTAGGAAT TTATGGTTTT TGGTTGAATA TCCGTTTGCA ACAATGGATT ACAAAACATA 2040

CCCACACACT TTCATAAAAA GAGGAAGGAT GAACTGAGTT ACGGTCCATC CTTCCTCTTT 2100

TTATGAATAT CTAACAGAGT GTTATCCTAC TCTTTGAAAA AACAAGGAAT TTATGCTATA 2160

CTTTGATAAG TATCAAGGGA TACTTATGTT ATTTCGGGGG CGTTACGGAT TCGACAGGCA 2220

TAGTTGAGCT TGAATTGCGT TTCGTAGGTT ACGGCTACGT TAAAACGTTA CAGTTAAATA 2280

TAACTGCTAA AAACGAAAAC AATTCTTTCG CTTTAGCTGC CTAAAAACCA GCTAGCGAAG 2340

ATCCTCCCGG CATCGCCCAT GTGCTCGGGT CAGGGTCCTA ATCGAAGTGG GATACGCTAA 2400

ATTTTTCCGT CTGTAAAATT TAGAGGAGCT TACCAGACTA GCAATACAGA ATGCCTGTCA 2460

CTCGGCACGC TGTAAAGCGA ACCTTTAAAT GAGTGTCTAT GAACGTAGAG ATTTAAGTGG 2520

CAATATGTTT GGACGCGGGT TCGACTCCCG CCGTCTCCAT TTTTAGAAAA GTAGAGAAAA 2580

ACAAAAAAGA GCTAAAACCT TATAAAATAA AGGTTTTAGC TCTTTTTTTC TATAAGCAAG 2640

AGAAGGCTCG TTACCTCCTC TTACACTTCC TATCACATTA TCTTTCTAAA ACCACTTCAT 2700

TTTCTTTTTC TATTGGTGAT TTTGGGTATT GCaTAGAACa TGGCTAAAAA TATTATTGTT 2760

ATCAAAAGAG TTTTTAAGAT TTCAATTTCT CCTGAATCTA TCCAAGTATA CACAAAAGCG 2820

ATGATAAAAG TAATTACTAT TTCCGTAACA TGATCTTTTA AATTCATTAG CTTAGCCCAC 2880

ATTGAACTGC CCATAAAGCT AGTTGTCCAG CTACAATAAT AGCATTCGTC TTTAACCCAT 2940

TGGCTTTAGC GAAGCGTAAT ACAGTCACTG CTAATTCTTT cCAGGCTTTC TTTTGTGCAG 3000

CTTTTACAAT TGCGCTGATA CTAATCATTG CAAAAAACTC ATCTTTGATT TTGTTAGCAA 3060

CGCAACTTCC TAATGCATTC CAATTAAAAC GTGCTTGAAC TACTGTTGAA GCATCAGAAA 3120

TATTTACATT TAATTCTTCC TCAACTGAAG AAATCTCTTG CTGAGCAAAA TTATTTTTGA 3180

CATAGTTTTT AAAGTTATCT TCCGTCACTC CATTCGCATA CATTTCTTCC AGTCCTTTTG 3240

CTACAGCCTC AACTTCTGAA TCCTCTAGCT GGTCAGATGC ATAAACACTT GTACCAGCGC 3300

TGATAAAAAT ACCTAATGTC ATTAGTAATG CCATCGCCAA ACAAAAAAAT TTAGCCTTTT 3360

TTGTCACCTG TAATTCCATA AATAAATCCC tACTTTCTTT TTCCTTGTTA TAATCATAAC 3420

AAGGAAAAAG ATAATTTATC ATCGAAATCC CTATTTGATG CTTATATTTT CCTATATATC 3480

CTAGAGCTGA TTTTTCTCAC TAGTGATCAC ATCAATTTGa ACTAATGaTT CAAAAAAAGA 3540

CACAGGTTAT TTGAGGTACG TTCGAGGTAA TGCTAGGTTC TATTTTGAAA ACAACGCAGC 3600

TCTCTTGaAA TATTTTTTGa GCTTGTCTaA ACCTTCTGAA ACGAATATAG AAAATAAAAA 3660

AAGAGTCCCC TTTCTCAAGG AAACTCTCAA TTGCTAAGTC AAAATAACTT CACGktATCT 3720

AAATCTTTkA AATCAATCCT ATGGTtCGCG GrACAAAGGa GCTAAAAaCT TAATAGAATA 3780

AATCGCACAA ATTCCGACAA TAATATAGAT AATTTTTGCT AATGGTGTGn TTGCACCGCC 3840

ACTAATTGTT GCAACTAAAT GCAAAATTA 3869

(2) INFORMATION FOR SEQ ID NO: 391:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1347 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 391:

GGAGGAAnCG AATGGTAAAA AACGATTAAT TATTGAAGCT GTGGATTCTG ATAATGATTT 60

aACAATTGAA CAATTGTTTG AAAATTATAA AGAAGAGaAG GTGCATGTAA CTATTCAAGA 120

TTTAGGAAAT GCGGTAGGAA ATGAAAAATG GTAAAAGTTC CTCATCAAGG TGATATTTTA 180

TTGTTGAATA CGGCTCCTAG ATCAGGACAT GAACAAACAG GTAAGAGACC GTATATTGTT 240

TTAAGTCATG ATATTATTGC AGATTACAGC AATGTTGTTA TTGTAGCTCC TATATCAAGT 300

ACAAAAAGAA ACTATCCTTT ATATGTTTCT ATTAATCCTT CTTACGGAAT GAAAACATCA 360

GGTAAAGTAT TATTAGATCA ATTAACCACC ATCGATTATG AGGCTAGACA ATGCGTGTTT 420

TTAGAAACAG CACACGAAAA ATTAATCGAT GAATTATTGC TAAAAGTTAG AACCGTATTT 480

CAAAAAGTAA ATAAAACAAA TAAATTTTAA CTAATAAAAA GGCGAAAAAA CGAACAAGTC 540

AAAAGTTTCC GACTTGTTCG TTTTTTTCTA TCGATTAAAT GCTCCCACCA GAAAGTAATG 600

ATACAGGTAA AATATAATCC TTACTTATCT TTTCTCCCTT AATCTTTTTC AACAAAACAT 660

CAACAATTAG TTTTGCAAGT TCATCAATAG GTTGTTTAAT TGTGGTCAAT TGTGAAAAAA 720

ATTGTTCAAT AAATGATGTT CCATCGTAAC CAATAATTTT CACATCTTCT GGAATATTTA 780

AGTTGAGTTT TTTTACAATT TTCATAGTAA GAATAGCAGT TAAATCATCT GAAATAAAAA 840

TGCCATCTGG TTTATTAGTT GaAATAATCG ATTTAATTTC CATCTCTCGA CGAATTGTAG 900

ACAAACTATT AGGAACTTTA AAGACTTTTC CTTCTGGAAT TTGGAAAGAA AATCCTAAAG 960

CACGTAAACC TGTAGGAGAA TCTGTATTAT CATTTCCAGT aATCATAATA ATATTTTGaC 1020

AACCATTTTT CCGTAAAAcG CTTGCTGcCA ATTTACCACC TTCaAAATTA TCTGATGCGA 1080

TGATAGGAAC ATTTGGTGCT AAATTTCTAT CAAAAGCAAT AATTGGTGCT TGAACACGCT 1140

CATAGTCAGC AATCCCAAGA TTATGACTAG AAGAAATAAT ACCGTCGACT TGATTTGCTT 1200

CAAGCATTTC AAGATATTCA CGTTCTTTTC CAGGATCATT TTCGCTATTA CAAATAATTG 1260

TTTTATATCC TTGATTAAAT aATTCTATTT CAAGGGTATT CAATGAGTTC CGCATAGAAT 1320

ATGTTACTGA TATTAGGGAA AATGAGA 1347

(2) INFORMATION FOR SEQ ID NO: 392:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5385 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 392:

TAACCTGATA ATGTAATCGG GTATCGTCTT GACGACCTTC TAAATCAAAG GCGCGCTGAT 60

AAAACGTCTT AACTGCTTCA TTTCCATAAA CAGTTAAACG ATTGGGAATA TCTAATGCAT 120

AACCAGACAT TCGATAAGCC AAATCCTCGC CGTAAAAATG ATCGGAATGC CAATGAGAGA 180

TTAGTAAAGA GTCAAAGTCT GCAAAACACA AGTCGTACTG CAACAGGTGT AAATaAGAAT 240

CCCCTGGAAA ATcAATCAGC AGGACATCAT CAATAATTAA TTGAGTTTGT GTACGAATTT 300

CTCGTCCTTT TTTCTCTCGA GCGTACTTAC ATAATTTACA ATTACAAAAA ATTCCTGGAA 360

TCCGTTCTGC TGCACCCGTG CCTAAATAGT TGATTTTCAT TGGGAGtACC TwACCtTTcT 420

TTGaCATTCT TCATACTTAG TAGATAACGC TGAACCTTAA AGTAAATGGG CATGGAAACT 480

gAAACaTTTG TTGTGTCCAT CTTACCTTTT TTAGAGAAGA ATTGAAATGG AATTTTTTAA 540

GTAAATATGC GTAAAATrGA AAAAGCTAGA CGCCTTTTTT GATAAGCTGA ATTTACTGCG 600

AATGCTCGTA CTATTTCCGA AAAAAAGTTT TACCAAAGGC ACTTTCATAT CCATCTGCTA 660

TTAGTTATAG AAAAATAACA AGATGGACAT AGCTTATTAG AGCAAAAATA ACTACGCTAT 720

TGATAAAGAG TGAGGCGTGT TTTGAACAGG TTGCGTGGGC AAATGAATTT GAATAAAAGG 780

GATGACAAAA GATGACTATT GAAAAATTAG AATATTTTTA TACCATTGCA AAATATAATA 840

GTATTTCCAA AGCAGCCTCT GAACTCCATG TTAGTAAATC AACCTTAAGT GCGTCATTAA 900

AAGATTTGGA AAGTGAATTG GGCCATTTAC TTTTTAACCG TAACGGTAAT TCTTTGACGT 960

TAAATTCCTA TGGTGATAAA ATCGTCCAAT CTGTATATAT CATTTTAAAC GAGGCTAAGA 1020

AAATGAAACT TAATTTGCAT GAAATGATTG AGAATCCAGT GATGCGTTTA GGTTTTGGAA 1080

ATACGTCCTT GATGTATAAA GTGACAGAAA ATGAGGATCA ACTGAATCGT TTTTGGGAAT 1140

GTTACCATGG TTCATCGTTT GAGTTATTGA ATAAATTGGA GAATCATGAG TTGGATTTTG 1200

TGATTACTTC TGCAGATGTG AATAGTCCAG TTTTAAAGAA AGAACAATTA ATTGAACTAA 1260

AAATGTATCT TTGTGTTTCA AGAGAAATTA AACAAGAAAT TGAGGAAGAA GGCTTCAGCT 1320

GTTTGaCCAA CTATCCATTT CTATTTTTGC CGCATCACTT AGATCATTTA GAAGCTACCA 1380

AGAGTGTGCT AGAAATGTTG CAATTAACAA GTCCGCTTGT TTGTTGCTAC GATACGTTGA 1440

TGCTTACTCG TCTCATTGAA AAAAGTAAAG GAGTTTATGC GGTGATCTCT CTTCGGAAAG 1500

AACAACTGCA GGAAATTGAT TCTAAACTCT TCTTTTTACC GATTGAACGA AAACAAAAAT 1560

TTTATTTATA CCGAAATGTT TCCAGTTCGG TTTTTGTCCA ACCTGGTCAA ATCAAGGCAA 1620

CTTTACAAAA GTTGCTAGAG ACTTAGCCAT TAAAATTACT TTGGAATAGT TACCAAAAAT 1680

AGCTTTATTC GCTTAGATTT ACTTCTGCTG CGGATAAGGC TATTTTTGTT GTTCATTGTT 1740

TAAAAGAGAT GCTCAATTAT GGGGCTGTTC GATATTTCTG AACAGAATGT TCAATTATTT 1800

AATAACTTAA GAATCGATAA ATCGTATCTA TTACGGTTCA ATTTAAAGGG GTTCTGGTGT 1860

ATAGTTTGGC TATCTAAAGA TTTACTTTTT TAAATATAAA ATAACTAAAT GGCAAGGAAT 1920

ATAAAGAACT GTTGATACGT GAGTACTTGC GATAACTGTG TCGTTAGAAG TTTAAGTGTC 1980

ATTGTTGCTT TTCCTGTTTT ATGTAGTtGC TAGAAAGGAC TGTTTTTATG ATCGATAGAA 2040

AAAAAGTTAT TTTGTATACT TGTTGCATGA GTTTGTTTGT AGTGACCATG GATGTTACGG 2100

TTGTCAATGT GGCGTTACCA TCCATTCAAA GTGATTTTCA CACGAATCTG TCTACATTAC 2160

AGTGGGTAAC AGATGGCTAC ACTTTAATGG TAGCATCCTT ATTATTGTTG TCTGGGTCTA 2220

CAGCAGATCG AATTGGCCGT AAACGAGTCC TTCAATTGGG CTTAGCCTGT TTTGGTTTAG 2280

CGTCTTTCCT ATGTGGGATT TCGCAAACCC CAGGGCAATT GATCGCGTTT CGCATGTTGC 2340

AAGGGATTGG TGGTTCTATG TTAAATCCGG TAGCAATGTC TATTATTACA CAGGTATTTA 2400

CCGAAAAGTT AGAAAGAGCG AAAGCAATTG GCTTGTGGGG CTCTGTTACA GGGATTTCCT 2460

TAGGCATGGG CCCGATTATT GGTGGACTGA TTGTTTCTTA TTTTAGCTGG CGGTATGTCT 2520

TTTTTGTAAA TGTACCGATC ATTGCTGCGG CAATCATCCT TACACAAAAG TTTGTACCTG 2580

AGTCAAAAGT AGAGAAGACG GTGAAAAATG ATTTTGTTGG TCAAGCATTG ATGATTCTTT 2640

TTCTATTTAG TTCTATCTAT TCCATTATCG GACTACCTAG AAAAGGGCTT TTCGCGCCAG 2700

ATATTTTAAG TACTGGGATA ATCGGCTGCT TAGCCATTGT TATTTTCTTT ATTTATGAAT 2760

ATAACATTGA CAATCCGTTA ATCAACCCGC GTTTCTTTTT ATCTATTCCA TTTACATCGG 2820

CTTCTTTTTT AGCTATTTTT GGCTTTATCA TATATAACGG CTATTTATTT TTAAACACGC 2880

TATATTTGCA AGAGATGAGA GGCTTCAGCC CGTTGGAAGC TGGCTTATCA ACCATTCCTT 2940

TGGCTTTTGT TAGTTTTCTT GTTGCACCGA GAGCCGGCGA AATGGTAGGG AGAATAGGGA 3000

CGAAACGTCC TATTATGCTT TGTGGTATTT CAATGTTGGC TGTTAGCTTT TTACAATTAT 3060

TTGTAACTAA AACAACGCCT ATGATTATTT TATTTATTAT TTATATCTTT TTAGGCATTG 3120

GGTTTGGGAT GTTAAATTCA CCGATTACCA TTACAGCGAT TGAAGGAATG CCACTTTCTC 3180

AGTCAGGAAC TGCAGCAGCC ATTGCGGTGA CATGTAAGCA AATTGGCAAT TCTTTAGGGG 3240

TGGCGCTACC AAGTCTTTTA ATTACAAAGC CTATTACTAG TTCGCTTACT CGAACACCTT 3300

TTACAAACGT ATGGCTTTaT TTGGATGTTG CGCCATTGCG ATTATCtTTT TAAGTTATTT 3360

aTCCAATTCT CCGCTAGCGA AAAAGTCATT GCGCAGAGTT AGGTTTTATT TTTAACAAAT 3420

CAAGATTTTT TTGAAGGGAA AGGAAGACAA ACTTAGATGA AAAAAATAAA TAACCGATTA 3480

GGACTCTTTT TATTTTTTGT TTTACTCTTT TGGGCAAAAA CGATTCTGGC GTATTACCTA 3540

GATTTCTCAT TGGGTGTTAA GGGCAGTCTC CAGCGAGAAA TTTTATGGTT CAATCCaGTT 3600

GCTACTACGC TGTTATTATT TGGTTTTAGT TTGTATTTTA AAAAGCATAA AACATTTATG 3660

CTTAGCTTAA CTCTTTtAAA TGTTTTAAAT ACATTAATTT TATATTTGAA TATTATTTTT 3720

TATCGTGAAT TCACTGACTT TATTACCATC CAATCTGTTT TTGGCTTTTC AAAAGTTTCA 3780

GAAGGAATTT CTGGTGATTT ACTTGCCTTA ATGAGGGTCC ACGATGTCTT TTATTGGCTA 3840

GATACTTTAG TGCTGCTCGT TTTGGTATTT TACTCTTTTA AACAGCGACG GAATCATCAA 3900

GTCTTTGTGA CCAAACCTAT CAACAGAAGA TTGGCGATCT CAATTAGTTG TTTAAGTGTC 3960

TTGCTCTTTT CGGTTAACTT ATTTTTAAGC GAAACCGATC GTCCTCAATT GCTTCAAAGA 4020

ATTTTTGACC GTAATTATAT TGTAAAATAT CTTGGCTTGG ATGCATTCAC AGCCATCGAT 4080

GGTGTCAAAA CAGCCATGTC AAATGAGGTA AGAGCAGAAG CTAGTTCAAA CGAATTACCT 4140

GAAATCATGA AATTTACCAA GCAAAACAAT TTACCGCTTA ATCCACAGTA TGCCGGGATT 4200

GCCAAAGGAA AAAATGTATT TGTGATTCAT TTAGAAAGTT TTCAACAATT TTTGATTAAT 4260

ATGAAAGTCC AAGGCCAAGA GGTTACACCG TTTTTAAATC AGCTTTATCA AAACAAAGCG 4320

ACAATCAGTT TTGATAATTT CTTTCATGAA GTAGGGCAAG GAAAAACGAG TGATGCGGAG 4380

AACATGTTGG AAACAGGAAC ATTCGGGTTA CCGCAAGGCT CTCTTTTTAC ACAATTAGGA 4440

TCAGACAATA CTTTCGAAGC GGCTCCCGCA GTTTTAGGAC AAAAAGGCTA TACmACCGCT 4500

GTTTTCCATG GAAATGTGGG GAGTTTTTGG AATCGAGACC ATGTGTATAA AAACTTAGGA 4560

TATCAATACT TTTTTGATCG GTCATATTTT TCAAAAGATG ACAAAATGCT AGGTTACGGT 4620

ATTTTAGATA AACAATTGTT GCGTGAATCA GCTAGTGAGC TTGAACATTT ACAGCAGCCA 4680

TTCTATGCGA AGTTTTTAAC TGTGACCAAT CATACCCCTT ATTATACGGA TGACCAAAAC 4740

TTCCCGTTTC CTAAATTGAA CACAGGCAAT TCGATTGTGG ATGATTATGT TCGAACAGCG 4800

CACTATTTGG ATCAATCGCT AGAACAATTT TTCCATTACT TGCAAACAAC AAGTTTATAC 4860

AATCATTCTA TTTTTGTTAT TTATGGTGAC CACTTCGGTA TTTCAGATGC GAACAACAAG 4920

GACTTATGtA AAgCTTTCAA CCGCGATCCA AAAACGTGGA CAAACTATGA CAATGCGCAA 4980

TTGCAAAGAG TACCTTTGAT GTTTTACATG CCAGGCTATA CGCAAGGGAA AATCATGCAC 5040

GAATATGGTG GGGAAATTGA CGTTTTACCG ACACTGTACC ATCTTTTAGG TGTCGATGAT 5100

AAAAATTATA TCCATTTTGG CACGGACCTT TTATCTCCTC AATATAAACA AGTGGTTCCT 5160

TTTAGAAATG GCGATTTTGT AACGCCTCAA TTTTCGTATt AGGTGGGGAA ATCTACAATA 5220

ATCAAGGGAA AAAATTAGAC CAAGTTCCAG CAGATTTAAA AGCAGAAGTC ACGAAAGATA 5280

ACGATTGGGT GAAAAAATCG TTAAGTTTnT CTGACAAATT AAATCGGGGA AACCTTTTAC 5340

GnTTTTATAA ACCACAAAGG GTTCCAAGAA GTGCAACCAn AAGGG 5385

(2) INFORMATION FOR SEQ ID NO: 393:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1146 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 393:

CGTGTGGTTA TCTCTGCTCC TGCAGGCGCA ATGAAAACAA TCGTTTACAA CGTCAATGAT 60

GATACACTTG ATGCCAACGA TAAAATCATT TCTGCTGGTT CATGTACGAC AAACTGTCTA 120

GCACCGATGG CTTACTTCTT AAATAATGAA TTCGGTATTG AAGTTGGTAC CATGACAACT 180

GTCCATGCCT ACACATCAAC ACAAATGCTA TTAGATGGCC CTGTTCGTGG CGGAAATCTT 240

CGTGCGGCAC GCTCTGCTGC AGACAATACT ATTCCCCATT CAACTGGTGC TGCTAAAGCA 300

ATTGGCTTAG TTATTCCAGA ACTTAATGGA AAATTACAAG GCCATGCCCA ACGTGTTCCC 360

GTTGTGGATG GATCCTTAAC AGAATTAGTT AGCATTTtAA AAACAAAAGT AACAGCTGAT 420

CAAGTAAATG AAGCAaTGAA AAAACACACC ATtGATAACC CtTCTTTCGG TTATGAwGAT 480

CGTGAAATTG TTTCTGGAGA CATAATCGGC ACGACAGAAG GTTCGATTTT TGACCCaACG 540

CAAACagAAg TAACGACaGC TgGCgATTTc CaAtTaGTgA AAAcTgTGGc TTgGtAtGAc 600

AATgAAtAtG GtTTTaCTTg CCaAAtGATT cgCTTatTGG aAAAAtTcGC tAATTTatAA 660

TTCCaATGaA taAAAAAcTg CCtTGcaTAA AtTaATgCaA gGCaGTTTTT TTAATGTTTA 720

CTTTGACAAC TTTGACAAAT ACCATGAAAA GTTAAGCGAT GATCTTTGAC TAAAAAATGA 780

TACCGACTTT CAACAATTTG CTCTACTTCA CCAAGCAAGT CTTCTTCCAC TTCCTCAATT 840

GTGCCACATT CTAAGCATAA CAAATGGTGG TGAAAGTGTT TGGCGCCTTC TTTTCTTAAA 900

TCATAGCGAG CTAAACCATC GTTAAAGCTC ACTTTATCAA CGACTTTTAA GTCTGTTAAA 960

ATTTCTAGCG TTCGATAGAC TGTTGCTAGA CCAATTTCCG GACTTTTTTG TTTTACCAAG 1020

AAGTAAATTT CTTCTGCTGA TAAATGATCT TTTTCAATTT CCAATAATAC AAGTACGGGT 1080

GCCTCCCGGT GTGGGGGTAA ATTnAAGnCn GAATCCnGGA ATGGGTTTTG GCCTTTTAAA 1140

GCCGCC 1146

(2) INFORMATION FOR SEQ ID NO: 394:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14141 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 394:

TGGCTATGTC AACATTGCgA CAaGgCCGTw AaGCTCTGGt GCGCCAcGGC GGAAGTGCCT 60

ATCGAAaCCG CCAwmAAATT GTTGAAAGAC ATGGGGGGCA GTTCCATTAA ATATTTCCCG 120

ATGAAAGGCT TGGCTCACAA AGAAGAATAC CAAGCAGTAG CAGCAGCTTG TGCCAAGTAT 180

GACTTTTATT TAGAGCCAAC AGGAGGCATT GATTTAGAAA ACTTTGAAGA AATTGTTCAG 240

ATTGCGGTGG ATGCCGGCGT TAAAAAAATC ATTCCTCATG TCTACAGCTC AATTATCGAT 300

CAAGAAACAG GGGACACTAG AACAGAAGAC GTGAAAACGT TACTAACCAT GATGAAAAAG 360

ACATTGAATA AGTAAAGGGC CAATCTTTAG TTGACTTTCT CCACCAGATT TGCTACTCTC 420

AAGTGGAGAA TAAATAAAAA CCAACAAATG AATGTGGGGT ATTTTGGTGA GTGTCGTAGG 480

TATCTAATTG ATAAGTTGAA TAACCAGTGA ACCAAGCAAG GTTTATTTGG TGTGCTTAGA 540

AAGTGATACC TGTCAGTCAT CTCAATTACT AAATACCTCC TTTAAAAGAG GTTTCTTTAG 600

TAACCTTTAT CCAATGATAA AGAGGACAGC GGTAGAACGC TGTCCTCTTT TTTCTGTAGA 660

AACTGGTTTT TATTTAAAGG CGCCAGGTAT TCTTTTCGTG ATAGAAAGGA AGAAAAAACA 720

AATGAATCAT GaAACCATTC AAAAATTACA ATTTACAACA ATTCTTAAAG AAGTTCAAAC 780

ACGTGCGATT GGCGAGTATA GTAAAGAACG GTTGGCGAAA ATGGTCCCTG CAACACGTTT 840

AGAAACAGTG CAGTCCCGCT TAACGGAAAC AACCGAGGCC CGCTTAATTA TTGATAGCGG 900

ACAACACGTC CCTTTCATGG GCTTGAGCCA AATTACCCGC TTGCTTCAAC AAGTGAAAAA 960

AGGCTTAATC CTTACACCAA ACGAATTGAT TGAGGTCGCT GACTTTTTAA GAAGTAGTCA 1020

GCGAATCCAG AAGTTTTTTG AAAAAAATCA ATACCAAACA CCACGCCTAT TTAGTTACAG 1080

CCAGCACTTA GCTGATTTCA GAGCAATTGA AGAACAAATT TATACGAAAA TCCATAATCA 1140

AAAAGTTGCC ACCGATGCTT CTCGACAACT TCGGAAAATC CGGCGTCAAA TCAATGAATG 1200

CCAACAAGAA ATTGAAACAA AAGTCACCAA ATTTTTACGT AATAAAAATA ATCAACTGTA 1260

TATTCAAGAA AATATGATGG TGAAAAAAGG GGAACATTAC ACGGTACCAA TCAAAGCTAG 1320

CTATAAAAAT AAGGTTTCAG GCACAATCAT TGAGCAATCC AATAAAGGAC AAACGGTCTT 1380

CATTGAACCT GTAGCAATTA GTAAATTAAA TGAACAATTG ACACTCTTAA AAGCGGAAGA 1440

AACAGCGGAA GAATATCAAA TTTTGGCAGA ACTAACAGGG TTAATCAATG AGCAAGAACG 1500

ACTAGTCGAT CAAGCCGTCG ATACAATGAC CACCTTAGAT ATTATTTTTG CACGAGGAAA 1560

GTATAGTCGC GAAATTGGTG GGATTACGCC AAAAGTAAAC AAAGAAGAAC gCTTGCGCAT 1620

TGTTCAAGGA AAACACCCTC TTTTATTAGA ACACGCAGTA CCATTGACGT TTTCTTTAGG 1680

AGCAGAGTAT CGAGGACTTG TCATTACAGG TGCCAATGCT GGTGGTAAAA CGGTTGTCTT 1740

GAAAACAGTT GGCTTATTGA CTGTCATGAC ACAATTTGGG TTGCAAATTC CAGCACAAGC 1800

AGGCACAGAG ATTCCAGTAT TAGATGAAAT TTTTGTCGAT ATTGGTGATC AGCAAAATCT 1860

GGAAAATGCG TTAAGCACTT TTTCTGGGCA CATGAAAAAT ATTGCGGAAA TGTTGCGTAA 1920

TGTTAAGCGA CACACGCTTG TTTTATTAGA TGAAATTGGC AGTGGTACCG AGCCGAACGA 1980

AGGAGCTGGC TTAGCGATTG CTATCATGGA AACGATGTAT CAAAAGGGAG CCTTGGTAGT 2040

TGCAACGACC CATTATGGTG AAATTAAACG TTTTGCCCAA GAGCATGACG ATTTTGTGCC 2100

CGCCGCAATG GCCTTCGATC GTGAAACATT AACGCCCAAA TATCAATTGA AAATTGGTGA 2160

AGTTGGCGAT AGTCAGGCGC TTTGGATTGC ACGTAAAATG AAGATGGAGC CACAGCTCAT 2220

TGAAAAGGCG GCTCATTATA TTCAGCGGAA AGAATACCCT ACTAATCGAA GTCTATTTAA 2280

ACCCCTGAAA GCAAAAGAAA CGACCTTTGC CACGTTTGTA GAAGAAACAC ACCGTTATCA 2340

AAAAGGTGAT CGTGTCCTAT TGACAGAAAC ACAACAAATT GGGTTGGTTT TTACAGATGA 2400

GGGTGAACAA GAAATCCAAG TATTTGTGAA TGATAAAATT GAAAATGTAC CACGAAGACG 2460

GTTGAAATTA CAAGCGAAAG CACAAGATCT ATATCCGCAA GACTATGATT TAGAAAGTTT 2520

ATTTACAGAT TTCCATGAAC GAAAGAAACT CAAAGATATT GAACGCGGCT CAAAAAAAGC 2580

GCAGAAACAA TTACGCAAAG AGGCCGAAAA AAGAGCAGCA AGACGCGATA AATAAGGAAC 2640

TTAACAGAAG CAATACAACA ACTTAACACT TTGTTTACTT GTTATTTATC AGAAATCAAC 2700

TAAGACTTGT TATAGTCAAT GTATGGGTAG ATATGaAGGA GGAAACAAGG AAATGAAGAA 2760

AAGAGCTTTG CTAGGGGTTA CCTTATTAAC ATTCACAACA TTAGCGGGTT GTACAAATTT 2820

ATCTGAACAG AAAAGCGGCG AAAAACAAAC AGAGGTTGCT GAAGCGAAGG CAACTGAATC 2880

TGAAAAAGCA TCAGTAAAAA ATGTTATTTT TATGATTGGA GATGGCATGG GGAATCCGTA 2940

TACAACGGGC TATCGCTATT TCAAAGCCAA TCACTCAGAC AAGCGTGTTC CCCAAACAGC 3000

TTTTGATACC TATTTGGTCG GACAGCAAGC CACTTATCCA GAAGATGAAG AAGAGAATGT 3060

CACCGATTCA GCTTCCGCAG CGACAGCGAt GGCTGCCGGA GTGAAAACCT ATAATAATGC 3120

TATTGCACTC GATAATGACA AGTCCAAAAC AGAAACAGTG CTCGAACGTG CGAAAAAAGT 3180

GGGGAAATCA ACGGGTCTTG TAGCAACATC TGAAATAACA CATGCAACCC CTGCTGCATA 3240

TGGCGCACAT AATGTTTCAC GCAAAAATAT GGCAGAAATC GCCGATGACT ATTTTGATGA 3300

TCAAATCGAC GGACAACACA AAGTCGATGT GTTACTTGGC GGCGGCTCCG AATTATTTGC 3360

CCGGAAAGAT CGTGATTTAG TCAAAGAATT TTCCCAAGCG GGTTATGGTC ATGTCACAGA 3420

CAAAAAGTCG TTAAATGAGA ACCAAGACGA CAAAATTTTA GGCTTGTTTG CACCAGGCGG 3480

GCTACCTAAA ATGATTGACC GAACGGAAGA AGTCCCTTCA TTAGCTGATA TGACAGAAGC 3540

GGCTCTTCAA CGGTTAGATA AAAATGAAAA AGGTTTCTTT TTAATGGTTG AAGGTAGTCA 3600

AATTGATTGG GCCGGGCATA GCAATGATAT TGTTGGCGCG ATGAGCgAAA TGCAAGACTT 3660

CGAAGCGGCG TTTGAAAAGG CCATCGATTT TGCCAAAAAA GATGGTGAAc ATTGGTGGTT 3720

ACAACTGcAG ATCATTCAAC AGGGGGCTTG TCTTTAGGCA AAGGAGATCA ATACAACTGG 3780

TTGACGGAGC CTTTaCATGC GGCAAAACGC ACGCCTGATT TCATGGCAGA AGAAATTATT 3840

AAAAATGGTA ATGTGGAAAA AACAGTGACT GAGTATATTG ATTTTCAATT AAGTGAGGCT 3900

GAATTGAAAG CAGTGAAAAC AGCGTCGGAG TCAAAAGATG TTGAAAAAAT CGCTCAGGCA 3960

TTAAGAAAGA TTTTTGATGA ACGTTCGAAT ACTGGTTGGA CTACTGGCGG ACACACAGGA 4020

GAAGATGTAA ATGTCTATGC TTATGGCCCA CAAGCAGAAG CTTTTTCAGG ACAAATTGAT 4080

AATACAGACC AAGCGAAGAT TATTTTTGGC TTAGTAGATG GCACCGGGCA AAAAGCTGAG 4140

ATTAAAGATA AAGGTATTGG CAAATAACAA AATGAAGAAA AAGCAGAACG TATAGAAGTT 4200

CTGCTTTTTC TTCATTGAAG GCATAAAAAT TTTATCTCTG TTATAATGAA GGTAACCAGA 4260

TAAAAGTGAG GTCAAATAAT GAAAAAAATG CATGTTGAAA GTGAATTTGC ACCATTGCGT 4320

TCAGTTGTTT TAGCACAATC GCAATTTTGT TTACCAGATA AATTTGATGA AGCGGATACG 4380

ACTTTTTTAA CAGAGGAAAA TGCCCGTTTA ACACAGAATA ATGAAGGGCG TGATTTAGCA 4440

GACTTTGCAC CAGAACAACA AGTGCGTTGG GAAAAAGAAA AAGAAGTCAT GCAAGGAGTA 4500

TTAGAATCTT ATGGCGTAGA AGTATTTCGT CCACGTTTAT TGACAACCTA TGAAAAAGAA 4560

CATGGCAAGG AATTGGGCGT AGGATATAGT AACTTCTTTT CACGAGACCC TTTTTTTACA 4620

ATTGGCAATC TCTTAATAGA AGGAAACTTA AGATTTCAAC ACCGTCGACA AGAAATTTTA 4680

CCAATTCGTC CGATTATTCA GCAGTGGACA CAAGAAGCAG AAGGCTATTA TTTTGctGCG 4740

CCACAACCTG ACATTTCTGA AGGGGCTTTA AGTGAAGCAG GTCCCTTTAT TGAAGGCGGC 4800

GATGTTTTAG TTCTAGGAAA AACAATTTTT GTGGGCTACT CTGGCTTGGC AAGTAACCTA 4860

GCAGGAATTC AGTGGTTAGC CAATATGATT GGTCATTTTG GTTATGAAGT GGTGCCGGTT 4920

CGGTTACATC CTCATATTCT CCACTTAGAT TGTGCGTTAA GCTTGTTGAG AGAAGGCTTG 4980

ATGATTGTTT GTGAAGAGGC CTTTTTGGAT GGACTTCCTG CGCAATTAGC AAACTGGGAA 5040

AAAATTCACG TTACTTTGCA AGAAGCAGCG TATTTAGTGA CAAATGGTCT ACCGCTTAAC 5100

GAAGAAACGT ATATAACGGA CCAATCGTTT ACAACTTTAA TTCCTCAAAT AGAAGCAAAA 5160

GGTATCAAAG TTGAGGCAAT TGATTACCAT GTTTCTAGAA TGTTAGGTGG TTCCTTCCGC 5220

TGTACAACGC AAGCTTTGAT TAGAGAATAG TCAGCGAGTG ATTATCAATT TTCGTAAGAA 5280

AAAATGTACG TTTTGTCTAT GAAAACGATA TAATTTTAGT TTTCACTATT TTTTTTAAAA 5340

ATATTGAGGA AATTTTGATT TAATAAATGC GAGAAAACGT CAAAATGTTA GTGTTTGTTA 5400

CTAATGTTTT GACGTTTTCT CGCTGTTTTT TTATAAAATT ACCACTTTAT ATATGAAAAA 5460

AAACAATATC TTTTTTGAAA ATTACGTTTT CAAAGACACA AAAAGTCGTT GGAACCTTGC 5520

TGTTATAGTA AATTCATTTT TTTTGCATGA GGTGTTTAAA ATTTGAAATC ACAAGCAGTC 5580

TTTTTTTCGC TATACTTAGT TTACGTTCTA CTGAGATAGA AAAGTTTTTC ACTTTTTTAT 5640

TCATTAATTA ACCAAGATAT AAAGGAGAGA GTAAAATGAA GAAAAAAACA TTGGTAACTT 5700

TATTTGCAGG AACGGCAATT TTAGGGGCAA GCATCGCTCC TGCAGCAGCA CATGCAACAA 5760

CGACTGGTAC AACACCAGCA AACGTAGAGA TGACAAGTGG TTCATTACCA GGCGGTACGG 5820

GTGATGGAGA TGACGGAAGC ACACAAGAAC CAGAACCAGG CTCGAATACA AATTTTGATT 5880

TGTTATATGT TCCAACAGAA TTCAAATTTG CTTCGACAGA AGTTAGTTCT GATTTATCAG 5940

CGATTTCGTT AGATGCGACA GGTACCCAAA CAAAACGTTA TGCCGTTGGG GATGTTCGTG 6000

GAACACAAGC GGGTTGGTCA GTTACTGCCG GCGTTGCTGA AATGAAAAAT GGCACTGCTA 6060

CTTTAGAAGG CTCAATTACC TTTGCACAAA CAGGAGCAGT AGCGAAATAC GATGAAACAG 6120

CAAAAACATA CTCACGTGAT GTaGCTGCCT TTGCAGCAGA TCCAGGAAGT CCTGAATTTG 6180

CTGGAACAAC TATTCCAGTA GGTGGAGCAG CTGTATCAAT TGCTACAGCA GCAGTTGGTA 6240

AAGGTCAAGG TACATGGGAT TCTGAATTAT CAAATGTAAA ATTAAACATT ACAACACCTA 6300

GTTCACAAAT TACAAATGGT GCGTATACAG GTAACGTAAC TTGGACATTA gTGcTGCGCC 6360

ATAAAAAGTT ACTAACGAGA ATACTGCATT ACTAGAAAAG GAGctGCGCA GTGGTGCACG 6420

GCTCCTTTTC GATACATAGG AGGAACAAAT AATGAAGACA ATGCGCTGGC TAATCGTCAC 6480

AGTTTTTATT TGTGTGATGG CATTAAGCAT CCCTTCTGCT GTACAAGCAG AAGAAGCGGA 6540

TGACAACATG AATACAAAAT TTACAGTTAA ACCCAATATT CCTGAGAATC AAGCGAAAAA 6600

TACGCAATCC TACTACGATC TCATTGTAAA ACCAGGACAG AAACAAGAAG TCGAATTATC 6660

ATTAATTAAT AAAACAGATG AAGAATTGAC GATTGATTTA AATTTGGCGA ATGCTGTGAC 6720

GAATGACAAT GGTTTAATTG TGTACAATGA TTTCGAGAAA AAACCAGATT CAAGTTTAAA 6780

AGTACCCTTG ACAACATTAA TTAAATTACC TGAGGAACAT GTAAAAGTTC CCGCTAAGAA 6840

AACCGTAACA GCCAAAATGA CTGTCGAAAT TCCAGCGAAT GGTTTCCAAG GATTAGTTTT 6900

GGGTGGTGTT TATGctTCGC TTGCAGAAGA TGAAAAGGAA GAAAAAGGAA AATCAACGGG 6960

TTTAACCAGT CGATATGGAT TTAATGTAGC AATTGCTTTA CGAATGTCAG AAGATACACC 7020

GATGTATCTA GCAGAGACTC TTCAATTAGT TAAATTAGTT CCGACCATTG CTTTAGGAAA 7080

TAAATCGGTC CAAGCGGTTA TTCAGAATCC GACATCCGCT ATTTTTCCTG AAGTACGTTT 7140

AGAAGGAAAA GTAACGAAAA AAGGCGAAAC GAAAGAATAT GCCAAACGGA TTTTACCAAG 7200

TGTTCGTTTT GCGCCAAATT CGACGATGAA CTTCCATTTA GATTTCGGGA AAGAAGATGT 7260

TCGTCCAGGA ACCTATATTT TTGAAGGAAA AGCGGTCTTG AAAGAAGATG AACAACAAGT 7320

GTGGCCGTTT AAGCAAGAAT TTACCATCAG TAGTCAAGAA GCCAAAAAAT TAAATAAAGA 7380

AGCAGTAGTG AAATTGGTCC TACCGACTTG GTGGAATCAA GCTTTTTATG GCTTGTTAGT 7440

TGCAACAATT GTGAGTCTAA TTCTAGCTAT TTGGCGTTTT ACGCAACAGA AAGCAACACT 7500

GAAACAACAA GCCTATTTAC AACAAGAAAA ACAGGCAGCC TTGAAAAATA GACAAGGAGC 7560

ACGCGATGAT GACAATGAAA AGTAAAGGGT CACTTCTGGT GACGTTGGGA ATACTTTTAA 7620

CCGTTGGCAT TGCGAGTCTA ATTGTTTCTT CTGAGAGTTT TGCAGAAGAA GTAGGGCAAA 7680

CGAATATCGG TGTAACGTTC TATGGAGGAA AAGAGCCACT AAAAACGGAA GGTGTCATTA 7740

AGCCAATAGA GCAACCAGTC ACTGATAAAG ATAAAAAAAC GTCACAACAA CAAGACAAAG 7800

TGAGCAGAAA AACCACTGCT AAAACGAATC CGACTAATGC ACAGACGTCA TTACCAAGGA 7860

```
CAGGTGAACG AAATAGCACG TGGCTTTACA GCCTTGGTAT TGCCTGTTTA CTCGTAGTAC     7920
TAACAAGTTT CTATTATTTG AATAAAAAAA GGAAAAAGGA AAAATAATGT GCTGTTTTTA     7980
TAAATAGAGT TAATGTAGAA GGAGGAATAT AGATGGAGAT TACGATGTGG ATTTGGATTT     8040
GTGCTGGCTT GGCCGTCCTT TGTTTTGCTT TATTTTTACT GTTCTTAGTA AATAGCATGA     8100
ATTTTGTACA TCTGGCAAAG CCTAAGCAGT CTTTAGTAAA ACAATCAATG TTTAGTTGGC     8160
TGCTTCTGGT GATGTGTGGC GGCTGGGTAG CGGTGGCTAT CTTACTGAAT ATCGCATTAC     8220
AAACACAATT AACATAAAGA AAACCCGTTG TAAGTTTTTA AGAAACTTGC AACGGGTTTT     8280
TGTTAACTTT AAAATAAAGA ACGGTACAAT TTATCAAAAA AGTACAGTAG ACATTTGTAC     8340
CCAACTTAGA TAGAAAATAA AGCGCTATCA TTATATGATG ATTACAGGAG GCGGACACAG     8400
CTGTGAATTA TAAAATAAGT TATCTCCTTG ACGAGTATGC AGGGCAAAAA ATTGCGCTCT     8460
CTTTGGCTCA ACAAATAACC GATCTTTCGC AACGAGAATT AGTGACGCAA CTTTCGACAA     8520
TAGGCGGAGT AGTTCAACAG AATACCGTCA GCATTCCAAA CCTTTCCGCA AATGAATGGG     8580
GAAAACAACT TTTTCAACAA CGTCAAGTGA TCTATAGCGA AACAGAAAGA CAAGCACTTA     8640
TTTTTTTGCT CACTTATTCA GAAATCGAAG AACTATCGGT GTATCATTAT CAAAATTTTT     8700
TAGACGTTTC GAAAGGCACT ATTTTAGCGG ACATCAAAAA AGTACGACAC CAACTAGCAA     8760
CTGAAAACAT TGCTTTATCC TACGAACGAA AACGAGGGTT TTACCTAGAA GGAGATGAAT     8820
TTCGGATTCG TCGTTTCGGG AAAAATTGGC TTGCCCGCCT CTTACAACAA AAGTCGGGCA     8880
CCTTCGCTTT ATTTTGCTGG CTCAGTCAGC ATCAAATGAG CCAATACGCA AAAATACGTG     8940
ACGGCATTCA ATTAGCCGTC CAGGAAGCAC AGCTTCAACT GGTACCCAGC AGATTGGATG     9000
AAATTAGTTA CTTTCTTGCT TTTTCTCAAA AAAGATTAGG CAAATATGGA ACAGTCGTCT     9060
GGCCTGAGAA CCAGTTAATT CAGTCCTTAC GCGCTTATCC CGTCAGTCAT CGTATTTTAA     9120
GTGATTTAGT TGGGCCAGTG GCGAATGTAG AAACTGAGAG TCTTTTTTTC ACGATTTGTT     9180
TAATGACAGC CTTGCAAGGA GAATTGCGAG ATACGAAGTT AGAGTTTTTA TTAGATATTT     9240
CTGGTGAAAT TATTCGGAAA ATGGAATTGT TGGCCGTTGT TCAATTTGAA CAGCCTCGTG     9300
AATTGCTCAT GAATGTGTAT TATCATCTGG TTCCTGCCTA TTTTCGGATT AGCTATGGCT     9360
TTTACTTGCC GAATGTTTTA ATCAAGGATA TTCGACAATC GTATCGGTCG CTTTATCATC     9420
TTTGTGAACA AGCGTTAgct CCTTTAGAAA AATTAACGAA ACGGTCAATC CCGTCAGAGG     9480
AAATCGGTTT TTTTACCATA TTATTTGGTG GAGAAATTTT CGGGCAAAAA AATGAACAGC     9540
GGCAAGAAAA GCTAAAAGCG CTAATTGTTT GTCCCAGTGG TATCAGTTCT TCGTTGATTT     9600
TGCAATCGGA ATTGCGTCGG CTATTTCCAA TGATTGAATT TAAAGAAACG AATGCGGTGA     9660
GGGAAATAGA GAATGTATCT GAAAAAAGCT ATGACATCAT TTTCTCAACG GTCCCTATTG     9720
AAACAGCCAA AAAGGTTTAT CTTACCAAGC CCATTATGAG CTCTTTGGAG AAAAATCAAT     9780
TGATGCACCA AGTTCAATCA GATTGGCTAT TGCCAGGTAT TTCCATGCCA GATGTTAAAG     9840
```

```
ATATTTTAGA TGCGTTAAAA CCTTATATTT CCTTAAAAAA AGGAGTCACA GAAGCGCAGT     9900
TATATCGTGT TTTACATCGT AAAATGAACA AAATTTTAGA AAAAGAGGAG GATCTGCGTC     9960
CAATGTTATC GGAATTATTA ACAACAGAAA CAGTCCAAGT CCTGCCTGAA GTTGCAGACT    10020
GGCGAGCGGG AATCACAGAA GCTGCCAAGC CGTTATTAAA AAATGGCAGT ATTACGGAAA    10080
AGTACATTGA TGCAATGATT CAAAAAGTCG AATCGTTTGG GCCATTTATC CATATCTGTC    10140
CAGACGTAGC ATTGCCTCAC GCGCGACCAG AAGATGGCGT TAATCAGTTA GGCATGTCAT    10200
TACTAAAAAT CCAAAAAAGT GTCGATTTAT TAGAAGATCC TAAGCATGAA ATTCACCTTT    10260
TTATTTGTTT AGCTGCTTCA GATAACGAAA CACATTTACG GGCCTTATCT AGTCTGACAA    10320
AAATTCTATC AAAAAAAGAA TCATTACATC GCTTAGTGGC GGCTGAAACC GTCCCAGAAA    10380
TACTATCAAT TATCCAAGAA GGAGAGAAAT AAGATGAAGA TTTTTACAGT TTGTCAAAGT    10440
GGTTTAGGAA CAAGTTTTAT GGTGCAAATG AATATCCAAC AAATTTTAGA GGAAGTCGGT    10500
GTTGATACAG ATAATTTTCA AATTGATCAC ACAGATGTCG GGAGTGCCAG TGGCGATATG    10560
GCTGATTACT TCTTTGTCGA AAAAACATTG CAAGATGCTT TAGGGTCAAT TCCTGCTGAA    10620
AAAATCATTC CTTTAAATGC GATTATTGAT TACGATGAAA CGAAGGAGAA AGTCTTGAAA    10680
GTTTTAGATG ACAATGGCAT TGCACATCAA TAAGAAGGAG GAAAAACATG GAAACTTTAT    10740
TGAATTTATT TGTTTCAATT GTTAGCCAGC CAGCGATTCT GGTTTCATTA ATTGCTTTGT    10800
TGGGATTAGT TTTACAGAAA AAGAAATTTT CTGATGTTGT TCAAGGAACG TTGAAAACAT    10860
TCGTCGGCTT TTTAGTTTTA ATTGGTGGCG CAGGGCTATT AGCGAACTCG TTAACGCCCT    10920
TTGCGGAGAT GTTCCAAACC GCGTTAAATA CACAAGGCGT GGTTCCTAGT AATGAAGCGG    10980
TTGTAGCGAT TGCCTTGCAA CAATATGGGA CACCAACTGC TCTGATTATG TTGGTGGGCA    11040
TGGTTGTTAA TATTCTGTTG GCGCGTTTTA CACGTTTTAA ATATATCTTT TTGACAGGTC    11100
AAGCGATGAT GTACGTTTCG TGTTTGACAG CCGTTATTTT AGTGGGGGCC GGTTTTAGCG    11160
TTAGTTTGCC AATGATTCTT TTAGGTGGTT TATTTGAAGG AACTTTATTA ACAGTGACAC    11220
CTGCTCTTTG TCAGCCCTTT ATGAAACAAA TTACTGGAAA TGATAAAGTA GCAATGGGGC    11280
ATACAGGGAA TATTGGTTAT GCGGCAAGTG GCTTTATTGG AAAAGTCTTT GGGAATAAAG    11340
AAAAGTCAAC GGAAGATATT AAAATTCCTA AAAGTTTTGG TTTTTTAAGA GACTCAACCA    11400
TCTCAATTAT GATTTTGATG TCGATTGTCT ATGTGATTTT GGCTTTGCTG GCTGGGACAG    11460
GCTATGTCGA GCATGAGTTA AGCAATGGCG AAAATGCGAT TATCTTTTCA TTAATTCAAG    11520
CAGGAACCTT TACGGCTGGT TTCGTAGTGG TTTTACAGGG TGTGCGGATG GTTTTAGGAG    11580
AAATTGTGCC AGCTTTCCAA GGAATTGCTA AAAAATTAGT GCCTAATTCT AAACCAGCAC    11640
TTGATGTTCC CATTATCTTT CCTTATGCGC CAAATGCTGT CTTGATTGGG TTCTTTGTGA    11700
GCTTTATTGT GGGAACTATT TCAATGTTAG CGATGGTTGG TCTGAATACC ACCGTGATTA    11760
TACCAGGCGT TGTCGGGCAC TTTTTCTGTG GTGCCGCGGC AGGTGTTTTT GGAAATTCCA    11820
```

CAGGGGGAAG TCGGGGAGCG ATTTTAGGGG CTGCGTTTAA TGGCTTATTA ATTAGTTGGT 11880

TACCGTTATT TATTTTACCA GTCTTGGGTG ATTTACAGTT AGCATCCTCA ACGTTTGCGG 11940

ATACAGATTA TTTAATCCCA GGCCTATTTT TAGGCAAATT AGGTGAGGCA GGTCCGAGTC 12000

TTTTAGTAGG AGGAATCATT GCTTTTGTAG TGGTTGTTCT TTTGGCTAGC GCTGTTTTGA 12060

CTATGCGGGA GAAAAAAGTC GCCGAAGAAA ACTAAGTGTT TTGTTCTAAA AAAAGGCAAG 12120

TGGGACAGAA GTCGTTTTGA CTTCTGTCCC mTTTGTCTTT TACCATTGTT CTTGGATAAA 12180

CTGTGCAATT TGTTTCCGCG CTTGTCGTCC TTCTGGCAAG TATTGTTTCA TTGCGAAAAC 12240

GTGTTGCATG CCAGGCACCA ATCGGCAAGT ATCTTGTGGT CGTGGCCCTA ATTGTTTTAA 12300

AAATTGGTGA TAGGAAGGGC AAAAAACTTC ATGTGTGCCG CCCACGAAAA AAAGTGGCGG 12360

AAATTTTTGA TAATCAAAGG ACAAGCAATT TAAAGAAAGT GTGGCGTGAT TGGCCGTGTC 12420

GTCAATCTCT TGGGCAATTA ATTGTAAATA GTTCATGCCA ATCATTTTGT CCTTTTTTTG 12480

ATAACTTTTG GCAACAGGAA TGTCTAAATG TCCATCTAAC CAAGGAGAAA TTAAAACAGC 12540

AGCTTTGGGT AAGCGTTCAG GGAAATCGGT GGCTAAATTC TGCAATACGT TTAAAACCAA 12600

TGTGCCACCA GCAGAATCAC CTAATAAAAT AACATCTTCG GGATCATAAT GCTGTAAAAT 12660

TTCTTGATAA GCAGCATAGC AATGATTATA AATATCAGCA AAAGACTTAT TTGGTGCTAA 12720

TGGATATACA GGGATAAAAT AATCTGTTGC aAAAGGGAGA TCTTCAATTA GTTTATAACT 12780

AAAGTCTAAT TGGCTATAGG CAAGATTCAA TGTAAAGCCC CCACCATGCA GGTACAGTAC 12840

AGCTTTCTTG GTAGGCGCAA TAACGGAGTT TGACATGCGA TACATCGTAG GAAATTGAGT 12900

CATTTCTTGA AGAGGAAAAA GATGACGTAA ATGTTTTGGC AAAGTTACTT GTTCATTTAG 12960

TGATCGTTGC CAATGATAAA AAGCCTTGGT GACTTTCGGA TTTCGTTCAA GGCGAGCAAT 13020

TTGGGAAAGA CGTAAAGAAT CTTCAAAAAG ACAGGCTTTT AAGGATTTTT GACGAGAACG 13080

TGCGTGAAAC AAATATGCGC TAGCTAACAA CAAACCAATC GATTTCTTTT TCAATTAAGG 13140

CCCTCCATTC AGCAAAAAGT AAATCATTTT TTCTACAGTA TAACATTTTA AAAGTAGAAC 13200

AACGAAATTT GTTGAAGGGG TGAAAGGACG GTTTAAACGA ATGTACGTAT TTATTTGTGG 13260

TTGCGTAAAG TAAATGATTG TAGTAAATTA AATGAGATTA AGAGGTCCGG TGAATGGGCG 13320

CGTTCACGGC AAGGAGAGAA GTAGGATGAC AAATAAAAAG GTTACAATTA AAGATGTTGC 13380

CAATGATTCA GGGGTCTCCA TTACTACCGT TTCACAAATT TTAAACGGGA ATGGGGCACG 13440

GTTTAGTGGG AAAACGGTCG AAAAAGTATT GGCTGCAAAA GAGCGATTAA ACTATCAGCC 13500

CGACTATTTT GCGCAACGTA TGGTGATGAA AAAAAGTAAG ACGATTGGCG TAATCGTACC 13560

TGACATTACC AATCCATTTT TCGCGCAATT GATTCGAGGA ATTGAAAGTG TTTTATACAA 13620

AGAAAATTTT ATACTGATGT TATGCAATGC GGACCAAGAT GTTACAAGAG AACATGAATA 13680

TTTAACAGAG TTGATTCGTC GTAGTGTCGA TGGCTTTGTG ATTGCAAGTT CTGAAATTTC 13740

CAATCAAACG ATTAATGAAA CGCTACGTGC GAAAAAGATT CCGTTTATTG TACTTGACCA 13800

AAAAAAAGCA GAAGGCTTTA GCGATGCTGT TTTGACAGAC GATTATAGAG GTGGGCAACT 13860

AGCAGCAAAG CATTTACAAG AGCAACGACA TGAACAAGTC ATCGTGGTAA TGCCGCCACA 13920

TGCGCCAGTC AATATTCAGC AACGGTTAAA AGGTTTTTGT TCTGTTTATA CGGAAAAGGT 13980

CCAGTTAATC GAAACAGAAT TATCAAAAAC CGGCGGCTAC CAAGCTGTCC CCGAAATTTT 14040

AAAAACAGAG AGCACCGGAA TTTTTGCTAT CAATGATGAA ATTGCTTTTG GTTTATATCG 14100

AGGATTAGCA GAAGCGGGCA AAAAAATTCC TGAGGATTAT A 14141

(2) INFORMATION FOR SEQ ID NO: 395:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5190 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 395:

ATnATACCTA TTATCCATTA GGCGACATAC GTCCTTTTCC GTTTAAAAnA AnnTAAATCC 60

ATATTTCCAT TTTTACCAAT ATCCGAAGTC TAAATTAAAg TAmGAAGATA ATcCGTAAAT 120

ATTCCATATT TTGAATGAAG GGTAATTCTT CTTTTAGCaT TGAATCaTTT AATAATAATT 180

CaGTTGCAAA TATATTTGCT TCATTTTCTA AAGTAGCTTC AAAAGCTAGG CTGTGCATAT 240

CTATACAGAA TACATTTAAA TGTTTGTGTA AGAAATAATG TCCTAGCTCA TGAGCAATTA 300

TTTTATCTTG TTTTGTTTCA TCAGTAATAT TGCTATTAAT GTGAATGATG TAAGTGTGAT 360

TATAGTATTG TAGAAAACCT TCTACATTAC CTAAGTTAGA TTTTAAAACA TGAATTCCTA 420

AGAATTCTGA AATTCTTTTT GGATCATTAG ATGAATAAAA ATTAATTATA CCTTCAACTt 480

CGCTTTTTAT TTTTGAATAG ATGCTACGTA TCGGACTCAC TCCTTAGTCT TTTTTGTCAT 540

ATCGTCAACA GTTTTTATTG CATTTTTCAT TGCAATTTTT AAAAGACGTG CAGTTTCTTC 600

GTCAATATCG CCGTTTTTTC CCATAAATAG AAGAGTATCT TCATTTTTTA ATTCTTCAAT 660

AGTTTCTTTA ATAGTACTAT CCAATTTAGT GTCATTTTTA TCTGGTATAT CTAGCAAGTA 720

ATTTGCGCTT ACATTAAAAT ATTCAGCAAT tTTTTTATTG TTTCTGGaTT AGGAGtAGAA 780

CGATTATTTT CATAGTCACT AATTGAACTT TGTCTTAAAT GAAGAATTTC TGCAAGTTCA 840

CTTTGCTTGA TTCCTTTTTT CGTTCGTAAC TGTTTTAATT TATCGGAGAA AGCCATAATA 900

TCACCTCTTT TCTGCAATAA TATCACTAAA TGTTATATGG GTTAACTTTA TTTGTAAAAA 960

AAGAATATTA GTTGTTGACA AAACGTAAAA TACGATATAT ACTAAAAaCG TAATAAACGA 1020

TATGGTTTAA AAAAAGGAGG TAAAAAACGT TATGAAAATT AAGTTTTTGA CaGCAAATGA 1080

AGTTAGTGaA TTAATGAATT GTTCAAAAGC TAAAGCCTAC GGAATTATCA AAGAATTAAA 1140

TAGCGAATTG AAAGAACAAG GAATGAAAAC TTCAGCAGGA AGAGTGAATG AACAATTCTT 1200

TGCTAAAAAA TTTGGTTTTG AGCATGAGTA ATAAAATTGG ATATTTTGTT TTATTAGAGT 1260

ATCTGATTAT ATGTCGGATA TTCTCATAAG GCAAAAAGCC TTAATAAAAT TAAATAAACA 1320

AAAAAGCGTT CACGAGAAGA AAATGAATTT CTGTTTCGCA AACACTTAAA GTAATACCTA 1380

GACAATATTA TTTTAAGCGA TTTGACsTAA TAATACAAGA ATGAAATTTT TGTTCGAGGT 1440

TATAAGTCTT GTTGAGGTGG TGTGTGCCAG CCGTTGAATG GAACACAAGG TGGCAGTAGT 1500

ATCTTTAACT AGAGCCATGT TGATGACAGA TTAGTGCCGA GCTGTCAGAT TATTTCCGAT 1560

AACCATGTCG TGATGACACG TTGGGAATAG TTGCTGAATT AGATTGAATG AAAAAATCTA 1620

TGCAAGCTTA GCAAGAGGAG GAGAGCTGTA TACCGTATGT CGTAGGTCTA TATAACGTGC 1680

CGTGATGTAA TTTTAATTAC TAGCCGTTGA AAAATGATTC TATGAAACGA ATGGATAAGG 1740

GGCTATTTCA TAGCTAAGGC GAACTTTTAG TCAGTTATCT GCAGAAATGC ACATTCTACT 1800

ACGATGAATT TTAGAATCTA AATTGCTATA TAAGTAAGAT TGTTATTTGT AAGAGTGGaA 1860

AAGATTTTTT CTAAAAGGAT AACCAATCTT AGTTTAAGTA TGCACAGCTG GATTACTCTG 1920

GTTGTGCTTT TTTATTTGAA AGAATAGAAT TTGTTATAAA AAATAAGGAG GATAGGAATG 1980

AAAGAAGAGT TATTACTTTT TGTTGAGAAG TTTGTTGCAC GAATGAAACG TCAGAAAAAA 2040

GCATTCTCTA TTACTGATAT AGAAAAAAGC TATAATTTGG AACGTAAGAA GTTAGGTAAG 2100

AGTGCTGTTA AGTTAACAAA CATGGAACGT TTAACAATTG AAAGTCGATT ATTAAAAAAT 2160

CAGATTTTAC AACGrACCTA TAAAATGACA GGyTATCATA ArCCTTGTCA AGTTGTGTTT 2220

TTTAGTTAGy AATAAGAATT GCATAAAGAG ArAAArGCAG TyAGTTGTTG ACTGTTTATT 2280

TTTrTAyAAA AAtTAAkTGA AAGAAGGAAA TTATTTTGGC AAATAGAATT CArGAAATTG 2340

CATTAAswGA wAAAAATyTA GGAAGTAAAA TTAAAGCTrT TTATAAAACA GAACGTGATA 2400

TTTATGTrAA TGGkGCrCCw GTwCGyGGTC GTGTrTTyAC yGCTTCAAGC GAAAArCAAC 2460

GTGATTTTGA AATTTTTGTC CCTCATTTAC GAAAyTCACA AATGTTGAAT GGTCGTGAAG 2520

AAGTTAAATT AATTGATCCA GTAGCACGAA TyAATGTGAA aCGAAATAAC ATTGGTAATA 2580

GTCGCTCaAT TGATATGGTA GTTTATGCaA AACGTTTAGA AGTAGTTGGA GGTAATAAAT 2640

AATGGGATTA ACATTTGAAA ATAATACrAT TGAAAATTTT GACGTACAAG syACATTTGG 2700

TAyGATGAAT TTyTTAGAAG TAGTkGCAAT TAATGArCGT GATGGwGAAG rmAATGTCAC 2760

TGAmGAAGTT CGTGAACArC GwGTGACkGT TTATTCTAGC AAATTAAATG ACCAAGTAGA 2820

AATTGyGATT GAyCCAGATT ATGATGTyAG TGGTATyAAy TATGATGATG ArATTGAATT 2880

GACTGGAAAT GTGACTGCTC GTGGTTGGCT TAATACTTTy AAAGGsTwyA ATGACAATGT 2940

rCArACwGAA CAAGCyTTTA AAGTACGTkC rGCwGGTATT AAAAArGTAG GGGCTGTTAA 3000

ACmAACrCAG rCACCTCArC CAAaGgAAAA ACAAGAAAAT AAAyAATAGA AgTAArTmAA 3060

TATTGTATTC GGACrAAtCG GTAGCGCCGT TTAACAGCTG TrAAGGGCAT TCTGACGAGT 3120

GCCCTTTTTA TTATTAATTT GTGTCAACAA GGTTTGTTAA TAGACATTAA TAATAAAAAA 3180

TGAGGTGGTT AGTTAAAAGA TGAATCGATT AGAGGAATTA ATCAAGAACC CAACAAAATT 3240

CAATCTATCA AATGAAGCGA TTGATTCATT AAGGGAATTA TTTGTTACTT TTGAAACGAA 3300

TCCATTTTTC CCAATGAGTC GATATGATTA TGCTCGCAGA TACTTAACAC AACTATATTT 3360

TGCAGGCTTT ATTAGTAGTG ATTTAGTTCA AAGCATTTTG AGTGAGTTCA AAAAGAGTGG 3420

GTAAAGTTAA TAAAAAATAG GAGTGATTGG ATGAAGCAAG GGAAACGCCC AGTTAGAAAA 3480

GAAAAAGAGT TGTTGAAAAG TCGCAGATTG AATGCGAAGA ATTGGTTAGT GGAAAGACGT 3540

TTGAAGAGTA GTGTTATTTT TCTTCATAAA GAAAGTGGTA ATTTGAAAGA GATTTTTTAT 3600

TAAGAAAGGA AAAATGTTAG TGGAAAAAAA TAATGAAAGT ACTAAATTAT TGCAAAGAAA 3660

AATTAGGTAT ATGTGTGCGG TTGAAGGTGA AATGGAGTTT TATGTGTTAC GTCCATTATT 3720

TACTGATGAT GTTAATGTTC AAGCAGTAGT TATGACATTT CAAGATGTTT ATGaTAATTC 3780

TTTTyTCTAT GAAGGTAGTG CAGAGGGATT ATATCAAACG ATTGTTCGGT GGATTGAAAA 3840

GAATATAGCT TAGTAATTAA ATGTAAACTG AGAGGGGGTG TCTTGTAAGT TAATGAGTAG 3900

TCTGTCGGAA AAAATGGAGC ATAAACAAGT AAGGTATCGA GCGTTTTTAG AAAGAAGATT 3960

TTATTCATAC AGGGGGTGGC AAAGTTTTAA TTATTACCGT GACTTGTACT TAAAATTATT 4020

TGATGAAACG AGTAATGGTC TTATTCAATT CTTATTACTA GATGATTCTT TTTTTGAAAG 4080

TGAACAAGCT GTTTTGAAAT TGTTAGATAG TTTTTTGGAT CAATTAGTTC GTGCGTACGA 4140

TTTAAAGTTC CATGAAGATT TTGAAAAGAA AGTGTATTTT GAAGAATATC CATTAGGAAT 4200

AAGTGAGGTA AATTGATTGT GATGGTTTGA TGGCTTATAA TGGCTTTAGA ATAGACGTTA 4260

GAAGAGGTGA GACAAATGGT AATGACAATG AATGAAATTA AAGAGCGTGT GAAGCCAATA 4320

GCTGAAAGAT ACGAAATTCC TGTGATATAT GTGTTTGGTT CTTATGCTAG AGATGAAGCG 4380

AAGGAATCAA GCGATATTGA TTTTTTGGTT GGCACTACTG GTTTGCCAGA AAAATATCGG 4440

TGGTCTGTGT ATAGTGATTT TTTCGATGAA TTGAAGGAAG CCGTTGAACA TGAGATCGAT 4500

TTAGTAGAGC TTGAAGCATT TGAACAACCA ATTGATTCAG AATACCAGAA AGAGTTTTAC 4560

GATACTATGA TGAAGGAGAA AGTTAAAGTA TTTGAAAGAG AAAAATAGAG ATATTCGTGT 4620

TCTCAAGGGA ACATTGATTT ATTGTGAAGA TGTAAAATCG GCACTCAATC GTTGTAGCAG 4680

AAGCTTTGAA GTTTTCTCTC ACGATAGAGT GtTTTTTCAT GCAATCAGTA TGTCTTGCAT 4740

GCAGGTAGGA GAACTTGCTA ACTCTTTTTC TGAAGGGTTT AAGGAAAGAA ATAGTGGTGC 4800

GGTTGATTGG CGACAATTAA GGTCAATTAG AAATATTTTT GCACATGCGT ATAGTAATAT 4860

TAATAAACGG ACTGTTTGGG AACTTGCAAC AGTCTTTTTT CCTGTTTACG AAGAATTTTG 4920

TAGACAAGAA TTAGCACGAT TACAAAAAGA AGGTAGGTAA ATGTTTAAAA AATTATTTAG 4980

ATATAGAGGA AGGCGTATCC GTTATTCTTC AAGAAATTTG TTGGCACTCn ATCGAGTGTT 5040

ATTTTTTATG CCAATTTTTG TTTGTTTAGG CTATTTTGTG GGCTATAAAT GGCTTTATCC 5100

ATTGTATCTC TCAAATCCAC CTGAtTGGAA AATTTATATT GTGCCAGCGT TAATCATTGT 5160

AGGTGTTTCG ATAGGTGCCA ATGGGGnAAT 5190

(2) INFORMATION FOR SEQ ID NO: 396:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1714 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 396:

```
CAGATAGAGA AAAATATTAC CCCGCATACA TTGCGGCATA GnTTTGCGAC ACATTTGTTG     60

GAAAATGGCG CAGATCTTAG AATCGTGCAA GAATTATTAG GAnATGCAGA TATCTCCACG    120

ACACAAATTT ATACACATAT TACCAAACAA CGAATGGCAG ATGTTTATAA GGAACATTTT    180

CCTCGGGCTT AATTTGCGAG AATGCCCACA AACGAATAAA TTTATGGTAA ACTAGGAAGT    240

CGAGAAAAGA AAGTAGGTGT CGTTTTTGCA AGAAATTAAT TTAAAACTAG ATGTCTTTGA    300

AGGACCACTG GATCTTCTAC TACATTTAAT TCAAAAATTA GAAATCGATA TCTATGACAT    360

TCCCATTACA GCAGTAACTG AAcAATACAT GTCTTATATT CATGCGATGC AAACCTTAGA    420

GTTAGAAGTC GCTGGAGAAT ATCTAGTCAT GGCTGCCACC TTAATGGCTA TTAAAAGTCA    480

GATGTTGCTT CCAAAACAAG AGCTTGAAAT CATTGATGAT GAAGATTTTT TTGAAGAGGA    540

AGACCCACGG GAAGCTTTAG TAGCTCAGTT ATTGGAGTAT CGGAAATTTA AATATGCAGC    600

GACTGTTTTA CATGAAAAAG AAGAAGAACG TAAACTCTAT TATACAAAAG AACCAATGGA    660

TATGGATGAT TATAAAGAGG AAGATACAAC ACTACCGCCT AATCAAATCA ATACAATTGA    720

TTTATTTCTA GCCTTTCATG CCATGTTAGA AAAGAAAAAA AATCGCCAGC CTGTTGAAAC    780

AACTGTTGCC AGTGACGATG TTTCCATCGA AGAAAAAATC TCCGCTATTT CAGAGCGTAT    840

GCGTCaAGTA CAGAAAGGGA AGGCAGTGTC GTTTGATTCT TTCTTCGATT CTTATTCCAA    900

ACAAGAAATT GTGACAACGT TTATGGCTCT TTTAGAATTA ATGAAAACGG GCgCGATTTA    960

TGCTGAACAA GAAAATAATT ATAGTGAAAT TTTGTTATTT AATACAGAAA CCCAGCAGGA   1020

AGATACAACA GAGGTGGAGG AAACACAGTG ACATTAGTTA GTCAAATTGA AGCAATTTTA   1080

TTTGTGGTTG GCGAaGAaGG AaTTGGCTtG GAaGAGCTtG CGTACTTATT AGAGAAGTCA   1140

ACAGCAAAAA CCTATGAAGA ATTAACGAAA TTGAAAGAAC ATTATGCTTC TGACAATAAA   1200

TCGGCGTTGA ATATTTTAGA AGTCGGCAAT CATTTTGTCT TAACAACTAA GAAAAAATAT   1260

GCCTCtTACT AAAAAAATAT GCGCAATCGC CTATGTCGAA CGCCTTATCA CAAGCGGCGC   1320

TAGAAACATT ATCCATTATT GCTTATAAGC AACCAATTTC GCGTATTGAA ATTGATGAAA   1380

TTCGTGGCGT ACAGACTTCT GGCTCTATCC AAAAGTTAGT TGCCCGTCAG TTGATAGAAG   1440

AAAAAGGCCG CGTAGATGGT CCTGGTAGAG CAATTTTGTA TGGCACAACG AAGTATTTTA   1500

TGGATTACTT TGGGTTGAAA AGTTTAGATG AATTACCAGA TATCCAACAA ATGGAAGATG   1560

AATTAGAAGA AGAACTACCA ATGGATCTAT kTTTTGATCG TaTCAAGAAm CGAACCCAAT   1620

GTCTGAAACA ACAGAAGGCG AGGAAGCTTA AACAaTGGaA AGtTACmAAa AggCAtTGCT   1680
```

CATGCTGGGG tGACTTCgCG TCGCAAAGCa GAGG 1714

(2) INFORMATION FOR SEQ ID NO: 397:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 885 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 397:

```
TTCAATCACT GCAGAATGCA TCTCGTGGAC TTCCATTACT TGGCCATCTT TTTGTTCATC     60

GATGACAAAG GAACCGTTGG ATGTTCGTTC ACTGATTGGA TAGTTTTTCG ACTGAATAAC    120

TTCTTGTGTG CCTTTTTGCG TAGTAACTAA TAAATCTAAA TCGGTACTTT CAGACATAAA    180

GACAACACGA TGTGGATTTT TCTTCAATTC CCGCAAGACC ATTAGTCCAC GTTTTGCTCG    240

TCCTAATTGT GTTAATTCTT GGGCTAACAT GCGTTTAACA CCGCCACGTT GGGTAACGAT    300

GACAATCGGT GtATCCCCTT CAGAATGAAC CAGTAAACCA TTAACGACAT AATCGTCTTC    360

TTTGAGGTTC ATTGATTTAA CTCCAGCCGC TTTACTACCC ACGACTGGTA CTTCATATAA    420

TGGATAACGT AAGCCAAAAC CGCGATTGCT AACGAGGAAA ACATCCAATA AGTCTTGATC    480

ATTTGTTAAA TAAACATTGG TAATTTCATC TTGGtCTGAT TTTAATTTCA TACAAGATGT    540

CGGGCGGCTC TTATAAGTGC GCCAAGGCTC AAAATCAGCC ATTTTGGTTT GTTTAATCAT    600

GCCAGCTTTT GtAATGAAAA CAAAGGTTTT TGTCGGACTT AACTCTTTaT AAGGATATAC    660

AGCAATAATC GATTCATCGA TAGCaAGATT TAAAATGGkT TGaGAAATAT GTTCGCCnAT    720

TTCyTTCCAG CGTAAATCTG GCAACTCATG GACGGGGCGA TAGATCATAT TCCCTTTATT    780

TGTTACTAGT AAAAGATGAT CTAATGTATT GACTTCGCCA GCATATAATA AAAAGTCGCC    840

TTCTTTCATG CCGATTCTTC TGGTTTTGAT GCTGTATAAG AACGA                    885
```

(2) INFORMATION FOR SEQ ID NO: 398:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2927 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 398:

```
GCAGCCACAA TATCAGCAGC CGCAACAACC GGTTGCGcAc AACTtGGAGC TCTTGTAAAA     60

GCGCACTGCG TTCCTTTTCT TTATTCTTTT GATCTTGAGA ATCTTCTAAA AATGCCGAAA    120

AGAAATGTTG GGAAGAGAGC GTAATCAGTT TAGAAATGCT CTTGATGGTA GCTTTATGTT    180

GATCCATTCT TCTGCCTCCT TTACGAATAA AATAGAATAA AACTCAAATG ACTAATTACC    240

TGTATTTTAC CTAATTTTGT GATAAAATTC AAGAAAATAT GTTCGCCTTC AATAATTATG    300

AGAATTTACC AACGTTTCTG TTTTTCGAAA AAGAAAGTAG TATAATGTGT GAGCATATAA    360
```

GAAATACATA AGATTATCGG ATGTAGTCAA TAGATAGAAA AGGATGAATA GTATGTCAAA 420

AAAACCATCA GGTTACCAAA AAAAAGTTCT CACGGTGTTG TTGGTATCGT TTGGCTTTGG 480

CGGGTTGTTA GCGCAGTTTT CGCCGATGAA ATCAGGTGGG TTTTCAAGTT GGTTTGATGA 540

TTCATATGTC AAAGCTAGTG CTGAATCAAG TAAAACAAAA GAACCAGCAC CAGTTAAAAT 600

TGAGAAAAAA GTCAAACCTT TAAGCTATGG ACAACAGGTC AATCAAGAAA TTGAAAAGAA 660

GCAATATGAT GGACATCTGG ATTTGCCGTT AGAATTGCAG ACAGATGCTA AATGGAAAGA 720

CACCGCATAC GGATTTGGCA ATGTGGATAA GCCGAATACA ATCGAAATTA ATGGCTGTGC 780

GATTGTATCG CTTGCAATGG TTGGTTCATA CATGGATCAC CAAGAAGTTA CCCCTCTGGA 840

TGTGTTAGCT TGGGCAAAAA ATGACTTCTT TATGGAAGGG CAAGGGACGG CGTGGTCTAT 900

TTTTAGTGCA TATGCTGAAA TGAAAGGCTA TAACTGTCAA GAAATTGGGG ATATTGAAAC 960

AGTGGCAGCT TTCTTGAAGG AAGGTCATCC AGTCATTATT TCTGTAAAAC CGGGCTATTT 1020

CACTACAACT GGTCACATTA TGGTGATGAG TGGTGTGGAT GAAAAAGGCG ATTTCTGGaT 1080

TAATGATCCA AACGATTCAG AAGAAAAGGG CCATTCAAAA CGGACATTTA CAGCCGAAGA 1140

AGTGATGAAT GAAGCGTTAA ACTTCTGGGC ATTTTATTAG GAATAAGAAA AGGTGTGGAA 1200

CAACAGGCAg TACCCTTTGT TCGCATCTTT TTTAGCAAAT AAAAGGAGGC AGGACAATGG 1260

AGCGATGTAC TTGGGCAACG AATACAACTG AAGAGATGCA AGCCTATCAT GmTGAyGAAT 1320

GGGGACGTCC AGTTCATGAA GAGCAACAGT TATTTGAGTT ATTGACCTTG GAAAGTATGC 1380

AAGCAGGCTT AAGCTGGGCG ATTATTTTGA ATAAACGAGA AACGCTTAGA GCAGCTTATG 1440

ATGCGTTTGA TTATCGGAAA ATTGCGCGAT ATGATGAGGA AAAAATTTTA GCTTTACTAG 1500

CAAATCCTGG TGTGATTCGT CATCGTTTAA AAATTCAAGC GACCATCACG AATGCGCAAG 1560

TTTTTCAAGA AGTTCAGGCA GAATTTGGGT CCTTTGATCG CTACCTCTGG AACTTTGTCG 1620

ATCAGCAACC AATCGTTAAT CACTGGCAGC ACCCAGAAGA GGTTCCAGCC TCCACGGAAT 1680

TATCGCAACA GATTAGTCGA GCGTTAAAAA AACGAGGATT TAAATTTTTA GGTGCCACGA 1740

CAGTCTATTC CTTTTTACAG GCGGCGGGAT TGGTGAATGA TCATTTAGAA ACGTGTCAGT 1800

ATAAATAAGT ATAGAAGAAT TAAAAAGGAA GGAACTAGCA AAACGAATGG CTAGTTCCTT 1860

CCTTTTTTTG CAGTGATTAT GGACACTCAT TTAACATCGT TGAAACAGTT TTCGCCAACT 1920

TTTCTATTAA GAATAGTGCA GGTAGTTGGG ACGTTACGTC TGCAATGGAA CTTTTCTCGG 1980

TTGTTATATA GTAGGGGATC GTGACTGTTG AAAGTTGAGC AATCGTGCTT TTTTCACTAT 2040

TTGTAATAGC AATTACATCA ATCCCCATTT CGATAAAATA ATCAATATTT TGCCGAACTT 2100

TTTGTGTTTC ACCCGAAACT GATAAGGCGA TAACACCCAC TTTTTTAGCC GTTTCTTCGG 2160

ATAATTTGCT AACGGGATGG GAAAAAAGGT GGGATACAGA AGCAGATAAA TTATAGATAC 2220

TGGAAAAATA AATTTCACCA AATTCAGCTA GGACTTTGGA CGTTCCTTCG CCAATAAAAA 2280

GAACAAATTC TTTTTTTGCC AGTAATTCCG CAGCTTGAAC AATCCGTCCT TCTAATAATT 2340

CTTCTGATGA GCGTTGCAAA AAATGAATCA GCATTGTTTC ATCTAAAGAT TTTGGTTGGG 2400

CTGTTTTTTT TCGTTCGCTT AGATAATTCT TCAATTTAAA TTTAAAATCT GTATAGCCTT 2460

CACACTCGAA TTTTTGACAA AAACGCCAGA TTGCCGTTTT ACTGAAATGA ATTTCATCTG 2520

CTAATTCTCG AATTGTCATA TAAGGAACTT TTTCAATATT GTGGCTAATA TACTTGTAAA 2580

TCTCTAATTC TGTTGGATTT AATTGTGGAT CGTAGGTTAA AAACAGCATG CTGATCCCTC 2640

CTGTCATTTG TGAACGTAAT GTCCTTTTTT AGGGACATGG TCTTATTTGT AGATTATAGC 2700

ACTCAATCAG AAAACTAGCT ATAAAAAgGC GCCACTTAGC TACAATCAAT TAGAGAAACA 2760

ACTAAAAATT CATGAAACAG AAGAGAGAGT GGGGAaAAAT GAGCAATAaA AAATtAATTA 2820

TCAACGCTGA TGATTTTGGT TACACGCCAG CGGTAACGCA AGGAATCATT GAAGCACACA 2880

AACGTGGCGT GGTTACTTCA ACAACTGCAT TACCAACTTC TCCTTAT 2927

(2) INFORMATION FOR SEQ ID NO: 399:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32768 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 399:

AAAAGTTATT TCTGCTTAAT GATGGTCTGG TCTTTTAAAC AGTCAATTTC TAAsAACCCT 60

TAGTTATCTC CATAATTTTT TGGAAGATGT GAAGATAACT AAGGGTTCTT TTTCTGTTTT 120

CAATAAAAGA AAGCAGGATT CTATAGAAAA GGCTACTAAA AAAgAGCGGa AATATGCTAT 180

TATTATTAAG AAGTAAAAAA ATATAGAAAC AAAGGGATGC CGATGAAAAA GTCCTACCGT 240

GTCAAGAAAG AAAAAGAATT TCAACAGGTG TTTAATAAAA AACAATCTTG TGGCAAATCG 300

TCGTTTCGTG GTGTACGTTT TAGAGAAACC ACaACAAGCC CATTTTCGAG TGGGGATTTC 360

TGTTGGGAAG AAAATTGGAA ACGCGGTCAC AAGAAATGCT GTGAAgCGGA AAATCCGCGC 420

GAGTTTATTT CAATTAAAAG ACCGTATCTC TCCAGAAATT GATTTTATCG TGATTGCACG 480

TCCAGGATTG GAAAAGTTAT CGTCTGAAGA AGTGAAAGCT AATTTAACAC ATGTGTTAAA 540

TTTAGCTAAA ATATTAGATG TAAGAGAGGG AATTGAGTAG TGAAGAAGTA TAAGCGCTTA 600

TTATTAATGG CTGGGTTAGT GACCTTAGTT TTTGTTTTAT CCGCCTGCGG AACCGCGCCC 660

GTAAGTGAAA GTAGTACAGG AATTTGGGaT CGATACATTG TTTATTATTT TGCCCAAGCG 720

ATTAAGTTCC TATCATTAGG CGGCAGTGTC GGAATTGGGA TTATCCTATT TACGTTAGTC 780

ATCCGGATTA TTTTATTGCC GTTGATGCAT TTCCAAACGA AGAGTATGCG TAAAACACAA 840

GAGTTGCAAC CGCAATTAAA GGCCTTGCAA CAAAAATACT CTTCAAAAGA TCCTGAAACA 900

CAACGCTTAT TCCGAGAAGA ACAGCAACGT TTGTATGCAG AGAATAATGT GAATCCTTAC 960

ATTGGCTGTT TACCACTGTT GGTTCAACTA CCAATTATGA TGGCTTTGTA TCAAGCCATT 1020

TCACGCGTGC CTGAGTTAAA AGAAGGAACG TTCTTGTGGT TAAGCCTCGA TAAGCCGGAT 1080

CCGTATTTAA TTTTACCGAT TTTAGCGGCT GTCTTTACGT TTGCTAGTAC GTATCTTTCA 1140

AGTATGAGCC AACTAGAAAC AAATGCCTCA CTAAAAATTA TGAATTACGT TATGCCAGCG 1200

ATGATCTTCT TTATGGGGAT TTCATTAGCC AGCAGTTTAT CTCTTTACTG GGTGGTCTCA 1260

AATGCCTTCC AAACAGGGCA AACATTGTTA TTAAACAATC CATTCAAAAT CCGTAAAGAG 1320

CGGGAGGAAG CAGCACGTCA AGCCAAAGCA CGGGAGCGAG CGCTTGAACG AGCAAAAAGT 1380

CCTAAGAAAA AAGGGAAGAA AAAATAAAGG AGGTCGTTCA GATGCCGATT TATGAAGGAA 1440

ATACAATCGA AGAAGCAACG CAGAAAGGCT TACAAGCACT TGGTTTAACC AAAGAGGATG 1500

TTACAATTGA TGTGTTAGAC GAAGGAAAAA AAGGATTTTT AGGTCTAGGA AAAAAACTTG 1560

CCCAAATTTC AATGGAACCA ACGATTAGTG AGCAAGTCAC AGAAGCGGTC GAAGAAACAG 1620

TGGAAGATAT TGTTGTTGCC GATGAAGCAA CAGCAGTGGA AGAAGCAGTT GAGGAATTAA 1680

CCGAAGCAAT TCCTAGTCTC TCAGAGGAAT CAACACATTC TTTAGAAAAT CTGGAAGACG 1740

AAGCTGCTAT TATTGAATTA GCGATGTATT TAACGAATAT TTCAAATGAA CTGAACGCCC 1800

CTGCCATGGT TCGCATAGCG CGCGAGAATG GCAATATTAT CTTCCATTTA GAAACAGAAA 1860

AACAAGGTAT CTTAATTGGA AAACACGGAA AAGTGTTAAA TGCGTTACAG TATTTAGCGC 1920

AAGTCTTTAT TCACCGCATT GCAAGCAATA AGCTATCGAT TGTTGTGAAT GTCGGAAATT 1980

ATCGTGAAAA ACGTCATGAA ATTTTGGAAC GTTTAGCGAA ACGCACCGCG GAAAAAGCCA 2040

AACGTACTGG GCGGCCAGTC TTCTTGGAAC CAATGCCGGC TTTTGAAAGA AAACAAATTC 2100

ATCACACGCT AAGTAAAGAT GAGCAAATCA AAACCCATTC TGAAGGGGAT GAACCTTATC 2160

GGTATTTAGT GGTTGAACCA GTTAAGAAGT ATTTTTAAAC AAACAAAGAA ACGTTGGCAA 2220

CTCTCAGCAG GGACTGAATA GTTGCCAACG TTTCTTTGTT AAATGATTTT TTAGCCGTCT 2280

CCGTTTCTTT TTTTATGTTA TGATAATCAG TAATTACTCC AGTAGTAAAT TCAAATGAAA 2340

AAGTTTTGTA AAGAGACGTA GGGGGAGCAA CAGGTATGCG TACGGAGAAA ATACTGGTTG 2400

TAGATGATGA TGCAGCAGTT CGTCGCTTGA TTTGGAAAGC ATTACAATCA ACAGGTATTC 2460

TTATTTACCA AAGTGAATCA ATTGAAAAAA CAAAAGATAT CGTGTCCAGA GTTACCTTTG 2520

ATTTGTTCAT TTTAGATGAT AACTTGGAAC ATGACAGCGA TGGTTACTAC TTAGTCCAAA 2580

TGATTAGAGA GAAATATCCA CTGGTGCCAA TTATCTTTTT TAGTAGCAAA AAAGAAGAAG 2640

AGGATATCAT TGCAGCCTTG GAAATGGGTG TGGATAGTTA TATAACGAAA CCTTTTTCCT 2700

TGAATGTTTT TAAGGCTCAG GTGATTGCTT CGTTGAATCG AGCTCGAATG ATTCGAAAAC 2760

AATTGAATGT TTATAAAAAA GAAGAGCTAC AGGTAGGCGA TTTCCGTTTT GATTATGGTC 2820

GTTATCAACT ATTTAAAAAG GATTTACCTA TTTCATTGTC TTCGAAAGAA GTTCAATTAA 2880

TTCAATTTTT TCTGGAGAAT CCCGAGCAAG TATTTTCTAA AGAACAAATT TACAGCAGTG 2940

TTTGGGGTAC TGGGGATGTC GATGCGAATA CGATTATGGT TTTTATTAAT CGTTTACGAA 3000

GTAAATTAGA AGAAAACCCG AAAGAACCTT GTTACcTTCG GACCGTTTGG GGCATTGGTT 3060

ATACCTTTAC GCCAGATGGT TGCGTACAAA AAAAATGAAA AAAGACCTGC AGTTGCACGA 3120

GCGACTGCAG GTCTTTTTTG CTTATTCAAT TTGCCGACGT TCCATTTCTT TCAAAATAAA 3180

TTGAAGCGAA CGATTTAGAT GTTCATGGGT TAAAAGTGTG ATCTGGTCAA TATTCTTTGT 3240

TTTCATTCCG CGGAAAATTA ACCAATGTTC TTCTAGGGCA ATACTACGAC GTTCTGAGGC 3300

ACGAATCGCA ATATCTCTGA AATAGATTAA GTAAGCATGG AGATCAGTCA CAATTTTTTT 3360

TAAACGTAGC ATCTGACTCT TAGTATAAAT AAAGGAATTA AAATCTGAAA AGTTCTGTAG 3420

TAAGTCATCT ACTTGGTTAT TTTTATTGTA TTGTTCTCCT TCGAGAAGTA AGGCTTCCAA 3480

TTCATCAAAA TCTTCTGGTG TCATGAGTTC CATGGCTTTA ATGGTTGCTA AAGTGTCCAA 3540

AGATTTACGA ATATCATAAA TTTCATAAGC ATCTTTCATA CTGATCCCTT TAACCACGAT 3600

ACCCACCATA GGTATATGTT CCACCAATTG TTCTTTAACT AATTCTTGTA AAGCAAAGCG 3660

AATGGGCGTC CGACTAATGT TTAATTGCTC GGAAAATTCT TTTTCGTTAA TCCGTTCACC 3720

TGCAGGAATA TCCCCTAAAA TAATGGTTTT TTTAAAGGCT TCATAGGTAC AAATTTTTAG 3780

AGGTTTACTG CTAGTTAAAT CTAAGTTCCG TTTTACTGCT TCAACCGTTG GATTCATATG 3840

AGGCACCTCC CTATAGGTTA TTTTTATAGA TAATAAATAT AATATTAAAA ATGAAAATAT 3900

AACGCAAGAA AAGACAAAAT GAAAAGAAAA CCGAGGAGAA AAGGTACTCC TACGGTTTCA 3960

GCGACGACGT TTAATTTTGT GGAATAAATA AAGGCATGTG ACCTAAAGCA ACAATTGTAA 4020

CAACGAAAAT AACGAAAATT CCTAAGGCAT ATTTTGCCGA TTCTTTTTGC CATTGGCCCA 4080

TATCTAAACC AGTTAAACGT AATAGCAAGT AAATAAATGC CACTAATGGA CTTAATAAGT 4140

GGAACGCTTG CCCCATTAAG GAAGCTAAAG CCATAGACAT ATTGTCAAAT CCATAAGAAC 4200

GTCCAGCTTC TGCTAGGACT GGTAAGACAC CGAAATAGAA ACCATCGTTT GAAATAAAGA 4260

ATGTTCCTGG TGCTGAAATG AACGCAATCA CTAAGCCCCA AGCGCCGGCT AATTGTTTTG 4320

GAATAATATT AACAAAGCTA TTTGCTAAGG CATTGGCCAT ACCAGAACCT TGGAATAAGC 4380

CCATGAAGAC GCCAGCAGCA AAAACTAAGA TAACTACTTG CACAGCATCA CCAGCATTTG 4440

CACTGATGCG GCTGCTTTGA TCTTTTAATT TTGGATAGTT AATCATTAAA GCGATACAAG 4500

TACCAACTAG GAACAATAAT AACGGTGGCA TTTCAATCCA CGAAGGCAAA CCTAATTGAC 4560

CAACGAATGA ACTTAGGACT AACCAACCGA TTAAGACAAT CGTTAAAATA CCATTGAACG 4620

CAAAATTCTT TGGACGACGA ATTTCTAATA ATTCTGGGTC GTTAACAATA GTCATTTCTT 4680

CCAATTCTTT ATCAGTAAGC GTTGTCACGC CTAAGCGAGC ACGTTCTTTT TTCCCCATGC 4740

GATAAGCAAC AAAGAAGATA ACATACAATA CAGATAAAAT CATACCAGGT GCTAAATAAG 4800

AAAGAATGTC GGCATCAACT TCTAGTACGG CCATGGCACG AGCTGTTGGA CCGCCCCAAG 4860

GTAATAAGTT CATAATGGTG TTTTGTAAGA TAACAAGCAC ACCTAAATTC ATCAATTTCA 4920

TATTAAGTTT TTTGTAGATG GGAATAAACG CAGAACAGCA AATCAAGGTC GTGGTTGTTC 4980

CATCTCCGTT TAAGGAAACA GCAGCAGCAA CTACCGCAGT AGCCATCAAG ACTTTCATTG 5040

GATCGCCTTT AGCAAAGTAA ATCATTTTTT TAGTGATTGG GTCAAACAAA CCAGCATCCA 5100

ACATAATAGA GAAGTAAAGA ATGGCAAATA ATAACATAAT ACCAGTTTTT GAAGTAGTTG 5160

AAATCCCATC TAAGACGAAA TTTCCAATGT TGACTTCAGG GACAACACCG TCAGCACCAG 5220

GTGGCGGACT CGCTACTCCA GTTGCAATTG CAATCAAAGT AAAAATAAGA GGAATAATTA 5280

CTAAAGATGT AAACGGTGAC AATTTCTTTT TCATGATTAC ATACATAAAG ATGATAATCA 5340

TACCATAAGC TAAAATAGTG AGTAACATGT TTTTCTTCCT TTCTTTTTTA GAACTTTTTT 5400

GTAAGCGTTA ACAACCGGTG TTAATAAAAA AGAATAGCCA TGAGAAATGT TTCACAGTAT 5460

TCTGGTTTAC AACAAAAAAT ATTTTTTGTG TTTGTCTATA GTGTGAGAAT ATACAAACTT 5520

TCGCAGAGAA TAAATAGCCT AAAACAAAAA GAACATCAAA AAAACAATTC AACAATCAAA 5580

TTCTTATTAA TAACTTAATA TAATATTAAT AGAATCTTAA CGGCTGATAA ATGCTTATAT 5640

ACCAATGTTT ACAGAGTTTT TTTCACAAAA TAAACAATAA AAGAACAATC TGAGCTAGTG 5700

TTTTCACTAG GTTTTTTAAC GGTTTTTTCT ATAATGAACG TATACCAAGA ATTAGTGAAA 5760

GGAGAGACAA TATGGGCCTA TTTAAGTCCT TATTCAAGAA AGAGCAACAC ACACAGGATA 5820

CCGAGATTAA TGAAAGCGTT ATCCAACGTG AAGCGGATAA CTTGGAAAAG CGATGGATAG 5880

ATATACCAGC ATATATTCCT GCTGAGCCAA AAGAATATCA ATTGGTAAGT ATTGTGGCCA 5940

CTGCCATCGC AGCAGGAGAG TTTCCAGAAA GTCGTTTTGT TATCAAAAAA ATCGCAAAAC 6000

GAAATCCAGA AGCAAAAGAA GTTGCCATTA TTGCAGCCAG TATAGCTGCA GAGCTTTCAG 6060

AAGAGTCTCA GCTAGTTGTC AAAAGAATTT CAAAACAAAA CCTTGAGGAG GATCATTATG 6120

TTACGCAAAT TTAAAATTTC AATTGATGGG AAAGAATATT TAGTCGAGAT GGAAGAAATC 6180

GGTGGCGTCC CACAACCAGC ACCTGTTGCG CCACAACCAA CAGCGCCAGT TGCTACGACA 6240

GAAACACCAG CCCCTGCTGT TGAAGAAACC CCAGCGTCAg CAGCACAGCC AGCCGCCCCA 6300

GCAGGAGCAG ATGCAATGCC TGCGCCAATG CCAGGAACTG TGTTAAAAGT ATTAGTCAAT 6360

GTTGGTGATA CGGTGTCCGA GAACCAACCA TTGTTAATTT TAGAAGCAAT GAAGATGGAA 6420

AATGAAATCG TTGCAGGAAA AGCAGGTACG GTAACAGGAA TTCATGTTAC CCAAGGCCAA 6480

ATTGTGAATC CAGGGGAACC ATTAATAACA ATTAACTAGA ATTAATTTAA GAAGAGGAAG 6540

TGATTTGTGT GGAGACACTG ATTCAAGGTG TTATAGGAAT GGGCCAAGAA CCAGGACGCA 6600

TTGTGATGAT GCTGATTGGT GGACTATTAA TGTATTTAGG AATAAAAAAA GAGTATGAAC 6660

CAACACTTTT AGTGCCGATG GGTTTAGGAA CGATTTTAGT GAACTTTCCT AATTCCGGCG 6720

TATTGAGCGC TGGCGGCGAG CCAGGTCCGT TTAATGTTCT TTTTGATTTT GGAATAAAAA 6780

CAGAATTATT CCCGTTATTA TTATTTATTG GGATTGGGGC AATGATTGAT TTCGGGCCAT 6840

TGTTACAAAA TCCATTTATG TTGATGTTTG GTGCCGCTGC ACAGTTTGGG ATTTTCTTTA 6900

CAGTTATTAT GGCTGTTTTA TTAGGTTTTG ATTTAAACGA TGCTGCTTCG ATTGGGATTA 6960

TTGGTGCCGC TGATGGCCCA ACGTCTATTT TCGTAGCAAA TACCTTACAT TCAAAATACA 7020

TGGGGGCCAT AATGGTTGCC GCTTACTCCT ACATGGCGTT AGTTCCGATT ATTCAACCAG 7080

TAGCGATTAA AGCAGTGACG ACGAAAGCCG AACGTCGTAT CCGTATGACT TATCGGGCCG 7140

GTGAAGTTTC TCAAACAGCG AAAATTCTTT TCCCGGTTGT TATTACGATT GTCGCAGGGT 7200

TAATTGCGCC GGTGTCATTA CCATTAGTTG GTTTCTTAAT GTTTGGGAAC TTATTAAGAG 7260

AATGTGGCGT TTTGGATCGT TTATCCATTA CAGCTCAAAA TGAATTGGTA AATATTATTA 7320

GTATCTTGCT AGGACTATCT ATTTCCGTTA GCATGAACTA TGCAGAGTTC TTAAAAGTCG 7380

ATACGTTGAT GGTTATTGGC TTAGGGTTAG TCGCCTTTAT TATGGACTCA GTCGGCGGCG 7440

TGCTTTTTGC TAAATTCTTG AATTTATTTA GAAAAGAAAA AATCAATCCA ATGATTGGTG 7500

CTGCGGGTAT TTCCGCTTTC CCAATGTCTA GTCGGGTTAT TCAAAAAATG GCTACGGATG 7560

AAGACCCACA AAACTTTATT TTAATGCATG CTGCTGGTGC CAACGTTTCA GGACAAATCG 7620

CTTCTGTAAT TGCTGGTGGC TTACTATTAG CCTTTTTAAC ATAGGACGGA AAGGAGAACG 7680

AATACATGTC TGTTAACATG GAAGATTTAA AAATTGCTTT TGAACTGTTA GGTTTTGGTT 7740

GGGGCGGCGT TTTCGTTGTT TTATTTATCA TTTATCTAGC ATCTAAATTA CTAACAAAAT 7800

TATTTCCAAT TAAAAAATAA AGGAGGTTAT TGCTGTGAGC AGCGTATATA CAATGAAAAG 7860

AATCTGGTTG CAAAAAGATA AAAAAGCTTA TCAAGCTTGG CGCGAATTAA TGGAAAAAGC 7920

CGAATTGCAA ACGACAGAAG CACTTGATTA CACCGTTGGC ATATATGATG ACACAGCATT 7980

AATCGCCTCT GGCTCGCTTT CAGGCAAAAC GATTAAGTGC GTGGCCGTTT GTAAAAAATA 8040

TCAATCTGAA AATTTGTTAA CCCAATTGGT GGTTCATCTA CTTGAAAAAC TGAGAGAAGA 8100

AAACCAATTA CACAGTTTTT TGTATACAAA ACCAAAAAAT GAACAAATTT TTCAGTCGCT 8160

AGGCTTTCAA AAAGTGGTTG CGAATCAAGA GGTGCTATTT ATGGAGCAAG GCAAACCAGA 8220

TTTTGCCGAT TATCTTGGTT ACCTGACACA GCATAAAAAA GCAGGGCCAG CCAGTAGCAT 8280

TGTCATGAAT GCAAATCCTT TTACGAAAGG CCATCAGTAC TTAGTCGAAA AAGCAGCCCA 8340

AGAAAGTCCG CACGTTTATG TGTTTGTTTT ATCTGAAGAT AAATCGCTAT TTTCAAAAGA 8400

AGCACGTTTC GCAATGGTGC AAAAAGGGGT CGCCCATTTA CCCAATGTGA CAGTCCTATC 8460

TACAGAAGAT TACTTAGTTT CGTCTGCGAC TTTTCCTACC TATTTTTTGA AAGAAAAAGC 8520

GCCATTAGAG GTAGCTGCAA TTCAAGCAAC GCTTGATGCA ACGTTATTTA AAGAAAGAAT 8580

TGCGCCAATA CTTGAAATTC AGCAACGTTA TGTGGGAGAA GAGCCTTATT CAGAAGTGAC 8640

CGCTGTCTAC AATCAAGCAA TGCAACAAGT TTTCGGACAA ACGATAACGC TAACCATTGT 8700

CCCGCGATTA GCGAGTGATG GAGAGCTGAT TAGTGCGACG AAAGTTCGCA AAGCGATGGC 8760

TGAAGGCGAC AAAGAGACAC TCAAGAAGTT TTTACCAGCA ACAAGTTATC AATACTTAGT 8820

AGAACACAAG AAACTAATGA GGTGAAGAAA GTGAAGATTA TCAAAAATGC TACTGCAGGG 8880

ACAGTAGAAT CAAGTGATAT TATGATTACC GTCCAACCCA TTGAATCAAC AGAAAATACC 8940

ATTGAATTAG AAAGTAGTGT CGAAAAACAA TTTGGCAATC AAATTCGTCA GGTGATTACG 9000

ACAACTTTGG AACATTTAGA TGTCCAAGGC GTTGCGGTCA AGGCGATTGA TAAAGGGGCA 9060

TTAGATTGTA CAATCAAAGC TCGGATGATT ACAGTTTTAC ATCGAGGCGC TGAAAAAGAA 9120

ACGTATGATT GGAAGGAGTT GGACTCATGG AACGTTTAAG AAGAACAATG ATGTTTGTCC 9180

CTGGAGCGAA CGCGGCAATG CTAAGAGATG CACCTTTGTA TGGGGCAGAT TCAATCATGT 9240

TTGATTTAGA AGATGCAGTT TCATTAAAAG AAAAAGACTC TGCCCGTGTT TTAGTCCATT 9300

CTGCTTTGAA AACTTTCGAT TATGGCAATA TTGAAATTGT TGTTCGGATC AATGCTTTAG 9360

ATGCTGGTGG TGCAGAAGAT ATTGAAGCAA TGGTCTTAGC TGGTGTCGAT GtGATTCGTT 9420

TACCAAAmAC AGAAACAGCC CAAGATATCA TTGATGTCGA GGCAGTCATT ACAGAAGTAG 9480

AACAGCAAAA TGACATTCCC GTAGGAACAA CGAAAATGAT GGCTGCCATC GAATCGGCCG 9540

AAGgTGTTCT AAATGCGCCC GCGATTGCCA AAAGTTCAAC TCGCTTGATT GGCATCGCGT 9600

TAGGCGCCGA AGACTATGTT ACCAATATGA AAACACGGCG CCATCCAGAT GGCCAAGAAC 9660

TATTCTTTGC TCGTAGCATG ATTTTACATG CCGCTCGAGC AGCTGGGATT GCGGCTATTG 9720

ATACGGTCTA TTCAGATGTA GACAATACGG AAGGCTTTGA AGCCGAAGTT CGTTTGATTA 9780

AACAGCTAGG TTTCGATGGA AAATCGGTGA TTAACCCTCG TCAAATCCCA TTAGTTAATA 9840

CTATTTATGC GCCAACAGAA AAAGAAATCC AGAATGCTAA GGAAGTAATC TGGGGCATTC 9900

GGGAAGCCGA AGCAAAAGGT TCCGGTGTTA TCTCCGTAAA TGGAAAAATG GTTGATAAAC 9960

CAATTGTTGA ACGAGCTGAA CGGGTCATTG CGTTAgcATT GGCGGCTAAA TTAATTACCG 10020

AGGAGGAAAT ATAGACATGG TAGTAAATAA AGTAGGCAAA GACATTCCTC AATCTTATGC 10080

AGAGCAATAC GGCATTTATG AAGGAGAATT AGCACACATT GGTGCTTATC AAGAAGCGAG 10140

CCGAGCAATT AAACCAGTGA AACCTCGGGA GACAAAACTT TTAGGAAGTA TTCGTGAAGC 10200

AATTGAAAAA ACAGGGTTGA AAGATGGCAT GACCATCTCT TTCCATCACC ATTTCCGTGA 10260

AGGAGACTAT GTGATGAATT TAGTTTTAGC CGAGATTGCC GCCATGGGGC TGAAAAACAT 10320

TTCGATTGCA CCTAGTTCAA TTGCAAATGT TCATGAACCT CTGATTGAAC ACATTAAAAA 10380

TGGAGTAGTC ACCAACATTA CTTCTTCTGG TTTGAGAGAT AAAGTAGGAG CCGCTATTTC 10440

AGCAGGTATC ATGGAAAATC CTGTGGTGAT TCGTTCACAT GGTGGCCgGG CACGGGCAGT 10500

AGCAGCAGGC GATATCCATA TTGATGTAGC TTTCTTAGGT GCACCAAGTT CAGACGCTTA 10560

TGGAAATGCG AATGGAACCA AAGGAAAAGC TACCTGTGGT TCTCTTGGAT ATGCGATGGT 10620

CGATGCAAAA TATGCGGATC AAGTGGTCAT CATTACTGAC ACATTAGTAC CATATCCTAA 10680

TACCCCTATC AGTATTCCAC AAACAGATGT TGATTATGTG GTAGAAATTG ATGCAATTGG 10740

TGATCCAGAT GGAATTGCCA AAGGCGCGAC ACGTTTTACA AAGAATCCTA AAGAATTATT 10800

GATTGCTGAA TATGCCGCAA AAATAATTAC ACATTCTCCT TACTACAAAG AAGGCTTCTC 10860

ATTCCAAACA GGAACAGGTG GCTCTTCACT AGCAGTTACG CGGTTCATGA GAGAACAAAT 10920

GCTTAAAGAT GGCATTAAAG CAAGTTTCGC TTTAGGCGGA ATTACGAATG CAATGGTGGA 10980

ACTTTTGGAA GAAGGCTTAG TAGAAAAAAT TATTGACGTC CAAGACTTTG ATCATCCATC 11040

AGCTATTTCT TTAGGGGAAA ATGCCAATCA TTATGAAATT GATGCCAATA TGTATGCCTC 11100

ACCGCTTAGC AAAGGGGCGG TTATTAATCA ATTAGACACA GCTATTTTAT CTGCTTTAGA 11160

AGTGGACACA GATTTTAATG TAAATGTCAT TACTGGTTCA GATGGTGTCA TCCGTGGTGC 11220

TTCTGGCGGA CATTCCGATA CAAGTATGGC TTGTAAAATG AGTCTAGTGA TTGCACCACT 11280

TGTTCGTGGA CGGATTCCAA CAATCGTTGA GCAAGTAAAT ACAGTCGTTA CACCAGGTAC 11340

GAGTGTCGAT GTCGTAGTAA CAGAAGTCGG CATTGCAATC AATCCAAAAA GAACAGACTT 11400

AATTGATTGT TTTAAAACCT TAGATGTTCC TCAGTTTACG TTAGAAGAGT TGAAAGACAA 11460

AGCCTACAAC ATTGTTGGCA CCCCGGAACC AATTAAATAT GGAGACAAAG TCGTTGCGTT 11520

AATTGAATAC CGAGATGGTA GTCTAATTGA TGTGGTACGC AATGTTTAAT GGGGAAAAAG 11580

TAACACTTTT AGAGATGTTG GATGCTCGCG AACAAAGAGC TGCCACACAG AAGGAGTTGT 11640

TAGAAACAGC ACCAGAAGCT AGTCTCTTAT CTGCCACAAT GAATATTCCT GGAGAAGTAA 11700

AAAATTCGCC CACCTTAACG GCTGTTTTTT TAGAAGTAAT TGACGAAGTC GAGCAACAAT 11760

TGCTGGATCA AGTACCGATT GTTAATTTTT ACCGGAATGA AAAAACAGGT CCTGAATATT 11820

ATTTGGCGGT TTCGTTAGCC CCACAAGAAT TAAAACAAAG AATGGTCAAA ATTGAAGAAA 11880

CACATCCTTA TGGTCGTTTA GTGGACCTAG ATGTTTTATG GGGAAATGAG GAGCTAAAAA 11940

GCCTTCATCG AGGAGATTTA GGCTTGCCAC CACGGCGCTG TTTTATTTGC CAAGAAGTGG 12000

CGAAAGTCTG TGGACGAAAT CGACGTCACA GTCTTGAAGC AATGCAAGAA AAAATAACAG 12060

AAATAATTTT GACACGAAAG GAGCAGCCAA GTGAGTAAAA AAATTCGTTT TACAGAAACG 12120

GTGCTACGGG ATGGACAACA AAGTTTAATT GCAACACGGA TGCCAACCTC AGATATGCTG 12180

CCGATTATTA AAACAATGGA CGAGGCAGGA TTTCATGCAT TAGAAATGTG GGGCGGGGCA 12240

ACGTTTGATT CCTGTGTCCG CTATTTAAAC GAGGATCCTT GGGAACGTCT TCGCCAAATT 12300

CGCAAAGAAG TTAAAAATAC CAAACTTCAA ATGTTATTAC GAGGACAAAA TTTACTAGGA 12360

TATCGTCATT ACGCTGATGA TGTGGTGCGG GCCTTTGTCG AAAAATCAGT TGAAAATGGG 12420

ATTGATATTA TTCGTATTTT CGATGCCTTG AATGATGTCC GTAATTTACA AACAGCAATT 12480

CAAACAACAA AAGAAGCTGG CGGACACTGT CAAGCAGCGA TTTCTTATAC TACGAGTGAA 12540

ATTCATACCA TCGATTACTT TGTAAAATTA GCCAAAGAAT TAAGTCAAAC AGGAGCGGAT 12600

TCCATTTGTA TTAAGGATAT GGCAGGCGTT TTGACGCCAC AAACTGGTTT TGAATTAGTG 12660

AGCAAAATGA AAGATGCCAT TGATTTACCT TTAGAGGTGC ATACGCATGC GACTAGTGGT 12720

ATTTCTGAGA TGACCTATTT AAAAGTTGCT GAAGCAGGTG CGGATATCAT CGATACAGCT 12780

ATCTCTTCTT TCGCTGGCGG TACGAGTCAA CCAGCCACAG AATCTGTAGC CATTGCGTTA 12840

GAAGATTTAG GCTTTGAAAC AGGTTTAAAT ATGGAAAAAG TAACAGAGAT TGCTGACTAT 12900

TTCAATCCGA TTCGTGATCG CTTTAGAAGT GAAGGCATTT TAAACCCTAA AGTTAAAGAT 12960

ACGGAACCGA AAACCTTGAT TTATCAAGTG CCTGGTGGTA TGTTATCAAA CTTATTAAGT 13020

CAATTAACGG AGCAAGGTTT ACAAGACAAA TATGAAGAAG TTTTAGCTGA AGTACCTAAA 13080

GTAAGAGCAG ATTTAGGGTA TCCACCACTA GTTACGCCGC TTTCACAAAT GGTTGGAACA 13140

CAAGCCTTAA TGAATGTGAT TTCTGGGGAA AGATATTCCT TAATACCGAA TGAAATCAAA 13200

GATTATGTGC GCGGAGAATA TGGCAAATCA CCAGCACCTA TCGCAGAAGA GATTAAGCAA 13260

AAAATCATTG GGGATGAAGA GGTAATCACT TGTCGACCAG CGGATTTATT GAAACCAGAG 13320

TTGCATAGTT TAGAAAAGGA AATCCAACAA TATGCAAAAT CTGAAGAGGA TGTCCTATTA 13380

TATGCGATGT TCCCGCAACA AGGAAAAGAC TTTTTAGGAC GTCGTGAAGA TCCTTTTTAC 13440

GATGTACCGT TACAGAAAGT GACAGTCAAA CTTGACTTGC CAGAATAGAA CGGGCAAGTG 13500

ctCCGCAAAT GTATAAAGAA GCAAGGGTTT GAGCGACGAT AACGGTTGTT CAAACCCTTG 13560

CTTCATCAAA TrGAAAAAAG GCAGCCCGTG AAAGGAAGTG ACAATGTGTT AGAAGAAGTT 13620

CTAGCAATTC ATCAAAAAAA TATTGGTGTG CTAAGTCTTG AAGCAACAGA AGCGGTGTTA 13680

AACCATCATG ATTTAGCCAA AATGTATACA CCTGGTGTCG CAGAATTGAG TTTAATGATT 13740

GCCCAAAATC ATGAATTAGC GCGCGAGTGG ACCATTAGTG GTAAGTTAAT TGCGGTTATT 13800

ACAGATGGAT CAGCTGTTCT GGGTCTCGGC AATATGGGCA CGCAAGCGGG CTTGCCAATT 13860

GTTGAAGGCA AAGCTTTACT TTATAAAAAT TTGGCGGGGG TCGATGCTAT TCCTTTAGCA 13920

CTGGAACAAA AATCAGTGGA CGAGATGGTT CAAACAATTG AAAATTTACA AAATAGTTTT 13980

GCAGGAATCC ATTTAGAAGA TATTTCTGCA CCAAAATGTT TTGAAATTGA AGAAAAACTT 14040

CAACAACGTT TGAACATTCC TGTGTATCAT GATGACCAAG AAGGCACAGC CATTGTCGTT 14100

TTAGCTGGGC TAATCAACGC AGCCAAAATT AAAGGAAAAC CTTTAAATGA ATTGCGTGTG 14160

GTGATTAACG GAGTTGGTGC TTCTGGTGTA GCGACCGCTA AGTTATGTAT TCAAGCGGGA 14220

ATTACGCATC TCACCTTAGT AGATCGCCAA GGTGTTTTAA GAGAAGAAGA CCCGACATTG 14280

AATCCTTACC AACGTGCTTT GCTTCGACAG GTAATAAAAC CATCAGTAGA AAACAAAGAT 14340

TTAGCCACCG CTGTCGTCAA TCAGGATGTT TTTTTAGGCT TATCAGAAGC GGATGTGctA 14400

ACGCCAGCCC TGATTAAATC AATGAACCAA GACCCAATTA TCTTTGCTTT AGCAAATCCT 14460

AAACCAGAAA TTGAACCAGC CCTCGCCCAA GCCAATGGCG TCCGTTTATT AGCAACTGGT 14520

TCTTCCAAAT ATCCGAATCA AGTGAACAAC ATTTTAGTTT TTCCAGGTCT TTTTAAAGGG 14580

TTATTGGCAG CGAAAGGCAG AAAAGTCGAT GTCGGTTTGC AAATGACAGT AGCACGAAGC 14640

TTGGCCGCTA TGATTTCAGA ACCCACAGTA GAAAAATTCA TACCTAATGT ATTTGATGGC 14700

GGCGTAGTTG ACACGGTTTT TAATGCTGTA CTAGACTATA TAAAACAAGA AAAGACCACG 14760

GCAGAAGAAG GGACGTAAAA GGCAATGACA AAGCAAGCAA GGAGTAAATT AGCGCAgcAA 14820

GCACAATTGG CACTACTCTA TGAAGTGACC TGTTTAAATA AACCAGGCTT GGTGGATCCA 14880

```
GTCGATGCAG GCGCCCATCA AGATATGGAT GTTTTCACTT TTTTAGAAAG TAGTGTGGTT    14940
TTGACTCCTT ATTTTGAAAC CTTCGTTCAA GCTGGGACAG AGCTACGTCA TTTGCCGATT    15000
GAGGCAACAT TTCGCACGAT TCGACAAGTT GGTTTAGAAG CAGAACGAGC GATGTTTGAG    15060
GCAACAGAAG GTATCAATAC ACATAAAGGA GCAATTTTTT CTCTAGGTAT TTTTTTAGCT    15120
GTCTGTGGTC GTCTAATGAT TTGGGAGACC CCTTGCACCT TACAATTGTT TCAACAAACG    15180
CTTCAAACAA TGACGCAAGA CTTATTGGCT GACTTTGAAA ACCTCGCTAC GAATGAGCTA    15240
GAAAAACTAA CTTGGGGTGA ACGACTATTT GTTCAGCATG GCTTGACTGG AATCCGTGGT    15300
GAGGCGCAAC AAGGTTATCC CGCTGTCTTT GAACAGGGAG TGCCTTATTA TCAAAAACAT    15360
CAAGGCACAC AACAACAAAA ACTCATTGAT ACGCTCCTTT ATTTATCGCT CTATGTTGAG    15420
GATACGAATT TGATTAAACG AAGTCAAAAT GTCCAAATTT TTGAAGTATA TCAACCGTTA    15480
ATCCAGGAAT ATTTTGCTTT AGGGGGGACA CAAACGTTAC GGGGAACCCT TTGCTTAGAA    15540
AGATTGAATC ACTTATTCAA AGAAAAAAAT TGGAGTATTG GTGGCAGTGC AGATGCTTTA    15600
ATTTTGGTGG TGTTTTTAGA TAAACTAATG AAATTAAATT GGCTGACGCC TTAATTAAAA    15660
AGATAGTTAC AACGCTGATT TGGTCAGTTA TTTTGACCTT ATATGCCTAT GGAAAGGTCT    15720
CAAGAGAATT CAAGAAAGCG ATTGACAAAT ACCACGAAAC ACAGTATATT TTATTACGAA    15780
ATCATTCTTA TCAGAGTATA CAAATACGAT GATAAACAGT TGAAGCGAAT CTTGATTTGC    15840
TTTTCCTAAA CCTGATTCTT TTAGAATCAG GTTTTTGTTT ATCTAAAAGG CGGCATTCGT    15900
TCAAAAACGG CTGTCGCCTT TTTTTATGAA ATTTTTTTAA TATAAAAAAT TTCATAAAAA    15960
AATCTTAAAA ACAGAAAAAA TTGTTATAGA TTAGAGTCAT TTTTTTGATG TAAAGTTTTT    16020
ATAACTTATG TCTATAAAAT TGATAAAAAC AGTCAAAGTA GTAGAAAAAA TAATAGTGGA    16080
ATTTCTCAAA AAAATAAAAA AATTTTCCGT TATTTAATTA TCGATTTATC AAACGATAAG    16140
TTATAATAAA ATAATTTATT TATCTTTAAT GATCGGGGAG GAACGTATAA ACTTCCTCaT    16200
TTTCCTCTCA TTGTATAATA ACTCACTAAT GAATTAGTGA GTTTTTTAAC ATATCCAGGA    16260
TGCTTGTTAT CAAATGAATT TTTATATAAA AATTCACAGC ATTATATTTT GAAAAGCTGA    16320
AGAAATGTAC TAAACTGCAG GTGTTAAATC CATTTAAATC GCTAAAGGGG GAGATGAACA    16380
ACTAGGGGAA GCTAATGATC TTGGTATTTA TCGTTTATTT TAAAGAAAAG AGGGACGATC    16440
AGATGAAAAA GAAAATTGTT GAGGATTTTA ATCGGAAAAG TCAGCATAAA AAATGGACAA    16500
AACGCAAGAT GCTTAATTTA GCAATATCAA GTGGTTTATT ATTTACGTCA TTAGCAATCC    16560
CTGTAAGTAT AGCTGTTACC TCTGGCACAA TCAGTGCATC AGCAGCGGTC TTGGATATCG    16620
AACTATTATC AAATGTTACG TCAAATAATG ACAGTGGCAC TTCAACGAGT AATCGTTGGA    16680
CAGCCGCAAA CCAAAATCAA CCAGTTAATT TCACGGTTTC TGGTGGCGCT TTAGCAGATG    16740
CTTCCGCTGT GTTTAGTGGA CAAAAACAAG CGGTGTTAGT GGTTCCTCCT GAGTTAAGAG    16800
GAAATGTAGC TGCAGCAGGC AGCGCAgcAA TCAATACCAA TGTCACGATT GATCTTTCAA    16860
```

AAGTTACTTT TTTGACTGCC GTTTTGAATG CAGCCAATGA TTTAACCAAT GTGATTACTC 16920

AAATTACCAG TGGGGCGTTA GGGAATTTAA CTGGTGTTGA TATTGATTTG ACGGAAGTGA 16980

ATCGTCAATT GGAATTAGTT AATAACATTG AAAACTTAGG TGCTGCTTCA TTTACAGCTC 17040

CGGAAACGTT AGCAGCTGAC GGCTCATACA TTAGTGCACC GATTAGTGAT GGTTTAGGGT 17100

TAGTTTTAGC CCAAAATGTT TCAAACATCT TACAAGATTT GAATGCGGCA GTTCAAGCTT 17160

TGGAGGCAAA AGGTACCAGT ATCCCAAGTA ATCTTGTCGC CGCAGCTATA AATGCAGCCT 17220

TGCTTCCTGT CAAAGGCACG GTAAACGTGG CTGTTTCAGG TGCTTTGCCT TTATTAGCGG 17280

TTGGTGGTTC AGGCGTAAAT GAGTTAGTGG ATGCTTCTTT ACTAGGCACA ACCACGGTTA 17340

CTTTACCAAC TACCGTTTCA ACACCTCAAA ATTTATCCAA TAATTTAGAT GCTCGTTTTG 17400

TAGGAACAGT CGTTCAAACA GATCTTTTAG ACGTTAATTT ATTAGCAACA GCAGACGGTG 17460

TATCCAACAT TTATTTTGCT GCAGGCACTA CTAGTGAAGT AACCGCACCA ACAATCACAG 17520

GAGTAACAGG TAATTCAACA GCAGGTTACG AAGTTAAAGG AACTGCCGAT GCCAATGCCA 17580

CGGTTGAAAT CCGAAATGCA GGAGGCACCG TAATAGGCAC AGGTACCGCT GATGGGACAG 17640

GAGCGTTTAC AGTTACCGTT CCCGCAGgTG AAGCAGGCGC CAATGAAACG TTAACCGCCG 17700

TAGCGAAAAA CGCCAGCGGy ACAGAAAGyA CGCCAACAAC GTTCCAAACr CCAGCGGATG 17760

AAGCAACCGT AACCGCACCA ACAATCACAG GAGTGACAGG TAATTCAACG GCAGGTTACG 17820

AAGTTAAAGG AACTGCCGAT GCCAATGCCA CGGTTGAAAT CCGAAATGCA GGAGGCACCG 17880

TAATAGGCAC AGGTACCGCT GATGGGACAG GAGCGTTTAC AGTTACCGTT CCCGCAgGTG 17940

AAGCAGGTGC CAATGAAACG TTAACCGCCG TAGCGAAAAA CGCCAGCGGC ACAGAAAGTA 18000

CGCCAACAAC GTTCCAAACA CCAGCGGATG AAGCAACCGT AACCGCACCA ACAATCACAG 18060

GAGTGACAGG TAATTCAACA GCAGGTTACG AAGTTAAAGG AACTGCCGAT GCCAATGCCA 18120

CGGTTGAGAT CCGAAATGCA GGAGGTGCCG TGATAGGTAC AGGTACTGCT GATGGGACAG 18180

GGGCATTTAC AGTTACCATT CCCGCAGgTG AAGCAGGTGC GAATGAAACG TTAACCGCCG 18240

TAGCGAAAAA CGCCAGCGGT ACAGAAAGTA CGCCAACAAC GTTCCAAACG CCAGCGGATC 18300

CTAATACGCC CGTGGCGACG CCAATTGTTG AGACTGTAAC AGGTAGTACA ACAAAAGGCT 18360

ATGAGGTCAA AGGGACTGCT GAAGTTGGCA CCACCATTGA GGTTCGCGAT GCAGCTGGCA 18420

CGGTCCTTGG TACTGCAACA ACTGGAACTG ACGGAAAATA TACAGTGACT TTAGATTCAG 18480

GAACAGCAAC AGCAAATCAA ACGCTGAGCG TTGTAGCGAA AAACGCTAGT GGCACGGAAA 18540

GTCAACCAGC AACGGCGACA ACACCAGCTG ATGTCACTGC ACCAACAGTT GATAACATCA 18600

CAGGCAACTC TGGTTCGGGT TATGAAATTA CAGGAACAGC AGACCCTAAC ACAACAATCG 18660

AAGTTCGTGA TCCATCTGGG GCAGTCATTG GTACAGGTAC CTCTGATGCG AATGGTGATT 18720

TTACTGTAAC GCTACCAACG GGAACGACCA ATCCTGGGGA TACGTTAACA GTGATTGGAA 18780

AGGATAACGC GGGAAATGAA AGTCAACCGA CTGAAGTCCT TGTTCCTGCT GATGCCACGG 18840

```
TTACAGCACC AACTGTAACA GGAGTAACAG GTAATTCAGT TGCTGGTTAT CAGGTGACAG     18900
GCACCGCTGA TCCGAATGCT ACCATCGAAA TTCGTGATGC AGATGGGAAC GTGATTGCAA     18960
CAGGGACTGC CGATGGGACT GGTTCCTTTG CTGTGAACCT TCCAGCTGGG ACGGCAAATG     19020
CGAATGAAAC ATTGACAGCG TTAgCCAAAG ATCCTGCTGG CAATACAAGT ACACCGACAA     19080
CCTTCCAAAC ACCAGCAGAT GAAGTAGTGG CACCGCCAAG TGTCGACAAA GTTACTGGGA     19140
ATACAACACA AGGATATCAA GTGACAGGTA CCGCTGAACT TGGCACCACC ATTGAAGTTC     19200
GTGCAACAGA CGGAACAGTT TTAGGCACCG CAACAACTGG ACCGACTGGC CAATATACTG     19260
TGACGTTAGC TTCAGGAAAA GCAACAGCTA AACAAACAGT GAATGTAGTT GCTAAAAATG     19320
ATACTGGACT TGAGAGTCAA CCAACTACAG CTATGACACC CGCTGATGTT ACCACACCAA     19380
CAATTGGTGA CATTACTGGA GATTCAACAA CTGGTTATGA AATCACTGGG ACGGCGGACC     19440
CTAATACCAC CATTGAAGTA CGGAACCCAG ATGGAACAAT TATTGGTACA ACGACAACGG     19500
ATGATCAAGG AAACTTTACT GTGGACCTTC CAGCGGGAGC CGCTAATCCT GGTGATACAT     19560
TAACAGTTGT TGGAAAAGAC GGTGACGGCA ATGAAAGTCA ACCAACGGAA GTGACGGTCC     19620
CTGAAGATGC AACCGTAGCA GCACCAACTG TGACGACTGT TACAGGAACA ACTGCCACTG     19680
GGTATCAAGT AACCGGCACG GCAGAGCCAA ATGTCACCAT TGAGATTCAC AATGAAGCAG     19740
GTTTAGTTAT TGCTACGGGA ACGACTGATG GTGCTGGCGC ATTTACAATC ACTCTTCCGA     19800
CGGGCACAGC AACAGCTAAC GAAGCCTTAA CTGCCATTGC GAAAGATGCT GCTGGGAAAG     19860
AAAGTAATCC GACTGCTTTC AAAACACCTG CTGATCCAGA TGCACCAGTC GCGACACCTA     19920
CTGTTGACAA AATCACTGGT AGCACGACAA ACGGCTATCA AGTAGTAGGA GCAGCAGAAG     19980
TTGGTACAAC AGTTGAGGTG CGTGACGCCG ATGGCACAGT CCTTGGCATG GCAACTACTG     20040
GAACTGATGG CAAATACACA GTGACTTTAG AGCCAGGGAA GGCCTCAGCT AACGAAACAA     20100
TAACTGTCGT AGCGAAAAAT GCAACAGGAA AAGAAAGTCA GCCAGCTACA GCAACTACAC     20160
CAGTCGACTT AGCCACACCA ACCATTGATT CTATTACCGG AAATTCTAGT AAAGGTTACG     20220
AAATCACTGG AACGGCGGAG CCAAAAACCA CTATTGATGT CCGTGACGCA GACGGAACCA     20280
TCATTGCTGC TACAACTGCT AACGAAACCG GCCAATATAC GGTGACTCTA CCAGCTGGCG     20340
TAGTGACACC AGGAGAAACG ATTACGATTA TTAGCAAAGA TGGCGCAGGT AATGAAAGTC     20400
AACCAGCTAC AGCCGTTATT CCAGCGGATG TTGTTTTAGC GGCGCCAACT ATTACGAAGG     20460
TTGAAGGAAA CAAAGCCAAT GGCTATACAG TCACTGGAAC TGCTGATCCA AATGTCACGG     20520
TTCAATTTTA CAATAGCAGT GAACAATTAT TGGCAAGTGG CAATACAACT ACTGGAGGTA     20580
CCTTCTCCGT TCATATTGCA GCAGGGTTAG CAACAGAAAA AGAAACGTTA ACCGCACTAA     20640
CCACAGATAC ACAAGGAAAT GTGAGTCCTA AAACCACATT TATGACGCCA GCCGATATTA     20700
CGGGAGAACC AGAGATTAAA ATTGCGGCAC CAACTGTTTC TTCAGTTTTA GGAACGTCTA     20760
AAGCCGGCTA CCTCATCAAA GGAACAGCTG AACCAAACCG AATCATTCAA ATTAGTAACC     20820
```

GACTATTAAG AAGTGTGATT GCTGTAGGTG CCACCGATGC TGAAGGCAAC TTCGCTATCC 20880

AATTAACAGC GGGACAAGCG ACTGCTCAAC AAAGTTTACT TGCGACAGCT ACCGATGGCG 20940

CAGGACATTA CAGTACGGCT ACAACCTTCA TGACGCCAGC CGACCCAACG AATCCTGGAG 21000

GAGGCAATGG TAACACTGGC GGAAATAACG GCAATACAGG CGGCAATACA GGAAACAATG 21060

GCGCAACTGG CGGGAATAAT GGGAATGGTT CAAACACAGG TTCAAATCCA AATGGAGGTT 21120

CTGGTTTAGG CACAACAGGT TCTGGCTTAG GTTCACTAGG CAATGGCCTC GGTACAAATG 21180

GTAGTGGCTA CCACCCTAAA CTAAGTACCA TCAGTTATGG CACTGGAAAT CACGGGAAAA 21240

CAGGCTACTT ACCTAGCACA GGTGAAAAAG AGTCTTCAGC CGTGACAACA AGTTTGTTTG 21300

GCGCCTTTGT CGCACTCCTT GCGAGCATGG GAATCATCAA ACGCAAACGT AAAAACTAGC 21360

TCTCTCCTTT AATTCCTATA GATGAGAAAA CGTATTGTTG AGAATTCCCC ACTCAATGAT 21420

ACGTTTTCTC GTTCCATTTA ATGAAGAAAA ATTGCCAAAT AGAATAAAAT GTGCTATTCT 21480

ATTTTCAGAA AATCGAAGGA ACAAGCAGTG AAAGTGCTAA ATCCAACCTG AATTTTTTAA 21540

TGCAGGAGGA GTGGGCGCTT TTTTTGTTGT TATGAAGAAA GGAAGCTAAT TATGCAATCA 21600

GCTACAATGG AATTTGATAC AATTGCGGCG ATTTCTACGC CGCCAGGTGA AGGTGCGATT 21660

AGTATTGTTC GTTTAAGTGG CGAGCAAGCA GTTGCTATTG CAAACAAAGT TTATCGGTCA 21720

GGCACTAAAG ATTTAGCTAA AGTCCCCACG CACACGATTC ATTATGGTCA TATTGTTGAT 21780

CCGCAAAACG ATCAATTGAT TGATGAAGTC ATGCTTTCTG TGATGCGGGC ACCGAAAACA 21840

TTTACACGAG AAGACGTGGT AGAAATTAAC TGTCATGGCG GCATTGTCGT GGTGAACCAA 21900

TTATTGCAAT TACTGTTACG TGAAGGCGCA CGAATGGCCG AACCTGGGGA ATTTACAAAA 21960

CGTGCTTTTC TAAATGGCCG AATGGATCTT TCACAAGCAG AAGCAGTGAT GGACTTAATC 22020

CGTGCCAAAA CGGATAAAGC GATGAATGTT GCATTGAACC AATTGGATGG TAACTTGTCG 22080

ACGTTAATTC GTTCATTAAG ACAAGAAATT TTAAATACTT TAGCACAAGT AGAAGTTAAT 22140

ATCGATTATC CAGAGTATGA TGATGTCGAA GAATTAACCA CGAAATTACT TTTAGAAAAA 22200

GCCGAATTTG TGAAAGCGCA GATTCAACAA TTATTAACGA CCGCTAAACA AGGAAAAATT 22260

TTGCGAGAAG GCCTAAGTAC AGCAATTATT GGCCGACCAA ATGTCGGGAA ATCCAGTTTG 22320

TTAAACCATC TGTTACGAGA AGAAAAAGCT ATCGTTACAG ATATTGCTGG AACGACACGT 22380

GATGTCATTG AAGAATACGT TAACGTTCGC GGCGTTCCTT TAAAATTAAT TGATACAGCG 22440

GGGATTCGCG AAACAGAAGA CATTGTGGAA CGAATTGGTG TCGAACGTAG CCGCAAAGCG 22500

TTAGCAGATT CGGATTTAAT TTTGTTAGTC TTAAATCAAA GCGAAGAATT GACAGAGGAA 22560

GATCGTCAAC TGTTGGAAGC AACGAAAGGC TTAAAACGCG TTATTCTTTT AAATAAAATG 22620

GACTTACCGA CAAAACTGGA TCCTAACGAA TTACAAGAAT TAGTCCCTGC AGAAGAAATT 22680

TTATCCGTTT CGGTGTTGTC AAACACAGGC TTAGACCAGT TAGAAGCGAA AATTGCTGAT 22740

CTTTTCTTTG GCGGCCAAAC AGGCGAAAAA GATGCTACCT ATATTTCTAA TACACGTCAT 22800

ATTGCTTTGC TGGATCAAGC AGCGCTTTCG TTACAAGAAG TGATTAACGG CATCGAAGCA 22860

GGTATGCCAG TCGATTTAGT TCAAATTGAT ATGACCCGTT GTTGGGATTA TTTAGGGGAA 22920

ATCGTCGGCG ATAGTGTCCA AGATGAATTA ATTACACAGT TATTTAGTCA ATTTTGTTTA 22980

GGAAAATAAG ACAAGAGAGG AATTTTTTAT GAATCAATAT CAAGCAGAAT CATATGATGT 23040

TATTGTCGTC GGTGCTGGAC ATGCTGGCTC TGAAGCAGCC TTAGCAGCCG CACGAATGGG 23100

CGTTAAAACG TTACTTTTAA CGATTAATTT AGATATGGTT GCTTTTATGC CATGTAACCC 23160

TTCTGTTGGT GGACCTGCCA AAGGAGTCGT TGTCCGAGAA ATCGATGCAT TAGGCGGCGA 23220

AATGGGTAAA AACATTGATA AAACCTACAT CCAAATGCGG ATGCTAAATA CTGGTAAAGG 23280

TCCAGCAGTG CGTGCGTTGC GCGCGCAAGC TGACAAACAT GCCTATGCGA CAGAAATGAA 23340

ACATACCATT GAAAAAGAAG AAAATTTAAC CCTTCGTCAA GGGATTGTTG AAGAATTAAT 23400

CGTGGAAGAT GGCGTTTGTC GTGGTGTTGT AACTTCAACA GGTGCCGCAT ACCGTAGTCA 23460

AGCAGTGGTG ATTACTGCCG GTACTGCCTT GCGTGGCGAA ATTATTATTG GTGAATTAAA 23520

ATATTCATCA GGACCAAACA ATTCGCAACC TTCCGTTGGT TTAGCCAATC ATTTAAAAGA 23580

ATTAGGTTTA GAAATCGATC GCTTTAAAAC AGGAACGCCG CCACGAGTAA AATCAAGTAC 23640

GATTGATTAT TCTGTGACAG AGGAACAACC AGGCGATAAA GAACCCAACC ACTTTAGTTA 23700

CAGCACGCCA GATAGTGCCT ACAATCAAAA CCAAGAGCCT TGTTGGTTAA CTTATACGAA 23760

TGAAACAACG CATGAAATTA TTCAAAAAAA TCTACACCGC GCCCCAATGT TTACTGGGAT 23820

TGTGGAAGGT GTGGGCGCTC GGTATTGTCC ATCCATCGAG GACAAAATTG TTCGTTTTGC 23880

AGACAAACCA AGACACCAAT TGTTCTTGGA ACCAGAAGGG TTGAATACCG AAGAAGTTTA 23940

CGTCCAAGGC TTATCTACTT CTTTACCAGA AGATGTCCAA ACAGAAATGC TTCATTCTAT 24000

TGAAGGATTA GAAAAAGTAG AAATGATGCG TACAGGTTAC GCCATTGAAT ACGATGTGGT 24060

AGTCCCTCAC CAATTACGCC CAACATTAGA AACAAAGGTC ATTGAAAATC TTTATACAGC 24120

AGGTCAAACG AACGGAACTA GTGGCTACGA AGAAGCTGCG GGTCAAGGCT TGATGGCAGG 24180

AATTAATGCT GCATTGAAAA TTCAAGGCAA AGAGCCTCTT GTTTTAAAAC GTAGTGATGG 24240

CTATATTGGC GTCATGATTG ACGATTTAGT TACAAAAGGC ACTAACGAAC CTTATCGCTT 24300

ATTAACGTCA CGTGCGGAGT ATCGTTTGAT TTTACGACAT GACAATGCGG ACTTACGTTT 24360

AACAGAAATG GGTCATGAAA TTGGCTTAGT CAAAGAAGAA CAATATGCCG CCTATCTAGT 24420

CAAAAAAGCG GCTGTTGAAG CGGAAATTGC CCGCTTAGGC AAACATCGGA TTAAACCAAC 24480

CAAAGAAGTG CAAGCATTTT TGGAAACTAA AGGCGCTGCT GGCTTAAAAG ATGGTATTTT 24540

AGCTAGAGAT TTTCTGAAAC GCCCAGAAAT CAGTTATCAA GAAGTCGCTC AATTTATTCC 24600

CGCACCTGAA GAGGCGCTCG ATCCGAAAGT AATTGAACAA GTAGAAATTC AAATCAAGTA 24660

CGAAGGCTAC ATTAAAAAAG CCATGGAAAA AGTTGAGAAA CTAAAACGCA TGGAAGCAAA 24720

ACGCATTCCA GAAAACATTG ATTACCAAGC GATTAACGGC TTGGCGACAG AAGCAAAACA 24780

AAAACTACAA AAAATCCAAC CAGAAACGAT TGCCCAAGCA AGCCGTATCA GCGGTGTGAA 24840

CCCCGCTGAC ATTAGCATTT TGATGGTCTA CATCGAACAA GGAAAAATCG CGAAAGTGCA 24900

AGGATAATAA AAATTTGAAG CAGTTGCTTC TTCACGGATG ATTGATTGTG AAGAAGCAAC 24960

TGCTTTTTTG TTTTGAATTA GTCTACTCTT TTTTCTAAAC TACAAAAAAA GACTGTCTTA 25020

AATAATTTCA CAAACTGACC TATACTGAAA GCGCAAACAA ATGATAGGAG TGGAACAAAT 25080

GACAGATTGG TTAGGAATTA AAGGAAAAGT AGTGATTGTT ACAGGAGGCT CATCTGGTAT 25140

TGGTCACAGC ATTGTTGAAA GCTTGTTAGC ACAAGAGGTC TATGTCGCTA ATTTTGATAT 25200

TTCAGCATCA GAAACGCAAC ATGAGCATTT ACTCTTTGTG AAAGTCGATG TAACTTCTCG 25260

AGCAGAAGTA GAAGCAGGTG TCGCTCAAGT GGTTGAAAAA TTTGGAACCG TAGATGGCCT 25320

AGTAAACAAT GCCGGCATCA ATATTCCACG TCTATTGGTG GATAAAAATC ATCCTCATGG 25380

GCCTTATGAA TTATCGGATG AGGTATTCGA CAAAATTGTA GCCATCAATC AAAAAGGGCT 25440

CTACCTAATG AGCCAAGCAG TGGGCCGGAT TCTTGTACAA AAACAAGCAG GTGTCATTAT 25500

TAATATGTCT TCAGAAGCTG GATTAGAAGG CTCCGAAGGG CAAAGTGcTT ACGCTGCAAC 25560

GAAAGCAGCG GTCAGCAGTT ATACACGTTC ATGGGCTAAA GAGTTAGGAA AATCAAACGT 25620

CCGTGTCGTG GGGATTGCGC CAGGTATTAT GGAAGAAACG GGctTGCGCA CTTTAGCATA 25680

CGAAGAAGCA TTGGCTTATA CACGAAATAT TTCTGTTGAG CAATTGCGAG CAGGTTATTC 25740

AAAAACAAGT ACGATTCCAT TAGGTCGTAG CGGAAAACTC CATGAAGTCG CTGACTTAGT 25800

ATGCTATTAT TTATCAGACC GTTCAAGCTA TATTACAGGC ATCACCACAA ACGTAGCTGG 25860

AGGTAAAACA AGAGGATAGA GGTGGGAAAA CTAAATGACG CTAGTAAATA GATGGTACCA 25920

AGTAATAAAA TTACTAGTTG ATCACAAAGG AATGAGCTTA CAGGAATTGC AAGAGAAATT 25980

AGCCGCAAGT CCTCAAACAG TTCGTAAAAA CATTGATACA TTGAATGATG AATTAATCGG 26040

TATCGCCCAA ATTATTCAGA AGGAAAATCT CTTTCAGTTA GAAATCAATA ACTTTGAAGG 26100

TTTTGAAGAG GTTTTGTCTG GACGGCTAAA AAGGGAATCT GATTTTAATT CTTCAAGCAA 26160

GCGTGTTTCC TACATTATCA AACGATTAAT AGAAGAAGAT CAATTTATTT CAACTTATGA 26220

TTTATCCGAA GAACTCGCCG TTAGTCGCGG CACAGTTAAT AAAGACATTA AGCGCATGAA 26280

AGAGCTGATT GCTCCTTGGC AGGTTGCGGT TGTAGGCACG CCAAACCGTG GCATTCATTT 26340

AGAAGGAAAA GAGTTTGATT TACGCTTACT TCATGTGAAT TACGTCCAAG AATATTTTGA 26400

AGAACAGTTT CTTCATGAAA CTACCAAACA AATGATACAA AAAATAATCA AAGAAACAAG 26460

ACTAGCGAAA CAAGATAGTT TTCTTTTACG AAGAGTGGTT TCAATTGTTT TGCAGCGTGT 26520

TTTAGCTGGA AAACTCATTA CGGAACTACC ACCAGAATAT GTTGATTATG TTCGTCATAA 26580

CGAGCAGATT GAAGAACTGA TGTATCACTT GGAAATAACC TATAATGTAA CGTTAAGCCA 26640

ATGGGAACGT TCCTTTATCA GCTTTCCTTT TAACATCAAC ACAAACCATA TTCGTAATTC 26700

TTTATTGGCT GATGAAGGAC TCTTAGCTGA CTATTTTCAA AAAATGATGA AAAAAATTCA 26760

TCATTCAGTT GTTGTTGAAT TTGATGAGGA CTTCTTGTTT TCAGAAATGA AAGATCATTT 26820

AAGAAACGTG ATGAATCGAC TCGTTTTTCA TGTGGAGTGT CACGATTTAT TTTACGGAGA 26880

AATTGAACGA CAATATCCTT TGGCCTATGA ATTAGCCAAG ATTGGCTTGC AAGAATTGGG 26940

CCGTTTGTTA AATCGTTGTG TGCCGACGGT TGAATTTGGG TATCTCGCTT TGTATTTTGA 27000

ACTAGCATTG AGAGGGAACT ATGAAGCGGG AGCCAAAAAA GAAATCGCTG TAATCTGCAG 27060

TACAGGTCAT GGAACGGCCT TGATCATTCA GCGGCAATTG GAAAAAGTTT TGGGTCCAGA 27120

CGTCCAGATT GCCCATTTTT CTGAAGAAGA GGCAGAGAGA AGAGAATTGA ATCAATATTT 27180

TGCTATTTTT ACAACCATTC CCTTAAAAAA TATTCGACCA CAAACACCGA TGATCCATTT 27240

AACTAATTTA TTTAACGATA CTTGGCTCAG AAATGAGTGG CAGAGAGCAA AAGAAATTCG 27300

TTCCGTGGAA TCTCAAAATA TTCATATGAC CTATCAATTA TTAGAGAATA CTCAAGGTTA 27360

CCGAGCGAAT ATCCAGAAAA TGGTTGCTGA TTTAGAAGCA GAAAAATTAG TAGATCCCCA 27420

ATTTATCGAA CGAATTTTTA CTAGAGAAGA TCAACAAACA ACTATTTTTG AATCGGGGAT 27480

TGCCTTTCCC CATACAATTA ATCAAGCCTT ACCAACGATT ATTTTATCGA TTGGGGTTTT 27540

TGAAGAATCA CTGGTTACAC CAGAAGGTAA AGTTGATGTT ATTTTGCTGT TGGGTATTCC 27600

AGAAGAGCTG ACGGCCACTG TTGAAGCAGA GTTACTACAA CTATATGACC GCTTGTTTAC 27660

AATCGCAGGA GACGAACGTT TACGTAATGA GTTACGGAAG CAACGAGATG TTTTAGCGGT 27720

TAAAGAGTGG ATGCAAAGGA AGGGGATTAT CATATGAGTT TTATCTACGT ATTCGGGGCG 27780

CTTTTTTTAA CAGCTTACTT CGTTCAGATA GCTATGGGAA TGAAACAGAT TAAACACTTT 27840

AATCAAAACT ACCAAGAATT GCGACAATTA GGGAAAGTCG CGATTGGTCG ACGTGCAGGG 27900

AAGTTAAAAG CAGGCACGAT TATCTTGTTT GCTGTGGATT CGCAAGGGAA AATCTTAGCC 27960

GCTAGAAAGA TGCAAGGTGT CACCGTTCTA GCTAAATTTC ATCCCTTACC TCAGTACGTG 28020

AACGAGGATA TTCATTATAT GGATCATTAC CATCCACTCG TACAAAAGGA AAATAAATTA 28080

ACCCAACAAG CGATGGAAAA TGCACGAGAA ATTTACTTAC GTGTTGAGCT TGGCAATTAT 28140

CAAGAAACAC AACCATTAAC GCCGCTTAAT GGAGCCAAAA TTCAATGGCA ATTATGGAAA 28200

AATAAAGTAC AAACAAAATT GAAAGGAAGT GTCGAGTAAT GTCTTATATT ACAAAATTTG 28260

CGGAAGGATT TATGCACCTT TTCCAAACAG GAGCAGAAAC ATTTATTTCT TGGATGACAG 28320

GCATCGTACC TGTTGTTTTG ATGTTAATGG TAGCTATGAA TACTTTAATT GCTTTTCTAG 28380

GAGAAGAAAG AGTAACCAAA GTAGCCAAAC TTTCTGCCAA AAATCCTGTT ACGCGGTATT 28440

TAGTGCTGCC TTTTTTATCC GCTTTTATGT TAGGCAATCC AATGTCCTTT ACGATGGCAC 28500

GATTTTTACC AGAATACTAT AAACCTAGCT ATTATGCTGC ACAAGCGCAA TTTTGTCATA 28560

CCAGTAATGG GGTTTTCCCT CATATTAATC CTGGTGAATT GTTCGTTTGG TTAGGGATTG 28620

CGCAAGGGAT TGATACGTTA GGACTAAATT CGATGGACCT AGCGATTCGC TATATGTTTG 28680

TAGGGATATT AATGAATTTT ATTGGCGGTT GGGTTACCGA TTTCACTACC GCCTATGTTT 28740

GCAAACAACA AGGAATTACC CTTAGTAAAA CAGTAGAGAT TGCTGTGGAT TAGAAAGGAG 28800

CAAGGTGAGA TGACGTACAA TAGTATTAAG GTAGTTAAAG GAAATGGCGG TTATGGTGGG 28860

CCCTTAGTGA TTACACCGAC AGAAGAAAAA CATAAATTTA TTTACATTAC AGGTGGCGGC 28920

GAGAAACCCG CAATCGTTGA TAAAATCGTA GAATTGACAG GAATGGAAGC GGTCAATGGT 28980

TTTAAGACAT CGATTCCAGA TGATGAAATC GCATTAGCCA TTATTGATTG TGGAGGCACT 29040

CTTCGTTGCG GAATTTATCC TAAAAAAGGA ATTCCTACTA TTAACGTTGT TCCAACTGGC 29100

AAAAGTGGTC CCTTAGCACA ATACATTAAT GAGACAATTT ATGTTTCCGC TGTAGGCCCG 29160

AAACAGATTA CATTGACATC AGAAGAAGCT TCTGCTGATA CAGAAAAGCC CACCTCGAAA 29220

AAAACAGAAT TTAAATATGA TACAAGTAAA AAAATTACTG AACAAAAAGC GGAACAACAA 29280

AGTTTTGTTG CAAAAAATTGG TATGGGTGCA GGAAAAGTTG TAGCGACATT CAACCAATCA 29340

GCCCGTGAAG GCGTACAAAC GATGATCAAC ACAGTGATTC CGTTTATGGC ATTTGTTTCT 29400

TTATTAATTG GTGTGATTCA AGGTTCAGGA ATTGGTGAAT TTTTTGCCAA TATCATGGCG 29460

CCATTGGCAG GAAACATTGT GGGGTTAACA ATTATTGGTT TTATTTGTTC TCTGCCGTTT 29520

CTTTCTCCTT TATTAGGTCC TGGAGCAGTA ATCAGCCAAG TAATTGGTAC GTTAATTGGT 29580

GTGGAAATTG GGAAAGGCAA CATTCCGCCA CAACTAGCCT TACCAGCATT GTTTGCTATT 29640

AATACACAAA ACGCGTGCGA CTTTATTCCC GTGGGTTTAG GCTTGGAAGA AGCTGAAGCT 29700

GAAACCGTTG AAGTCGGGGT CCCTTCTGTT TTATATTCTC GTTTCTTAAA TGGTGCCCCT 29760

CGAGTACTAG TTGCTTGGTT AGCAAGTTTT GGATTGTATA GTTAAAAAAT TTTTTTTGGA 29820

GGATGGGAAA TGTCAGTATT TACAACAAAG GtTCAGTCTA TTGGACCAGA AGCCGAATTA 29880

TTTAAAGAAG AAAAAATGGT CATCTTATTT GGAAAAGATG CACCAGATGC TTTAGCAGAT 29940

TATTGTTATA ACATTGAGGT TCAGCCAGTC ACAGAAGCGA TTACGGACAA GCAAACGTTA 30000

GTGATTGATG AACAAACTTA CCAGATTACC GCTGTAGGCG AGGTAGTATT AACTAATTTA 30060

GACACATTGG GTCACATTAC TATTAAATTT GATGGCGCAA CAACACCTGA ATTACCAGGT 30120

ACATTATACG TTGAAGAAAA AGCTATTCCT GAAATAACTG TTGGTACAAC AATCACTATT 30180

CTATAATTAT CGTTTAAGGG AGGAACGTAA AAGATGGAAT TTATGCTAGA TACAATTAAT 30240

CTTGAGGCTA TTCGAAAGTA TCAGAAAATT TTACCACTGG CAGGCGTAAC GTCTAATCCC 30300

AGTATTGTTA AACAAGCAGG AAAAATTGAC TTCTTTGCGC AGATGAAAGA AATCAAAAAA 30360

ACGATTGGAC AAGCGAGTTT GCACGTCCAA GTGGTCGGTC AAACGACAGA AGAGATGCTA 30420

GAAGATGCCC AAACCATTGT ACAACAATTA GGCCAAGAAA CATTCATTAA AATCCCAGTC 30480

AATGAAGCAG GCTTGGCAGC CATCAAACAA CTAAAGCAGG CAAATTATCG GATAACAGCA 30540

ACCGCTATTT ATACAGAATT TCAAGGTTAT CTAGCGATTG CGGCAGGTGC AGATTATTTA 30600

GCGCCATATT ATAATAGGAT GGAGAATTTA ACAATTGATT CTCAAAAGGT TATTGAGCAT 30660

TTAGCTGCAG AAATTAAACG GACGAACGCA AAGAGTAAGA TCTTAGCTGC TAGTTTTAAA 30720

AACGTGGCAC AAATTAATCA AGCCTGTCAA ATGGGGGCGC AAGCGGTCAC TATCGCACCA 30780

GAACTCGTGA CACAAGGCTT GGCAATGCCC GCCATTCAAA AAGCAGTAAC TGATTTTCAG 30840

GAAGATTGGG TTGCTGTTTT TGGAGTAGAA ACGGTAAACG AATTAGCATA AAAAGAACGA 30900

GAAGAAAAAG CGGCCATTTT TCTTCTCGTT CTTTTTTTGT TTCACATGAA ACATCTACGA 30960

TTATTCACTT GCTTAAAAAT TAGACAATTT TCTTTATTTG TCAAAGAATT TCTTAAAAAA 31020

AAGGGTTATC TTTTGGGTAT TTTTTTTGGT AGCAGTATGA TGAGGGTATC AAATGATAGG 31080

AGGAAACGAG AATGAGATAC ACAAATGGAA ATTATGAAGC ATTTGCTCGA CCAAGAAAAC 31140

CAAAAGGTGT AGATGACAAG AGTGCCTATA TCGTTGGTGG TGGCCTAGCA GGTTTAGCAA 31200

CGGCTGTTTT CTTAATTCGT GACGGTCAAA TGAAGGGCGA AAATATTCAC ATTTTTGAAG 31260

AACTCACGCT GTCTGGCGGT TCTTTAGATG GTAAATTTAT TCCCCATGAT GGCTTTGTCA 31320

CTCGCGGCGG ACGCGAAATG GAAAACCATT TCGAGTGTTT GTGGGATCTA TTTCGTTCGG 31380

TACCATCATT AGAAGTAGAA GATGCCTCTG TCCTAGATGA GTTTTACCGG TTAGATTTAG 31440

ATGATCCGAA TTCTTCAAAT TGCCGCATTA TTCATAACCG TGGCGAACGT GTCCCCGATG 31500

ATGGACAATT TACTTTATCC AAAAAAGCAC AAAAAGAAAT CGTTGATTTA TTTATGACTT 31560

CTGAAGATCA ATTAATTGGC AAAAAAATTG AAGATGTTTT CGGCGAAGAA TTTTTTGAAT 31620

CCAACTTTTG GATCTATTGG TGTTCAATGT TTGCTTTTGA AAAATGGCAC TCTGCCATTG 31680

AAATGCGCCG TTACATCATG CGCTTTATCC ACCACATCAA AGGTTTGCCT GACTTTACAG 31740

CCTTGAAGTT TACCAAATAC AATCAATACG AATCGTTAGT GAAmCCATTA TTAAGCTTTT 31800

TAACCAATCA AGGGGTAGAC TTTCAATATG AAACAACGAT CAATGATATC CAAGTCGACA 31860

TCAAAGGAGC TACTAAAGTT GCTCGGCGAT TATTGTTGAC ACAAGCGGGC AAAACCAAAG 31920

AAATTCCTTT AACTGAAAAT GATTTAGTCT TCGTAACGAA TGGCTCAATT ACTGAAAGTT 31980

CGACGCAAGG AGATCATCAT ACACCAGCAC CAATTACTCA TGAACTAGGG GGAAGTTGGA 32040

ATCTCTGGAA AAACTTAGCA AAACAATCCC CAGAATTTGG GCACCCAGAA GTTTTCTGTG 32100

AAAATTTACC AGATGAAAGT TGGTTTGTCT CAGCAACGAT TACGTGGGAA AATTTCGATA 32160

TTGAACCGTA TCTTTCTCGG TTAACACATC GTAAATTACG CACAGGAAAA ATTGTCACAG 32220

GTGGTATTAT TACTATCAAA GACTCTAATT GGATGATGAG TTTCGCTATG CATAGACAAC 32280

CGCATTTTAA AGAACAAAAT GATCAACAAT CTATAACTTG GGTATATGGA TTGCTTTCTA 32340

ATAAACCAGG AAATTATATC AAAAAACCAA TCGAACAATG TACAGGCCAA GAAaTTACGC 32400

AAGAGTTACT TTATCATTTA GGGGTCCCGG AAGGGGaAAT TGAACGTATT TCAGAAGAAT 32460

CGGCTACCAC TATTCCGGTT TATATGCCGT TTAtCACTTC TTACTTTATG TTAAGAGAAC 32520

CAGGCGATCG aCCATTAGTT GTACCAAACG GCTCTAAAAA TCTTGCTTTC ATTGGGAACT 32580

TCGCAGATAC TGAACGAGAT ACGGTGTTTA CAACAGAGTA TTCTGTACGG ACAGCGATGG 32640

AAGCTGTTTA TCAATTGCTA GAGGTTGAAC GCGGAGTACC AGAAGTGTTT GCTTCTGCAT 32700

ACGACTTGCG CGTCTTAGCA AACTCTGTTT ATTATTTATC TGACAAGAAA AAATTAACAG 32760

AAATGGAT 32768

(2) INFORMATION FOR SEQ ID NO: 400:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5960 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 400:

ACCCCTTTTT AGTTCGTTCG ACAAACCCAn GGCTATATCC CAGGnTCCAA TGACCAAGTA 60

TTGGTAACGA ATTAACCnCC TATCCAATAn AATTGAAGGA AATGAACCTA CAACAATTAA 120

AAAAGCTGTA TCAAAAAGTC TTAAAAAAGC AAAAGGGTTG TTGAAATAGA AAGATTTTAA 180

AGGAGATTAA AAATGAACGA TTTATGGAAT TTTTTAGATA ATCATCCTGA AGAGAAAGAA 240

TTGATAAAGG CAAAAGCATT ACAAGGTTAT ACAGTTGATG GAATTTTAAA TGATCAACTT 300

AATTTTGTTG CTCTAGATAC TCTATCTATT GGTCTGGAAA AATTTGGAAA TATTCATTAT 360

AGTATCCAAA TCCCAGAATT TAATGAATTA CTAAGTAAGA AGATGGTAGT ACTTTTTGTT 420

CCTGCGGATG GAGTGGGACA CTCTGGCGCC AAAGAAAGAT ATTTTGGATT GAGACAATGG 480

AATACCCTGT CTCAATCTGT TCGAAAAAAT ACCATCATTC TTAGAATAGC TGATGTTGGC 540

TTTGTGTCAG GATCCTATTA CACATTATTT GAGCAAGAAA TTTTGGATTT AATTGAAGCA 600

AAGAAAACAG AATTTGGACT GAAGCATCAA GATATAATCT TCTATGGAAA CTCTAGAGGC 660

GGAACGGGAG CGTTATTATT AGGTATTACA GGTAATTACA AAACAGTTAG TGTAGATCCA 720

GTAATTGATC GAGAGCCTTG GGTAAACAGC GCGGATGGTA GAATAGATCG CCAGTTAATG 780

TTTGATTTCG TAGAAATGGA TTTTTCAAAC CGAATAAAAG ATAAGATGGA ACAGTCAAAC 840

TATGAATTAG AAAACATTAC GGTGATTGCT TCTGAAGCAA ATAAAAAAAT GTATGAGAAA 900

ATTTACCCCT TATCTAATGG TTTTACCTTG AAAAAATTAG AATTAATGAT TGATTCTGAA 960

AATACAATTG GTATACACGG CCTTACGATA ATGGAAAGTA TTCATTTACA ATTAGCGTTA 1020

ATAAATAATT ATTTATTTGA TTAAGTAACA GAGAAAGTAG GAGCgGAAAA ATGAAAAAAG 1080

CAGCmATTAT GGGCTGGTGG TATAACCaAA ATTATGGATC AaTCTTAACA TACTATGCTC 1140

TAAACAAATA TGTACAGAAT AAAGGGTACG AGACGGTTAT GATTGATGGA CCACTAGGTT 1200

ATAAAAACAG AAGTAACTTT CGAGCATGGA TGCCTTTAGC GTATAATTTT TTTAAGAAAA 1260

ATAATATGCC CTATACTGAA CAATTAACAA AAGAAACGTT ACCGACATTG AACGATTTAG 1320

AAAATCTAGA TACCTTTATT TTAGGCTCTG ATCAAATGTG GAATCCCTGG AATGGTTGGG 1380

TAGACGACGA TGATTTCTTA GATTTTGTTT ATCCTCGTAA TAAAACAATT GCTTATTCTG 1440

TTTCATTAGG AAAAGCAGAT ACAAGTAAAT ATGATCCTAA ATGGGTTGCA AACCGCAAAA 1500

AAGATATTAC ACAGTTTAAT CATGTTTCTA TGAGAGAAGA TTTTTCTGTA CAAATTATGA 1560

```
AAGATATTTT TCAAGAAGAA GTTATTCAGG CTTTGGATCC AACCTATCTT GTTGAAAGAA 1620
GTGAATATGA TTCTTTAGCT GATCAAGCTA CGTTTAAACG ACAAGAATCA GGAGATTATA 1680
TGGCTGTCTT TTTCTTGGAT ATTAACAAGG AAAAAGTTCG TGTAGCACTA GCGATCGCAG 1740
AAAAATTGGG ACTGAAATTA GTGGTCCTGC CTAATCCTTT AAAAGGACGT GAACTAGCGA 1800
AAAAAACGTT TCCTAATACA GTAGAATATG TTTCTGAAGA TACACCAGAA AATTTCTTAT 1860
CTGTTTATCG AGAAGCAAAA TATATTATTA CTGATTCGTT TCATGGCACT GTTTTTGCGA 1920
CTATTTTCCA AAAACCGTTT TCAATTTTTT ATAATGAACA ACGTGGAATT GACCGCTTTA 1980
ATAGCTTGAT GAACCTGTTT GAGCTTGGCG AAACTAGACG TGTCTATGAG CaAAACACAA 2040
AACAAGAAAT TGAAGAAAAT AAAAATATTA GTTTAGCACT TGACTATCAA AAGACCGCTG 2100
GAAAGATTGC aGAGAAAAAG AGATTTCAGA TAAGTGGTTA ACCGATGCGT TGACAAGTAA 2160
ATTTAGCCCA GATGAAGACG AAGTATATGA TGAATTTAGA CGATATATTA TCAATAAACC 2220
TTTGGATTTT TATCGAGCTA AAACAACTGT TCCGATTTCC GCAGCAATTG TTTTATCGCC 2280
AAATGGTACG ATAAAAAATT CCAATCCTAA TGAAAGCTTT TGGAAGCTAA CAGATAAAGA 2340
ATTAATTTTT ATGAATAATC ATCGTGAAGT CACTACAAGA TTTAATCGTG AAAAGTTTAA 2400
GGGAAAGGAA ACGTTCAGTC CTTTTAGAAT CGTCGGTAAA TATGTGAAAG ATGAAAAAAT 2460
CGAACATGTT ATGTATTTGC ATCCGATACA AAAACCTAAG CCTAAACCGA AACAAGGCAC 2520
aGAAAAAAAA GCATTAGCAA TTCCCGAGAA TCTTTTAACA AGAAAACAAT TAATTATGAA 2580
AATTGTGCAA AATGAGAGAA TTCGTTCGTG GAAACAAGAT CTTTTAGAAG AATTACCAGA 2640
TTTAGAGTTA TTTGTAGGTC GTGAATTAAA AGAATATGCT TCGAATTTTG TTGGTGGTCC 2700
AGCAGATCTT TTAATTTTTC CAAAAACGAT AGAAGAAGTT GCTTTAATTG TAAAATATGC 2760
TAAAAATAAT CAGATTCCTT TAACTGTTAT CGGTAAAGGt TCAAATATTT TAGTCCGTGA 2820
TGGTGGTATC CGTGGGATTA CTTTAAATAT GACTGCCTTG AATTATCGTA AAATTACGGG 2880
GAATGTTTTA ACAGTTTCGG CAGGTGCAAG TTTAATTGAA ACATCTTATT ATCTACTAGA 2940
ACACTTGAAA TGCGGTTTAG AATGGGCAGA TAATATTCCA GGAACCATTG GTGGCGCAGT 3000
TTATATGGaAT GCTGGGACAG TTAAGGATAT TAATAGCATG TTTCTTGAAG CTACGATTGT 3060
AGATGAAAAT GGAGAAATTA AAGTTTTAAA TAAAGAAGAT GTTCAATTTA GTCATCGCTA 3120
TAGTTCATTT ATGGATCATC CTGAGTGGAT TATTCTAGAA ACTAAATTAC AGATTTCAGA 3180
AGGTAATTTA GAAAATATGG TAAATGACAT GGTAGGAACT GTCGAAATTA GAGAAAGGAT 3240
GCACCCACTA ACACATCCAA ATCATGGGTC AACTTTTACT TGGGGACGGG CGCCTCGATT 3300
AATTCAACAA GCTGGCTTAG TTGGTACTAG AATTGGCGGA GTAAAAGTAT CTGAAAAACA 3360
TCCAGGTTTC TTTATCAATG TTGAACAAGC CTCTGCTCAA GACTATGAAG CATTAATTTA 3420
TTTAATTATT GCAAAAGTTT ATGAATTTTC TGGATTTTTA TTAAAACCAG AAGTGAGAAT 3480
TTTAGGTGCA AATATGTGGG AAGCAAGTTT AACTTTTAAA TAAACAGGGG AGTGTGTGAC 3540
```

ATGAAGAAAA TAATCACAAT AATTGTGTCT TTAGCAATTG GTATCGGTTG TGGTTTTTGG 3600

TCAGGATATT ACATCAGTAA AAAATATTCA GAAAACCAAG AGACAACTAT TTCCAAAAGT 3660

ACTAACTCAG AACAACCTTT GATTACAAAT GGAGAGTATC AAGTAACAGC TAATTGGGAT 3720

TgGAAAACAT TAGTTGTTAA AGATAATGTG TGGACATTGG ATGGCAAACG TATTTATGAT 3780

GTTTCCACTA CAAATGATAA TATTGTGATT TTAAAGGGAA AAAATTTGAA ATCACTTGTA 3840

GTTTATAAAA TAGAAAAAGA TAACGAAAAA ATCAACTTAT ATTCTTACAC AAAAGAAGGC 3900

GTTGGGAAAA AAGTATTAGC TACTCTAGTG AAAAAATAAA TAATGCATTT GGTGTTTGAA 3960

ATAAATTGAA AGTAACTTTG TTAGGAGAAG TTAGTAAATG AAATACAATA CGAAAAATTA 4020

TGACTATTTA ATTGTTGGTT CGGGACCTTT TGGGGCAACG TTTGCTTGTG AAGCAGCTAA 4080

ACGAGGAAAA AAAGTATTAG TTATTGAGAA ACGTGCTCAT ATTGGTGGCA ATATGTATAC 4140

ACATAAAGAA AACGGGATTA ATGTCCATGA CTATGGTGCC CATATTTTTC ATACTGACAA 4200

TAAAGAAGTA TGGTCTTATG TCAATCAATT GACGGAATTC AATGGCTACA TCAATCAAGT 4260

AGTTGCTAAT TACAAAGGGG AGCTCTACAA CCTACCATTC AATATGAACA CGTTCTACCA 4320

AATGTGGGGG GTAAAAACAC CAGCTGAAGC ACAAGCAAAA ATTTCAGAGC AGCGGGAGGC 4380

AGCTGGAATT AAAGGCGCTC CmAAAAATTT AGAAGAACAA GCGATTCGTT TGATAGGAAC 4440

GGATATCTAT GAAAAATTAA TTAAAGGCTA TACAGAAAAA CAGTGGGGAC GTAAGGCTAC 4500

GGAATTACCA AGTTTTATTA TTCGCCGTTT GCCAACTCGT TTTACTTTTG ATAATAATTA 4560

CTTTAACCAT CGCTATCAAG GCGTwCCTGT CGACGGTTAT ACAGCTATGT TTGATAAATT 4620

GTTAGAGAGT GAATTAATTG ATATTCAATT GAATAAAGAC TTTTTGGCAG AGAAAGAAAC 4680

TTACTTATCA GAGTTTCCAA AAATTGTTTA TACAGGGATG ATTGATGCAT TCTTTGATTA 4740

TTCTTATGGC GAATTAGAAT ATCGTTCAGT ACGTTTTGAA AATGAAACAG TAGATTCAAA 4800

CAATGCACAA GGCAATGCCG TAATTAATTA CACAGATGCG GAAACTCCCT ATACACGGAT 4860

AATGGAATGG CGACATTTTG ATCAAAAGGC CGATGAAAAC AAAACGATTT TGACGAAAGA 4920

ATACCCACAA GATTGGGATC GTTCAAAAGA AGCCTATTAT CCTGTTAATG ATGAGAAAAA 4980

TTCACTTTTG TTTAGAAAAT ACAAAGAAAA AGCAAACGCT CAAGAAAAAG TTATTTTTGG 5040

TGGACGTTTA GCCAATTATC AATACTATGA TATGGACCAA GTTTTCGGAG CAGCTTTGAA 5100

AGCAGTCGCT AAAGAATTTG GTGAAATGGC CAACTAGTGA ACGGTTAGAC TCGGAGGAAC 5160

AAATGGAAAA AGAACTGAAA GTTAGAACAA AATTACTTAC CAAAGAATAC AGTTTAGCGC 5220

AAACACGTAT TGACAAATTG AAGACACTTT TTAGTGTTTT TCAAAACAAA GTTCCCACAT 5280

TTTGGGCGTT AAAAGGTGTT AGCCTCGACG TATATTCTGG AGAAACAATT GGTATTATTG 5340

GTTTAAATGG TTCAGGCAAA TCCACTTTAT CAAATATTAT TTCAGGCATC ACCCCTCAAA 5400

CGTCAGGAGA ACTTGAAATA AATGGTGAAG TTTCAATTAT TTCGATTGGT GCAGGACTAA 5460

ATAATAATTT AACAGGACGA GAAAATATTC GAATGAAATG TTTAATGTTG GGGGAAAAGA 5520

ATAAAGAAAT TGATGCCAAA ATCGATGATA TTATTGAATT CTCAGAATTA GGCGTTTTTA 5580

TCGACCAGCC TGTTAAAACT TATTCTAGCG GTATGCGTGC TAAGCTTGGT TTTTCAATTG 5640

CGGTTCATCA AAATCCAGAT ATTTTAGTGA TCGATGAAGC GTTATCTGTA GGAGATCAAA 5700

CCTTCTATAA TAAAGGCTTA AAAAAAATGC TTGCATTTAA AGAGCAAGGC AAAACTATCT 5760

TTTTTGTCTC TCATtCTATT CAACAAGTAG AGCAAATTTG TGATCGTGTC GCTTGGATGC 5820

ATTATGGTGA CCTACGTGCA TTkGGkGAAA CTAAGCATAT TTaAAGGAtA TCGTGCTTTC 5880

TgCAcCGTaT AAyCATTTCm CTGAACCGCA AAAAGAATCG TATCAGAGAG AGGCAAGGCC 5940

AAACAGAGAA ATTTTCCTTG 5960

(2) INFORMATION FOR SEQ ID NO: 401:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1268 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 401:

AAAAGTTCAA AACCTTTTTG TTTAATCACA TTTTGTATGG CGGCGACGAC TTCTTCAATC 60

GTCAAGCCTG TTGTATCGAT TTTCACAGCA TCACTTGCTT GGACAAGGGG TGATACTTCT 120

CTGGTTGAAT CTAAATAATC CCGACGTTCA ATCTCCGCCT TCAAGGTTTC AAAATCCGTC 180

TCAATCCCTT TTTCTTGATT TTCTTTAAAA CGACGTTCTG CACGTTCTTC AACGCTAGCC 240

ACCAAGAATA TCTTTACTTC GGCCTTTGGT AAAACTGCAG TACCAATATC TCGACCATCC 300

ATGACAACTC CGCCGGCTTG ACCAATTTTC TGTTGTAACG CCACCATTTC TTCTCTAACT 360

TTTGCATGTT TTGAAACAGC TGAAACAGCA TTCGTTACAT CTGGCTGGCG AATCGCTTCG 420

GTCACTTCAT GACCGTCGAT AAAAACACGT TGGCCATTCT CTGCTTGCTG AAAAGAAATA 480

GTGTGGTTGA CACAAAGCGC TACTAATGGC TCTTCCGCTT GAATGTCGAT TTGATTTTGC 540

AAGGCCAGAT ACGTAATCGC TCGGTACATG GCACCCGTGT CACAATACAC ATAATTCAAT 600

TGTTTTGCTA AAATTTTTGC AACTGTACTT TTGCCTGACG AAGCTGGTCC GTCAATAGCA 660

ATACTGATTT TCCCCATGAA AAAACTCCTT AACTGATATG CATCTTTTTA AATGCGCTTC 720

TTCTTTTATA TTAAAATCAG AAGAAAAGAA CTTGGATTTA AACTGATGAC GAGTCTCTAA 780

CGAAAAAGCA ATAAGCCGAA ATCCCTCGAA AGACGCAAAA CATCTTTCGG AAATTTCTGC 840

TTATTGGCAT TCTCATGTAA TCATCCGTCg CGGTTTAGCC CTTGAAAGTG CTGTCCAAAC 900

CTCTTTTTCT GGATTATTTA ATACGTAATG GTTGTCCTGG ATAAAACACA GATGTATCAA 960

TCCCAGGATT TAACGCTAAA ATTTGTTCTA ATGTCATTCC GTTACGTTCG GCAACTTGAC 1020

GTGCCCCTTC GCCACTTTGA ACGGTATCAT AAACAGGTTG TTGGTTCTGC GCTTGTTGCT 1080

GTTGTTgATT TTGAGCTTCT TGTTGTTGTT GCTGCTCTTG TTGTTGCTGT TGCTGTTGTT 1140

GTGCnTGTTG GTCCTGTTGT TGCTGTGCAG CnATGCTTGA ACTTGAAGGT GTTGTATTTT 1200

CAGCTGGTTG GCTGCTTGCT GGTTCTGAGC TTTCCACTGA TTTTGGACGT ACTTTCTTTG 1260

TTTCnTTC 1268

(2) INFORMATION FOR SEQ ID NO: 402:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1204 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 402:

CTTTTGTGCC ACAATCCGTT CCTCCCACCT TGGCGTGCCC AAACAAGGTT AATGATTTGT 60

ATAAACAATT TCCAACATCA ATAGTTGGGA CCACTAAAAT ATCTGCATCC CCCATGATAG 120

GACCGCTGTA ACGTTTATGC GCAACCGCTT CTTCAGAGGT CGCTAAATCA AGCGAAAGTG 180

GCCCAAAAAC CGTAGCCTCT TGTTGATCAT TAAAATGTGC CGTGACTTCT TTTGCTAAAA 240

CAGACGAAGG CATTTTAGGA TTGAAATTTT CCGCTGCGCT TAACAAAGCA ATTTTCGGGT 300

GATGCAGTCC CAATTTTTGG GCGACTTCTT TAGCATTTTC AACAATTTCA ATGAGGGTCG 360

CTTGAGTGGG GGCGATATTC ATCGCACAAT CGGTTAACAA GAAGGTTTTT CCCGCAGGCA 420

GCTCCACCAT TGCTACATGG GAAAGAATCG GTTTATTTTT TAATTGATGC TCACTTTTCA 480

ACATTTCTTT CAGTAATGTG TGGGTCTGAA TAATTCCTTT CAATAAAATT tGTGCTTGAC 540

CGGTTGCAAC TAAACTGACA GCTTCCTGTG CTACCGCGGC CTCATCTGAG CAATGAACAT 600

ATTTCCAAAG ATTTTCAGTA TCAAGATTTT CATTTGTATC AAATACAATA AACTGCAACG 660

GTTGCTCGGC TTCTTTTAGT GCTTTTTTAA CTAACTGTAA AATCTCTGGT TGTGAACCTC 720

CTGCAATTGA AACAGTGATC ATTTTCTCAC CTCATTTTTT CTTCTTTACA CTTTAAATCT 780

ACCTTAGTAT GTAAACGCAT ACAAATTGTT AATTTTTATG TCCTCATTAA TATTTTTTAT 840

AGAGAAGGAA AAGCTTAGTA AATCAAGCTT TTCTTTCTCT AAACAATTGT TATCTAATTT 900

ACTTTTCATG AAAAAACTTC TTTTCTAAGA ATGAAAGGAA AACTTATGCT TCGGCACAGC 960

TGTTTATTTA TGTCAGATCC TAAAGCATGT TATACTTTAG ATACTTATTT TTAGGAGGAa 1020

TTAAATGGCA CTACTACTAT TCTTTTTGTT TATCGCCCTA CTAGGATTTG GTATATTAAA 1080

AATTAATAAC CGCAGtATCC TCGGCGGTAT CACtCTGGCT tCCGGCACCT TATtGTCATt 1140

AGtCaCCTtA CTATTTATCG GATTAGACAA AATTTATTTA CATTTTAAGA ATGGCGACCT 1200

AATT 1204

(2) INFORMATION FOR SEQ ID NO: 403:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4996 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 403:

ATAACAAAAC ACTACTCTTC TATCmATAAT TGaCTTAATA aTTGAATTTC TTGrTTCTTA 60

GkTAAGTCAA TTGtTTTCAT kGTAmCCGCT GTGGCTAAmC CTAAATAGAT TACCACCATT 120

TTAGGAAATA ATnCGACAAA TTTyCCTTkG yCCATTtGkA CATATTTTTt ATCCAATAAT 180

AAACGCTCTT CAAAAAAAGT TTGAAGCATT TGCTGCAATT CTTCTTCATA ACCATGTGCT 240

TTCATAAATA ACGCCGTATA GATTTTTTGA TGTTTTTGGG CAAATAATAT AAAGTTCAGA 300

CCCAATGCAA TCACTGGGTC TGAGGCACCG ACATCTGCTT TATAAAATAA TTGCATACGA 360

TGAAATAACG ATTTAATTAA GGTTCGTTTT AAATCGTCCA TGTTGACAAA TTCTAAATAA 420

ATTGGCTGCG TAGAAATCCC CATAGCATCT GCAATATTTC TAGCAGTTAA ATTAGAAAAA 480

CCTTTATCTA CTAAAATTTT TTCAGCTACG TTTAAGATTT TTTCCTTTGT ATATACCTTC 540

GTTCTAACCA TTTTCATACC CCTTTTCTCT TTTAGTTAAC AATAACTTTA TTATGTTCAC 600

ACAAAAGAAA ACCATTTTCA TTATATATAT AATATTTTAC ACTTTTTTAT GCGTTTCTTT 660

TCTCAAAATT TTCTATACAT TTCAAAAAAA TTGAGAATTT TCATCGCCTA TAAACAAATA 720

GGTCCAGAAC AAAAGGTCAT CACCTTTTGT TCTGGACACT AAACGGCTGC ACTCTTTATT 780

CTTCTATTTG TTTGCCTTGT TCTTTGGCAA GCTCATTTTT ATCATAAGCT ACAAAGAACG 840

GATAGTAAAT CAAAATTGCC ACTAGAATTA AAACAATATT CAAAACAGCA GCACGCCAAT 900

CGCCGCCTGT CGCCATAAAG GCGCCAATTG GTCCTGGTAA GGTCCAAGGA GCGGTAAAGA 960

CAACTTCACT TACTAAACCT AATTTTGTGG CAAACCAAGC AATCGTTGCT AAAACCATTG 1020

GTGCAAAGAT AAAAGGAATC ATTAAAATCG GATTTAAGAC AATCGGTACA CCGAAAATAA 1080

CTGGCTCATT AATATTAAAA ATAGACGGTG CTAAAATAGC TTTTCCTAGT TTTGAACCAT 1140

ATTCTGATTT TGCCCGGAAG GCTAATAATA TAGCTAAACC AATTGTGGCT CCAGCACCAC 1200

CAATCCAAAT AAACCATTGG TAAAAAGGTT CTGCCGCAAT CCGAGGTAAT GcTTCGCCAG 1260

CTGCTTTAGC AGTCGTGTTT TCCTCTAATA GTTGTAGCCA TAGAGGACGC GCCAACGATC 1320

CTACGATTGA AGCGCCGTGA ATCCCAAAGA ACCAGAAAAA TGTGGTTAAA AAGACTAACA 1380

GCAAGACACT CAATAAAGAG TCCGCTGCTT GAACTAGTGG TCCCACTAAG TTCCCTACAA 1440

ATGTGTGCCA ATCAAAGCCT AAATAATAAG TGATTGAACC AATCAATAAC ATGACAATTA 1500

AAGTCGGAGT CAATGATTCA AATGAACGAG CAACAGCAGG TGGAACGTTa TCAGGCATTG 1560

TAATTTTGAA CTTAGACTTA TCTGTCACTC GATAGATTTC GACAGCAATA ATTGACGTCA 1620

CAATCCCGAC AAACATCCCT GCACCACCCA AATTAGCCAT TGGTAAAACA AACCCrGrAA 1680

CGCcAGCTGC TTTACTGGCT tCTTCAGGAa TCGCTACAGG TACTAATGTT AATAAAAAAG 1740

CAATCGTAGC CAAAATCCCT CCAGAGACTT GATCAAGATT ATATGATTTC GCTAAACTAG 1800

CGCCAATGCC AAAAGTCGCA TATAACGCCA TAATGTACAT TGTCATTCGG TAAGGCAATA 1860

AGATTGTAGC AGCATTACTT GTTAAAAATT GAGTAATGCC CCAAGATTCT GGTAACGGCG 1920
GAAACGCAAT AATTAGAAAA AAAGAACCGA CAATAATTAA AGGTAATGTC GCAATAATCC 1980
CATCCCGCAC CGCACGTAAA TGGCGCTGAT TGGCAATTTT AGCCATTGGT CCAGCTAGAT 2040
GATTTTCAAT CCAAACAGTC AGTTTATCTA ACATAAATAA CACTCTCCTC TGTTTTTCTC 2100
TTAAAAAATA ACTCCTTTAA TCAGCCATAm ATAAGAGATA AGCTCTAATA AAATAAGAGA 2160
rATTAArAAC AACCAAATAA AACGCTTACA AAAACCTGTT TTCkTTTcAA TAAAATAATG 2220
TTCaTCGGCA ACGGCCGCCA CTTGTGGCAA GTGCGTGATA CACAACACCT GCGAATTTTC 2280
TGAAATTkGA TAAATTTTAT CGGCAATCGC CTGTGCTACT CGGCCACTaA CTCCTGTATC 2340
CACTTCATCA AAAACAATAC TAGTGATCCC TTGCGTTTGA GAAAAGATTG TTTTCATAGC 2400
CAACATCACT CGCGAAAGTT CTCCGCCAGA AGCCACCCGA ACTAACGGTT TTAATGGTTC 2460
CCCTGGGTTA GTAGTAATAT AAAATTCTAC TCCGTCTAAG CCATTTTCTT GTAAATGCTC 2520
AAGTTCTGTA AAGCGGACTT CAAATTCAGT TCGCTCCAAA TATAGTTCTT TTAATTCGGT 2580
TAAAATTTGT TGTTCGAGCT CTTTGGCTAG GCGCTTTCGT TCTTTTCGTA AAGcTAACGC 2640
TTGTTGATGA GCCGCTTGTT GTTTCTCTGC AAGCAACGCT TCTAAGTCAC CTGTACCGCC 2700
TTCCAAAAAA TCAGCCTCTG cTAACTCTTT GGTGATTTCT TCATAGTAAG ATAAAATCGT 2760
TTCGATTGaA TCGCCATATT TACGTTTCAT TTGACGGATT AATTCCAAAC GATTTTCTAC 2820
TTCATTCAAG CGGCCTTCGT CTAGTTCTAA GCCATCAATC AACCTAGAAA GATCTCCACT 2880
GGCTTCTTGT AGTAAGTAGT AAGCATTTTG AACAGTATCT GACAATGTTT TATATTCTGA 2940
ATCAAGGGAT TCAATCGAAG CGAGTTCATT CATACTTGTT CCGATTTtAT CCAAACTACT 3000
ATCGTCTTCA CCATTTAGCG CGGcATAACT AATCGTcAGT GgCATCAGCA ATCTTTTGAA 3060
AATtGTTCAG TTwATtGCGT tCTtCyAACA ATTGTTCTTC TTCGCCAGCG ACTAACTGTG 3120
CACTAGCAAT TtCATCACTT TGAAAaTGAa GCATGTCCAT tCTTTGGGCA AATTCTTTTT 3180
CATTTTTTtG tCGCTTTCTG ACTTTGGCTT CGAGTGCGCG ATACTCTTGA TACGCCTGTG 3240
TATATTTTTC TTTGACTGCT AAAAGTTTTT TCCCACCAAA TTCATCTAAC ATATCAATAT 3300
GGCGTTCACT TTGCATCAAT TCTTGATGTT CGTTTTGGCC ATGAATATCT ACTAAATATT 3360
CCCCAATTCT TTTTAAATTA GTAATGTTGA CAATCCGTCC GTTGACACGG CAAACATTTT 3420
TACCAGAAGC GGAAATATCT CGTTGAATCA CTAAAGAATC TTCTTCTGTT TCAATACCTA 3480
ATTCTTCTAA TAATTTCTTT AATTCTTGAC TTTTCGGCAT TGAAAAAAGT CCTTCTAAGG 3540
TGCATTTATT TGCTCCTTGA CGAATATAGT CACTGGAGCC GCGTCCGCCT GTGAGTAATC 3600
CCATTGCATC AATAATGATG GATTTCCCCG CACCCGTTTC TCCCGTTAAA ACGGTCATAC 3660
CCATTTGAAA CTCTAATTGT AACGAAGAGA TAATCGCAAA ATTTTTCACG GAAAGTTCTT 3720
GTAACATCTT GTTCACCTCA TTCTTTATAG ATAGCTTAAC AGCTCTTTAT GAAGTTTTTC 3780
TGCTGCTTCC TCAGTTCGTG CAATCATGAG GACACTATCA TCATCGTTGA TGACGGCAAA 3840

AAGTTTCTCT TGGTAATTCT TTTCGATTAA ACTACCACAC GCCGCCGCAT TGCCAGGAAT 3900

GGTTCGTAAT ACGACATATT TTTCCATTTG TTCTGCAGCT ACAAAGGCAT CTTTTAAAAG 3960

CTTTGCAAGT TTAGCACTCG TATTAGCTTG TGTTTCTAAT GGAAGGCTAT ATCGATAACC 4020

CCCTTCTGCA GATGGAACTT TGATCAGTTT CATATCTTTA ATGTCGCGAG AAATTGTTGC 4080

TTGTGTTaCA GTACGCCCTT TTCTTGTAAG TAGTTGACAA AATCTTCTTG CTTTTGAATA 4140

TTTTTTTCAG TCAATAAACG TGTAATTAAA CGGTGTCTAT CTTGCTTTCT CATGGCGCCG 4200

CTCTCCTTTT CAGTTTACTC TTGCCCTCAT CATAGCTTAT TTTTATGCAT TTTAAAAGAT 4260

GAATTACTAG ATTTATACAA AAAAGAAGCT TTCACCTATA GGCGTCCACT TCTTTCTTCT 4320

AACTTTTCTC ATCATTCCGC TAATTTTTGA TGCGCATGCG CGaCAACTTC TTCAATGGAT 4380

TCTGGCGCTT GATAGGATCC TTCCCCTTCA ACAGAAACTA AATGTGCCAA AAATTCAATG 4440

TTTCCTTCGC CACCAGTGAT TGGTGAAAAA TCTAAGTTTT TCACATCATA ACCGTTATTC 4500

ATGGCAAAGC GTGTAATTTC CTCCACAACC ATTTGATGGG TTTCTGGATC ACGCACGATT 4560

CCTTTTTTCC CAACAAACTC TTTGCCAGCT TCGAATTGTG GCTTAATTAA CGCCACCACA 4620

CTACCGCCTT TTTTCAAAAT ATCATGAAGA GGTGGCAAAA TTAAGCGTAA TGAAATAAAG 4680

GAAACATCAA TCGTTGCAAT ATCCGGCACC CCTTCAGTGA AATCTTCAGG TTTACTATAG 4740

CGAAAATTGG TACGTTCCAT TACCACCACG CGTTCATCTT GCCGAATTTT CCATGCTAGT 4800

TGGTTATAGC CAACATCTAA AGCGTAACTC AAACGAGCCC CATTTTGTAA AGCAACATCG 4860

GTAAAACCAC CTGTGGAgGA aCaATATCTA ACATTGTTTT GCCTTGAACG TTGATTGCAA 4920

AAACATTTAG TGCCTTCTCT AGTTTTAAGC CCCCACGAGA AACATAAGGC ATTACTTGGC 4980

CTTAnTATGT AAGGGT 4996

(2) INFORMATION FOR SEQ ID NO: 404:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1926 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 404:

CACCGTGGnG AACTCTTTTT TGGGTAACCn GGATGGTTTT TTTTGGACAT ACTATTCCnC 60

CCTTGTCTAT CTAATTGACT ACCnTCCGAT AAATCTTATG TATTTTCTTA TATGCTTCAC 120

ACATTATACT ACTTTCTTTT TCGAAAAACA GAAACGTTGG TAAATTCTCA TAATTATTGA 180

AGGCGAACAT ATTTTCTTGA ATTTTATCAC AAAATTAGGT AAAATACAGG TAATTAGTCA 240

TTTGAGTTTT ATTCTATTTT ATTCGTAAAG GAGGCAGAAG AATGGATCAA CATAAgCTAC 300

CATCAAGAGC ATTTCTAAAC TGATTACGCT CTCTTCCCAA CATTTCTTTT CGGCATTTTT 360

AGAAGATTCT CAAGATCAAA AGAATAAAGA AAAGGAACGC AGTGCGCTTT TACAAGAGCT 420

CCAAGTTGct GCCGCCCAAA AATCCTTAGT GGTCTTACAA TTGAAAGAAG TCCCTACAGC 480

GCATAAATTC GAAACGGTGA TTGGCTGGAT TGTCTCAAAA AATATTTCGG ACAATATCGT 540

TGTTCGTTTA CAAACGGATG AACAACAGTT GCGCATGATT CCTGTTGCTT CTGTCATGAA 600

AGTCAGCACG TTAGCCAATC GTCACCAACG TCAAATTCAC TAGAATATTG AAGTCTTTAT 660

AAAATAAATG GAAACCACCA ACCCTGGCGA TTGCAGGTTG GTGGTTTCTT TTATTTATGA 720

CTTTAGTCAA AATCCCATGA AACTTCTGAC TTAGGACTCA ACTTATTTCT TAGGAACAAA 780

TAATTGCGGT GCTACTTTTG CCAATAAGCC TAGAACTAGA CTTACATACA AACAATTGAC 840

GACGGTACTT GTACCGTTTA GAACAAmGGA ATAGAGATAT GGGCTCATTC CTTCTGGTGC 900

ATACATGcCC CAAACCAATA CACCCGCCCA AAAATGCCAA AACCAGCGGG CGAAGACTGC 960

CACAACACTT CCTAAAATAA TCGTACGAAT GGCTGATTTT AAACGATTCG CTTGAAAATA 1020

CAGTTGAATT TTTCTAGCGA ATACGCCACC CATTCCTCCA AACGCAAAAG CAAAGGGATA 1080

TTCAAAAATA ATCTGTGGAA CGGACAAGAA ATTTTTCATG GCTGTACCGA GAATAATATT 1140

TAACATGCCC CAAACAAAGC CTGCTGCAAC TCCAGGAAGC AACCCACGAC GAAACGAATA 1200

AAGAACTAAT GGGACCATTC CCAACGATAG GTCCAAGCCA CTATTGGCAA ATTCAATCGG 1260

TAAAAAAGAG AGGGCCATTG CCATTGCTGC GACGACTGTT CCTTCTACCC AAATACGTAA 1320

ATCCATTTTT CTAGCCAAAC AAATTTCCTC CTATGTtCTG ACGCTCTCAA CCACTAATAC 1380

ATCGTTATTT TACCATAGGG ACTTTAGAAA AACACGTTTG TCGAACAAGC AAAAAAATAG 1440

CTGATCAGCA TTTAAGGGAA TCAACTCCTA CTCAAACGCA TGCCAGCTAC TTTTTATCCA 1500

TTAATCGTCC CGCTCAGTTG TCAAAATTTT ACCAGAATGA GCATCGATTT TTACCTCTAT 1560

TTTTTGTTGG CCCTCTTTCA CTTTCACTTC CCAATAGGTG ATTCCTAATT CTTTCTCTAA 1620

CTTCCAATCT GTTGCCTGGC CGACTTTCGC TGCTTTTTCA GCCAAGGTTG TGGCTTGTTC 1680

TCTGGAAATG ATATTTGTTA AATCAAGTGC CTCTTCTTGC ATTTTAACGC CATTTTGTTC 1740

ATCGGCATCC AAGGTTTCGA CTTTTTCTTT TGTGAATTCG CCTGTTTCTG CGTTCACTTC 1800

AACTTGATAT TCTTTTTGCT GATCCACGCC TTCTATTTCA TAAAAATAAC GTCCAAAATC 1860

TGTATCTAAC TGCAGACTTG TTATTTTAGC TTCTGGATAT TTTTTCTCAA ATGCTGTCGT 1920

TACTTT 1926

(2) INFORMATION FOR SEQ ID NO: 405:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1479 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 405:

GATTTCTGCT CACTCTTTTT GATGGGAAAA GTCATCaGGA CACGGTTCGT TACATCAATC 60

TCGAAGGAGG CGAAAATTTT GAGTAGTGCT GTGGAAACTG TAACAAAGTT AGTCACACCA 120

ATCTTAGAAG AACAAAATTT TGAACTTGTG GAAGTAGAGT TTGTTAAGGA AGGAAAAAAT 180

TGGTTTTTAC GAGTATTTAT CGATAAAGAA GGTGGCATTG ATATTGAGGA ATGTGCCTTT 240

GTCAGCGAAA AATTAAGTGA AAAACTAGAT GCAATGGATC CAGATCCAAT TCCGCAAGCG 300

TACTTTCTGG AAGTTTCTTC TCCGGGAGCA GAACGTCCGT TAAAGAAAGA AAGCGATTAT 360

GAACAAGCGG TAGGTAAGTA CATTCATATT TCTCTTTATC AAGCAGTTGA TGGAGAAAAA 420

CAGATTGAGG GTACCTTAGT CCATCTTGAT TCAGAACAaT TAACATTAAG TGTAAAAATT 480

AAAACCAGAG TGAAGGAAAT GACTTTTGAA CGTAAAAATA TCGCAAAAGC TCGTTTAGCC 540

ATTCAATTTT AATCAACAGA CTCGAATGTA AAAAAGGGTC CACTAAAAGT TTTACTTTTA 600

GCTAGGCCTG TGACAAGTTT TAACGAGAAG TTTTTCCCTT GGAAGAAGAA CCACTCGTTC 660

AGCTTTAACG AATCCATAGG AGGAAAAGAA AAACAATGAG CAAAGAAATG TTAAATGCAT 720

TAGACGTATT AGAAGCTGAA AAAGGGATTC CTAAAGATAC GGTGATTGAA GCGTTACAAG 780

CAGCTTTAGT TTCTGCATAT AAAAGACATT ATGGACAAGC AAGCAACGTC GAAGTTGAAT 840

TTGAACCAAA AAAAGGCAAT ATTCATGTGT ATGCTGTCAA AGAAGTAACA GAAGAAGTAT 900

TTGATTCACA ATTGGAAGTT TCATTAAAAG ATGCCGTTGC ATTAAATGGT GCATATGAAA 960

TCGGCGACAA AATTCGTTTT GAAGTTACGC CAAAAGACTT TGGCCGCATT GCTGCCCAAA 1020

CAGCAAAGCA AGTAATTTTA CAACGTGTCC GTGAAGCAGA AAGAACAATT ATTTATAATG 1080

AATTTAGCGA ATATGAAAAT GACATCATGC AAGGaATTGT TGAACGTCAA GATCGTCGTT 1140

ATATTTATGT AAACTTAGGT AAAATCGAAG CAGTCTTATC AAAACAAGAT CAAATGCCAA 1200

ATGAATTTTA CCAACCGCAT GrTCGGrTTA AAGTTTACGT ATCACGTGTT GAAAACaaTt 1260

CAAAAGGcCC ACAAGTTTTT GtAAGCCGTA GTCaTCCAGa TTTaCTAAAA CGkCTTTTTG 1320

AACAAGAAAT TCCAGAAGTT TATGATGGAA TTGTTGAAAT CGTAAGCGTC GCTCGTGGAA 1380

GnGGGCGATC GTTCAAAAGT AGnCCGTTCG TTCAACAGAT CCnAATATTG GAnCCCGTGG 1440

GGTACCCGTG GTTGGTCCAA AGGGCCACGT GTCCAGGCG 1479

(2) INFORMATION FOR SEQ ID NO: 406:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5454 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 406:

CAGGTGCAGT AGGAATTGCT TTTTTTGCTG GAAATATGAn ACAAGACAAA AGGATAGCAG 60

ATAGGCAAAA CAAaaAAAGT GAGAAGAaGT ArGGAGAATA TTTATGCCAG TTAATAAAAA 120

TGGTTTACCT TTTTATCTCG ACAGAAAATT AGATGTAATA CGCTTTGGAG TTGATCCGGA 180

TAATCATGTG AAAGAACAAC AACTTTTTAA ATATCCGGTT AAAACGGATA AGCCAAATAA 240

AAAAGAACGT TTCATTGAAA AAGAAAGTGA ATCTTTAATT CAATTAGGAT ATAAAATTGT 300

AGTTGATACA GTTACTGGTG TTAATTATTT AGAGAGCCCT AGAGGAGGAA TTACACCTTT 360

GCTTGATGAA ATTGGGAAAG TTTGCATCGA TAAAATTAGT AAGTAAATAA AACCCACCCC 420

TGAAAATCTT TTTTTGGGGG TTATTCTATA TTCAATTATT AATGATAATt ATaGTGATGA 480

GAGAGCAATT CTAAGCCATT TTCACAAATA TTGATCGGTA GTTATGCAAT GAAGAGATTT 540

TAACGTTCGC TCTTACAGGA TTAAATATGA TATGTAGATG AAAATAAAAA aaTCTCTAkT 600

TAAaTACGAA TAATTGTTCA tACCAGGCAt GTGGCCAaCC ATCAAAAGAC GAATAACTGG 660

TTTCGCTTTG CTCAACkGGC CTTCAACGTT TTGTATAGTT TCCCTCTCTG CGTTCAGTCC 720

AACTATGCCA AAAGTTTACA GGTGTATAAA AAAATACGGA AAACTGAATT TTACTAGTTG 780

CAAAAGAGAA CATTCGTTCG TATAaTGGGT TAGTAGAAAG ArGGGATTTk GGATGaATTT 840

GACGTTCGAT TATACGAArG AACCCAGTCG TGATGTGTTT TGTATCGACG TAAAAAGTTT 900

TTATGCTAGC GTGGAATGCG TAGaACGTGG ACTTGATCCA TTGAAGACAA TGCTTGTGGT 960

CATGAGCAAT TCTGgwAAAT TCAGGTGGCC TGGTACTCGC TGCTTCCCCT ATGGCAAAAA 1020

AAGTATTAGG TATTTCCAAT GTTACAAGGA AAAATGAAGT TCCGGACCAC CCAAACTTAA 1080

TTATTGTACC TCCACGCATG AAATTATACA TGAAGAAAAA TCAAGAAATT AACAATTTAT 1140

ATAAACGCTT TGTTTCTGAT GAAGATCATT CTGTATTCAG TGTCGATGAA TCGTTTCTTG 1200

ATGTGACTGC TTCGCTGACC TATTTTAAGT GTGACACCGC CTATAAACTG GCCAAGATTA 1260

TTCAACGTGT GATTTATAAC CATATGGGAT TGTATGTGAC AATCGGAATT GGGGAAAATC 1320

CGTTGCTGGC CAAGTTAGCA TTGGATAATG AAGCAAAGAA TGCACCAGGC TTTGTGGCTG 1380

AATGGCGCTA TGAAGATGTG CCAGAAAAAG TTTGGCCAAT CTCCCCTCTT ACAGAATTTT 1440

GTGGGATAGG AAATCGCATG GCTGCTCGCT TAAAAAAGCT AGGTATTCGG TCCATTTATG 1500

ACCTAGCGCA CATTGAACCT TACATGTTAA AAGAGCGTTT TGGTATCATG GGGCTACAGC 1560

TGTACGCGCA CTCATGGGGC ATAGATCGTA GTTTTTTAGG GCAAAAAGCG GGACGGCCAA 1620

CAGAAAAATC ATTTGGCAAT AGTCAGGTTT TACCCAAAGA TTATGCAAAT AAAGAACAAA 1680

TAAAATTGGT TTTGAAGGAA TTGTCGGACC AGGTTGCTAG TCGcTTACGC ATGGCATCGT 1740

GTCAGACAAC CTGTGTCTCT TTGTTTGTTG GCTACTCAAA AGGTCAGACA GACAAATACG 1800

GCCAAACAGG CTGGCGTAGA CAAATGAAAG TAGAACCAAG TAATAATACA AAAGTTTTGA 1860

CTGAACATGT ACTGCGTCTT TTTGAAGAGA ATTACGCACC AGGGGTCGAC GTACGAAATT 1920

TAGGGGTATC TTATGGCCGT CTTGTATGGA ACAAAAATTT ACAGCTGGAT TTATTTTCTG 1980

TTCCGGAGGA GCAAATTCAT GAAACGGATA TGTATTTTCT TATTGATAAA ATACGTCAGA 2040

AATTTGGGTT TAAAGCTCTG ATACATGCTT CTTCACTGAT GGAAGGCGCC ACAGCAATTT 2100

CACGAGCAAG CCTTGTTGGT GGCCATGCAG GTGGAACTGT TGGCTTAGGT ACAACCAAAT 2160

GAAGCAAAAT AGAAAAGAAT TTTCTTCTTA TTTTTCTCGC TCAATCAAAC AAAATAAGCC 2220

ATTATATTTA TTACTTATGT CTTCAGAAAC CAATCCTTTT CCTATACCAG TtATAGGAAC 2280

TTTTCGAGGA TATGCGGAAG AAAACAAAAT AATAATTGGT GAAGATTCCT ATAGTATAGA 2340

AGAAATTGAC TCCGTGTGTA TGTATCGTCA TTCTAGCAGA ATAAAACAAG AAGAAAGAGT 2400

AAATTCAAAC CCTTTTcAAG GTTTGGTCTC CCCTCCTTTT TcAAAAAAAG AGCAGAAAAA 2460

GTTAGATGAG TATTTTAGCG ATTGGGGGTT TCGTAATGAA GGCATATGAT TATAATGGTA 2520

CCATTATTCC TCTAGGAGAT AGCGCCCGCC AAGTGGCTTA TGTAGAATGC AGCTGTGGCT 2580

GTCTAGCAAG CAGAGTTACG AATGACTCAA GCAAGTATAA ATGTTCCTGG TGTAAACGAG 2640

TTTACACGCT TAGAAAAGAT CATAAAACAG CTAGATAAAT TGTTGAAGGT TTTATTATTA 2700

AATTGGCAGA ATTTCAAATT ATGATATAAT TmATGCGGCA GCTCGCCTCG ATTGGAGGTG 2760

TGTTATTTGT GAAAGATTTA ATGTCGTTGG TTATCGCACC AATCTTTGTA GGATTGGTTC 2820

TGGAAATGAT TTCTCGTGTG TTGGACGAGG AAGACGATAG CCGAAAGTAA GCTGCTATCA 2880

ACACACACGC TAGAAGTCGC AACTAGTGTA AAAAAAAGCA ATCCTATTCG CCGTAGGATT 2940

GCTTTTTGTG TTATCTGTAC GATTTAATGT CGTTTCGCAC TTTTAGTATA GCATATTTTT 3000

ATTTTGGGTC AAGTTTTGTG ACTATGCAGG AATTGGTAAA gAATACAGTG GTAGCAATTT 3060

TCATCGATGC TATTTTATTA ATAAAATAGT AGTAGAAAAA TATATTTATT GATAAACTTA 3120

TAGTTATGAA TCTGTATAGT TAGTTATAAT AATTGGTATT TTTTTAGGAA AATTTGAGCT 3180

TTTGAATTGA ATAAGAAGGA GTGATTTTAT GGATTTAAAG TACAATGTTT TTGGTAATTC 3240

AATGTATTCT TTGAAAGAAA TGGAGCTAAT TCAACTAGCT TCACAATGGT TTTATGATGC 3300

AGGAAGGTTG CCTAGTAGTG CATGGGCGCA AGTAACCTAT GGGGATGTCG CTGTTGATAG 3360

TATACGGAAA ATTTTATCTA GTCAAAGtGA GCCGTATCAC TTTGGAGGAG TTGATAATAT 3420

TTTTCCTCCC TATTTACATG CTATTAGATC AGCAGTAGAA ATGTATAATG ATTTTTGTAG 3480

ACATGCACCA CCGTATTTCT TTAATAAGAC TTCAGATGAA GAAATAGAAA AAGTAATGAA 3540

TATTTCAGCA AATTTTGGTA TTGAGTTAAA GAAAGAATAT TTTACACGTT TCACAATGCT 3600

AGAGGAAATA TTAGTTAAAA TAAAAGATGA TATTGATTTC TTTGAAAATA ATTCAAAACT 3660

GATGGAATAA TAAAGTGCTA ATTATACGGA CAATACTTTA TAATTAATTA TACAGAGCTT 3720

TTAAGGCTAT TTCATAAAAA AACAAGTAGT AAGGGGCACA AGCATTTCCT TTATTGCTTG 3780

TTTTTTGATA ATATATTACT GCAAACGCAT TAATATATTA AGAAAGGAAA CTTACCCACC 3840

ATGCCTAGAA AAACCTTTGT ATATCCACAC CCAATCAATA CTTATATCAT TAAGCAACTA 3900

GGGATTACTG TAGAGGAATT TTGCGAACTA CACTCCTTCT CTCAAGGAAC TATTTCTTCT 3960

TGGATCACTC GCAACAAAAA AATGGAAACT TTACCAGTCT CCTTTATCTA CTCTCTTTCC 4020

TTATCTGCTA GCAAGACTAT GGACCAAGTC TATTCAGAGC TTTTAAAACT GCAAGACGAT 4080

TATTTGCTAC ACTTAGAACG CCACAGACGA ACTAAAAAAA TTATTGAATG AAAATACTTT 4140

TTTACTATCT TACTATTTTA CTACAAAAAA TAAAAACACA AGAGCAATGT TTGTACAGCA 4200

TTTCTTGTGT TTTTATTTTA TTAAACAGTA TCTTTTATTT TTGTTTCTTT TTGTCTCGTG 4260

ATTCATAAGC TTTCATATAC ATTTCAAACA TTGTACGTTG TTCTGGTTCT AATGTGTTAT 4320

CTAATCTATC TAACATGAAC GATACAATGT CGTCTTGTGT TTCAAAATCT AAAGTATTCT 4380

GAAACGCATT TATTCTATTG ATGTTACTCT TTTGTATCCT AATTGTACTA TATACTTTAT 4440

TTTTAGCAGG TCGCCCAATT CGTTTTGATC CTGAATCACT TGCTGAACTT TCTTCTTTTG 4500

AAAAACTACT TAAATCAAAA GTAGTTTCAG GTTCTGAAAT TTCAGTTGCk TTAATTTCTT 4560

TTGATTGTTT ACGTTCAGGA TCCTTATGAA ATGTATACTT ACTCATTAGC AGTCGTCCCT 4620

TCTGTTCGkT CAATTATTTC ATCAGATAAT TTGTkATACA ATtCATGAAC TTTTCTATCA 4680

TGAAAATCAT ATAATTCCGT ATATCCTTCT TCTGATATAC CTTTACGATC ATAACGTTTT 4740

AACCGTTCCA TGTGTCTAAC AATTGTATTA AACAAAGTCT CATCACCAAA AGCGTCTTTG 4800

GCATCTTTAA TTATTTGATT ATCAATTCCA GAATCGTTTT yCAACAGCAC TGGAAGTACC 4860

CCAGCGATAT CAAAGTCAAT GTTTTTATAA GTATCATACA AGGTTTGCAA GTATTCCCCA 4920

AAATGCTTCT GCACCATCAA GGGAGCGTTG CTGCGTTTGC AAAACAATTA CAATGTAATT 4980

AGAAGAATAA AGAGCTGTAT CGGTAAATAC TGAAAGTGTA GGAGGAACAT CAAATATAAT 5040

GTAATCATAA TCATTTTCAA TTGGTTTAAG TAATTCACTA AAAAATGCTA TTCGTCTCTC 5100

TTTGTAGTTT TTCTCTGTCG GCATAATTGT TAGCTCAAGA AAATCAGGAT AATTAACAAA 5160

GTCTTTGTGT GAAGGCAATA AATACAAATT AGGCATTACA TTTACAACAG CTTTTCCTAA 5220

ATTTCCTTCT TGAATTGCTA CCATCATCGT TTCTTTTATA GGTAGTTCTT TATTATTTTG 5280

TAATCCATAT GTTCTTCGCA ACAATTGCGT TGAATTTGCT TGTGGGTCTA AGTCACAAAC 5340

TAAAACTCTA AAACCTTTTT TAGCTAATTC ATAGGCTACC ATTACAGAAT TAGTAGTCTT 5400

CCCAACGCCG CCTTTGAAAT TTCCAAATAC AATTTTTTTA ACCATACAAT GACT 5454

(2) INFORMATION FOR SEQ ID NO: 407:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4053 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 407:

ATCCCGGGGA ATGAGCCAAT CTATCAGCAA GCTGCAGTAG CGTTACCGCT TGGATGAAAG 60

AAAATCACTA TGATCTAGTT TATGGCGGTG GCAATGTTGG TCTAATGGGT ACGGTGGCTG 120

ACACGCTTTT AGCAGAAGGT GGAGAGGTAA TTGGTGTCAT GCCCACTTTT TTAATGGAAC 180

GAGAGATTGC ACATAATGGC ATTACAAAGA TGCATACAGT CAGTGATATG CATGCGCGTA 240

AGAAAAAGAT GATTGAATTG GCAGATGTGT ATTTAGCTTT ACCCGGTGGT CCCGGCACGC 300

TAGAAGAAAT TAGCGAGGTG ATCTCTTGGG GCCGAGTCGG TGAACATATG AATCCTTGTA 360

TTTTATACAA CGTGAATGGT TACTATGATT TACTAGCTGC TTTTTTTGAT AAAATGGTAG 420

AAACTAATTT CTTAACAGAA GCGGATCGAG CTAAAATTTT CATTTCAGAT TCCTTAGAAG 480

AAATTGGTGC ATTTATTGAT TCCTATGAAC CGCCAATGAT TCGTCAATAC AAAAAATAAC 540

AAAAGAGGCA GTCACTCTAA CCAGAATGAC TGCCTCTTTT TTACTAGTTT TTCGTTAAAA 600

TCCGATGTAA GAACTCTTTT GTCCGAGCTT CTTTTGGATG GACAAAAATA TCTTCAGCGG 660

TCCCTTCTTC GGCAATGACA CCTTTATCCA TGAAAATAAC ACGATCGGAT ACGTCCCGCG 720

CAAATTCCAT TTCATGTGTT ACAATCACCA TTGTTAGCCC AGTATGGGCA AGATTTTTCA 780

TCGTTTTTAG CACTTCGCCA ACCATTTCAG GATCCAATGC AGAAGTTGGT TCATCGAAAA 840

GTAAAACATC TGGATCCATT GATAATGCTC GAGCAATGGC TACCCGTTGC TTTTGTCCAC 900

CAGACAATTG aGCAGGTTTG GCATCAATAA AACGAGCCAT GCCAACTTTT TCTAAATTTT 960

CTAAGGCAAT TTTTTTGGCT TCTTCTTTGT TTCTTTTTAG GACAGTCATC TGACCAGAAG 1020

TACAGTTTTC TAAGACATTC ATGTTATTAA ATAAATTGAA TGATTGGAAA ACCATACCTA 1080

AATGGGTCCG ATATTTGGGC AAACTATAAC CACGTTCTAA AACATTTTTA CCATTATAGA 1140

TGATTTTGCC ATCTGTTGGT TTTTCTAACA AATTAATACA ACGTAGTAAC GTTGATTTAC 1200

CAGAACCAGA AGAACCGATA ATGGTGACAA CTTCTCCTTT GTTGACGTTC ATATTGATAT 1260

CTTTTAATAC TTCGTTTTCG CCAAAGCTTT TTCGTAAATG TTCAATTTCA ATAATTGCAC 1320

TCATCTGTTT GCCTCCTAGC TTTCTTTTCC TTCTTCAGTT AATTCTTCTA CTTTAACGTA 1380

AGCAGAAGGG CCATCCATTT TACGTTCCAC AACACGTAAG ATTCGTGTGA TCACAAAAGT 1440

CATTACCAAG TACATGATGC CGACAATTGT AAAGGTTTGG AAGAACTGGA AGTTTGCGCC 1500

AGAAGCAGCA TTCCCTTGGA AGAATAAATC TGCAACACCG ATAACACTTA ACACGGCCGT 1560

ATCTTTGATA TTAATGACAA ATTCATTCCC TGTTGCAGGT AAAATATTAC GTAATACTTG 1620

GGGaATAACG ACTTTACGCA TTGTTTGTCC ATGGGTCATC CCAATCGCTT GTGCTGCCTC 1680

GAATTGTCCT TTATCTACTG CGAAAATACC TCCGCGAaCA ATTTCAGACA TATAGGCACC 1740

AGTATTCACA GAAACGATAA ATAATGCGGC CACTGTTCGG TCTAAAGAGA TGCCAAAAGC 1800

TAGAGCTAAA CCATAGAAAA TAACCATTGC TTGAACCATC ATTGGTGTAC CACGGAAAAC 1860

TTCAATATAA ATAGAAAGAA TTAAATTTCC TAATTTTTGG AAGAAACGAG CGACTGGGTT 1920

TTCAGAATCT GGAATTGTAC GGAAGACACC AATyAGTAAA CCAAGGGTTG TTCCGACAAC 1980

TGTCCCAATT AACGCGATAA ATAGCGTTAA ACCAGCACCA CGTAAAAACA TGTCGCCATA 2040

TTGATTCCAA ATATTTTTGA AATCGTTAAT AAGTCCTGTT TTTTGTTCAT CACTATCTGT 2100

TGCTGCAGGT TGATCCTTGA TTGCTTGGTC CATAATTTTT GTCCGTTCAT CTTGAGAAAT 2160

ACCTGCTAAA ATTTTGTTTA CTTTTTCAAT GTCTGGATCG CCTTTACGCA TGCCGACTGC 2220

TACTTGCGAG TCTTCAGCGC TAGCATCAAA GCCCTTTTCT TTTGGAAATc TAACATTTTT 2280

AATTCTTTAT TGACACTAGT TGCTGTTATG CCTTCAGGGC GTTCACTAAC ATACCCGTCA 2340

ATCATGCCAG AAGCAAGTGC TGTTCGCATA GCTGAAAAGT TATCCATTGC TTGCTGTTTA 2400

TTGACATTTG GAATTTGATC AATTAATCCG TAGTGGAATG TATTTAGTTG AGCCGTAATT 2460

TTAGCATCAG CTAAATCTTT CAAGCTTTTT GCATTAGCGT ATTTGCCATC TTTTTTTACA 2520

ATTACGACAA ATTGTGATTC ATAATAAGGG TTTGTAAAAG CAATTTCTTT TTTTCGTTCA 2580

GCGGTTGGGC TCATTCCGGC GATAATGGCA TCAATTTTTC CAGATTGTAA GGCGGGCGCC 2640

AAGCCATCCC ATTTGGTTTG GACGATTACT AATTTACGAC CTAAACCATC GGCAATTTTT 2700

TTAGAAATTT GAACATCATA GCCGCCTGCA TAAGAGTTGC CTTGAATAGG CACTGCGCCA 2760

TGTGCGTCAT TTTTcTGTGA CCAGTTAAAA GGGcGTAGCC AGCTTCCATG CCGACACGAA 2820

ATTCACCATT TTCCTCCGCC GAAGCGTTGG TTGTCATGCC AAGTAACGAA AATAGCGTGA 2880

CAAGTAGTAG ACTGAATGAA AAAACTTTTT TGTTCATCTT TTTCTCTCCT TTAAmTTTCT 2940

TGTCTGATTG ATCGTTTTGT CTTTTTTCAA TCAACAAGCT TTGATAATCA GCTGAGcAAA 3000

AGCGATTCTG TCTTTTTAAA GGAGAATACT TCTGTCACAA CCTATTTTTG AATACAAAAA 3060

AACAACCAAA CAGATGTAAG GTTGCTAGGT AATGTCGAAA AAAATTCCTC AACGAAACAT 3120

AGCACAACTT AGCATCACTA CTAAGACAGT CCGCAAGCTC TTAACTTACG TCCCAACACA 3180

CAACAGAAGA GGCTCCGTTA TGCATTTCGG CGATTGATCC TAAACTTGCT TCCCAGTGTC 3240

CTCTCACTCC ACAAGCGTAA TAAAAACCGC GACCTCTACC TCTTTTTTGA GGCACTATTC 3300

AGTTATTAAA TTCAGTTATT AATTTTaGGA TAAATCCAGA AGAAAGTCAA TATGTTCCAG 3360

CAAAGGAATT TATCAAGAGG ACACAAACAC CATGAAtTAA CGTAACATTT TCCCTACGAA 3420

ATAGAAACTA ATACGAAAAA AGCAACCATT CCTAAAGGAA CAGTTGCTTT TTACTTAAGC 3480

TTTGACAGCT TTTGGTTCAC GACCTAAAAT AGCTTGAAGA ACAGTCAATA CACCTAAAGA 3540

AGCTAAAATT GTATAAATTA GACTTGAGTA AGTTAAGATA CCGAAATCCC AAGTTTTGTA 3600

AACTAATAAA ACAATTGCTA AGGCGATAAT AGCGATAACG TTGAAGAACG CAAAAGCCCA 3660

AAGATTTCCG TTTTCTTTAC GAGAAAGATA AAAAACAGAG AACCAAATGC TTAAAACGCC 3720

CCAAGCGATG CAGGCAATAA TTGCAGCCCA ATAAATAATA AATGCGACCA TTGTTGTTTC 3780

ACCTCACTTT TACTCATGCA TATTTTACCA CGAGTTGAGA AATTTGTCAC AAACTCTTTC 3840

AAGATAAATC AATATATAGG TAAGAnGTTT tCmAAAAGTG GAAAAaCTTT tGGTGCTAAT 3900

GTTAAAATAC GTTATAATAT AGAAGATTAG GAGCTGAATC GATATGCAAG ATATCACGCA 3960

AGCCGAGGCA ATTCATTGGT TGGCAACCCA ACAAAAAATC ACATTTnTAA CAGGTGCGGG 4020

TATCTCGACA GCCTCAGGTA TGCCAAATAA TGC 4053

(2) INFORMATION FOR SEQ ID NO: 408:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1758 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 408:

TAAACATATT CTTTACTTGT TAAAGGATCT GTTATTTTTA TTTAAAAATT GTTAGCTAGT 60

TAGAGCAAAT GGTTCGTAAT CATTCGGTTG ATTGTTCGAA TTTATCATAG TAGTTAAAAC 120

GAAATTATCT TGTGTATATA GAAAACACCT ACTGAAGACG CACAATACTT CCCCAAGTTG 180

TGTATGTTTA GTAGGTGTTT ATTTATAATT GTATATTTGA ATTATAAATA AAATACAATT 240

GTTTTGTAAA ATGAAAACTt GAAATTTAAA AAaGAAAGCG GTAAAATATC TTaCCAAACA 300

ATTTTATTTT TCATTTTATT ACTACCTCTT GTCGCCTTAC CCCAATGAGG CGGcTTTTTT 360

TACATAAAAA TATTGAACTC AAAAAATAGA TATTGTAAAA TATCTATACA TATTTAAACT 420

CTTATTTTTT CACGCAGACC GCCTTTTCTC AATAAGGTGG TCGTTTTTTT GTTGAAAATT 480

GAAAATCAGT AAAGTAAAAT TGTTATATAT TAAGCTATAA CTAGTAAGAA CTATTTTATC 540

CCCAATAAGC TAGTTATTAC TTGGACCATT AGCTCAGCTG GTTAGAGCAA ACGGCTCATG 600

ACTATTCGGT CGATGGTTCG AATTCATCAT GGTCGATAAT AGAATTAGAC AGGTAGTGAA 660

TTATTTTGAA AAGAACAGGA ATTCTCTTAG TGAATTTAGG AACACCAAAA GATTCTTCCA 720

AGACGGAAGT AAGGAAGTAT TTAAAAACTT TTTTATCAGA TAGAAGAGTA ATAAAAATAC 780

ATCCTATAAT TTGGAAACCG ATTTTGAACG GTATTATCTT GAATATACGT CCAAAAAAAT 840

CAGCAAAACT CTATCAAAAA ATTTGTACTG AAAATGGATT TCCTCTTTTG GAATATACTG 900

AGAAACAAAT GGAAAATTTG AAAAATATAT GTCCAGAAGT GGAAGTTACA ATCGGAATGT 960

CATATAGTGA ACCGAGTATA GAAACTGCTC TAGATACATT ATTATCmAAA GAGATTGAAG 1020

AACTCAATGT CATACCAATG tATCCGCAAT ACTCTGGGAC AACGGTAGGA TCGGTGTTTG 1080

ATTCCGTTAT GAATTATTTT ATAAAAAGTG ATAGGATAGT GGATATTAAA TTTATTCGAT 1140

CATTTTACAA TAATCCACAG TATATAGATT ACTTTTCAAA AAAAATAAAT GAAGCTTTGA 1200

ATGAAAGTCC AATAGATGCT ATCGTTTTTT CATACCACGG AATTCCTATG TCTTATGTAA 1260

AAGATGGGGA TAACTACCCG AAGGAATGTA CTAAAACAAC AAAATTAATA ATGGATAAAT 1320

TAGGAGATAT TCGTTATTAT CAAACCTATC AATCGAAATT TGGGCCATCT GAATGGTTAA 1380

AACCAGCAAC TGATGATACC TTGAAAAAAT TACCATCAAA AGGTATTAAA AATATATTAA 1440

TTGTTGCmCC AGGTTTTGTT GTAGACTGTC TAGAAACAAT AGAAGAATTG GAACACGAAA 1500

ACAGAAACTA TTTTTTAGAA AATGGCGGGG AAGTCTACAA ATATGTTCAT CCATTTAATG 1560

GAGATATTGA AtTTGCmAAA TTAgTAAAGG aCATTAtTTC TTTGTAGTAA TTTGTTTTAG 1620

TkTTTTATTT TAGTTTGAAT TTGTTAAACA ATTTATTCAG GGTTTGCACG AATATATnTA 1680

CTGGnATACA TATAGGTAGA ATTAAGCCAA AATTAAACGG CCTATCCCCC GGTCCAGTTT 1740

AnGGAATAAG CCGCCTAA 1758

(2) INFORMATION FOR SEQ ID NO: 409:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1091 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 409:

ATGCTGGCGT TTCAGCTACT AAATTTCTAG CAAAGAAATT TAAAAAAGAT ACGGACGTAA 60

CAATCACATT GATTGATCGC CATTCTTACC ACACTATGAT GACCGAATTG CATGAAGTAG 120

CAGGCGGACG TGTTGAACCA GAAGCAATTC AATACGACTT ACAACGTTTG TTCTCACGCA 180

AGAAAAAACGT AAAATTGGTA ACAGATACTG TTACTGGAAT CGACAAAGAA AACAAAGTTG 240

TGAAAACTTT AGCAGGTTCT TATCCATTTG ACCAGTTAAT TCTGGGTATG GGTGGCGAAC 300

CAAACGACTT TGGAACACCA GGCGTGAAAG AAAATGGTTT CACCTTATGG TCATTTGATG 360

ATGCAGTGAA AATTCGTCAT CACATTGAAG CAACCGTTGC TAAAGCAGCC ATCGAACCAG 420

ATGCTGAAGT ACGCAAAGCA ATGTTAACAT TCGTTGTCTG TGGTTCTGGT TTTACTGGGa 480

TCGAAATGGT CGGCGAATTA ATCGACTGGA AAGATCGTTT AGCGAAAGAT GCCAAAATTG 540

ATCCAGACGA AATTACATTA ATGGTTGTTG AAGCAATGCC AACCATCTTA AACaTGTTAT 600

CTCGTAACGA TGCTGCCAAA GCAGAACGTT ATCTAGAAAA GAAAAATGTT CAATTATTAk 660

kGAACTCACC AATCGTTGAA GTTGCAGCAG ATCACATTAA ATTAAAAGAT GGTTCTGAAG 720

TACCAACGCA TACGTTAATC TGGACAGCTG GTGTGAAAGC AACTTCTGAT GCCGCTGACT 780

TTGGTTTAGA AGCAGCACGT GGTAGTCGTT TAGTTGCGAA TGAATACATG CAAGCAAAAG 840

GTTACGAAGA CAAAAACATT TATATCATTG GTGACTTGGT TTATTACGAG GAAACACCAA 900

ACACACCAAC TCCTCAAATC GTTCAAGCAG CTGGAACAAA CAGGTCATAC TGCGGCTGCC 960

AACATCGTTG CAGATATTAA AGGCGGCGAA AAAACACGCC TTTTAAGGGG AACTACCCAG 1020

GGATTCCTGG GTTTTCCATC GGCGGCGAAA TGGGGGCGTT GGCCAACTTA TTCnGATnAA 1080

AATTCCATTT G 1091

(2) INFORMATION FOR SEQ ID NO: 410:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1229 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 410:

TCATCTATTA ACTCTCCATA GGTTTCATTT AAGGATTTTG GGGAAACTTT ATAATTTTGA 60

ACTTTTCTTT CTATTACTAA TTTCATGACC CAGTCAACTG CTTCTAACAC TTGACTGTTC 120

TCTAATTCCA TTTCTATCAT tAATkCCTCA TCGTTTATGT TAGTCTGAAT TGCACTCGGT 180

ACTATTmCAC TCTTCAAATC ACTTAGATTA AGAAAGTTAA CAGTCATATA ATTTCGATCG 240

TTTATTTTTA TAGACTGTTT TTTTAAGTTC TTATCCTCAA GAACCTTTCC ATGATTCATT 300

AAAATATATC TATCAGCATA TTTTTCAACT TCTAAAATAT TGTGTGTGAC CACGATGATT 360

ATATGCCCCT TATTAGATAA TTTTTTCAGA TATCTCCATA TTTTTTGCCG TCTAATAGGA 420

TCCACATCAT TAGTAGGTTC GTCAAGGACT ATTATTGGAG AAGGATATAC AACGGCCATT 480

GCAAATGAAG TGAGCCTTCT AAGTCCGCCT GAAAGTTTCT GTCCTGGTAT ATCTTTCCAT 540

CCATCTATTT TTAATTCTTT TAAAATTTTA ATACACTCTT CTTCAGCTAT TTTCTGACTG 600

CCTCCTCTTA TTCTTATTAT TGACTCGACA GATTGAGCCA TCGTTATACC TGCCAAGGGA 660

GCATGTAGTT GAGGCATCAT AGAGACATAT TCTCTTGcAA TTTTACTATC TCTCGCAAAT 720

GATAGACCAT TAAATGTTAT AGAGCCCTTT GTTGATTTCG TTATACCTAT AATCTGATTC 780

AACATTGTTG TTTTACCAGC GCCGTTATGT CCAGCTATGG CAGTTATTTC ACCTTTATTA 840

AAGCGCAAAT TTATATTATC ATTTGCATAG CTCTCTGATT TCTTGTACTT TTTACTTAGC 900

GAATGTAATT CTAACTGTTG ATTCAAAGCA AACCTCCTAT ATTTAGACTT AATTAAATTG 960

ATTGATAACA GGTTGGAACC ATAACCAAAT TATCATTACT AGCACACTTA TGCTTATTAC 1020

TTTAATAATG ATTAATTTCA TAATTGCGAC CTCTTCAAaG TAGAAaGGAT TATCAACCAC 1080

AAAAGAAATC AAAaTGTTAA TATCTGTTTT TTGTATAATG ATAAATTGCC ACAATATCCT 1140

TGTTATAAGT ATTGATGCCG CAATAGAATT AACTCCAACA CTTTTATTAT TCATAATCAA 1200

AAAAACTAAC AAGAATATTG CATCCGACG 1229

(2) INFORMATION FOR SEQ ID NO: 411:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3721 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 411:

CATCAGTATT AGTTGGTGTA GGnCTTGTTT TTGGGGCAAC TGGAATAGTT AATGCACAAA 60

TGGGTGAAGG AAGATTAGTA AATTATTCTG CTTCAGGAAA TACGTTTCAA GAAAATCCGG 120

GATATACGAA GAATTATAAT TTCTCGGATT TACAATTCAA CCCTAAAGCA ATAACTGGTG 180

ATGTGTTACA AGGAAATACA ATTGATTTTG AGGTTTATGG GAAACATAAT ATTGCAGCTT 240

CAACTGCAAA CTGGGAAATT CGTCTTCAAT TAGATGAACG ATTGGCCCAG TATGTTGAAA 300

AAATTCAAGT TGATCCGAAG AAGGGCGTAG GAAATAGTAG ACGAACTTTT GTAAGAATTA 360

ATGATTCGCT TGGCAGACCT ACAAACATTT GGAAGGTTAA TTACATTCGA GCAAATGATG 420

GACTATTTGC TGGGGCAGAA ACAACTGATA CACAAACTGC TCCTAACGGT GTGATTACAT 480

TTGAmAAAAA TTTAGATGaA ATTTTTAAAG AAATTGGtGc AGATAATCTT aAAAGCGACC 540

GTTTaATGTA TCGTATCTAT TTgGtAAGTC ATCAAGATGA CGATAAAATT GTACCTGGAA 600

TAGAAAGCAC TGGTTACTTT TTGACAGATC AAGATGATTT CTATAATAAG TTAGATGTAT 660

CCGAAAATAA TTCTGATCAA TTTAAGCATG GATCGGTAAA TACTAAATAT GAAGAAGCTA 720

ATATCCAAAC AAAAGACGGT TCTGGGTCTA CTGGTGCTAA TGGAGCTATA ATTTTAGATC 780

ATAAGTTAAC GAAAGAAAAG AATTTTTCTT ATTCAACTTC TGCAAAGGGA ACGCCTTGGT 840

ATGCTAACTA CAAAATCGAT GAACGTTTAG TACCATATGT ATCTGGAATA CAAATGCATA 900

TGGTTCAGGC TGATAAAGTG GCATATAATG TTGCTTTTGA ATCAGGAAAA AAAGTGGCTG 960

ATTTAGCAAT TGAGCGACGT GAAGGTCATG AGAATTATGG TATGGGTTCA ATCACTGACA 1020

ATGATTTAAC TAAACTTATT GATTTTGCTA ATGCTAGTCC ACGACCTATC GTTGTTAGAT 1080

ATGTTTTACA ACTGACTAAA CCTTTAGATG AAATCTTAGA AGAAATGAAA GCAGCGGATA 1140

AAATCGAAGA AAATGCGCCA TTTGGTGAAG ATTTCATCTT TGATTCTTGG TTGTCGGATA 1200

CGAATAAAAA ATTAATCCAG AACACTTATG GAACAGGTTA TTATTATTTG CAAGATATTG 1260

ATGGTGATGG AAACCCTGAC GATAAAGAAG AGAGCGGAGA CACGAATCCA TATATCGGGA 1320

AACCTGAATT AGAAGAAGTA TATGATGTTG ACACAACAGT TAAGGGGAAA GTATTCATCC 1380

ACGAGTTAGC GGGAACAGGT CACAAAGCGC AACTTGTTGA TAAAGAAGGT ACTGTATTAG 1440

CAGAAAAAAC TATCGCTCCA AATGAAAAAG ATGGGGCTCC AATTTCAGAT ACTGTAGAAT 1500

TTGAATTTAC GGGTGTAGAT TCAAGTAAAC TAATCGCGAA AGATGAATTA AAAATCCAAA 1560

TCGTTTCTCC AGGTTTTGAT AAACCAGAAG AAGGTTCAAC CGTTATTAAG GAATCACCAA 1620

AAGCGGTTGA TAAACAAACC GTGGTAGTTG GATTTAAACC AGATGCTAAA GAATCAATTC 1680

GGAATAATAA AAACTTACCT GAAGATGCAG AGTATTCATG GAAAACAGAG CCTGATACTT 1740

CTAACGTTAC TGATAGTACG AAAGGTATTG TAACTGTTAA GATAGGAAAT CGAACTTTCG 1800

ACGTGGATGT AGAGTTTGCT GTAAAAGCTT CTCAAGCTAT GGAAAATGAT GCAACATACG 1860

TACCTATAAC AACAACCCCA GAAACGACAA TTCAAAGTGG TAAACCTACA TTTGATAAAC 1920

CAGATGTTCC TCTAGCTAAC GATGCCTTTT CAGTTTTAGA TGTTTATAAT AAGGACTTCG 1980

GCAATGCAAG TGTTGACGCA AATACTGGTA TTGTTACATT CACTCCAGCT AAAGGTGTAG 2040

GAGAATCGGA GCCGATTACT GGAACAATTC CTATTAAAAT TGTTTACCAA GATGGTTCTG 2100

TAGGCACGAC CGATTTAGCA GTAACTGTAA GTAAAAATAT TTATGAAAAT CCAGGAGAAA 2160

ACATTCCTGC AGGCTACCAC AAaGTAACCT TCACCGCAGg AGAAGGAACA AGTATTGAAA 2220

GTGGAACAAC AGTCTTTGCA GTGAAAGACG GCGTAAGCTT ACCAGAAGAT AAACTTCCGG 2280

TGTTGAAAGC AAAAGATGGT TATACAGATG CGAAATGGCC AGAAGAAGCA ACGCAACCAA 2340

TTAAAGCAGA TGATACAGAA TTTGTATCAA GTGCAACAAA ATTGGATGAT ATCATTGAAA 2400

ACCCAGGGGA TAACATTCCT GCAGGCTACC ACAAAGTAAC CTTCACCGCA GGAGAAGGAA 2460

CAAGTATTGA AAGTGGAACA ACAGTCTTTG CAGTGAAAGA CGGCGTAAgC TTACCAGAAG 2520

ATAAACTTCC GGTGTTGAAA GCAAAAGATG GTTATACAGA TGCGAAATGG CCAGGAGAAG 2580

CAACGCAACC AATTAAAGCA GATGATACAG AATTTGTATC AAGTGCAACA AAATTGGATG 2640

ATAAATCTGA TGCTGACAAA TATACTCCTG AAGGTCAAAA AGTGACTACA GAATTGAATA 2700

AAGAACCTGA CGCATCTGAG GGAATTAAAA ATAAGAAAGA TTTACCAAAA GATGCTAAGT 2760

ATACTTGGAA AGAAAAAGTA GATATTAGTA CAGCTGGAAA TAAAAAAGGT ACGGTTGTAG 2820

TGACATATTC AGATGGATCA TCTGATGAAG TTGAAGTAGA TGTCACAGTA ACAGACAATC 2880

GCTCTGACGC TGATAAATAT GAGCCAACAG TAGAAGGTGA AAAgTAGAAA TCGGTGGTAA 2940

AGTrGATTTA ACAGATAaCG TTACTAACTT ACCAACGTTA CCACAAGGAA CAACAGTAAC 3000

AGATGTTACT CCTGGTGGTA CAATCGATAC TAATACACCA GGTAATTACG AAGGTGTCAT 3060

TGAAGTAACG TATCCAGATG GTACAAAAGA TACAGTAAAA GTTCCAGTAG AAGTAACAGA 3120

CAATCGCTCT GACGCTGATA AATATGAGCC AACAGTAGAA GGTGAAAAAG TAGAAATCGG 3180

TGGTAAAGTA GATTTAACAG ATAACGTTAC TAACTTACCA ACGTTACCAC AAGGAACAAC 3240

AGTAACAGAT GTTACTCCTG GTGGTACAAT CGATACTAAT ACACCAGGTA ATTACGAAGG 3300

TGTCATTGAA GTAACGTATC CAGATGGTAC AAAAGATACA GTAAAAGTTC CAGTAGAAGT 3360

AACAGACAAT CGCTCTGACG CTGATAAATA TGAGCCAACA GTAGAAGGTG AAAAAGTAGA 3420

AATyGGTGGT AAAGTAGATT TAACAGATAA CGTTACTAAC TTACCAACGT TACCACAAGG 3480

AACAACAGTa ACAGATGTTA CTCCTGGTGG TACAATCGAT ACTAATACAC CAGGTAATTA 3540

CGAgGTGTCA TTGAAGTAAC GTATCCAGAT GGTACAArAG ATACmGTAAA AGTTCCgTAG 3600

rAGTAACAGA CAATCGCTCT GACGCTGATw AATATacaCs TArAGGTGAA AAGTAGAATC 3660

GGTGGTAAGT GnTTTACnGA TACGTTACTA CTTACCAACG TTACCACAGG nACCACGTAA 3720

C 3721

(2) INFORMATION FOR SEQ ID NO: 412:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1531 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 412:

nCAAAnGAGG GCTAATTAAT GGAnAAAAAT GGATTGAaAA AAACmAATAT TTTGTTGGAG 60

CAATTGGAGT TATTCTCaTT GCTCTTTTTT CATTCCtTCT TTTTGAAAAC CAACAAAAAG 120

AAGAAGTGAA AGAAGCTATA AATACAGTGG TATCaAGTGG CTTAGAGGAA GCGGAGAAGC 180

CAAAAAAACA TGTAGAGACA AAATCTAGTG AAACTGTTCA AAGTAATCAA GTAGATGTTA 240

TTACTGGTAC GGAAGCACAA GAAGAaACmC CTATTTCAAC AGTGATTTCT TTAACCGAAG 300

GCGAACGTCC GTCTGTTAAA ACGATTCAAA ATGCCCAAGA AAATTTTAAT GAAGTGTCGG 360

ATTCTGTTTT AGGGTCTATT GAAATTGATT CTATTCAATT AAAATTACCT ATTTTAATGG 420

GTGATAGTTT TGACAATATG CTTTATGGAG CGTGTACTGT TTTATCAAAA CAAACGATGG 480

GAAAAGGAAA TTATGCTTTA GCTTCTCACA ATGCTGGGTA TGAAGGATTA CTTTTTACCT 540

CGTTAAATAA AGTTTCTGTA GGAGATTTGG TCAAACTTAA TGATCGTGAA GGACATTCGT 600

TCATATATAA GGTAAAAGAG CAAAAGCATG TTGATATGAC TGATACAACT ATGTTGAATC 660

TAACAAGAAA GCCAACATTA ACTTTAATAA CTTGTGACCA AGCCACAAAA ACAACAGGAA 720

GAATTATTGT GATTGCGGAA TTAGTTTAGA GGAGGGAAAA AGTTGGTTGA AAAAATAAGG 780

GAATTGTCAA AAGAAGAATT TGTTAATCAA TACATGTTTG ATTTTAAGGT ATTGATTTTT 840

AAATCTAAAA ATCCTGAAAA AATAACTGCC AAGCAAGAAA GACTTTTAGA ATGTGAAAAG 900

AAGATTGCTG TGTTGTTTTA TGAAACGTAT AAACGAAATA AAGGTTATCT GCCTGATGAA 960

AAAGAACTGG GGAGGATTGT TCAAAGAAAT TTTCTTGACC GTTTAAAGTT ATTTCGAGTG 1020

GAATATGATG TGATtAGTGA AGAAAAATTT tGTGGTCTAC ACGTACAAAT GGTCAAACAG 1080

CAAATGCCAT TAGAAAAGTA TCGAACAGAT GATTtATCTT ATATTTTAGG TCGAGAAAAG 1140

AAGATAGCTA TTAATTATTT TATCGCACAT GATGATTTTC CTATGGGTTA TGAAGAACTA 1200

ATGATAAGTC GTTCTAAACA AGCAGTTACC CAAGGcTTGG AAGAGTTAAG GGGAGAATTt 1260

ATGGAAAGAT ACCATAAATA TTACAGAAAA ATGGAAAGGA GTTGTATTTT ATGAATCAAC 1320

ATATTATGCC AATTTTTAGA GAGCATAGTA AATTTTATCA AAGGTATTCA AATGTACCAG 1380

TTCCTGAAAT GCTTCGCTTG CATGGATTCA TAGTTGATGA AATGCCAATT GATGAAGCAG 1440

ATATTGTTTA TATGGAAGAA TATAAAAATT CTTTACTTAA AATGGGGATT ATTATTGATG 1500

AAGTGCCTTT TAAGAAGGAA TTTGTTGCGC A 1531

(2) INFORMATION FOR SEQ ID NO: 413:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2760 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 413:

ATTGTAGGAG ATAATGGTGG CTTGGCACGG GTCTCTGAAT CGTTTATTTT AGGTTCGGTC 60

TCATTAATTT TTTGGACATT GACTATCTTA ACCACGATTA AATATGTAGT TATTGCTTTA 120

AATGCTGATA ACCATGGAGA AGGTkGGATT TTCTCTCTTT ATACATTGGT TCGTAAAAAA 180

AGTAAGTATT TGATTATTCC cGCAATGATT GGCGGTGCCG CTTnacTGgc TGATGGCGkk 240

TtGACCCCAG CTGTGACAGT GACGACAGCG ATCGAAGGCT TACGGGGGAT TCCAGCATTT 300

TTTGAACGTT TTGGCAATGA CCAAACGATT ATAGTGGTTA TTACGTTGAC GATTATTTTA 360

ATTCTTTTTT CTGTTCAACG ATTTGGAACA GAATTGGTCG GGAAAGCTTT TGGGCCGATT 420

ATGTTCTTGT GGTTTACTTT TTTAGGCATC ATTGGTTTAA TGaATTTTAG CCAAGATTGG 480

ACGGTTATTC GTGCaTTAAA tCCATATTAT GcTTTGcAAT TACTAGTCAG TCCAGAAAAT 540

AAATTAGGTC TATTTATTTT AGGGAATATT TTCTTGGCTA CAACTGGTGC TGAAGCTTTA 600

TATTCTGATT TAGGGCATGT AGGAAAAATG AATATCCGCA TTAGCTGGCC GTATATTAAG 660

ATTTGTTTGA TTCTTAATTA TTTAGGGCAA GCTGCGTGGT TGTTAACGGT CAAAGAAAAC 720

CCTGAAATGC AGGCATTAGC TGAAATTAAC CCATTTTTCC AAATGaTTCC ACGAGGTATT 780

CTCGTATTTG GCGTTGTTTT TGCaACAATC GCCGCAGTGa TTGCTTCGCA AGCGTTAATA 840

TCAGGTTCGT ATACTTTAGT TTCTGAAGCG ATTAAATTGA AGTTATTACC AAGACTTAAA 900

ATTATTTATC CTGGGAGCAA TATCGGTCAA ATGTATATTC CAGCGGTAAA TTTAATTTTG 960

TGGTTAGCTT GTTCAGCAAT CGTGCTGGCT TTTAGAACAT CGACACACAT GGAAGCTGCG 1020

TATGGTCTAT CGATTACGAT TACGATGTTA ATGACAACCA TTTTATTGTT GTTTtATCTT 1080

TtGGATAAAA TTCCAGCTTG GTCTGCGTAT CTAATTTCCT TGTTCTTTGC TGCCATTGAA 1140

GTTGTCTTTT TCTTTTCAAG TGCTGCGAnT TTTTCCATGG TGGATATGTT GCTGTGGGAA 1200

TGGCAGTCTT CTTGTTATGC ATCATGATTA TTTGGGAACG CGGCAACGAG ATTAAGGAAG 1260

CGACGGCGGa ACAAGTGTCG TTGAAAAAAT ATGTACCACA ACTTAAAGCT TTGAAAGAAG 1320

ATACATCTGT ACCTATGTAT CAAACCAACG TGGTCTTTTT AACCTCCGAT CGAGTGGATG 1380

GCGAAATTAA TCGTAATATT ATTTATTCGA TTTTGGACAA ACAGCCAAAA CGAGCGAACG 1440

TTTATTGGTT TGTTAATGTT CAGGTAACAG ATGAaCCCTT TACTCAAGAG TATTCTGTTG 1500

ACATGCTAGG AACTGATTTT ATTGTGCAAG TTCAGCTATA TTTAGGCTTC CATATTTCTC 1560

AAGAAGTGAA TGTATATTTA AGACAAATTG TTCATGACTT GATGAAAACA GGCCGCTTGC 1620

CAAAACAGCC ACAACGTTAT TCATTAACGC CAGGAAGAGA AGTTGGCGAT TTCCAATTTG 1680

TATTAATTCA AGAAGAATTA TCTAACGTTT CTGAACTGAA AAAATGGGAT CGACAAATTA 1740

TGCAGGCAAA ATTAGCTATC AAAAACTTAA CTACCTCTCC TGAAAGTTGG TTTGGCTTGG 1800

AATACAGTGA AGTCAAATAT GAATCCGTTC CTTTAATCAT TGGCCCACAA CGGAAAACGC 1860

ACTTAGTCGA ACGAAAGAAT CGTTCATAAA TAGAAGAAAT TTAACTAAAA GATTGAGCCT 1920

AAGTCCTTGA GATACATGGC TCAATCTTTT TTGTAGAATC AGAGTAAATA CTGTAATTCC 1980

ACAAAGAATT CATTTAAAAG CGTTTGGTCT TTAAAGGAAT TCATCCGCTT AAAGATGGCC 2040

AGTAAGACTT CTTTTTGAGG TTCAGTTAAT TGATCTACTT TTAACAAAGG AAGAAAAGCG 2100

ATAATCAGAG AATAATCCAA TTCTGGATTT TGAGAATGTG TAAATAATTC ATGGAGTTCT 2160

GACAAAATAA ACGGATGATA ATAAGATTGG TCAGACTGTT CTTCTGCAGT TGTTGGTTCA 2220

TAAATAGAAA TACTCTCTTG GGAATGAGAG GGGGCTTCTA CTGTTAAAGG AAAAGCATAA 2280

AAGGACTGCT GTAACTTTTG CGCTTGAATC ATAATTTTTA AGGAATCGTG ATAGGATAGC 2340

TGTTCTTCAG ATAAACGTCC CATCAACTGA GTAAAAAAAG GTGGTTTTAG CAGTTGTTTA 2400

TTTTGAAAAT TTtCTTTTAA TAGCTGGCGC ACAAATCGAT CGAAATCACC ATAAAGTAGC 2460

CAATCTACTG TTGTGTCACA GGCTTTAGCG ATTTTTTTTA AGTATCGTTT ATtGGGTAAA 2520

TTCAAACCTA GTTCCCmGTT ACTAATkGkA CyTCCACTrG ATAAmCCAaT TTTATkGGCT 2580

AAATAACGTA AAGTGTACTT CTTCTTTTTT CGAGCTTCAC GCATGCGTGC CCCGACTTGC 2640

TTTGTAGTAA TTTTCATTGA AACACTCCTT TTGTAATAAA TATGAGGATA AGGTGTTCTT 2700

AAAATATACA TACGCAAAAT TCGGCTCTTT TTnTTCTAAA CnAAAAAAAG TnATTTATTC 2760

(2) INFORMATION FOR SEQ ID NO: 414:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1142 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 414:

TTCGGCTACC CCTAATTTTA TCGGCtTAGA tAGAaATAGT GAATTAAGAG GTTAGCAGAC 60

AACAAAATAA GAACCGACTG kCaTAGGACG GGAATCCTAG AGGACAGTTA AACAATTCtG 120

ATACAGCATG AACTATACTt CGGTTCTCGT TTGTTGCAAC ATTAGTTACA ACGTATAGTA 180

TAACAATTTT TAATGTAAAA TTCTAGGCTT TTTTAAACTC CTTTATTGGC CTAGGTAAAG 240

TTTTTTACAA TGAATTGTTT TAATTAGTTG TATAAATGTT GGGACAGCGG GGAATGTATA 300

CAGTTCATGT ATATGTTCCC CGCTTTTTTG TTGTCTGTTA ACTTGTTGAA AAAATAAAAA 360

AACCAAGTAG CAAGCATTCT CATTTTCTAA AAATGGCCCG CTTGTGGAAT AATTCAATGC 420

TTTCACGCAT TGTTTGTACT GGAACTACTC GTAAGAGCCG TGCAAACAAC GTGTTGCATG 480

GCTCTTTTTC TACTTGGTTC AGAAAGTAGG TGAATCAATG ATTGAACTGA AAGCAACAGA 540

TTTGCACCGT ATTACGTGTC AACTGACCTT GTTTAATCGA CAAAGGTTTA AATTGTATAA 600

CGGCACAACA GAACGTATCG TATATGATTT TTCAGGAAGA AATCTGTTGA TCAATCCTGT 660

GAGTTTTGAA ACGGAGCAAG ACATGGAACG TTTCTTTCGA CAAGTAAAAC TGATTTATTT 720

TGATGGAAAA GTGGTCGGTT ATCGTGGCTG CTCGGAACTT CCGTTAAAAC TTGAATGTAG 780

AGCGTtCAAA AAAATGGnTA AcGACaAAAG ATGTTACTTA ATGCACCGTT CAACTATAAG 840

GTTTTTGAAG AAAATGAATT TAGGGAATGG TGCAAGAAAA TGAGAAGCTA TAAGTAGTGA 900

ATTTGGTTTT TTATACTTGG TCTTTTTGTT TGTTAGTTTT AACTAACATG TGCAATTTAC 960

CAATAGCTAG GTTAGCTATC GAAAAAGGAT GCAGAAAAGC ACACGGACAC GAAGGTGCAA 1020

ACTGGCTACA CGTCAACTTT TTTGCAGATT TGGGAGATCC TAAAAAATGA ATACCACAGG 1080

nCTTATCCAG TCCATCAGCG TGCCTTAAAA GTCTAGGnAA GGGnAACCAn CCTAAGAGTC 1140

CG 1142

(2) INFORMATION FOR SEQ ID NO: 415:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6285 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 415:

CTTTAGTTCA GGCTGGGCCT TTTGAAAATT TAGATGAAGG ATTTGATGTC TACGCCACTC 60

AAGCCTTTAA TGCGCCAGAT GGTCGTGCAC TTGCGGTCAG TTGGATTGGG TTGCCAGAAA 120

TCACTTACCC AAGTGATGTG GAGGGTTGGG CAAATGGCTT AAGTCTGGTT AAAGAACTCA 180

CAATTCACAA CGGGAAACTA TTTCAATATC CAGTTTCTGA AACAGAAATG CTTCGTCAAT 240

CCGCTACTAC TTTATCAAAT GGCTGCCATT TCTTATCTAC TGCTTCTTTT GAATTAGAAG 300

TGGATATTCC CAAAAATGAG ATTGCTTTTA TTCGGCTTTT AGCGAACGAA ACGGGTTCAA 360

AAGGACTTTT AATTACAATT GrTACGATTC ATGGTAAAAT AACCCTTGAT CgAACATTTG 420

CTGGCCAATC TTTTGCTGAA AAGTATGGCA CAATTCGTGA AACTAAAATT AGGAAAAATA 480

AGTCAGTTCA GTTAACTATT TTTGTTGATT GCTCTGTTGC AGAAATCTAT GTAAATAAAG 540

GTGAAAAAAC GATGACTGGT CGCTTCTTTC CAGATAAAGC GCAACAGTAT CTTCATCTAT 600

CCAAGACGGC AAAAGCTTGT TTTTATGAGC TGGAAAATAC GAATAATTAG GAATGATGGT 660

GAATTTTGAT GGTGGTTAAA TTAACGGATG TAGCAAAGCT TGCTGGGGTG AGCCCGACAA 720

CGGTAAGCCG CGTGATTAAT AATTATGGTT ATCTTAGTCA AAAAACAATT GATAAAGTTC 780

ATCAAGCGAT GGAAGAATTA AATTATCAAC CTAATGGATT AGCCAGAAGC CTCCAAGGAA 840

AAAGTACGCA GCTGATTGGT TTAGTCTTCC CTTCTGTTAG TCATCCATTT TTTGGTGAAT 900

TAATTGAAAC ACTGGAAAGA AAGCTCTTTG TTCAAGGATA TAAAGTGATT TTATGTGATA 960

GTGAAAAAGA TCCAGAAAAA GAGCGCGCCT ATTTACGAAT GCTCGCTGCA AATAAAGTGG 1020

ACGGTGTAAT CACTGGTAGC CATAACTTAG CTATTAACGA ATATGAAAAT GTTTCACTAC 1080

CTATTGTTTC CTTTGACCGT TTCTTGGCAC CTGGCATTCC AATTGTCTCT TCGCAAAACT 1140

TTCAAGGGGG CCAAAAAGCC ACTGAAGCCT TATTTGCAAG TGGTGCACAA AAGATTGCAA 1200

TTATTACTGG TGCTAATAAC ACAGGCGCAC CTAGCGATTA TCGATTGGCT GGTTATAAAC 1260

AAACAATGGA AAAATATGGC GCAGAAAAAA CGATTCTACA AATTGATAAT GGGACCTCAA 1320

CAACATTAAA AAATCTAGAA ATCGAACGTT TGCTTCAAAA TAAAACTGTA GACGGCATCT 1380

TTTGTACAGA TGATTTGACA GCAATTACAG TTATGAATAT TGCTCAAAAA TTGAAGATAT 1440

CCATTCCTGA AGAATTAAAA GTAATTGGTT ATGATGGGaC AAAATTAATC AAAAGAATTG 1500

CCCCACAACT ATCAACCATT GTGCAGCCAA TCGACGAGAT GTGTGACGTT ATGATTGACT 1560

TACTGCTTCG TAGAATGAAG GATCCTGATG TTGCACTTGA GGAAAATTAT CCTATTCCGA 1620

TTCAGCTATC ATTGTCTGAA TCCTGTTAAA AAAGACACCT TTTCAGGTGT CTTTTTTTAT 1680

TCTTAATTCC TCATTTGCCT ATTTTCCCTT AATTCCAAAG CACAGAGATT AAATAATACA 1740

ACAACATGAT CCCTACAAAT ACACCGAGCA TTAGTAAAAC ATACCACCAA TTTTTGGCGC 1800

TATAGGTTTC ATTTTTATTT TTTTCGGTTC GTTTCCCGGC AGCCCATTTC CATAAATCGT 1860

AATTATTCAT CTCTCTCACC ACCTTTAATG GGTTTATTAA CTTAAGCCTA ACACGGAATG 1920

AGAGAACTTG AAAACAAAAG GATCACCCGC TTATGCAATG GATGATCCTT TTTTGGGCAC 1980

TATTCTTCTG ATAACCCATT ATTTTTAATA ACTGATTGGT ACCAATAAAA ACTATCTTTT 2040

TTAATGCGTC GTAAATCTTT TAACTCATGG TCTTCACGAT TAACGTAAAT AAAACCGTAG 2100

CGTTTTTTGA ATCCTTGATG GGAACTTAAA ATATCCATGA CCGACCAAGG GCAATAGCCA 2160

AACAACTCAA CACCGTCAGA AATAGCCGCA TGACAAGCAG CAATATGATC ATGCAAATAC 2220

GCAATTCGAT AATCATCATG AATTTTTCCG TCCTCTGTTA GGTGATCTGG TGTGCCTAAG 2280

CCGTTTTCTG TAATGATTAA CGGCAAACGG TATTGACGAT AATAATCATT TAAAACTAAA 2340

CGTAGACCAG TTGGATCAAT TTGGGCACCA TACTTAGAGG TTTTTAAATG TTGATTTTTT 2400

TCGATTTTAA AATAACCGTA TAGATCAAAA TCAATGTCCT TTTCTTTCGT ACCCAATGGG 2460

TGTTGCTCAT CTGTCGGTAA ATAGCTAGCG ACTAACGTCC GATAATAATT AAGAGCGATA 2520

AAATCTGGTT TAGCGGCTTT CAAAATTGCT TGATCTTCTG GCTGTGTTTC GGGATAAATG 2580

CCGCATTTGG TTAAATACTC TTGATAGTAA CCAGGGTATT CGCCAAAACA ATACATTTCT 2640

AGAGCGTAAT TCGTTTTAAA GTTATCATTC ATACGAGCTG CCCAAACATC TTCAGGCTTA 2700

TTAGAACTAG GATAAGTCAT CGTTGAGGAA ACAGAGGGAC CGACTTTTCC CCCAGCGACC 2760

AATTGGTGAC AATCATTCGT CGcTAACGCA TGGGCAACAA ACATATGGTA ATCCATTTGT 2820

GCACGAATTT TTTCAACATT TTCTGGTGCA ACATCGTACA TATTCATTCG TTCATCAACA 2880

CGAATCATTA AGTTTTGTTC GTTCGTAACT TGCCAATTTT TGACTCGATC ACCAAATGCT 2940

TGATAACAGA CTTGCGCATA GCGTTGAAAT GCGGAAACGC AGCGACGATC CGCCCAACCA 3000

TTGTATTTTT CCACCAACGC AAAAGGTAAA TCAAAATGAT AAAGTGTGAC AAACGGCAAA 3060

ATATCGTTTT CTAAAAGTGT ATCAATGACA GCATTATAAA AATCGATGCC GGCTTGATTG 3120

ATTTCGCCAT CACCTGTTGG AATGATTCGT GCCCATGAAA GAGAAAAACG ATAGGTTTTT 3180

AAGCCTAATT CTTTCATTAA CGCAATGTCC TCTTTGAAAC GATGGTAAAA ATCGCTGGCG 3240

ACTTTCGTAT CTGCTTGTAC AGCTGACTTT TTAAAAGAAT TGTAATCAGC GACTGTTAGG 3300

CCTTTGCCGT CCTCATTCCA AGCGCCCTCA ATTTGAAAAG CGGAGGAAGA AGCGCCCCAT 3360

AAAAAATCGT TCTTAAATTG TTTAACCATT AGTTTTGCTC CTTTACTTGC ATGAAAAATA 3420

AATTTGTTTG TGGTGTAATC TCTGTTTGTT TGGTTGTCTC AATTACAGCC ATTTCCGCAG 3480

TATTTGTTAC GATGACTAAG ACGGTTGGAT CATAACCAGC AGCTTGGATG ATTTGTTTAT 3540

CAAAAGAGAG CAGCTTTTGT CCTAGCCTGA CTCTTTCTCC TACTTTTACG AACGTTTCAA 3600

ATGCCCGCCC TTTTAATTCG ACGGTATTGA TTCCAATGTG AAGGAGGATT TCAGCACCAT 3660

TGTCTAACTT TAAGCCAATT GCATGATGAG TTGGATATAA AGCCGTCACT TGCCCATTTC 3720

CTGGTGAAAC AATTTGTTCC TCTGTCGGAT AAATAGCTAG CGTTTTTCCC ATTTGTTGAG 3780

AAGCAAAGAC TTCATCGTTG ATTTGGTCTA AAGATACGAG TTGTCCTTGA ACAGGTGCTT 3840

GAATCCGTTG TTTTTCTTTG ACTGTCCCTG CCTCTGTCTT GAAGCTGGGT TCTTGACGTG 3900

ACTCGTCAAG CGGTAAATCA TTAAAGCCTA GAAAAATCGT TAATAGAAAT GCACCAAAAA 3960

AGCAATACCA CAACCTAATA AATAACAAAT AAAGGCTTTG GTTCCTGCTT CTGCGTAAAC 4020

TGGGATGGTC AAGAGTCCTT GGTTAGCGAA GGCTGTCCCG TAGGAACCGC CCCAGCCCAT 4080

TAAGGCCCCG CCTGCTGcTC CGCTAGCCAC CGCATAAATC ATCGGTTTTT TTAGCCGCAA 4140

GTTTGCACCA TAAATGGCTG GTTCTGTGAT ACCAAATAAT GCAGTCACTG TTGCAGAGGC 4200

AGCAACAGCT TTCAAGTCTT TATTTTTTGT GCGGACAAAC ACGCCAAATG CGGCGCCCGC 4260

TTGAGAAAAG TTTGCTGCAC CAGCAAACGC TAGCAAATTT TGCCGTCCTG TTCGAGCGAC 4320

ATCATTAATC CCAATTGGCA CAAGCCCACG GTGAGCCCCA AAAATGACTA AGACACACCA 4380

AATTCCCCCG ATAAAACCGC CTAACAAGAT GGGACTCAAA TTCATTATAT AATAATACAA 4440

CCAATTCACG CCGTCACCTA AATAAATCCC TATCGGTCCA AATAATAATA AAGTGGCTGG 4500

AATCATTAAA AGTAATGACA GCGTAGGAAC CATAATAATT TTTAGAACTT CTGGAATGAC 4560

TTTTTCCAAA AAGCGTTGCA CATAAGCTAA CATAAAAATT GCAAGAATAA TAGGAATTAC 4620

TGTGGAAACA TAATTGGCTT TGGTCACATG TAATCCAAAC AAGGTAACCG CTGATTCTTC 4680

AGTCATTAAT GAAACGATGG TTGGATAACA TAAGGTGGCA CCAATTACCA TGGCGATATA 4740

TTCGTTTGTT TTGAAAACCT TAGCAGCTGA ATAACCTAAA ATAACTGGCA TAAAATAAAA 4800

AAGTGCATCG GAAGCGGCGC TTAAAATAAT ATACGTATTC AAAGGTTTTA AATCGACCTG 4860

GTAAAAATTG TTTGCGACCA TCACAGCAAT AGCTAGAATT CCTTTTAACA TGCCTGAAGC 4920

GGCAATTGCT GGTACGACAG GTGTGAAAAT CGCCGCAACA GTATTTAAAA TTCGATTACC 4980

TAATGAATTT TTTTCTTTTG TAGATGTTTT GGTCGTAAGT TGCTGAGCCA ACAAAGGTTC 5040

TAGAGCATTA TAGACATGCG CGACTTTATT GCCTAAAACC ACTTGATACT GCCCGCCGCT 5100

TTCAACAACA GAAATGACAC CTTCTGTTTG TAGTAAGGCT TCTTTGTTCG CTTGTTTCGT 5160

ATCTTTCAAT ACGAAACGCA ATCGTGTATA ACAATGTGTG ACATCCACGA TATTGTCTTT 5220

GCCACCTACG TCTTTCAAAA TTTCTTTTGC TATTGCTTGG TAGTCCATGG TGACCACTCC 5280

TTATCTATAA AATAACAGCT CAATaAAAAA GACCAAATCA CACATGCACG TTTCAGCAAA 5340

CACACTCAAT ATTTGCTCGT CGTGAATGTC CCTTTGATCT GATCTACGTA ATAACTATGT 5400

TTCTCACGAT AGAATAGCTG ATTGAATTTA TTTTTAATTT ACCATTATTT AGAGGGATTA 5460

TCAATGAATT AGGTAGAGCA ACTCAGTGTT TTGTTACTAA AATCAGCAGT GAAAATAGTA 5520

GTTTACTACA TTCACTGCTG ATTTATTTGA TTACAAGACT GGGGACAACA AGCGTGAAAA 5580

GTTTTGCTTG AATCGTAGCC AGTACCCTTG TTTTTTAATT ATCTCGTCGG TTAGTAAGAC 5640

ACTTTTTTCT ATATCTTCTT CAAAGACTTG CGCTAACATC CACGCAATGT CGGGATTGTA 5700

AATAAACGTA CTTACCTCAA AATTCAATGA ATAACTGCGA ATATCTTGGT TTGTTGTCCC 5760

AACCATAGCT AATTCATCAT CAATCACCAT TGTTTTTGCA TGAATAAACC CACTATCATA 5820

AATATAAATT TTAATGCCGC GTTTGTGTAA GTAATTTGCG TAATATTGTG TCGCGCGGTA 5880

AATAAAGGGA TGGTCAGGCA TACAAGGAAT CATAATTCGC ACGTCGACAC CTGAACGAAC 5940

CGCCACTAAC AACGCATTAA TCATACTATC ATCAGGGATT AAATACGGTG TTTGAATCCA 6000

GACACTCTTT TCTGCAGAGA AAATCATCCG AACAAACCCG CTTTTCAAAA TTTCTTCTTC 6060

TGAATCAGGA CCATCTGAAA CAACTTGTAA GGACAAATGC TCCACTTCAT GATCTTCTGG 6120

TAATTTAAAA TATTTTTCAT GATACGTCAT ACGTTCTTCA GGATATTTGA CTGAAGCATT 6180

CCAATCCATG ATAAATATCT CTTGCAATGA AAATACGGCC GTCCCGACCA AACGAATATG 6240

GGTATCTCGC CAGTAACCAA ATTTTTCGGT GACATTCAAA TATTG 6285

(2) INFORMATION FOR SEQ ID NO: 416:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 787 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 416:

AAATTATTnG AnTTTCTTAC CCGCCATTAT CCATAATCCT TTAGTTCTCT CTTGGTTTCg 60

GGATTCGAAG TAGGCGTATC aCAACTACTA AAaTGGTTAG tATGGAATGG AGGACTgCAG 120

CATTATTTAC ATGTTTTACT TTAGCTAATC CTTTAACAAT TTTAACTTCT TTTATAATGG 180

CTCCAGTTGG CGCACTAAGT CCsGTATTAT CmGTAGGAAT GTTTTCAGCC CTAATGGAAG 240

CTAGTATAAA AAAACCTACG GTCAATGATT TTATGAATGC ACAAGACGAT ATTTCATCAA 300

TTAAATCGAT TTACAAAAAT AGATTGTTAA AGGTGGGTTT AATCTTTGTG TTAGCTAGTG 360

CAGGTGGGGC AATAGGTAAT ATTATCGGGG GAATAGAATT ATTTAAGAAT TTAATTTAAT 420

TAATAAACTA GGAGGAATTT AATTTATGAA GAAGAGAACG ATGCTTTCAT TTTTTTGTGT 480

AAGTGTGCTA GTTTTGATTT TAGGAGCTTG TAATGGAAAA GAAGGAACAA AAAATAACTC 540

TAAATCAAAA GAAGTTGAGC AAGTAGCTAC TTTTTCAGCT ATGACACCTG TTCAAAGTTT 600

AGACCCAGCA ACAGCAGTGG ATCAAACAAG TATTACATTG TTGGCCAACG TAATGGAAGG 660

GTTGTACCGA TTAGATGAAA AAAaTCmrCC mCmACCmGCm ATTGCmGCyG GGgCAmCCAA 720

AAttCCmATA AtGGTAAAAC ATATACAATT GTGATTCGAG ACGGTGCAAn ATGGGCCGAT 780

GGTACAG 787

(2) INFORMATION FOR SEQ ID NO: 417:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1092 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 417:

CTACGTTAAA TGAATGGAAA AAACGATTTC CGGACATTAA GGACACCATA nAAAGAGGTA 60

AAGAAGTTGT AGACCGCCAA GTGGAAAaTG CaTTGTTTAA AmGTGCGGTG GGTTACGArT 120

ACACTGrAAT TACTAAAGAG TTAACTGACT CgGCATGCAA AGTAACTAAA CGAGTAACAA 180

AACAAGTGGC CCCTAACCCA ACATCTGCTA TTTTCTGGTT GAAAAATAGA AAACCAGATG 240

AGTGGAGAGA TAAGAAAGAA ACTCAAGTTT CTGGTGAAAT GTCTGTTAAC AATCCTTTTG 300

CTGGTTTGTC TGAAAAAGAA TTAAGAAAGT TAGCTGAAAG TGATGAATAA GATTGTTTTG 360

GGTGCGAAGT TAGAATTATC CCGTCGTTAT TTTTGGGAAT ATTGCAAATT AACGGCATCT 420

GACTTTTATA AGCAAGACAG AGAGTACTTA AAAGAGTTAT GTGATGACTT GCAAGAATTC 480

ATTTATGATA GCGATAATGA CGTTCTAGTT ATCAACGAAC CGCCGAGACA CGGGAAATCG 540

AGAACTGCTG GTAAATTCGT AGAATGGTTG TTAGGGAATG ACACTCGAAA AAAAATAATG 600

ACGGGATCAT ATAACGAAAC GTTATCTACA ACATTTTCTA AAAGCGTAAG GAATACTATT 660

CAAGAAATAA AAGCTGATAA AAACAGAGTT GTATTTTCAG ATGTATTCCC TGGTGTAGAG 720

ATAAAGTCTG GTGATGGAGC CATGAACTTA TGGAGTTTGA CTGGCGGATA TAACAATTAT 780

CTAGCAACGT CACCAACTGG CACaGCTACA GGGTTTGGTG CAGACATTAT TATCATTGAT 840

GATTTAATTA AAAATGCTGA AGAAGCAAAT AACGCTATGG TATTAGAGAA ACACTGGGAG 900

TGGTTTGTTA ATACGATGCT ATCTCGTTAG GAACCnGGCG GCAAAATCAT CATCATTATG 960

ACCAGATGGG ATTCnAATGA TTTAGCCnGT AAAAGCTTAA AAGGATTGCC GCAATCCAGG 1020

CTATAAAGTA AAACCTTATT TAGCCTGGAA GGCATATTAC CGAGGAAACC GGCCACCAnG 1080

CCTTGTGGAG TC 1092

(2) INFORMATION FOR SEQ ID NO: 418:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3709 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 418:

GTTTTCGCAA TGTCCATGTT TCCCCAAGTG TCATACCAAA TATAGGTATG CGCACGTTTA 60

CCTTGGGCAA TAGCACTTGC TACTTGCGTT TTATATGTGT ATTGTTCATA AATACCGCTA 120

GCATTGTAGC CGCCAATCTG AGCAATGGCG AATTTATCAT GTGCATAACC AAAACGACCC 180

TGTTCGCCTT GATAAATCGC CCAGTCAACG CCTTGGTCAC CTTTTGCAGC AAATACATTT 240

AAAGGCATAA AAAATAGAGC GACAAGCGCT CCTACTAAAA TTTTCTTTTT CATTTTTACT 300

TCTCCTTGTC TTTTAAATTA TATGCTGACA CACCTGTTAC TACTCCTAAA AAAGTTGCAA 360

TGGCATTGAT AGTTAAAACA GCCATATCTG TTTGTTGCCA ACCGTAGGCT TTACCTAGCG 420

TGGCAATCAA AACAGaACTT GCAGGAAGCA CAGTAAGCAC TCCCCACTTG ATGATTTTGT 480

AATACTTGTC TGGTAATATC ATTTCTAAAT TCCTCCTAAG TATTTCGTGA TTAAATAGAC 540

AGCAACAGAA ACCCCAATTC CTGCAATTGT TCGCCACGTC CACTTTTGAT TCTCTTTTAT 600

TTCCGCAATA TCGCCTTCAT TGTTTTTGGC CATTGAGAGC GCTATGTCTG CTTTCTCTCT 660

TAATTGTTCA TGATTATCCA ACTTTGTTTC AATCCGTGCC AAACGATCGA CGATTTCAAT 720

TAAAGGATCA TCTTTCAAGT TATCGTCTCC ATCCCTCTAA CAAAAGAACC GCCTAGCTTT 780

TGCTAAACGG TTCACCTGTC ATTTTAGTAA ATTCATCTTC TGTAATACAA CTAGGCACAA 840

ATTCTGCAAC CTGCTCTGGA GTAAATAGTC CCCAATCATA CATCAGTTTA ATGTCATCAT 900

ATGAATACAT TATTTTGCAC CTCCGATTTG TTCTTTAATA GCATCAATTT CTCTTGTATT 960

TTGAAGAGAA GTAAGCATCG TCTTAGAATT GATTTGTGCA AGTGATTCTG CTTTAGCAGT 1020

TAGCTTTTCA TTTGCTTGTT TTAACTCACT ATTTGAAACT TCTAAGCCAT TAGCAAGATT 1080

TTCTAACAAA TTCAATTTTT TTGTATAATC CTGTGTCACT GCTTCTTCCC ATTTTTGTTC 1140

TGAAAAGTTA AAGAATTGGG ATTGTTCATT CACTAAATTT TCTAGTGGTT TCTCCTCTAC 1200

AAATGGCAAG GCTACAACAT AATCATCTTG AACTTCGAAA ATTTGAAATC CTACTGGGTA 1260

GAGCACTTTA TATATTTTTT TCATTTTATT ATCCACTCCT ATTTACGACG CATAACGCCG 1320

ATTTCTCTAA AAACAATTTT TTTATTTGCT TCATTTAATG TGTTTGCAGT ATCTCCTGTA 1380

ATCACTAAAG ATGTACCAGA TTTAGAAATA GATATTTTTT TATACGCTAT TGTTGGGTCT 1440

GTTGAAGCAG AACCTACCCA TGCAACTGCT GGTAATTCGT ATTTTAATTT CGTAATATTC 1500

GGAGTAACTG GTATTGAGTG AACAATACCG CCACCGTTAT CGTTGTAACG CGAGAATATC 1560

AAAACAATTT CATCCACATT GCTGCTATTT GAAATTGTTA CTGATTGAGA TTCTGTCAAA 1620

TAAGCTCCAG TTCCATCCCA GTAGTCTTTC ACCACCTCAT AGGTCGGTTT TGCATTAACC 1680

GATACAGATT TACCACCGAT TTGAATCCCA TCTTGAAAAT TCTTTGTCCC TAAAACAGTC 1740

TCATTTCCAA CGGCCTTTAC TAATTTTCCT TCCACGCCGT CAATAGCATC TGCATGTGTT 1800

TTCATATACT TTACAACACC ATTTTCTTTT AGCTGAACGA TATCTGCCAT TACGCTTCAC 1860

CTACCTTTTC AAATGTAAAA ACTGGTAATG CATCCAATTT TGCTTTATCC GTTTTAGACA 1920

TTAAACCGTC TTTTTCAGAA GTGGCATTGC TAGGAAGCGT TGGAATAACA GTTGTGTCTG 1980

GCAGTGCTTT TACATCAGAA GCAGTTAAAA TAACTTCACC TGTATGACCA TTTACAGACG 2040

AGACAGTGCC TGCTTCAGCA CCACTAATTT TTCCATCAAC AAATTCATTT AATCCAACAA 2100

CGCCAGCTGT ACTAGTTTGT ACATCAATAG CTACGCCGTC TTTTTTCACT ACATATAAAT 2160

CAGGCATTTA TTTCTTCATC TCCTTTTACT TTTTCAAACT CAACACCAGA ACCACCTAGT 2220

TTTCCATCTT CATAATCGGC TATGATTTTT AACATTTTGT CATACTCCTG TTTCGAAATC 2280

ATAATCCCAT CAATAGGTAA ATCTAGGTCT GCACGCGTAA TAATGACTGC GCCTGTATGT 2340

CCATTTACTG AAGATACTTT TGAATTACCA GCCATTATCT CTGTTAATCC AAGGATTGCC 2400

GATACGTGTG TCATAGGAAA AAACTGACGT TTAATCCCAT TTTCATCAGT TTCCATCATT 2460

CTTTTAGCAT CAACCATTGT CTACACCTTC AATCGTAAAA ACATTTTGTT TTGAATCATC 2520

AACTGTCGCT ATAATTAACG CCCCTTCTTC aATTGGATGA TTAACTGTTC CCAATACTTC 2580

AACTTCATGA TTCTTAGAAA ATGAATCATC CTCTAAAATT TCTAACGTGT TTACATTACC 2640

GTATTTGATG GTGTATAGCC GTTCCTCTAA CCGATGATAC AAATAGTTCA TATCAGCCAA 2700

TAAACGCTCA GAAAGTGAAT TATGACGTAC TCCTTGTATG TCTACACGTG CATCCATTAG 2760

CTCGGCTAAC ATTGTTCCGC CTGGATCAAT AGTTTTTAAA ATATCTTTGA TTGATTCGAA 2820

CCATTTTAGA TAATCTGTTT CTTGGCCGTT CCGCCAAGCT TCAAAAGTAT CTTGTTGATT 2880

TTTACGCCAT TTTTCAAACT CTTCTTTTCT AGCGTTCATC CACGCTGTGA AATCGCCCTT 2940

GTTTTCATTA ATAAAAGCAG TCATATCAGC TATCAAATCC TCGATTGACT GCCAATATGA 3000

CCCCATTTCA CCTTCTGTTT TAGAAGCAGC ATTCACAACA AAGTAAGAAA AGTTCTGCGT 3060

TGAGCCAATT AGATTGTCGC CTTTATGAAT ACTGAAGTAT GCTTCCTGTC TATGCAATGA 3120

CTGCATAGAA TATTCATCAA AGGTATACTG GATAATCCCT TTTTTGGCAT TCACAATTTT 3180

TGCTGCTCGT TGAATCGGAT ATTTATTATC AATAACTGAT TCAAAAAAAA CTTCGCAACC 3240

TGTTAAATCA AGTGGCAAAG CATTTTCAAC TAGTATGGCT TCTAAGACTT CTGTGTTTCG 3300

ATTTCCTTGT CGTACATTCT GAATCCCAAT GTAATTGTAA GGTTCAGTTG TACTTAGCGT 3360

TGCTTGCCAT TTAACCATTG AAAAATCCTC CTCTCGTTAT TTTGGTGGAA TAACAATCGA 3420

TTGAATAGAA TTAGCAAAAT ATAATCGGTC ATATTTTGCG ACAATTTGTC CTTGCTCGGC 3480

ATTCTGTTCT ATCGTTTGGA TACGTCCGTT ATTTAAGCCG TAAATCACGC CCGTGTGACC 3540

ATATGTTGGG TCTACTGTCC AACCTGgTTC CCCATTGGnC ACTCGTCTAA TATTGACGAT 3600

GCCCCTACnA CTAAAnCTGA TACGTGGATT GGATACTCGC CAACTACGnA TCCAATCAAA 3660

GCTCACCATA TTGCAGCAGA GATGATCACC ATACAGTGAA AGCATAATG 3709

(2) INFORMATION FOR SEQ ID NO: 419:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1649 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 419:

AAAATAAATT CCCTGTAGGT TCATCTGCAA AAATAATTTG CGGATTATTA ACTAAAGCAC 60

GCGCTATACA TACCCGTTGT TTTTGGCCAC CAGATAAATT ATTCGCTTTA TTTTGAACTT 120

TATCTTCCAA ACCGACTGCT TTTAACGCAT CTAAAGCCAT CTTTTTCCTC TTGCTACCGG 180

AAATCCCACC AATTTTTAAT GGTAAAAGAA CATTATTTAA AACTGTATCC TTGGCATTCA 240

TAAAGAATTG TTGAAAAACA AAACCAAATT CTTCATTTCT CGTTTTATTC AGCACTTTCT 300

TTCTAATACT AGTAACATTT TTACCATTTA AATAAATATC TCCTGAAGTT GGTTGATCTA 360

ACAATGCCAA AATGTGCATA AATGTTGATT TCCCAGAGCC ACTTTTACCA ATTATGGCCA 420

CAGATTCACC TTTTTCAACT TTTAAATCAA CACCTTTTAA CGCGTCAAAT TTTGTtTCGT 480

tCTTTCCATA ACTTTTTTTG ATATTTTTtG CTtCAATTAC CGCCATGTTA CCCCTCCATT 540

TTTCATTCTT TGTCGTTTAC TTATTTACAA AGTTTTTTAC TCTTTTCATA CTCCTTTCTT 600

CTGTTTCTAT TTTGACGGAA GCCCTTTCCT TTGTCATGGT ACTTCCGATG TATCTCTCAA 660

TTAAAACAAT AAGAACATGT TCCAAATTCA CCTCATTGAA CATGTTCTTA TTAATTAGTT 720

ATGCATTTAC GCGTTTAATT TCACGATGTT TATTCAATAA ATAGGCCAAG CCAAGATTAA 780

TACCAACAAA GCAAcTGATT GTCAGAAGAA GTGGTAGGAC AGTAAAACCA GTGGATTGAT 840

AACCAACCAT TGTTTGCAAG ATGTGGAGCA TTCTCAGGCG CATTGGTGCC CCCCCTAATT 900

GTGCAATCGT GCCCAGTAAA ATAGGAATCA TTAATAACAG CAATAAAACA ATGAGTTTTT 960

TTACGCCATC TATCCGATCA TTAAAGACAC CTGCTAACAA GCCAAGTGAG CCACTAAGCA 1020

TTAGTAAGAT AAAAAAGAGC AACCACGACA TGAAAAAGTT TCCTTGTGCA TAGACATCAA 1080

TTAATTTCAT TGaTAATTgA AAATGAGAAA TGAAATTGcC GCTAAATACT TTAATCAGTA 1140

CAAGTACAGC CAAGGrTCCA ACCAAAGATA rAaTwGCATT aCTTACmAAA TTCmCCAAAA 1200

AAATTGTCCA GCGGGAAAGC CCGTTTTGAA TAAACAGTTT AAAATCTGTG TTCATCCCCA 1260

AAAATGATAG TATCCCCATA AAAACCAGAC AAGGAATTAC TGCATCAGAG CTGACAGTGT 1320

TCACGTCGTT AGAAAAGAGC AACCCAATAA GAGGAAAAAG AATGCCGAAC AAAGCmTAAA 1380

AGCCAAAATA AATTGCTAAG GrACGGrCTT GATAAATAAC ACGGTACCGC AAAGCTGTTT 1440

TAAATTTCAT TTTACTTTTC CyTCTTTCTA TTGGTTAATk GTmCAAAATA TGTTTGTAAA 1500

TTTAGCGGCG CAATTTGAAC CTCTGCTTTT TCAGTCGGCA AGTCGCCGTA CACGTAAGCC 1560

GTAACTGCTC CACCTAAAGT GTCCATGCCT AAGATTTCCA ATGACTGAGT ATAGCGCTCT 1620

ACTnGTTCTT TCGGCCCnGA AACAATGCG 1649

(2) INFORMATION FOR SEQ ID NO: 420:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4114 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 420:

AGGGTAGCAC AGAAAAGAGG GAGGTATGAC AATGGAACAT CATCAACTAA AAGCATTTCC 60

GCAAGATTTT TTATGGGGCT CTGCTTCGGC AGCTTATCAA GTGGAAGGTG CTTGGAATGA 120

AgACGGCAAA GGCGCATCTG TTtGGGGATG AATTTGTTCG GCTACCAGnA AAAACGTTCA 180

AAGAGACAAC TGGCGATTTA GCCGTCGATC ATTATCATCG TTTCAAAGAA GATGTGGCTT 240

TAATGAAACA ACAAGGTTTA AAAGCATATC GTTTCTCGAT TGCATGGACA CGAATTTTAC 300

CTGAAGGGCG TGGTCGAGTA AATCAAGCGG GACTGAAATT TTATTCAGAT TTAATTGATG 360

AGTTATTAGC GGCAGGAATT GAACCAATGG TAACTTTGTA TCATTGGGAT TTGCCAGCTG 420

TTTTGCAAAA AGAATATGGT GGCTGGGAAT CAAGAAAAAT TATTGATGAT TTTGTGGCCT 480

ATGCGAAAAT TTTATTTGAC GCTTTCCGTG GCAAAGTTCG CTATTGGATT AGTTTAAATG 540

```
AACAAAATGT GTTTACTTCG TTAGGTTATC AATTAGCCGT GCATCCTCCC GGCGTAACAG      600
ATAATAAACG GATGTATGAG GCCAATCATA TTGCTAATTT AGCCAATGCA GCAGTGATTA      660
ATGCGTTTCA TGAATTAGAA ATGCCTGGCA AAATTGGCCC TAGCTTTGCT TATACGCCTG      720
TTTATGCCGT TGATGCGCAT CCTGCCAATG TTTTGGCGGC AGAAAATGCA GAAGATTTGT      780
TGTCGCATTT TTGGTTAGAT GTCTATCTTT GGGGAGAATA CCCAATTGCC GCATTAAATT      840
ACTTACAGGA GCAAGGTGTG GCCCCAACGA TAAAAGAGGG CGATTTAGCG TTACTACGTT      900
CAGCCAAACC TGATTTTCTA GGTATTAATT ATTATCGGAC GGATACGGTT GCTGCTAATC      960
CATTAGATGG CGTCGGAATA GGGAAGATGA ATACAACGGG TGAAAAGGGT TCTGAAACGG     1020
AATCTGGGGT ACCGGGTTTA TTTAAAAAAG TAAATAATCC ATATGTTGAA CGGACGAACT     1080
GGGATTGGGC CATTGACCCG CAAGGTTTGA GAATTGCCTT ACGACGATTA GCTAGTCGCT     1140
ATCAAGTCCC TATTTTAATT ACTGAAAATG GGTTAGGCGA ATATGATACC TTAACAGAAG     1200
ATAAACAAAT TCATGATACG TATCGCATTG ATTATTTGCG AgCCATaTCC AAGCGATTCA     1260
AGAGGCGATT ACGGATGGTG TTTCAGTTAT TGGTTATTGT ACATGGAGTT ACACCGATTT     1320
ACTGTCGTGG CTCAACGGCT ATCAAAAGCG TTATGGTTTT GTCTATGTGG ATCAAGATGA     1380
AACACAAAAA GGTTCCTTAG AAAGAATCCC GAAAGACAGT TATTATTGGT ATCAAAAAGT     1440
CATTGAAACA AATGGACTTA TCTAATTATA AGAAAAGAGT GTGGAAAATA TGACAGTAAA     1500
AATGGGGTTT ATTGGATTTG GTAAAAGTGC CAACCGCTAC CATTTACCTT ACGTAATGAT     1560
TCGAGAAACA CTAGAAGTGA AAACGATTTT CGATTTACAT GTTAATGAAA AAGCAGCAGC     1620
GCCCTTTAAA GAAAAAGGCG TAAATTTTAC CGCTGACCTT AATGAATTAT TGACGGATCC     1680
AGAAATTGAA TTGATCACAA TCTGTACACC TGCGCATACA CATTATGATT TAGCCAAACA     1740
AGCCATTTTA GCTGGTAAAT CAGTGATTGT GGAAAAACCA TTTTGTGATA CGTTAGAACA     1800
TGCAGAGGAA TTATTTGCTT TAGGGCAAGA AAAAGGGGTA GTAGTGATGC CTTATCAAAA     1860
TCGACGTTTC GACGGCGATT ACCTAGCGAT GAAGCAAGTA GTAGAGCAAG GCTTTTTAGG     1920
AGAAATCAAT GAAGTGGAAA CACATATTGA TTATTATCGA CCAGGAAGTA TCACGGAACA     1980
AGGACCTAAA GAAAATGGTT CGTTTTACGG ACTAGGCATT CATTTAATGG aTCGCATGAT     2040
TGCTCTTTTC GGGCGTCCAG ATCAAGTCAC TTATGATATT CGTAATAATG AAGTGAGCGA     2100
GGCAGTCGAT AATTATTTTG ACGTTGATTT ACATTATGGT TCaAAATTAA AAGTAAAAGT     2160
TAAAACGAAT CaTTCTGTCG CTAGTCCgTA TCCaCGTTTT ATTGTTCATG GCAGCAATGG     2220
CTCGTTCaTC AAATATGGTG AAGATCAACA AGAAAATGAT TTAAAAGCCG GAATCATGCC     2280
AGATGCACCA GGCTTTGGAG AAGATTCACC AATGTATTAT GGAGAAGTGA CCTACCGTAA     2340
TGGCAATGGC GACTGGATTA AAAAACAAAT TAAAACGCCT GTAGGTGATT ACGGTCGTTA     2400
TTATGATGCC GTTTATGAAA CGTTGAAAAA TGGTGCACCG CAACTAGTAA CTAAAGAGCA     2460
AGCATTAACT AACATCGAAA TTTTAGAAGC AGGTTTTCTT AACCCAAGTC CAAGTGTTTA     2520
```

TCATTTGAAA GAAAACTAAA GCGTATTTAA ATGACTTGAA AAGsGCCTGC CCTACAGTCC 2580

AGACAACTGT GGGGCAGGCG CTTTTATTTA TAAAGAATTG TGAATTTTAA TATAAGAAAC 2640

GTATATTGTT AGTAAAAATA AaAAAAGGAA AGACATCTTA GTAGCACTAC TTCCCCAAGA 2700

ATTGCTACTA AAATGTCTTT TGTGATGCTC GCTCAATTGA AGAGCCTTAA ATAGGATATC 2760

GAATATCCAG TATAAATGCA AGTAAACTAC TTTATAGTTA ATATTAGTAT ACTTCATAAA 2820

ATAATTTTTC AAAGAGAAAA AGGCTTGTTA TGAGAAAAAA TTAATTTTTA ACTCGTTTTT 2880

TCTAAGCAAG ATTCGGGTGA AAAAGACAAT AATTACTGAA GCTGAATGGT AATTAAAGAG 2940

GTTCTTTGCT AAAATAGGAA AGACAAGAAA GTTTGGTTAA CGAAAGAACC AAATCAGAAA 3000

GGAGGACAAC CAGTGGAGAG TTGGCTTTTT TTACTACTGA TTGCCTTGAT TGCGTTTGTT 3060

GCAAAAAATC AAAGTTTATT AATTGCGTCA GTTGTTGTGC TTGCTTTAAA AGCTTTACCA 3120

AATAGTGCGA AAATCATGTC GTGGTTGTCA GATAAAGGCA TTAATTTAGG CGTCACAATT 3180

ATTTCAATTA CCATTTTAGT CCCGATTGCT ACAGGACAAA TTGGCTTAAA AGATTTAATC 3240

CAATCATTCA AAACACCGAT GGGATGGCTG GGGATACTTT GTGGAATTCT GGTGGCTGTT 3300

TTATCCTCAA AAGGAGTAGG TTTAATTAAT CAATCGCCTG AGATTACGGT GGCCTTAGTT 3360

TTTGGAACGA TTCTTGGCGT CGTTTTTTTA AAAGGAATTG CAGCTGGTCC GATTATTGCA 3420

TCGGGGATGA TGTACGTAAT CATTACTACC TTTCAAGCCT TTCAACACTA AAAAAGCAAA 3480

CAAAACGCGT AAACGTTTTG TTTGCTTTTT TAGTGTTGAT AAATGTCCAC ATTACTATCT 3540

AATATGCCAC TTTCATTTTT TAAAGTAATG GTTCCCGCAT TCGGATGAAT CACTAGCTGT 3600

AAATGCCCTG TTAAAAACAT TTGGTCATAC GTCGAAGTAT AGACGTGAAA AGAATCATTA 3660

GCAATTGTTT CGGTATAAGG CGCTTCTGCT TGATCAACAG GAACGATAAT GAAACGATCG 3720

CCAATCATAC AACAACCTTC TGCGGAAGAA TAAGGATTAG AGCCGTCATT TGCTACCAAC 3780

AAGAAGGTGT TCGTTTCAGT TAAGAAAGGG GCTAAACGTT GTTTTACTTC TTCAGAAATA 3840

GTCAGAGACA TTTGCCTCAC TCCTTTAAAT AGTATGTAAA AGGAAGAGGG TAGCTTGTCG 3900

CTATGCCGCT CTTCAAAGGT TTTGTTATTT GCTTGTTTCT TTACTGCTTT TTTTGAAAAG 3960

AGCAATAGAA AAGGCGACAA CTAAGACAAC TGTTGCAGCT AAAACAAATG GATTTTTTGT 4020

TGCTGaTGTG AACCCTAGGa CTGTTGAAcC AATCGCACCT ACGTAAGATA AAATAATCGC 4080

TGCTTyCTTt GGkGTTGAAA GTTCTGTCAA AGTG 4114

(2) INFORMATION FOR SEQ ID NO: 421:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2285 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 421:

```
ATGnAGGCCA GATCCCCTCC AAAAAAAATG GTGGCCnnnT TACATTACCC CGCGGCATTT     60

GCCTTTGACT ATCAGAACTG AAGAACTTCA AGGAATTTTA GTCCTGTTGC CGACTCACCC    120

ACGGACATTG GAAATCGTAA ACATCGCTGT CTCAGAAGAA AGCCGAGGGC GGGGCATCGG    180

GCAAGAATTA TTACATTTTG CCATTGATTT TGCCAAAAAA GAAAAGTATG ATTCCTTGGA    240

GATTGGGACT GGGAGCACAG GCTTTCAACA GCTATATCTC TACCAAAAAG TCGGTTTTCG    300

AATGACACAT ATTGAACCAG ATTTTTTTAT CCATCATTAT GATGAGCCTA TTATAGAAAA    360

CGGCTTGCCT TTAAAGGATA TGGTTCGCTT ACGGTTACAT TTAAGTTAAA AAGAGGTCGG    420

GACAGAACCG TTTAGCTCCG AGAAATAAGA AGAAATTTCC GAAAATTGCT CTTTAATTTT    480

TGGAGAATTT CGGCTTATTT CCGAAGaGTT GGTTCTGTTC CCGCCGTTTA TCAGTTTTTG    540

AGCGTGGAGC AAAAATCCAA AGTGATTTTT GTCCCACGCT CTTTTTAATG CCCAAGAATT    600

GGTAACAACT CATCTAGCTG TTGAATCGTA TAATTGGGTT GAATGGTTGG TGTTGCTGGA    660

GCGGAGGTCG GATTGAGCCA AACAGTTTGA ATTTTCGCTT GTTGACCTCC TTGAATATCT    720

GAATGCAAAG AATCGCCAAT AATCATTGTT TTTTCTCGAC TAAAATTAGG AATCCGTTCA    780

AAAACAAAAT CAAAATACTC TTTCATCGGT TTTTGATAAC CCACTTCTTC TGAAACAAAA    840

ATATCTTTAA AATAAGGCAT TAATTTCGAA TCTTCCAACC GACGATACTG TGTTTTTGCC    900

ACTCCGTTTG TCACAATATA TAAATCATAA TGGGGTGCTA ATTTTTCCAA GATACGTTGA    960

CTATTGCCAA GCAAGTCATG TCCTTGACTT AAATATTCTC GGTAGTTTTC TCCCATTTTA   1020

ACGCTATCAA CTTCTCGATT AAACGCGTCA AACAATAAAC CAAACCGTGT ATCAAGGAGA   1080

ACATCTTTTG TTATTTTTCC TTCTTCAAAT GAACGCCATA ATCCTTGATT AATTCGGTGG   1140

TATGTTGACT CGACTTCAGA TGTAAATGGT ACATTTTCTT CTTCAAATAA GGCATGTAGC   1200

GCCTTCTTTT CAGCCAATTT AAAATCTAAC AATGTATCAT CAACATCAAA AAGCAATGTT   1260

TTCATTTCCG TTCCTCTTCT CTTTTCAAAT TACTCGGTAA CGCTTTGCTT GTTTGTTGTT   1320

TTTCAATAAA AAAATAAGGC AATAGCAATT GTAACTCTTC TCTTGTTAAT GAACGATAAG   1380

CACCTAAAGG TAATTGCGGA TCTAGCCGCA ACGGCCCCAT AGTTTGGCGT TTTAAAGCAG   1440

TCACTTTTTT CCCCACGGAT AAAAACATTT TTTTCACTTG GTGAAATTTT CCTTCTTGAA   1500

TAGTCAGGAA AACATGACTC TCCGTTTCAG TCGCTGCAAG AACAGTTATT TTTGCTGGCT   1560

TACATTGAAT GCCTCCTTGC AAAACAATCC CATCTTTAAA TTTAGCACAA TCTTCTGCAG   1620

AGACCAAACC ATTGACTTTC ACTTCATAAC ATTTCGCTAC TTCTTTATTT GGCCGAATTA   1680

ATTGATACCC TAATTGTCCA TTATCTGTTA ATAGAAGTAA CCCTTCCGTA TCGCCATCGA   1740

GTCTGCCCAC TGGATAAAGA CCAGGGCGTC TGTCTTGTGG CGCAATCAAA TCAATAACTG   1800

TTTTTTTAGA TGCATCTGAG ACAGCGCTAA CCACACCTTT TGGCTTATTC AACATATAGT   1860

ACACATGGGT CTTATGCGAA ATTAATTGTC CTTTTACTTT AATCATTTGT AATTGCGCAT   1920

CGACATTTTG GCTTTCATTT AAAGTCACGA TGCCATCCAC AGTGACCTGC TTGCCTTTGA   1980
```

TTAATCGTTT CACCGTTTTG CGTGATCCAA ACTGAAGTTC TTCTAGTAAT TTATCTAATC 2040

GCATTGTTCA GCTCCATTCT GGTATAAGTA TAGTGGAATC CGTTAAAGAT GAAAAGTGGA 2100

AAAAGGTAAT TCCTTGACGT ATAAAAATAT AAGTGTTACG GTTAATAAGT AAAGTAAATA 2160

AAGGTTcTCA aAAACAGCGT AATGACCTGG GAGAGAACac GATAACatGA ACCATTCCCa 2220

TGAGTGTTGC CTGGCAATAT TTGTGGGGAT TTTGTATGTA CATTTCAGCA GACACGAAAA 2280

GGTCT 2285

(2) INFORMATION FOR SEQ ID NO: 422:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2974 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 422:

ATCGgTCAtT GAAAcGGTTG ATCAGrATGa TCAAACmACG CTGTGCCTAA CTCAGATTCC 60

ATTTTTTTCA ATGCTAAAGA AAGTGTTGGT TGGGTGACAA ATAAATTGCG AGCGGCCGCC 120

GATAAACTGC GGGTATTAAC AATTTCaATG AAGTACTTTA AAAGCTGAAT ATTCATCAGA 180

AACTCCCTTT CTTTTTAAAA GCGCTTACAC AAATGTGCTT GGATAAAAAA TAACAAATAG 240

ATCACTTATA ATATAACATA TGTTTTGCCT TTTTATTGCT ATAAAAGTAA GAGATAGAGA 300

TTTTCTTAGC AGATTTTTAC CTAAAAATAA AAAATTGAAC GGAGGAAAAG AAATGAAGAT 360

TATTATCGCA GGTGCTGGCG CTATGGGCAG TCGCTTTGGA CTAATGTTAC ATCAAGCAGG 420

AAATGATGTT ATTTTTGTCG ATGGTTGGCC AGAACATATT GAAGCAATAA GAAAAAACGG 480

GTTAGTTGCT GATTTTAATG GCGAAGTGGT GACCGCTAAA ATTCCTATCT ATCCTCAGAC 540

AGAAGTAGCA AGCATCGATT TTACAGCAGA CTTAGTTATC CTATTTACAA AAGCAATGCA 600

ACTGGCAGGA ATGCTTGAAG CATTAGAACA TTTCCTTAAT AAAAATACGA AAGTTTTATG 660

CTTGTTAAAT GGAATTGGGC ACGAAGATAT TATCAAAAAA TACGTGCCTT ACGAAAATAT 720

TTTATTAGGA AACACAATGT GGACCGCTGG GTTAGAAGGA CCGGGCAAAG TAAAATTATT 780

TGGTAATGGT TCGATTGATT TACAAAATTT AGGGGCGCAA AGTGAACAAG CAGCAGAAGC 840

AGTCATTAAG GTATTAAATA ACGCTGGCTT AAATGCTCGT TATTCTGAGA ATATTTTATC 900

ATCCATTTAT AAGAAGGCCT GTGTGAATGG AACAATGAAT GGTTTATGTA CTTTGTTGGA 960

TTGTAATATG GCAGACTTTG GCGAAACAGA GCAAGCAGAT CAAATTGTTC AACAAATAGT 1020

TAATGAATTT TTAGCAGTCG CTGAAGCGGA ACAAGTTACG TTAGATAAAC AAGCAGTTCT 1080

TCAACAAATC CGCTCGTGTT ATAATCGAGA AACAATTGGT TTGCACCATC CATCGATGTA 1140

TCAGGACTTA ATAACCAATC ATCGTTTAAC CGAAATTGAT TATATCAATG GAGCAATTGT 1200

TCGCAAAGGG AAAGTCTATG GGATTGCTAC ACCGTACTGT CAGTTTTTAA CAGAGTTAAT 1260

TCATTGTAAA GAAAATTTAC TTCATGCTAA ATAACAAGTA AAAACAGAAA TACTCTTGAG 1320

TCGGTCTTGA GAAGAGCTTG ACCTACTTCT TTCTATACTT TGCCAGTGTA CAAGGCTGAG 1380

ACAAAATCCA AAGTGAGTTt GGCTCAGCTT TTTTGTTTTT TAAAGAACAC TGCAGAAAAT 1440

AAATCAAATC CGTGaAAmAT ATGTTAAACT AGAGAAGGTG TTTTTTCGCA AGAnTCTACG 1500

CGGaACACAG GACTCGTTGA AGAATTGAAT CAAAGAAACA ATGTGTTAAA GAAGGGATTT 1560

GAACAGATTA TGACCATTCA ATGGTTTCCC GGCCATATGG CCAAAGCAAG AAGAGAAGTT 1620

TCTGAAAAAA TCAAATATGT CGATATTGTA TTTGAACTAG TTGATGCGCG GTTACCGCTT 1680

TCTTCCCGTA ACCCAATGAT GGACCAAATT GTCCAACAAA AACCACGATT AGTCCTTTTA 1740

AATAAAGGGG ACTTAGCAGA TAACGACCAA AATAAAAAAT GGCAACATTA CTTTCAAAAG 1800

AAAGGGTACC AAACCCTAGT TATCAATGCC CAACAAAATA AAGGAATCAA TAAAATTGTG 1860

CCAGCAGCCA AGGAAGCGTT AAAAGAAAAA TTGGCTCGTG AACAGGCGAA AGGCTTGAAA 1920

CCGCGTGCGA TTCGCGCCAT GTGTATTGGG ATCCCCAATG TTGGTAAGTC AACCTTAATG 1980

AATCGTTTAG TCGGAAAAAA AATTGCGCAA ACTGGGAACA AGCCAGGTGT GACAAAAGGA 2040

CAACAATGGT TACGCTCTGG TTCCCAATTG GAATTACTGG ATACACCAGG GATTCTCTGG 2100

CCAAAATTTG AAGATGAAGA AATTGGAAAA AAATTAGCTT TAACTGGCGC TATAAAAGAT 2160

CAATTATTGC ATCTAGATGA TTTGGCGATT TATGGATTAG AATTTTTTGC ACGTTTTTAT 2220

CCGCAACGTT TGACGGAACG ATATGGTTTA ACAGAGGAAG AGCTATTTTT ACCTGCCCCA 2280

GAACAGTTGA TGTTAATTAG CCAAAAACGT GGTTTTAGAG ATGATTATAA TCGAGCCAGC 2340

GAAATGATTA TTTTAGAAAT TCGCAGTGGA AAATTAGGTA CGTATACTTT GGACCGCTGG 2400

GAAGAACTAG GAGACGAATA ACATGGCAAA AGAGTCGATT CAGGCAATCA AAGAAGCGTT 2460

GTTAACTGTC ACTGATGCAA CCGATGAACG ATTAGCAAGC TGGCGAGAAG ATGAACGTTC 2520

AGGAGTGCAA CAAGCGATCA AACAATGGGA ACGTCGCCAG TTAGCCAAAG AAAAAGAGTG 2580

GCAACTTTTT AAAGAAATGA GTCAATTTGA GGAAGCGGCC TATAAAAAAG GCCATCGTTT 2640

AATTGCAGGA ATTGATGAAG TAGGCCGCGG ACCATTAGCT GGCCCTGTTG TGACAGCTGC 2700

AGTTATTTTA CCGAAAGATT GTCAGTTATT AGGCTTAAAT GATTCAAAAA AACTATCTGC 2760

TAAAAAACGG GAAaCGTTAT ATAACCAGAT TCmAGAACAG GCTGTCGCTA TCGGGTTAGG 2820

TATCGCTGAT CmAGGAGTCA TTGATCmAtT AATATTTATC CAGCCACTTA ACCAGCGATG 2880

GAATGGGCAT TGACGATTTA GCTTTTTACC CnGATTATTT ATTAATTGAT GCCATGCnGT 2940

TAGCGTTCCC CCACCnCAGA AGTTTGATTA AGGG 2974

(2) INFORMATION FOR SEQ ID NO: 423:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3078 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 423:

GCCCTGTTTG ACGCTCTATA GAnCTTTGGC TCTCTTGTnA CGTGTCCATA AAACCTCCTT 60

GGTGTGAAAA GAGCCGCAGC TATTCCGTCT GCGGCTCtAk GTAAGTTTaC ATCaGCCCtT 120

AAgCGGTAGG cCAGTTTGTG TCCGCCCCGC TGCTGGCTGT TGTACTTTTC CAGAATGACG 180

CGGGCAAAGC GCAGGGCGGC TTTATTGGTG GAGAAGTCCA TCTTGCCGCG CCGGATGATT 240

TCTTCCGGGT CAACCGCTGA CAGCCGCTTA ATGAAAATGC GGTCGTTCAG TTCGGTTTCA 300

TAGGTTTTCA GAAACAGCGC CATGCCGGAG AGCATCGAAG CACGGAGGGA AATGGATGCG 360

CCGTGCCAGG TGCTGTCTAA GAGTACCAGC AGGCGGGAAA AAGCTGTGCC GCCAAGCAGA 420

CGGTAGGCGT TGATGATGGC GCGGGTAGAA AGGATTTCGT AAGCGTCGCC GCTCTTTTTA 480

TCCAGTATCC AAACAAAGCC CGCAGCTTCC ACTAGCCGCT TGATTTCCAG AACCTCCGCA 540

TCCGTGCCGG ACTCCAGCAG GGCTTTTGTG GCGTGGGACA GACGAAGCCG CCCTTTGGAC 600

TGATCCAGCT TAAAATACAT CTCGGCTTCC TGCTCATAGG TCATGCCGGT GTAAATCAGG 660

CAGGGGACGG TCACgTTgCC GCCGTCTGCC ATTTTCCGCA TTGnGGCGAt CCGGTGCTGG 720

CCGTCCACCA CATTAAACTT TCCGTCGCGG AAnCTGACCA CCAGCGGGGT CaGgGCGGTC 780

GTCCCATTTG GCAATCAGCC GGTCAACGTC CTCCGGCTCC ACGGGCCGCT GGTAGGGCAG 840

GCCGGAAGTA AGCTGGTTGC TGGATAAGTC GCGGGGGACG CCGGGGTTGG CGTACTGGAT 900

TTCTGTTAAA AGGGTTTGTT GATCCGTCCG CTCGTTACCT GGCGGACGGC GTTTTTTTCG 960

ATTGCTCATG TCGTTATTGT CCTTTCCATT TGATTTAATA ATTCCTCCAT CGCAGTAGAT 1020

ATGCTTTCAA AGCGCTGCCG GACAAATTCA AACTGTACCT TGGAAATGCG CGGGAACACG 1080

ATAGTACAGA ACGGGTCGTG ATACCATGCG AACCCTTGTG AAAGGTCTGC ACAAAGCCGT 1140

CGATGTCTGC CAGAAGCGTG TCCGGCGTGT AGCTGCAATC CTTATCCGGG TTTTTCAGGT 1200

CGGCTATGGA TTCCTCCAAA GTGTCATAGT GCTGATCGCy cwGCGTATAG GGAACGGAGG 1260

GCTGCGAGGG AAGTTTTGTA ATCGGCTCTG AGGTCTGCCC TTGTTCCTGC ACCGGTTCCG 1320

TTTGTTCTGG TTTCTGTGTC AGTTTGGACT GTTCTGGTTT CTGCGTTTGC TCAGCCTGcT 1380

CCGGGATTGG TTCGGGCTGT TTCGGTTCCG GCATTGGCTT GGGGTACTTT GGCTCCTGCC 1440

TTGGTTCCGC CTGTTTCAGT TTCGGTGCTG GTTTGGATTG TTCCAGTTTC CGCGCCGGTT 1500

TGGGAAGTCC CGGAGGTTCT GCCGGTTGAA ACTCGTCTGC GGACTTGATT TTACCAGCCG 1560

CCAACTGTGC GGCGGCGTGT TCCTGCTGTT CCGGCTCCAT GCGGGACAGC TTCAGCAAGT 1620

CCTTTTTCTT GATTTTGCGG TCGGAACCCT GAATGACTTC CTGCGCTTTk GGGGTtATAT 1680

CCCTGGCAAT CTGGATATTT yCTtCCACCG AACGGGCGGA AATGCCCAGC TTTTTTGCTG 1740

TGTCCCGTGT AAAGGACTTT GACGTGGACG CGCCGCGCTC GCTCACGTTG TTGCCCTGGG 1800

CGCGGTTCAT TGCGTTTGCC TGCGACACGC CGCACCTTGT TTCCGGATGC AGGATTTCAT 1860

AGATTTCTTT CCGCCGCATC AGCAGCTTTC CAAACTCCAC CGTTTCCAGA TCGGTGCGTA 1920

CAAAGTTTTC GTCAATCTCA GCAAGTTCCG CCTGCAAGCC CTCCAATGTG CTGATATGGC 1980

ACTCGATTTC CGACCAGCCA AGCAGCTTTG CCGCTTCCAG CCGGTGCAGT CCGGCAATCA 2040

GGGTATAGTC CCGGTCAACG GTGATCGGAT TTAGCAGCCC TACCGCCGAA ATACTGTCCG 2100

CCAGTTCGCG CACGGCTTCC GGCAGGGCTT CCCGCCGCCC GGAATTGACC TGGATTGCGT 2160

CAATGTTGAT CTGCATTTTC ATCACCTCTT TCGGCCCTGG CCGCCTTATT TGTTTTCTCC 2220

TGCCGGATGA AACACGTTGC GCCGCAGGAA TACTACGGCG TAAGGTCTTG CATCCCGCAG 2280

GAAAGGGAAA GCCCGCGCCG GTTTTGTTTT GCGCCCTcAT TCTGTGAGGA AATATGACCG 2340

GCGTGGGTTT CCGCTTACTG CATTTCTTAA AAATCTGTGC CGTATTATCA GGCGGCATGA 2400

CCGCCTTACT GCCCACCCCC GGCACGGGAA TATGTAAGCC CAACATGGCA CGTCGGCGCG 2460

ATACCAACAG ACGGCGGCGG CAGGTGACGA AGCTGCCCaT ACCAATCCCT CAGTATCTCC 2520

CTTTTACCCT GCGATACGGG TTTTTGCGGT GCATCCCGCG TCTGCGGCGC TGCCATGCGC 2580

CGGACAGCCA TGCCGGTATC GCTCGTGCCT TTGATGTGAT CGGTTTGACG GGCAGGGACA 2640

GAGGTTTACA CCCCCTTTCG CTGGCCCTGT CCGCAGGCAG TCCTGGCGCG GAATGGCTGA 2700

TGCAGGATGC AATCACTTAC ATACTGGATA TGAACAGGCG CTTGTCCTGC TGTTTTTCAA 2760

GGTGCTTTGG TTCTTACGCT CAGGGAGCAA ATTGGTGGAG AATACTGTCC AGTGTGGCAA 2820

GGTCATCACC GGTAAGTTGT GCGCTCAATT TGACAAGCAG CGCAGCCTGC CGTGCCGTCA 2880

ATTTCCTGCG ATAAAGGTAA AACCCATCCG CAACCTTGAC GCCGCCGCCA TACCGTCCGC 2940

GTACCGTTTC AATGGGATAG GAACACATAA GCACGGCAAT ATCcTGACGG ATGGTTTCTC 3000

TGGAAACTTG AAATTCAAAT GCCAGATTGC GGTAGGTATC GTGCyGCCTT ATGCACAGCA 3060

CTTCCAGAAT CGCTTACT 3078

(2) INFORMATION FOR SEQ ID NO: 424:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1935 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 424:

AAATATTACG ATaTTCmAGG GGCTTAAAAG CCCcTTATyC AtGCaAGTTA TTAAGTCGCt 60

AAAGGACAAG TGGTGTGGGT ATTkGCACAm CCTGTTAGTT TAyCCTATCT CCTAATTCTA 120

TTCTTAGGGA TTGTGTTAAC TGGCTTGTTT TGGAAAGAAG TACCATTACC ACAAATTTTA 180

GGAATTAACT TAAATTTGTT AATCATGGTG TTGGTGCCGA ATAAAGTTGC TCGTTGGTAT 240

TCCGAAAAGG AATTCGAAGG AAAAACAGGC TTTGGTTTTT TAAAAGATGG TTTTGTCTAT 300

GTGAAAAATT ATGTGTTAGA TAGTAGACCA ATTGTAGCTT TCGAACGAGT AAAAGAAATA 360

GAAGAGTTTT CATTTAAACG TTAGTACAAA AAGCGCACTA TATCTATGAA TATAAAAAAA 420

CAAGGAGCGC ATAGCATTAG CTCCTTATTT TTTATGTCTA AATTATGAAT TTTTATATTT 480

TAAATAAAGC ATGACTGCAC AGACTATTAT TCCGATGACT AAAAGTAGCA CAGTAACAAT 540

ACTTTCATTT TTATTTATTT TTCTTTCAGG TCTATAGCTA GGATTAATTT GTTTTGCGTT 600

AAAATCAGTA ATTCCAAAGT CTAAAGTAGG AAGTAAAAAC AGTAATAAAA TAATATATAT 660

AATTCCCGCT ATAAGGAAAA TAGGAGCGTT GGTGAAAAGT GAAATAATGG TACCTATGAT 720

TATAGGAGTT AAAATGTATA AAAAGGATTT GATAACATTT TCTTTTCTCA TTTTAATTAC 780

CAAATTTTGT TATTCATGTT TCCCATATTT CCTAAAGGAA CACAATAACC ACCGCTCATA 840

TTTTTTCCAC CTGTAGACCA ACCACCTTTT TTAAAATAAA AACCACCAGG GCAAGAGCTA 900

CCAAATTCAG CTTCTCTGGC ACTAGCTTCT GAACTAAAAA TAAAAGGTGT TACCATTGCG 960

AAAATAGTTA AAGCTAAAAA AACTTTTTTC TTCATACACT TTCCTTCTTT CCAATTATTA 1020

TTTAAAAGCT ATATTAAATA TAGCTTTATT TTATAATAAC TGAATGAAAA AAAGTTGTCC 1080

AATAATATAT AGAAATTTCT GAAAAATCAT TTTTAATGTT AATTACTAAA ATAAAATCAA 1140

CAAGAAAGGA AAAAATATGC AATTAGAATA TCCATTAACG AATCATATGA AAGGAAATAT 1200

TGTTTTAACA CAAGAAAAAA CAGCGATAgC TTACTATCGT GTTCGTTCTG AAACGGTAAT 1260

GTTGACTGAT GTTGAAAAAA AATTAAAAAC GAAAAGAAAA GTTGCTCGTG CGTTAAATCG 1320

CTTAAAAGCA AATGATGGAT TTGAAATTAA CCTTTTGCCA GTTAATGCCG ACATTCGAGG 1380

GAAAATGGGT GCTATGCGTT CTTTAATTGA TTCTAATAAT TATGAAGTAG GCGTAGATAA 1440

ATTGATGAAA ACGGCAATGG TTTTAGAACA AGAAATGGGT ATGGTTTACG AATATACGTG 1500

GATTATTGGT GTGCCTTTAA TCAAGAAAGA TCTCGCTGTT GATTTAAAAG AAACCGTTAC 1560

TAATTCAATT AGCAACGTAT CGGAAAAATT TGTTAAAGGG CTAGGTTTTG ATGTGGAACT 1620

AGCAGAATAT TGGGAAAAGG ATTACCGAGA ACAAGAATTA GAAGTTTACC AAAATCTATC 1680

TGAATTGTTA ATTGAACGAT TAACAGAAGA TGAATTGTAT TATTACCAAG CGTACCAATT 1740

CTTGaAAAAT ATTCCTCATG AGCGACAAGA TGTTCTTTCT AGTCAAAATT TAGaTAATTT 1800

aATGGCmACa AAAATTAAAC CGTTATTTGG CgGGGGATTG CATTTAAATT GTGATTATGG 1860

CGAAAGCTAT ATATCCTATC TTCCAATGGG AGATGTAGGT GTTTTTCTTG ATGGTAATCA 1920

TTTTTTAGAG GTCGT 1935

(2) INFORMATION FOR SEQ ID NO: 425:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 767 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 425:

TTTGTAAAAC TTTTCTTTTG GCAAGCAAAC TATTTTGTAG AAAGCGGTTC ACTCTGTTAT 60

CATAATAGTT GTATGAAAGA CGTATGTTAC ATACTAAAAA ATAGGGGGAT TTAAAAAATG 120

ACTTACACAT TACCAGAATT ACCATATGCT TATGATGCAT TAGAACCTTA CATTGACGTG 180

GAAACAATGC ACTTACACCA TGATAAACAC CACAACACTT ATGTGACTAA CTTAAACGCA 240

GCGATTGAAA AACATCCAGA ATTAGGCGAA AAATCTGTAG AAAACCTAAT TTCAGATATG 300

AATGCTATTC CTGAAGATAT CCGCACAGCT GTTCGTAACA ATGGTGGCGG TCACGCAAAC 360

CATACATTCT TCTGGGAAAT TATGGCACCA AATGCTGGTG GACAACCAAC TGGCGCTATT 420

AAAGAAGCAA TCGATGAAAC ATTTGGTAGC TTTGaTGAAt GAAAGCTGCT TTCAAAACAG 480

CTGCAACTGG CCGCTTTGGT TCAGGTTGGG CTTGGTTAGT TGTGAATAAC GGTAAATTAG 540

AAATCACTTC AACACCAAAC CAAGATTCAC CaTTAATGGA TGGCCAAACA CCTGTTTTAG 600

GTCTTGACGT TTGGGaACAT GCGTATTACT TAAAATATAA AAACGTTCGT CCAGATTACA 660

TTGAAGCATT CTGGaACGTk GTTAACTGGG aTAAAGTAAA CGAACTTTTk GCtGCAGCAA 720

AATAAGTTTG CCAAtGGTAA AAGGAAAGAG TCAAATGACG GCCGnCA 767

(2) INFORMATION FOR SEQ ID NO: 426:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2114 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 426:

GCAGGTACTG GGATGGCCGT TTTCTTCGCA TTACAATTAT ACCTTGACTT CATCAACTTA 60

TTCTTAGCTT TCTTAAGAAT TTTTGGTAAA AACAACTAAC ATGAACGTAA AAGAGGTnGG 120

GACAGAAGCG TTTAACTCCG AGAAATAAGA AGGAATTTCC GAAAATTGTT CTTTAATTTT 180

TGGATAATTC CGGCTTATTT CCGAAGGAGT TGCTTCTGTT CCCGCCGTTT ATCAGTTTTT 240

GAGCGTGGAG CAAAAATCCA AAGTGATTTT TGTCCCACGC TCTTTTTTTA ACCCATTCCA 300

ATAGCGAAAT TCAACAGTTT ATCGGAGTAT GTTGGGTTTt CyAAATCGGT TTTCATTGTT 360

TCGTTCAGTA GACTGAGCTG ATCGGTAATG ATGCCATCGA CGCCATAAAA AATCATTCGT 420

GTCATAACAT CTTCATCATT TGCTGTCCAA GCATAAACCT TTTTCCCATC ATGATGAGCG 480

GCATTTACAA AGTTACGATT CATTGTACTG TATTCCATAG TAAAAAAGTC AACTGGCATT 540

TTAGGCGGTC CCACAATGTT GAAAGGAATA ACGTAGCCTA CATAAAAGCG TGGTTCTTCT 600

TTCTTTAAGG CCATCGCCGT ATCATAGGTT AAGGTATGAA GGATATGGCC GTTTTCAAAA 660

ATATTTTCAC GGTATTGTTT GACAAAGCGC TGTACTAAAT CAGGGCTATC TTGAGGCGTG 720

GTTTTGATTT CAATTAATAA ACGTTGTTTT AATTGATTGG CTTTGGCTAA ATAATCATCG 780

AAAGAAACCA TTTTAGCAGT TTGACCATTT TCTGTGACAT TCATATTCGT TAATTCTTTA 840

AGTGTTAATT GATTCGGTCG TTTATTAACG CCCGTTAACG CCCGTAAATT AAAATCGTGC 900

ATGACAATAA ATTGATGGTC TTTGGTTTCT TGAATATCCA TTTCAACATA ATCAGGTTTT 960

GCCTTATTGG TCTCGATTAA AGAGTCCAAG GTGTTTTGAA CGCCATTGTT ACCATTAACA 1020

CCACGATGGG AGATGGTAAC CGGTTTTCgG TCAGGTGTAT GACTTAGGTA ATcGTATTAT 1080

AgGTGCCTAC ACCAATnCCG AAGATGAnTG CAATCaAACm AAAAAGAcGA CtCtGAAGGn 1140

GGtCCATCnA CCcGAmCcyG ATTTCGGGGT GAACCATTCT GGCGTGTCAG GCAAAATCCC 1200

TTCATCTTCC ATATCATCna CaATwaTGtA AAAaATAGCm ACaCTwGaCA ACACCAAGTT 1260

TaAGAGCATC ATAmCCTGAA GGAGTGTCAT AGCGATGACa GCGCTAGGTA AAGACCACTC 1320

AGGTTTAAAC GTTTCGACTG CGGTTTGACC TAAAAATATC AAGAAATAGC CAGCCGTAAA 1380

AATCCCTAAA ACAGTTCCTC CAATGATAAT AAATTGCCCT AAAATCTGGA AAAAGCGCCG 1440

CTTTGTAATT TGCCAACTAC GTTTTAAAGA CTCACGAAAc CAACATCACG TAAAATCAAC 1500

TCTGGAAGGG CAAAAGCAAG TCGAATCGCT AAATATAAGA TAATCAGTTC TAACAATACG 1560

AAACCAATAA TAATAATGAC GCGATTTGCA AAAATAAAAT CTAAGATAAA GGCTGGAATC 1620

TTAACTCTTG TTAATAAATA AGAGCGGAAG GTTAACCCAC TTAATGGTAA AATCAAAAAG 1680

AAATATGCCA GGAAAAACAG AAAGGTAATT CCTCGAACTT TTTTGATCTG AAGCAAGGTG 1740

CCGTGCAATA ATTGCGTTAG AGTAATTGGT TGCTTTTTCT TAATAAAATA GACACTAAGC 1800

AATAAAAAGG TAAATTCAAA AAAGAGCGCC AGTAAAATTA ACAGCCCAAT TAGTAGAAAA 1860

GTCAAAAGGA CATAGGGATG TTGCGTAGca ATTGCCCCTA AGTTATCGTA GGAAATATAA 1920

TCAATCCCAC CTTCTTTGAG TACCAATTTG GaCAAACTTG TCAATAAAGG CAATAGAACG 1980

AATAACATAA ATCCGTGCAT CAAGAGCACA TCGCGaAAAT ACGCCGTTGT CCCTTTAAGA 2040

AAATCTAGCG TATTCACmAT ACTATTTTTA AAATAACTAA ACATCmAAAG TCCCCCAGTT 2100

TCtACCmAAA TTCC 2114

(2) INFORMATION FOR SEQ ID NO: 427:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1065 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 427:

CAATTGTCTG AACACATTTA GCGCCGTCCT GAATTAATTC ATGGCAGCCT GTTGAAAAAG 60

ACTGGAAAAT ACTTCCCGGT ACGGCAAATA CCTCGCGCCC GTAATCCATC GCCGCTTGTG 120

CAGTAATTAA AGAGCCACTA TTCTTACGAG CTTCCATCAC GCAGACCCCT AAAGATAATC 180

CAGCAATAAT GCGATTACGT GCAGGAAAAT GATATTTTTT CGGCCCACTC CCATTCACAT 240

ATTCTGTTAA AACAAGCTGA TTCTGCTTCA TAGTTTGCTG TAATTCTTTC TTTTCATACG 300

GATAATATAC ATCTAATCCA GTGCCTAAAA TTCCAATTGT CTGACCCCCG TTCTGGATGG 360

CCATCTCATG GCTACGACTA TCGATACCTT TGGCAAGACC ACTGACGATT GTGAACCCTT 420

CTTTGACTAT TTTAGGAACC AGAGCTTCAG TCACTCTTAA CCCATACGAA GTGGCATAGC 480

GAGCGCCGAC GATGCCAATT TTCCTTCTAC TTAACAATTG TAAATTTCCT CTATAAAAAA 540

GAATtGCTGG TGGATTATAA ATTTCTCGTA ATAATTTAGG ATAGACGGCA TCTAAAATTG 600

TCATAAATTG GTGTTGGCAT TGTCGTATAT ATAACGCTTC ACTATGAATC GTATGGTGTA 660

ACCAAGATTG TTCAAATAAT TCACGATATT TCTTGATTTC AGCAATTCTA ATAATTTCTG 720

TACTGGAAAA ATCAACACGC TTTTGTGCGA TAGCTTCTTG TAACACTTTT AAGATCCCTA 780

AATTGCCAAT TCCTTGGCAT ACCGCTAATT TAAATAATAA TTGTCGTTGT AAAATTTCCA 840

ACAAAAAACA CCTCTTTCAC ATCTTCTTAA AATAAGAATA CGTAAAAAAA GTGTTTTATT 900

TTTAAATCAT ATCTTTTATT GGCGCAAACG TACGTCGATG AATGGGCGTT ACACCGTATT 960

TTTCTAACCC GAGAgAkGTT CTTtTGTTCC aTAACCTGCA TTGynTTTAA ATCCATATCC 1020

AGGATATAAT CCATCATATC CTTCCATTAA TCGTCTCTGG TACTT 1065

(2) INFORMATION FOR SEQ ID NO: 428:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3212 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 428:

CTGACnAGTG CACACAGTTG TTGGCGCGAT GCTGGACGCC CtCGtTTTTG awcTTaAgyT 60

TgGGCtCtct TctGtTkGtT AGCCAGTnCA GCTCTTTTGC TGCTGGTcTT TTTCaGCCTT 120

tGCTTTTycC rCTyCCGcAt TTTtCGcTAA tyCTTmAAgC TTGGCTTGTT CTTGtTTCGC 180

AGCTGTTAcT GCTTGTTGAT CTTTATCATT TTGyACTTTA GCAGCATTCG TTTCTTTTGC 240

TTTTTCTTCG ACAACAGCTT GTTGTTGGTT AACGTCTGTT TGAGCTTGAT CyACTACTTT 300

TTGTTGGTCA TCAACwGCCT GTGTATcCGT AGCCACTTGC TCTTTGGCTT TTTCAATGGC 360

TGAAGGTGTT GCTTCATCCA CGACTTTTTT AGCTTCGTCT ACAACTGCTT GTTGGTCAGT 420

CACTGCTTGT TGACTTTGAT CCAATGCGTC TTTGTTTTGA TCGACCACTG CTTGTTGGTC 480

TTTAACAGAT TGATCAATGG CGTCTTTTTC TTTTTTCGCA GTGTCAGCGA CTTGTTGTTT 540

TTGATCGACA ATTGCTTGTT TCTCTGTTAC TTGTTGCTGT TTTTCAGTAA TAGCTTGCTC 600

CGTTGTTTGT GTAGCTTTCA CTGTTGGTTG TTCTGTAGAA CTGTTTTCAG GTGTTTTTGG 660

TTGCGCTTGT TCTGCTGCCT GTACTTCATT TCCCCCAATA CCTGCTACCC CTGTTGCGAC 720

TAAAATACTT AACCCTGyAC TTGCGATCTT TTTCATTTTA TTTTCCTCAC TTCCGATTAT 780

TTTTGTTGTA AATGGGATAC CCATTGTCTT TACGCCATTG TTGTAACGCT GTAGTTGTTA 840

TrTGGAACTC TTCCGCTATT TCTTTAAACT GTAArCCCAT TyTTTGATAG GCTTCAAATC 900

TTTCCTTGCT AATATGAGAA GAAAATTGTC TGGTGCTTGC TGTTTCTATC TTTTTTGCAA 960

GTTCTTGTAT TCGTTTAAAA CGCTCGGTAT CTTCATTCCT ATACCAGTCT GCGTCTTCAT 1020

TCATCAATGC CAACATCTCT TTACGCCAAA TACTTTGCTC TTTTTTTGTC ATACCCCTCT 1080

CACCTCAAAT TAATAGTAAA AAAATAACAT TTATTAAAAA AATACTATCC TTCTTTTGTC 1140

TtCTTGGACT TACCAAAAAA TAAAAATATT TTCGACTAGT TTATGGCGTC CGGAACTAGA 1200

AACCCTAATA GACAACGGTA AATGTAACAA AATTATTGTT ACAAGTGCCT ATATGCATTG 1260

CAGCACACAA CATATCTTTT ACAGATATGC GCTCTCCCAT TACGAAGAGC CTATTAAATA 1320

TTACTTAATA ATAAGTGTTT TAGCCTAAAT CCCTTAGTAG CTGCTCTGAT TAATTTAGGC 1380

GCTATAAATT GCTTCCTCTC TGGTTCAAAA TGTTTGTTTA CTAACTATCG TTATCATAGT 1440

TATTCATTTG CAAAAGCTTA TGATTACGAC GTTTGCTTAC ACTAACCAAT TTTTTGTTGT 1500

TAATGTACTA AAACAATCCT TCGTCCGAAA TAAAGGAGGG CCACTCCACC GCTGAATCAT 1560

TAAGCCTATC TTAATCTTCC ACTCTCTCCA CATTCTTCTT ACAGAGATTT CACAATTGTT 1620

TTTTTAAGCA AGTCCAAGAA GACmAAATCT ATCTTATGAT TTTTTATCGT AGTCCAAAAA 1680

TTTACTCATT CCTCTACCGG TAACTTTTAA CGGACCACTA GGAGAATTTC TTCTAGCGAT 1740

TAATACCCCA TTATGGTCAT AAACATGTTT AATTTGTTTA TAGACACTGC GCAAAGGAAC 1800

TTGACCAACT TCTCTTTCGC ATAACTTTTT CCCTTTAAAA TACTTCTCTA TTTGAACAGC 1860

AAACATCCCC TCAATTCTAT ATGTTTTTGT CTTTCTAACA AACTTTGATG TAGCATTGGG 1920

TAACTTACGC TTCTCTCGTT GACCAAAAAC AATAATTCGG TTGCCGTAGT ACCCAACATG 1980

TTCAATTCGT AATACTGAAA AGCAACTCAT TTACATTCCC CCAATCGAAT ATTTTTTAAA 2040

TAAAAATAAG AGCCATGCAA CACGTTGTTT GCACGGCTCT TACGAGTAGT TCCaGTACAA 2100

ACAATGCGTG AAAGCATTGA ATTATTCCAC AAGCGGGTCA TTTTTAGAAA ATGAGCGTGC 2160

TTGCTACTTG GTTTTTTATT TTTTCAACAA GTTAACAGAC AACAAAAAAG CGGGGAATAT 2220

ATACATGAAC TGTATACaTT CCCCGCTGCT CCAACATTTA TACAACTAAT TAAAACAATT 2280

CACTGTAAAA ACTTTTCCTA GACAAATAAA GGAGTTTAAA AAAGTCTAGA ATTTTACATA 2340

AAAAATTGTT ATACTATACG TTGTAACTAA TGTTGCAACA AACGAGAACC GAGTATAGTT 2400

CATGGGGTTC TGTTGCAAAG TTTTAAATAA AGAATAAAAT CCCTTACGGT ATCTATGATT 2460

TAAGCTGGGA TTCCCAATAA TACCTTGATT TCAGTACAGA CCGAAAACCC GAAGAGAGTG 2520

CCTTCTTTTC GGGTTTTCTT ATATAATCCT CGAATGGCTT CCATGCCTTT AATCGTGGTA 2580

GAGGCAGTGC GTAAACTTCG ATAGAATTTA TTGCGTCTCT TTACTGGACG ATGGTCTTGT 2640

TCAATCAAAT TATTCAGGTA TTTAATGGTA CGATGTTCTG TCCCTTGATA AAAGCCGTAT 2700

TCTTTTAGTT TCTTAAAGGC ACTTGTAATA GAGGGGGCTT TATCTGTGAC TAcAACtTCG 2760

GTTCATCAAA CTGCTTCACT AACCGCTTAA GAAAAGCATA GGCTGCTTGT GTGTCCCGTT 2820

TTTTACGTAA CCAAATATCC AAGGTTAAAC CATCTGCATC GATGGCTCGA TACAAATAAT 2880

GCCATTTTCC TTTAATTTTG ATGTACGTTT CATCCATTTT CCATGAATAA AAGGATTTTT 2940

TATTTTTCTT TTTCCAAATT TGATAGAGTA GTTTGCCATA TTCTTGCACC CAACGATAAA 3000

TCGTCGTATG AGAAACGTTA ATGCCACGAT CATATAAGAT TTCTTGAACT TCACGATAGC 3060

TAAGGTTATA ACGAAGATAG TAGCCCACGG CTACAATAAT CACATCCTGC TGAAATTGCT 3120

TTCCTTTAAA ATGATTCATC GTCATTCCTC CTGCTATCTT TTTCTATTAn ACCACACAAT 3180

ATATTTCCAA TTACCTCTTT CTAATACTGG CA 3212

(2) INFORMATION FOR SEQ ID NO: 429:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1488 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 429:

ATTGAGTGAG GTATTACnGG AnATTATGGT TCGCTATGGC GAATTATCAA CTAAAGGAnn 60

AAATCGTAAA ACGTTTATCA TGCAGTTAGC ACAAAATGTG AAACGTGCGT TAGCTGATTT 120

TCCAGCATTG AAAATCCATG CAGATCGAGA CAGAATGCaT ATTTTaTTAA ATGGGGAAGA 180

TAGTGAAGAA GTGATTCCGA AACTAAGTAA AGTGTTTGGA ATTCAAAATT TTTCGCCAAG 240

TATTCGTATT GAAAAAGAAA TGCCAGCGAT TCGTGCAATG GTTCAAGAGG TTGTTCGGGA 300

AGTTTATACG CCAGGTAAGA CGTTTAAAAT TACTGCTAAA CGCTCGGACC ATTCATTTGA 360

ACTTGATTCA AATGGCTTGA ATCAAGAATT AGGTGGGGCC GTTATTGAAG CCATTCCGGA 420

AATTCAAGTA CAAATGAAAA AGCCTGATAT TAACTTAAGA ATTGAAATTC GTAAAGATGC 480

GGCGTATTTA TCTTATGAAA CGATTCGCGG TGCGGGTGGA TTGCCAGTGG GCACCAGTGG 540

TCGTGGTATG TTGATGTTGT CAGGAGGCAT TGATTCACCA GTTGCTGGTT ATTTAGCTAT 600

GAAACGTGGC GTGGAAGTGG AAGCTGTCCA TTTTGCTAGT CCGCCTTATA CAAGTGAACA 660

AGCGTTACAA AAGGCTAAAG ATTTAGCTGA AAAATTAGTC CCTTATGTAG GAACTATTCA 720

ATTTATTGAA GTGCCCTTTA CAGAAATTCA AGAAGAAATA AAACGAGTTG TTCCCCAAGG 780

CTACTTAATG ACCATTACAC GACGTTTAAT GTTACGCTTA ACAGATGCAA TTCGTGAAAT 840

GCGGAAAGGT TTAGTCATTA TTAATGGTGA ATCGTTAGCA CAAGTGGcTT CGCAAACACT 900

TCAAAGTATG GTAGCAATTA ATGAGGTTAC CTCAACACCG ATAATTCGAC CCGTTGTTTC 960

AATGGATAAG ACAGAAATTA TCGAAATTGC CGAAAAAATT GACACGTTTG AATTAGCCAT 1020

CCAACCCTTC GAAGATTGTT GTACAATTTT TGCGCCACCG CAACCGAAAA CACGACCTCG 1080

CTTGGATAAA GCGCAAGAGT ATGAAGCGCG CTTAGACTTG GAAGGTTTAA TGGCTCGGGC 1140

CTTAGAAGGA CTCAAAATTA CAGAGATTTC AGCAGAAACA GCGAAAGATA AACAAGAAGA 1200

CGAATTTGCA GATTTCCTAT AAGArATAAA TACAGTGACG GTCGCTTAGA CGTCACTGTA 1260

TTTATTTTGT ACTTATTTCA CmAACAAGTT GCTCGGAGAG GGGAACCGTG ATAAACTGAA 1320

TATGTGAAGT AGGTAACAAG GAGGACAGAG GATGGAGAAA AAAATGCCTA AAGCAACGGC 1380

AAAACGTTTG CCGGTTTATT TACGTTATTT AAAAATGTTA GGGgATTCTG GTGTTAAaCG 1440

CATTAAaTCA CGAGaGTTTA GTGAAATGAT CCAGATTCCT TCTGCAAC 1488

(2) INFORMATION FOR SEQ ID NO: 430:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1145 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 430:

ACGATGGTAT GTGCCCTTTT CCGACGACCT GTCAACnCnA AAACCTTGTA TCCTAGCAGA 60

CTGTTCGTCT TTTCCGATCT GATCTCTTGA TTGTACTTGC CATTCTTTCG TACTTCCTCG 120

TTCATGGCTA AAGATCAATG TATTTnCCTT TTAGTTTTTC TGTATCTAAC TGAGAACTAT 180

CGATAGCACC GTCAGTCATT GAnCTGCGAC AGCGACAATA TATTCCATCA TATCTGTTGC 240

TAAATAGAGT TTTGGCTCTT TGCCCTCGTT ATCTAACGAT CCAACCAATG AAATCGGTAC 300

ATCCATTCCC ATCGCTTTTA ATTTCATCTC TAAGTTGAAT GCTTTTTCTT TTTTGGCATC 360

TACTTGAATC TTCCCCAGAA AGCGATGCGT CTTTGATTTG CGTGATGAGC ATCCCAATCA 420

TAGGATTCGT TTGTGCACCA TCTTCTTGTG AAAATTTCAT GTCACTAATT GACATAGAGA 480

AATCyCAAaC CCCAGATTCT TGTGCGTTyT GATTGCCGAT TCCTTGAATG AACTCATCTT 540

TCGCACTTTT GCCACAACCT GTCAGTAACA AAGCTGAACA GAAAACGCCT GCGACAATCC 600

CCACTTTTCT TTTTAACACA CTTTTTCCCC CTCATTATAT AAAAAATTAA TTCCCTTTTA 660

TTAAGAATAG CAAACATTGC TTATAAATAG CAATGTTTAA TTTTAGATTT CCTCTCTTTT 720

AATAAATGGG ATTGATTTAA TCTTTTAGGT TTGAATCATT AAGCTGTATT ATTTTTAAAA 780

TAAAATTGCT TTTCGCCAAA AAATACTGGG AAAAGTTCTG CTAAACTGCC AGCTAAAGGA 840

CCTTCTCGCA ATTCTTTTTC AGTGACCCAA AACAAGGGTT GTTCCGTTTC TTCTGGTAGT 900

TCATCCGCTG TTTCTGCAAT GAATAATGTG GCTAATTCCC GCGCACCTGA ACAATCATTC 960

AGCCATTCTG CAACACCTAC TAATTGTGGC TTCAGCACTA ATCCTGTTTC TTCAAGGAGT 1020

TCGCGAACCA TGGCTGTTTC CATGcCCTCG TTTTTTTCAA CATGTCCGCC AGGGAACGTC 1080

CAACCAGGCC AATCCTTTTT TTGTCGTTCT TGCACAAGGA TTTCATTTTT CTGATTCTAA 1140

TCATA 1145

(2) INFORMATION FOR SEQ ID NO: 431:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4628 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 431:

CAGAAACACT CTTTTTTTGT AAAGGATTAA ACTTTAAAGA AATCTATTTA GAAAATGCTG 60

TAAAAGACGG CCTAGAAATC TATGTTTCTG AGGTGCCTTA TGAGGTTCCG GCGCAACTAG 120

GAATTATCGT GACTGACATT AAAAAAGCAA TGGCTGTGTT AAGCATGGCT TTTTATGATT 180

ATCCACAAAA TAAATTAAAA CTGATTGGCT TTACTGGAAC CAAGGGTAAA ACAACAGCGG 240

CTTATTTCAC CAAATATATT TTAGATGTTG CTACACAACA AAAAACAGCC TTGCTTTCTA 300

CCATGAATTC GACATTAGAT GGCAAAACAT TCTTTAAATC TGCCTTGACT ACACCCGAAT 360

CTTTGGACTT GTATCGAATG ATGGCAACCG CTGTTGCTAA CGGGATGACT CATTTTATTA 420

TGGAAGTTTC TTCCCAAGCC TACAAAACTA ATCGTGTCTA CAAGTTATTT TTTGATGTCG 480

GCATTTTTTT AAACATTACC CCCGACCATA TCAGTCCAAT TGAACATCCA ACCTTTGACG 540

ATTACTTTTA TTGCAAACGT CAATTGATTA CACATTCTAA AGTCATTGTG TTGAATCACG 600

AAGCTGATTA TTTTCCATTA CTTAAAGAAA CTGCACAACA GCAAAAAGTC CCAGCCATTG 660

TCTATGGCAG TCAGCCGGCT CCAGAAGTAG ACTATTCATT CGCTGTTTCT TCAGAGGATT 720

CCTTACGTTT TATCGTTGAA AGTCCTGCAG ATGCGTTAGG ATTGGCCGGC TCTTATCACC 780

TTCGTTTAGG TGGTGATTTT AACAAAGGTA ATGCGTTAAG TGCGGCGATT GCTAGTGTCC 840

TTGTGGGAGC TTCAAAAGAA GAGTGTCAAC AAGGAATCGC AGCAACCACC GTCCCTGGCC 900

GGATGGAATC TTTAACGAAT ACCAACGGTG CAACCGTCTA TGTTGATTAC GCTCATAACT 960

ATGACAGTTT GAAGAATCTT TTAACCTTTG TCCGTGAAGA ACATCCTGAC GGCCGCTTAA 1020

TTGTTCTTGT CGGTAGCACG GGAGACAAAG CGATTTCTCG CCGCAAAGAC TTTGGTAGGG 1080

TTCTTTCTGA ATTGGCTGAT GTTGCAGTCC TAACAACCGA TGACCCAGCC AGTGAAGATC 1140

CAGCAAAAAT TTGCCAAGAA ATTCAAGCCC ATATTACCAA AGAAATGCCT GTTTACACCG 1200

TTTTAGATCG TGGCGAAnAA TTGCTCACGC ATTATCTTTA AGCACAACGG CTGATGATGC 1260

GATTGTCTTA GCTGGTAAAG GCGCCGACCT TTATCAAAAA GTAAACGGCG TAGATGAACC 1320

TTACGCaGGA GACTTTGCCC TTGCGGAAGC TTTCATTAAT AAAAAGrACT AAATCATTTA 1380

AaCaCACATG AAAAAAACAA GCTGCAACCA CTTATCAGCA AAGGTTGCAG CTTGTTTTTT 1440

AAAGAACTTC TTTTATTTTC GTTACTTCTT CTTCTGATAA GCCATGACCG CCATCCACTA 1500

GTAAAAAGTC GACTGGAAAA TGTTTCTTTA ATTGATTAGC TAAAGAAAGA ACTTGTCCCG 1560

GAATACTCAA CTCATCCTGA GCCCCTGTTG TGACAATCAC TTTTGTCGCA AACTCACCAG 1620

ATACATATTG ATAGCCTAAA TTGCTTGGAT GCAGTAAAAT AACAGTGTTA GCAATAGTTG 1680

GGTTCTTTTC TAACAGGCCT AAAATAAAAT TGGCACCGTT AGAGTACCCT ATAAAAATAA 1740

CTTCATCGTA TAACTTTTGC GGTTGCCAAA CTTCTTCTAG AAACGCCGTC ACTCGCTCAT 1800

CAAAGTTAGC ACGAGATAAT GCTCCATTGA CTAATGGCGC AAAAAAGCGT CGCTGTTCGC 1860

CTACACCTTC TGTCCCTAAG TAACTTAGAA CACTCGCTTC TGGATAAAGG GTTGCGATGG 1920

TGGTTAATAA TTGATATTCA TTGCCACCAG TACCATGGAA AGCGACAACT AATTGATTAG 1980

GTTTCGACCC TTTTTGAAAA AATTGTTTCA TTTAATAAAC CTCCAAGAAA GGAATTTTAC 2040

TATGTATGCA CAAATTCATC ACATCTCTAT TTTAAATCGA TTGATAAAGC CTACATTTGA 2100

CTTCTATCAT AATACACTGG GATTAAAATT ACTAATGAAA ACGATCAACC AAGATGACCA 2160

CACCATGTAT CACTTGTTTT TCTCAGATAA TGAGCAACGA ACAGGTACCG AATTAACTTT 2220

TTTTGAATTG AATGATGGCC AAGATCAATC CTTTGGGACA AATACGATTG AACGAACTAT 2280

TTTAAAAGTA CCAACGCGCG CCTCTTTGGA TTTTTGGGCA GAACGTCTTG AGCAAGCGGG 2340

TATCTGCCAC TATGGTGTGG AAACATTTAA CCAACACCCA ATTTTGCGTT TTGAAGCGCC 2400

CGATAATACG CAAATGGCGC TTGTTCCGTT ACGTGAATTT GAAAATGCGG AAGATTATTT 2460

TCCAGCTGAG CATTCCGAAA TTCCAGTAGA ACATGCAATT TTAGGCATTG ATGCAATTCA 2520

ATTACGCGTC CAATATGCCG TAGCTACCAA ACAACCTTTA GTGGATTATT TACATTGGCA 2580

TGAAAAAGAA ACAGTACCTT TTTTTGAAAC TGCGCATGAA GTAACCGTTC TGGAAAATCA 2640

TCATCCACAG TTTTATCAAG AGGTGCATAT CATTGATGAC CGTACCAATC CGCTTGCCAT 2700

TGAAGGAATT GGTGGCGTTC ATCACGTCGC CTTTGGGGTG GAAGATACGC AAGAATTAAC 2760

GCAAGTTGAT AGCCTGTTAG AGGAAAAGAA TTTTACAAAC TCAGGAATTA AAGACCGCGA 2820

ATTTTTCAAA TCATTGTATT TTAGAGAACC GAACCACTTG TTATTTGAAG TAGCCACACA 2880

AAAAGGCCTG CTAGATGCCG AAGCCTATGA AAATCAAAGC TCAAACTTTG ATGAAATCCC 2940

CCTTTATTTG CCTCACTTTT TGGCAGATCA AAGAGAACGA ATCGAAGCGA TTTTAGCACA 3000

ACAACGCCAT GAGTAAACAA AAATAGCGGT TCTAACCGCT ATTTTTGTTT TTTAATATCC 3060

TATTATTGCC CTTTTTTGAA AAAAAGGAGT ACACTAATAG TAGATGAACT GCGCAAGGAG 3120

GTGAGTACAT GTTATCTAAG GAAGAAGTTC TACACTTATT GAACGAAGCA AAAAAAGAAG 3180

TCGATCGTTT AGAAACAAAC CGCCAAGAAG ATTTAGGTAA TTCAATCAAC TATATTGAAA 3240

ATGAATTACA GCTCCAACGT GTCTTATCTC AAGTGGAAGC TTATGAAAAA GTACTTGGTT 3300

AATTGATTAC GCATTAAATG AGAGAGGAGC AACCAGATTC AGGTGAATGA ATCTGGTTGC 3360

TCCTCTCTCa TTTTTTACTA GACTAATTCG TAAGCAAGTT CTCGTGCAAT AATTTTTGCG 3420

GTTGTTTCAG CGATTTCTCC AGTAAAGCCA ATGCATTGTT CCATATGTTC AGGTGAACCC 3480

ATCGTCATTT TACTTGTCAA AATACTGCAG CAAACCACTC GATGCTTTTG TCGAAAACTT 3540

TGATGCAATT CGTTGGCTAA AGCCATCGTT TTATTTACTT GTTCTCCTTT GGCTTCTGTC 3600

CGACCAAAAA AGTAGCCCAG ACACATAACC CCGCCTGAAA GAGCACCACA AATACACTTC 3660

GCTCCTCCAA CACCTACAGG AAACCCAGAA GCCATGGCCA CAGCTTCAAT GGGAATCTCC 3720

GCATCGAAGT GTTGCCGAAT GGTCGCCACA ATCGCTTCCG AGCAATAAAA ATCACCGTTA 3780

CGATAATAGG CTTCCGCTTC TTCCCTGATT GCTTGTAAAC AAATTCTCTT TTTCAAAATG 3840

CCCTCCATTC TGTATCCGCT TTCATTTTAT GAATAAAAGG ATGCCAAGCA TCCCTTTATT 3900

TTATAGATTT TCTGCAGAGG TAAATACTTC TGACACATCA TCATCGTCTT CTAACTTGTC 3960

AACCAAGCGC TCTAATTGCG CTTTTTGTTC ATCGTTCAAG GTTAGTAAGG TTTGCGGAAC 4020

CATTGTTAAT TCAGCTTGTG CTAATGAATA ACCTGCTTGT TCTAAAGCAT CGCGAACCGC 4080

TGTAAAATCT TCTGGTGCTG TGTAAATTTC AAATACTTCT GGTGATGTTT CTATATCCTC 4140

GCCGCCTGCT TCCAAAACGA CTTCTAGCAT GTCGTCTTCC TCAATTGCAT GATCTTCTCG 4200

CTTAATGACA ATATAGCCTT TACGATCAAA CATATAATTC ACCGAACCCG TTTCACCTAA 4260

AGAGCCACCA TTTCGAGTAA ACGCGACGCG AACATTGGTG GCCGTTCGGT TGCGGTTATC 4320

TGTTAGTGCA TGAACTAAAA CTGCCACGCC ACCTGGCCCG TATCCTTCAT AAGTTACCTC 4380

GTCATAATGT TCCCCTTCAC CAGCACTACT GGCTTTTTTT ATGGCGCGCG CAATATTATC 4440

ATTGGGCATA TTGGCTGACT TAGCTTTATC GACTGCTAAA CGTAAAGcAG GATTCATTGC 4500

GGGGTCAGGG CCACCTGCTT TTGCCGCCAT ATAGATTTCC CTTGAAACTT TTTGGAAAAT 4560

TTTCCCACGT TTGGCATCTT GCGCATTTTT ACGCCCTTGA ATGTTGCTCC ATTTGCTGTG 4620

TCCTGACA 4628

(2) INFORMATION FOR SEQ ID NO: 432:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2651 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 432:

AAATTATTGA AGTTATTACA GAnAAATTTG AGAGTAGTGG AGaAATCGAT ATATTATTTA 60

ATAGCTTTGA TTTAGTCTCA TATTGGTTGG wACAATACTT TTCAATAmAT GAACTTTCTG 120

AAAATCATGT AATATATTwr CGCATAGAAG AAATTTTAgA AAAATATTAT TTAGTGGATG 180

AAGTTAATAC TGAGGAACCC GCAATAAATA TAAGCAAGAC mGrAAACaTT ATTATATGgA 240

ATATTACTyC ACCaATTaTC mAAAAATwGG rATgGCAAAA ACAAGTAGTT ATGAACGGAg 300

CTCATTAGTG AACTCATTGA TACGCTTGAA ACTTCTGATT TGGATACTTT AGAtCGAATA 360

TTTTATCCTG AATATAAGAA AAAAATGACT GGTTTGCTTA TTGaTGATAA TGGTAAACTT 420

AGAGTTCCTA CTCACGGATT TCAACTTTTA AAAATTAGTG AGTACGAGaC GAACCAGTTG 480

TTGGATGAGT TAGGAGAATA CTTGAAGGGA CAAGGCTATG TATATGGTGC GATTCCTAAA 540

AAAGACAATC TACAATTTTG TAATAAAATA GTTGGTTTTT TGTATTCTAT TTTAGAAAGA 600

GAAGCAAATG TATTCAATAA AAATCAATTG TTAAAACTTT TAATAGCTCA AATTGAAACC 660

CTTCTTCCTG TACAATTAAG AGGAGAATCA TCCTATAATA ATGATATAGC TCTTTCAGTT 720

CAAGAGAAAG ATCGTTTTTT TGAGCAGTTA AATGAGGATA ATAGAAATTC AATAGCGACT 780

AAATTTTTGT TAGAATATGT AGTTGCATCA CCAATTACTG GAGAACAAAA TGTAGGAAAG 840

TGGGAGATTG AGCGACTTTT AGCAATATGT TCATTAATAA TTGAATGGGC ACACCGGAGT 900

GATTATTTTA AATATAATTT TGTAGATACT ACAATGAATT TTTTACAATC AAATAGAATT 960

GGTATTAAGA AGAAAGACTT TAAAAATGTA AATTCGGCAA TGTTAGCTTC TAGGAATCTT 1020

CAATTGGCAA ACTCTAATTT ACCTATTTCG GAGAACAGAA GATATGTTGA AAGAGTGAAC 1080

CAACTTTTTA AATCTAAGTT AGATAGTGCA TTTGTTGAAG CATTTGGATA TTCATATGAG 1140

GAATTTAATT TAGTGATAGG AGGTCTAAtA GATACACACA ATAATTTAGA GAAAATAGTA 1200

TGGATTGAAG AGGAAGAGGA GTTAGTTCGC AAAATATTTA ATGATTTAGA CAAAAAAATT 1260

AGTGAGATTA AAATAGTATC AGTATTAAAT TCTATTGCTC TTGTAGAAAG AGAAAAATAT 1320

TTAACTCCAC CAATAGGATT TGAAAAACTA GATATTTTCC CTTGGAGATT CAATCGTCGA 1380

TTGTCATTCA TTAGAAGACC TCTAGTAAAG TATGAAAATA AGTATATGTT TGGAATTAGG 1440

AATGTAATAC ACGCTCATAA ATATTTAATG CGACTGATTT GGGGTGGATG TTTAAAAACA 1500

AATAGTCGTA AGATGAAAGA TGTTATGTCT AAATTACGAA ATATACTAGG TGATGAATTT 1560

AATGATCGGG TAAAAAGTAT ACTCGAAACA TATCCTAACT TACAAGTAAT GAAAGGGGTA 1620

TCTAAAATTG GAAAGAAACG TATTTCTGGT GAAAATAATA ATACATTGGG CGATATAGAT 1680

GTTTTCGCTA TTAATCGAAA ACAAAAAAAA TTATTTGTTA TCGAAACCAA AGATTTTAGC 1740

TTTTCAAGAA ATCCCTATGA ATTAGCTATG GAGCAAGAAA AAGTATTTAT AGGAGATAAA 1800

GCTTTTTTAA ACAAGCACTT AAAAAGAAGT GATTGGATTC GCAAAAATTT AAAATATGTG 1860

ATTGATAATT ATGATTTAGA AGATGGAAAG TGGGAAATTG TTACTATGTT TGTTGTTTCA 1920

GAACATTTGA TAACAAAAGA CTTGGTAGAT ACTAAAGGTG TTAAATTTAT TTCATTGAAA 1980

GAGCTAGATT ATAATATGTT CCATTAATTG GAAACTAGTA AAGTTTTAAA GTTATAAAAA 2040

ATAGTATTGT TCTGATTCAC GCATATTTTT GGGTACGTCT TTAGGTACGA TGGTCTTAAA 2100

ATGTGGTGAA ATTCATTAAA ATCGTACCCG TTGAAAAGTT ACTCAGTTTA TTGCTGTTAA 2160

AAGGTTTTCT AACTCGTTCA AAATCTCTGT CACAGAACCA GGTCTTAAGA AAGCTCGTAA 2220

AGCTTCACAG TTCTCAAAAC GTTAATTATT TGGAGTCGTT GCGAACGTGT TCGCTTACGA 2280

ATCAATTACA ATGTTTCAAG ACACTTTCCA TTTTTTGGAA GGTGTCTTTT TTTATGGTAG 2340

CGACAGTGTG TGTTCTGTCA TAAACCAGTA TATAAAAATT CACATACTGA CAATATGTTT 2400

ATCTTTCTTA CCAAGAGAAT AATTGATTCC GATTTCAGCA AGTATCCTCA TTCCGAATAA 2460

CGCACTTACG CAAATGATAC ATTTCTTTTT AGTTGATTTC CTCATTTTGA TCGCCGCTAT 2520

TTCTTTTTTC TTTGAATTCA TTATACGTTT TTGGGAGATA TTTCTTTATT AACTTTTTCT 2580

CTCTTGCGTT TTGGAGGGAG TAAGTTGCGG TTTTTTGAGA GATTAGTATG TTTAAACTGA 2640

AGAGTTATTA A 2651

(2) INFORMATION FOR SEQ ID NO: 433:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1966 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 433:

nAAGGCCCGT GTGGTTGGAA AGCCGnTACG CATTGCGAAT TAGTCGnACA AAAGGCAAAT 60

ATTTTGACGT ATATCAAAAc aATTAGGTCT GACAAwGGAA GGTGAGTCGC GTGGCTAGAA 120

ATACGTTTGA TGTAGATGAG ACTTTAGAAA AAGAATTTAA TTGGTCCCAC TATAAACGAT 180

TAGGCGCCTA TATTAAGCCT TACAAAAAAG CTGTTTTTAA AACATTATTT GTCATCATAT 240

TAGCTAATCT GGCAAGTATG TTAGGCCCTT ACTTTACTAA AATTGCGATT GACCAAGTAA 300

TTCCACAAAA GAATCTCTCG TTACTCTTGA TACTAGGGGC TATTTTCCTA TTTTCCTTGG 360

TGATTATTGG TTGGTGTATG CGTTATCGGA TATATGCCAT CACAGAAATT GGACAAGATA 420

TTTTAAAAGA TATGCGTTTT TCAATTTTTG AGCATCTACA AAAATTACCA TTTTCTTACT 480

TTGACAGTCG GCCGCATGGA AAAATCTTAA TTCGGGTGGT CAATTACATC AATACGTTAA 540

GTGATTTATT AAGTAACGGA TTGATCAATT TAATTTCTGA TTTATTTAAT GTGATTATCA 600

CCTTAATCTT CATGTTATTT ATTGATGTGA AATTAACCTT ATACAGTTTG CTGCTATTGC 660

CTGTGTTGTT CGTGATGGTT CTATTTATCC AAGGGAAACA ACGGAAAGCT TACCAAGAGC 720

TTAGCAATAA ACAATCAAAC TTAAATGCCT ATATTCATGA AAGTATTTCA GGAATTAAAA 780

TTACACAATC TTTTGCACGA GAAGATGAAA ATTTCCAGAT TTTTAACGAG GTAAGCGAAG 840

AATATCGTCA ATCCTTTATG AAAGCTGTGC GTGTCCAGTA TTTATTATGG CCAGCAGTGC 900

AAAATATATC CGTCATTACG ACTTGTTTTA TCTATTTCGT AGGAATCCGT CAATTAGGTG 960

TGTCCGTTAC CACTGGGACG CTAATTGCTT TTATCGGTTA TATCAATAAC TTTTGGAATC 1020

CAGTCATCAA TATTGGTAAC TTCTACAATT CTCTGATTAC TGCAACTGCC TATTTGGAAA 1080

GGATTTTTGA AACGATGGAT GTTGTTCCAG AAATTCAAGA TGCGCCGCAC GCTATTGAGT 1140

TGCCGCCAAT TAAAGGAACC GTCGACTTTC AGCATGTTTA TTTCCGTTAT GAAGAAGGAA 1200

AAAATATTTT GACCGATGTT AGTTTTCATA TTGAGCCAGG ACAAACAATC GCCTTGGTTG 1260

GGCCAACAGG TGCGGGTAAA ACAACAATCA TCAATTTGTT AAGTCGTTTT TATGATGTGA 1320

ATGAAGGGGC CGTTAAAATT GATGGTTATG ATGTTCGCGA TGTGaCACTC CGTTCGTTAA 1380

GAAAACAAAT GGGGGTAATG CTTCAAGATA CGTTTATTTT TTCAGGAACG ATTATTGAAA 1440

ACATTCGGTA CGGAAATTTA GCAGCCACAG AGGAAGAAGT CATCCAAGCC GCAAAAAATTG 1500

TTCGTGCGCA CGACTTTATC AAGGATTTAA AAGATGGCTA TGAAACAGTT GTGGAAGAGC 1560

GGGGTAGTAC ACTCTCGGCA GGACAACGCC AATTAATTTC ATTTGCTCGT GCTTTACTGG 1620

CAGATCCCAA AATTTTAATT TTAGACGAAG CAACCTCCAG TATTGATACA AAAACAGAAG 1680

AATTGTTACA AGAAGGACTA CAACAACTTC TGAAAGGACG GACATCGTTT ATTATTGCTC 1740

ATCGACTTTC TACCATCAAA AATAGCGATA AAATTTTCTA TATTGATGGT GGTCGTATTG 1800

TGGAAGAAGG ATCTCATGAT CAACTAATGG CAAAACATGC GTTATATCAT CATTTATATC 1860

AATCGCAATA CGACTTATTA mAGAGCTAGA TGATAAAACA GAGAAAATTC TTAGTTGAAT 1920

CGTTCTCTGn nTTTCTGTAT CGTAACCTTA ACAAAGGnnA AAATGG 1966

(2) INFORMATION FOR SEQ ID NO: 434:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2809 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 434:

CCTTCTATAA ATATATAGTC AAACTTAAAg TTcTGaCCTG CTTGTTnnaG TTAGAAATAT 60

AACGCAACTA GCTTGTGTCA TATTCAACAT ACCATGACTG TACTCGAAGA ACTGTTTAAA 120

AAGATTTGGT CTACTCTTTA ACCCAAGCTT CCGTTTCAAC TTCTCCaTAA TTGCTCGCCA 180

CAATGTATTC ATAAGGTAGG CTGTCTATAA TGTTAGGAAT ATCGCGTACT TTTATTACTT 240

CCGTTCCATT AAATTTTGTA GAAACAATAA AGAATTCTAT TTCTTCATTG TAATCATTAA 300

GAAGTTGTAA CATATCTTTA TCTTTACTAA AAACATAGTG ATCAAAAATT TGGGGTTCTC 360

CATGATTGAC ATCTGATACA GAGATTACTT CAGCCATTGA TAATGCAATT TCTTCATTGT 420

ATTCTATTAA AAATATTCCA TGAGCTTTAT TACCCTTGTA AATTTTTAAA TTTTGAATAT 480

GAGTATCTAT GTTTTTCTTA CACGACTTTA ATAGGTTTTC CTTTAAATGA TTATTAGAAT 540

ACGTTAATCT ATGTCTAGCA AAAGTATCTT CATCGATCAG TCCTTCTTTT ATCTCATTAC 600

TTCGACTCTC GAACATTCTT TGTCTTTTTA TATGCTCGGC TCCTTTTTTC AAAGTTCTAC 660

ATTCTTGTTT TGTTTGTTGT TCTTCATTAT TCATTTTCTT AAACTATTCT CCTGTCTTTT 720

TTTAATATTA TGCTTTTACT TTATTTTGCA CGATAATCTT ACTCCACAAC AAATTGTTAA 780

ATTAGCATTA TTAAATGTAA AAATTGAGTA TACTCATATT TTACACAAAT GACGTATGCC 840

CAACTTGAAT GTCCTACTGT TTACCTAGTA ACAATTGGAA TATGAAAAAr AACCAACACG 900

GCATACCCTG TTGGkTCTTT TAGTTAATAT ATTTTAATAG AAAATTTATT TATTACATCA 960

TGAAATACTG TATAAAAATC GTAGCAAGCA CACATCACTT TTTTTAGCCT ATCATTTTAC 1020

GTCGACAGTC TAGCTGGTTT CTTATCTCAA TTTTGAACCC ATTGTTTTTC TGCTATAATC 1080

ATAGTTGACG TTTGATCTGA ATAATTTTTw ATGATATTGy TAAATTyTGT CATTGATTGG 1140

CTTAyTTCAT CATCAGCACT TTTTGAkCAT TTAAAATTTA GATATTTAAT TCTAGATCTA 1200

CTAATTGAGT CTAGATTTTT TAAATTAATT TTTTTAAAAG TTTGTTTATT rTCTACAGAT 1260

AAGATATATT CTATCATATA CAAATTAGCT AAATAAAAAT TCTGTAAATA ACTTTTATTA 1320

TTCAGTATCT TGTCTTCTTT ACTAAGTATT GGGTATATTA GTTTTAATTT CCTATTATTT 1380

TTGTCTATTA AATACTCATA TATAACATAC AGTAAGAATG TATTTAGCAC ATATAATGGT 1440

AATTCTTTAT TTTTTTTTAT TATTTTGCAT ACCTCATTGC TATTTAAGAA AGTTGACTGA 1500

TTATATTTAT TAAGTAAGCT ATTTGATACA ATATAATACA TTTCGAAAAT ATAATCCTTA 1560

GTAACTAATT TATATATATT GTATCCATCA TTCATAAACC CGTTAAAATT AAAAGGAACT 1620

ACAGTAATCA AGAATATTGA AAAATTAAAT AAAGAAAAAT ATTGAAAAAA GTCAATTTTT 1680

ACAAAAAAGG TTAATGTAAA AAAAAAGACT AAATATTGGT CCTCCGGTTA ACAACCTCAG 1740

AATTTTCTTA TTATTAAACG TAGTTGGCTT ATATCTAAAT AATCCACCAA AAAATCCCCA 1800

CTGTGTATTT AATCTAAyCT TAAACTGwTT TTCCCACGAA AGTTTTATAA ACCCAAATAT 1860

TAATGACTCT GGTCTAACTT TATTAAATAA TCCAAAArCT AAATGACCTA ACTCATGTAC 1920

AAAGATAGAC AATAACAAGG AAAAGTATGC AGTAGACACT AAATTTAATT CAAAAACAAG 1980

AGTAAAGATT ACGAGAAGTA ACGTAAAGAT CATCGGAGAT ATAAATTTCA TCTATTTTAT 2040

TCTCACCTCT TTGTATTTAA GCATGCTATT TTTATCTATA AAAGAATCAG CATTTTCAAA 2100

TAAAGTATTT AATACTTCAT TTGAATCTTT AGAATTATCT ACATTCTTTG ACATAATAGC 2160

TGCTAAAGtG CGCTTACTTC TGGAGTTGCT AAACTTGTTC CAAAAGAAAG TGTATwACCA 2220

TCTGGgAAAG TCAGCAGCTT TGCCTAGTGG AGAAACTAGC GATGTAGGGT AATAAGTCAT 2280

CATCATTtCA CGAGCATCAA TTTGtCCTGT TAwTtTGtAA TkGTCACCAT ATCCTCCTGc 2340

AGGACCATAT ATwGAgTrCA TTAGACCCAT AATTAGAATA GTCAGCAATA TCACCACTCT 2400

TTTTTGTCGC TCCGACGGTA ATTACAGACT CTAGTCCTCC TGGTATATGT TTTTCATTAC 2460

CAGTGCTTAT ATCACGCGAC TCATTTCCTG CTGATGCAAC AATTAGAATG TTATTTTTTC 2520

TTGCATAGTT AACAGCTTTT CTGAATGCTT CTACAGTAAA TCTTTCGTCG TCTATTTCCA 2580

TATTTTTATA TGATCCAAGA CTCACATTTA TTATATCTAC TTGATCATTT GTAGCATCAA 2640

CTATAGCTTT AAGCATATTT ATAGAGTTTC CATCTGTCCC ATCCATCACC TTATAAGAAT 2700

TTAAATTTAC TCTTGGAGGC AATCGTGTCT ATTACTCCAG CAACTTGTGT ACCATGACCA 2760

TATTCATCTA ACTCAATATC ATTAACATAn GTTTTTTAAT CTTAnGnTA 2809

(2) INFORMATION FOR SEQ ID NO: 435:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4149 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 435:

TTAATATCTn TCTTATTTAT TGACATAAAT TTATTACCCn TCGCAAGTAC AGTTAGCGTC 60

TCAAGAACAT TTTCTTTTAA TATTAATGAA CTTTTTATAG AGTCTTCTAG AAAAACTTTG 120

GCTTTTATTT CTTGATATTG TGTATTTTGA AGCAGCTCTC CAGATAACAT CATTGTATCA 180

ATAATAAATG TATAGATAAC TGGATCAATT TTAGTACTAA TTTCTTTTTC ATTTTGTAGT 240

TCTTTAACAG TACTCCAGTA AACGTCAAAA AAACTTCTAT TATTATAGAA TTGTTTAACC 300

CATTCAGTTA TCATTTTTTT ATCACTAGGA ATTTTATTAT AAAATATGAT GAGTTCAGGA 360

TGTTGCTTTC CTAATTTTAA AATCATTTCA AACATTATTT CTAAACGATC TTTTAGTTTT 420

AAAGGAGCGT CGTTTATAGT ACTAACaTAT TCATCTAAAA TAACAAAAAC GTGTTCTATC 480

GCGACTGTTA ACATACCTTC TTTATTGTTG AAATACTGAT ATAATGATCC TACACTTAAA 540

TCTGCTTTAC TGGCAATTTT ACGAACAGAA GTTGCTTCAT AACCAACTCT CCCAAATTCA 600

TCTAAGATGT AGCATAAAAT AAGTTGCTTT TTTTCTAATT TTAGTCGATT AAATAATTCA 660

GTTGcTGTTT TATACATAAC TATCTCCTCA AAAAAGATAT TkTAATTTAG TaTACCATAT 720

TTATTATAtA AAAACGACCG CTCaTTTTTT TGCGATCCTT AATTGAAtCT TrGcTTGayA 780

AAAATTAAAT TACATATAAA ACAGTTAACC TCTGTAAGAG AAAATTGCCA AACTATCTAC 840

CTTCTCAAAA GTCTATTATC CTTTATCAAA AATAGCTATA TATTTATACT TCTTTATAGA 900

AGAATAGTAG AAGTAATAGC CTAACTGATA TGATAGATTT AAATGACTAA ATTTGTAGAT 960

TGCGTCATTT TTGCGCAATT TTAAGTCTGC TAGTCCATCA ACCAGTTTTT CCGTATATTC 1020

CGTATATAAG GATACTATTA CCTTAATAAT TATATTTTTA GAGATTTTAA TATACAAAAT 1080

GGAATGGTAA CGTTCTTTTT GACAAAAATT ATAAGGTGTT GTGAATGCCG TGATGTTATT 1140

GCTCTGATTG CTTTTGACAA TGAGCCTCAT CGGATCTGAT TCTCTCCGGC AGAAGCGCAA 1200

TCATTTAATT ACGCTGCCTC TCCAAAGATA TCAGACCCGA GCTCATTATC AAGGGCTCGC 1260

TTTGCCAATC TCTGTTGTCG AAATTGTATC GGTGTTTCAT ACCCAATGTC CCATGCAATC 1320

GCAGGTAGTT CCACCAATGC ACATAATCAA ACAATTCTAG GCGTAGTTGT GCCAATGTTT 1380

TAAATTGGTA TTGATGCACG AATTCTACTT TGACAAATTT ATACGTtGAT TCCGCAACAG 1440

CGTTATCATA CGGAAAGCCT TTCTTACTCA AAGAGCGCGT GATCCCAAAC CCATTTAAAA 1500

TATCATCGAT GGATTGATTA TCAAATTTTT TTCCATGATC TGCATGAAAA AGTTTGACGT 1560

CTGTTAAAGG ATAGGGAATC CGACCAAAGG CTTCTTTTAC CAATGCGGTA TCTTTTTTTC 1620

TCACCACAAG AATAACCGAT GATTTCTCGA TTAAATAAAT CAAGAATTAA GAAGATATAA 1680

TGCCATTTTT TCCCTACGCG AACATAGGTA AGATCGGTGA CAATAGCCTC CAATGACCGT 1740

TTCTTGTGTA AATGTCCAAT CTAATACAAT TTCAATACAT TCGTTGTCTC CGCTTCATTA 1800

CAAGCTGTTC GTTGGGCATT AAAATGAGCG ATCGTATAGG TGAATTTCAA CCTGCGTTGT 1860

TTCATGATCC GACCGATTCT GCGACGACTG ATTTGAAGCG CACGCTTGGC TAAGTATTTC 1920

TTTAATTTCG GGGTGCCGTA AGCTTTTCGG TTACGGATAA ATTCTTCGTG AATAAGTTCT 1980

TTCAGATCTT CTTCGCTTTT AATGGGTGTT GCCTGATTAA TAATAGGTCT GAAGAGAAAT 2040

CTTTAAGATT TTATACAGCG CTGATATCGA GTATTTGTGT TTGTTGACAT CAATCACTTG 2100

TCTTTTCGTC CGAATATCAA CGACGCTTGC TTTAAAATAT CATTCTCCAT TTCAAGCTAC 2160

TTATTTTTCT TTCGTAGTGC GATCAATTCT ACTTATTCCG GGGTTAAGGT ATCTTTTTTC 2220

TTGAACGATT CTGTTGAATG TGTTTGCTTC ACCCATTTAT CGAAGGAAGA AGGCGTAAGC 2280

TCATGTTCTC GAATGATTTC TACACGCGGC TTGCCTGCAA GGTAAAGATC GACGACGTGT 2340

TGTTTGAATT CTTTTGAATA GGTTCTTCGT TGACGTCTTG ACATAAAAAA TCCTCCAGTG 2400

TGTTTTTTAT TATTCTACAC ATCTTATTTT TTCTTTCTAG TGTAGCCGAT TCAGAAAAAG 2460

AACCGATAGT TAATTCGACC ATATTCATTA TAGTTAAAGA CATGTAAATG AAATGATCTC 2520

TCAAAGAATA TGTATTCCAT TTTTCTCTTA ATGAGAACTT ACTAAATAAA AATTGACAGA 2580

TTGTCTCATT GCATTACTTT ATTGTACATC TTAAAATTAT CTGCCTTATT TTGTTTGACC 2640

ATTCACTTTT TTTTTAGATC CTAAATATAC TTTTATCCGT TCATAAGAGC AGCCTGAATA 2700

AACACAACCT ACAATAATAC CCAAAACTAC TGCTCCAATA AGATTGTAGA AAAAGCCAAT 2760

TTTTTTTACA AATGTTTGAT AAAAATTTGT CATTATTATC ATCATGACTA CACTGTTACA 2820

CAAGTCATAC ATAAAAAATA ATTTAAGAAA GAGAAAAGAG ATCATTCCTT TTTGATCAAT 2880

TCTTCTATAA AAATTATATA TCCAATACAA ATAAAGAAAC ATACTTCCTA CCAATAAAAG 2940

TAATTTTAGT ACAAACCACC ACGaAAAATG TGGATAGAAT AAAAAATCAA TCAAACATAA 3000

TAACGGAATA AAATAAAGGA ATTGATGAGA GTGGGTTAGA TATTTCCAAA AGGTTTCTTG 3060

TTTTAATCGT TCCATTTGTT GATAAAATTT TTGGTCATCT ATTATTTCCA TTTATTTACT 3120

CCTTATCTTA ATATCATACA TAGCAACATG TACAACCTTA TAAAAATAAA TAATTTACTT 3180

ATTTATTTTC ATCGcAATCG ACATTCATTA GTAGATATGT TkGTTAGTAT TAAAAGAACT 3240

TGTGATTATT AAATTAGTAA ATGTTTTATC ATATAGAACT TATTGATGGT TATTAAAAAA 3300

AAGCTTCTTT TCTATAATTT AAACACATCT ATAGGTAGAT AGATATAGAA ATTTATACAA 3360

CATTCAAGTA AGAATTTATT TATTAAAAAT GTGGTCTTTC TAGTAAAAGA AACAAAAAAC 3420

CAGTTAAGCA TTCTATCTAA CTGGTTTTTT ACTAAATTTC AGTTGTATCT TTCTATATTA 3480

ATTAAGTTGT TATTTTACGC GAGGTTCAGT TTCTTCTTTT ACGTTTTGTA CACCATCACT 3540

TACTTTTTCA CTTGTACGCC TAAAGCTTCT TTTGATTTTT CAAATtGTTT kGAAGCAAAT 3600

TCTCCAGTAG CAGAAGCTGC ATCGGAAAGA TGATCTTGTA GAGTAACTGA ATCATTTTCA 3660

TGTTGTTCTT TCGTTTTTAC ATCTACGACG TTTACATTAA yCyCAATTAC ATCTAGGCCA 3720

GTCATTTTCT TAACTTCATT TGAAATAACT TGCTTCATTT TATCATAAAT TGTAGTTACA 3780

TCTTTACCAT ATTCAGCCAC TATTGAAAGA TCTACTGCGA CTTGTTTTTT TCCTACTTCA 3840

ACATCCACTC CAGAAGTTGT GTTATCCGTA TTTACTAGTT TTCCAGCTAT ATTTGAGAAG 3900

AAGCCTCCAT CAATCGTTAG AAGACCATCA ATCTCATCTA ATGCAATACC AATTATTTTT 3960

TGAATTACTT TATCTTCAAA AAATAGTTCT CCTTTAATAG TCTCAAAAGT TTTTGACATG 4020

TCCkTTTTTT TGAkCTCTGT TTTTGTATTA TTTGTTTTGT TTTCCATTAT TTTTCCTCCT 4080

AAAAATATAA AATTTATTTT TACCGTAAAA ACGGTTATTA ACTAATTTTG TTTGTCAAAA 4140

AAACCAACA 4149

(2) INFORMATION FOR SEQ ID NO: 436:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1585 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 436:

CCTCTATGGA ATATTTTTTT CCTCAAGAAA TCATTTTTTT ATAAAAACCA TCCTTAAATA 60

TGATACCAGT GGCTGGCTTT ATTTTAAATA GACAAAAAAG AGTATGAGAC AAAAATCACT 120

TTAGATTTTT GTCTCATACT CAAAATCTGA AAAACAGTTC GATCCCGACT TCTTCGTTTC 180

TATTGTTTGG TTCTGGCAAC AACTAACAAG TAACTAATTT TACGAAATCC ATGACAATAT 240

TCATTATTGA GCATATTCAA TCCTTCTACT AATTTTCGTA TAAGATAAAT ATTGATTGAA 300

TGTGTTGACA CAACGAAAGG CCGTCCTATT CGCATATTGA TCATAATAAC GCTTGGTATA 360

CCACTGAAAT TCTTTCGTCC AAGGTTTGAC ATGGCCAGTA AAGTGGATAA TGGATGGTGC 420

TCTTCTCGTT TCTTCATATT GTTTTTCTCC ATAAATAGTT GGATGTTTTT TCGCTTTTGA 480

CAGAATATAA CCTTGGGCGT TCCACTTAGG ATGCAACAGG GTCCAACGGT CGTGTAAAAC 540

CGCATTCAAT GCATCTTGAT CATGGAAACG TAATTTATCT GGATTTTCTT CAATAAATCT 600

TAAGACTTTA GTGGTTACAT CTAAATTCAA CCATTTTTTA ACATCGATTA GCAACAAGCC 660

TGAGTTAAAA TAACACATAG ATTCTGCAGG AATCGCCATT TTTTCCaAAC GTTGATGGAA 720

CCCTGCATCT TCCaCAGCGG CGATGATATT TTCACCTAAA TCAACTGTCC ATAATTTTGC 780

GACATCATCT AAGGCAATCA TGTCACAATC CATATACAAC AGCCGTTCGA TCTGGCTTCC 840

TCTAAACAGC TCTGGAATAG CAATCCTGTA GTAAGCTGTT TTGGGAATTC GTTCACTGGT 900

TACAACATTT TTGAAAAAAt GGGGGTTGAT TTTTAAAAAT GTTAACTCTG CATTGATTTG 960

GGTATGTTTA ATCGTAAAAT ACAGTAGTTG CTTGCTTTCA AAATTGATAT TGTCATCAAT 1020

CACATAAAAA TGAACAGCTG CCGCTGAAGG ACTATTTTCT AAAATTGAAA CAAACATAGC 1080

AGCAAGATGC GGTACAAAGT TGGTATTGCA ACAGGAGACA ATCGCTAATT CTTTTCTATT 1140

TTCCATTTCT TGCACCTACT TTCTTTAACT TGTTTAGATA AAGAGCTGTA TAAGGATGAT 1200

TTTCTAACGT ATTCCAAGGT TTATCGTGTC CAGTAAAATG AATAATCGCT GGCTCTTGAT 1260

TTCCCGCTGT ATATAATTTT TGATAGGCTA CATTAGGCGC TTCATGACGA TTAAAAACTA 1320

AAGACGTCTG CATATTCCAT CTTGGTTCTA GAGCTAGCCA ATCTTCATAA AGAACAGCGT 1380

TTAAAGCATC TTGATCATGA TAAACTATGC GGTCCCCATT TTCTTCTAGA TACTGAATTG 1440

TTTTTTGGGT GATTGCTTTC TCGTTCCATC GGTCAATATC AATAATCATA ACTCCTGAAT 1500

TAAAATAATA ATGCGGAGGA AkGGATGCCC AAACGTTCAA nGCATAAGCT GACCAGGATC 1560

ATAACCGCGC CAATCGTTGG TTCAC 1585

(2) INFORMATION FOR SEQ ID NO: 437:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1838 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 437:

CGAGCTCGGC CCCCCGATGT nAAAGTGGGC GACACTGTTA AAGTGAAATT TAGTGTTGAT 60

GCTTGGGCnA CTGGCGAAAG CTATTCCGCA ATGGGTAAAA GGAAACAGCT ATAAAGTACA 120

AGAAGTAACT GAGAGCAGAG TATTGCTTGA AGGTATCTTG TCATGGATCA GCAAAGGCAA 180

TATTGAATTA TTGCCAGATG CGGCAACTGT TCCTGATAAA CAACCAGAAG CGACTCATGT 240

GGTACAATAC GGCGAAACAT TATCAAGCAT TGCTTATCAA TACGGAACAA ACTATCAAAG 300

ATTGGCTGCA TTAAATGGAT TGACAAATCa AATCTTATTT ACCCTGGCCA AATTTTGAAA 360

GTAAATGGAT CAGTAGTAAG CAACATTTAC ACAGTTCAAT ACGGTGATAA TTTATCAAGT 420

wTTGCAGCTw AGCTTGGTAC GACTTATCAA aCCTTAgCTG CATTAAACGG ATTAGCAAAT 480

CCTAACTTGA TTTATTCTGG TCmAACaTtG AGCTATtGGA TTTTTkGCCA AAaTCGGkTG 540

aAATTTAGAG GtACCTCATT tACaATAGAG TTACCTTTTC ATATTATGAC TCTTTTCGAT 600

TTAGAAAAGA GGTGCTCCTA TCTTATGcCc AAGTCCTTAG GTAGGAGCAT TTTAATTTTA 660

CTCATACGTA GCTTAGTTTA AAGCTATTGT ATTACATAAG AAAACACCTA CCACAGATGC 720

ACGTACTTCC CCAAGCAGTT GTCTGTGCGG TAGGTGTTTT TTTAGATTTT ACTTGTTAAT 780

CATAACCGAT TTTCTTTTTT TATAACAAGG TTAAGCTTGA ATTTATGAAA AGGATAATGT 840

GTACGGAATT TATCCTAATA AAAAATAGAA TTAGCCTATT TCTAAGCTAC TATTACTTGA 900

GAACAGTTTC CTTTTAAATA AGTGGAAAAT TGTAGTACAG CTACTTTAGC GATATTCCTT 960

AATATCCCAG TTCTAATTTA AAGTCAAACA CAATTTTGTT AATAAGTTGT GAATAATTTT 1020

ATTGGTTGAA TTAACAAAAC TAATTTGCTA GAATAATAAA GCAATTACTT GTAGTAAAAG 1080

ACTACAGGsC CCTTGTATCA TAATTGTGCA aGGGGGTTTT TGTATCGTAT ATTTATATTT 1140

CATGTCTAGA GTAGATGGAT TATAACCTTT TTATAGTATT GGTcCCATCA TGtTTTAATA 1200

CACTCATTtG CTTTATAGGA AaTAGTATGG tATTATAAAT AAGTAATCAA TTTTGACATG 1260

TAACGAGAGC GCTATACATA AGACTATAAT TCTCCTTTTG CTTATCAAAC ATTTTTTGCA 1320

ACAAATCGCA TACAGGGTAT GCACAAGGAG GAAACATAyA TGAACAACGG TACAGTAAAG 1380

TGGTTTAACT CAGACAAAGG TTTTGGATTT ATTACAGGTG AAGATGGCAA TGACGTATTT 1440

GCTCATTTTT CAGCAATTCA AGCGGATGGT TTTAAGTCAT TAGAAGAAGG ACAAGCAGTA 1500

ACTTTTGATA TTGAAGAAGG ACAGCGCGGT ATGCAAGCTG TTAATATTTT TAAAGCATAA 1560

TAAAGGTTTT GGTAAAACGT TCCAATTTAA TTGGAACGTT TTTTTTATAA GATTAAATTG 1620

ATAGGCTCTG AATGTTGAAA GCGAAAGGAG ATCATCATAT GGGATTTGGA AAATGGAAAG 1680

AAaTTATACC ATCAGATAAA AGTAAACAGT TTGAAGAAGG ACAGAAAGTA TCATTTATTT 1740

ACAGAAAgAA tTAGAgGkaT AATAGCgTTT TATTAATTAA TcAGCAGaAT TAATATTATA 1800

GACAGTTCTA AGCCAGAAAA CnnAAAAGAn AAAACAGT 1838

(2) INFORMATION FOR SEQ ID NO: 438:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1239 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 438:

GGATTTGGGG TCTTAAGCGT TCTCTATTTT GTGGAACCCT TTGAGAACAA TCTCTTGGTA 60

TTGTATGTCG GATCGGCTGT TTTAGTGACG ATTTTAGAAT ATGTCACTAG TTATGGCTTG 120

GAAAAATTGT TCCATGCCTC TTGGTGGGAT TATCATGATG TACCTTTTAA TTTAAACGGT 180

CGCGTGGCTT TACCGGTTTC TCTTTTTTGG GGATTGGGGT GTGTCTTGAT TGTCAAAGTT 240

ATTCAACCAG AAATCGCCAA AGTTGTCAGC TTTTTACAAG CAACCTTTGG CAACTACTTA 300

GCACTTTGCA TAGTGATTGT AATGGGGCTT GACTTAATTT ATACGTTGCT GAACATGCAA 360

GGTTTCAAAA AACTTATTAC CGAGATGGGC CAAACGTTGG AAACGAACCA ACAAGAATTT 420

AAGCAACGGT TGAACACAAA AATGGAAATC GCTACAGCTG ATTGGCAGAA GTTAAAAGAA 480

CGAACCAAAA AAGAGCATTT CCAAAATCGG TTAAATTTCC AACAACGTCG TTTTATTAAC 540

AATTACCCAA AACTAACTTT AAAAAACATT AAAAATTCCA AAGAAGTTCG GCATCTCTTA 600

GAAGATTTGA AGAAAAAGGG CCAATAATTA ACAAAAAATA TTACAAAATA TTTCAATCAA 660

TGGTCTAATG GGATGAAAAA AAGAAAAATC GTCAAAAGCT AGTTTTTAAk GCGTCTTAAG 720

AAATTCTTAA AGAATTGAGA TAGAAATGGA CTAATGAATA AACGCAGGAT AAAAAAAGAG 780

TTGCATTTAA ATTGTAACAA ATTGTAACAT AATTGAAATA AAACAAGCCC TCCTGACTAG 840

TATAATGAGT AGGTGTTTTA GATAAGTTAG GAAAGGTGTT GGATTTAGTG AAGGTTAAAA 900

AGTTGATACC ATTTTTAACA ACAATGATGC TTATCAATTT TGGACTTCCA GTCGCTGCGG 960

CAGCAGAATC ACTTGATTCG TTGCAAGAAA AAGAAGCGCA AgCGrAACAA TCTGGGGAGT 1020

TAATTAGTAA AGAaATCAAT TCAGCCTTAG ACGATGTAAA TGCTAAATAT GCGGAAATCG 1080

AACGTTTAAA AACGCAAATT TCAAAAGCGG AAGATACGTT GAAAACTTCG AAAGAGGAAA 1140

TAGAmAAmAC AGAArAAAAT ATTTCTCGTC GTAAAGAAGC TGTTGGCGAT CGTaTGwaGG 1200

ATaTTCAATT AAaTGGCGAC CAAcGTACTT GGCcAgCCT 1239

(2) INFORMATION FOR SEQ ID NO: 439:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2025 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 439:

TTGAAAAATC CCAAATTAAA GTGGAACCAG CAGGCTTTGA CTGCGACTAT GTGGACATCC 60

AATGGTGATT AAACTGGGTC GCTACGGAAA ATTTTATGCA TGTAGTAATT TCCCAGATTG 120

TCGGAATACA AAAGCGATCG TTAAAGrAAT TGGCGTkACg TGTCCTGTGT GTCACGAAGG 180

ACAAGTTATC GAACGCAAAT CAAAGAAAAA TCGGATTTTC TATGGCTGTA GTCGGTACCC 240

AGAGTGTGAC TTCACTTCTT GGGACAAmCC aGTAGGCCGA CCATGTCCmA AATGTGGGCA 300

aTATTTAGTA GAGAAAAAAG TTAAAgGCGG CAAACAAGTC GTTkGTATCa ATGGcGATTA 360

TGAAGAAAAC GTCCAAAAtT AAGAAAAATT GAGACCGAAG AATTTACTTC GGTCTTCTCT 420

TTTAGTAAAA TGAATTGAGT TTGTAAGAGA CAGCTCTTAA CTTCTTGATA TAGAAAGGAA 480

TTTTTATACA TGACAAAAAA AGTCACAATT ATTGGTGCCG GTTTAGCCGG AmGTGAAGCG 540

GCATGGCAAG TGGCCAATGC TGGGGTACCA GTTGATTTAT ATGAAATGCG CCCTGTCAAA 600

AAGACACCAG CGCATCAAAC AGAAAATTTT GCAGAGTTAG TTTGTTCTAA TTCATTACGT 660

GGAaATAGTT TAACCAATGC GGTAGGTGTT TTGAAAGAAG AAATGCGCCG TTTAAATTCA 720

ATCATTATTG GTAGTGCAGA TCAAACCGCT GTACCAGCTG GCGGCGCCTT AGCAGTTGAC 780

CGTGATTCTT TTTCTGAAAC AATCACTGAA AAAATTAAAA GTCATCCATT AATTACAATC 840

AAAAATGAAG AGATCACCGA TATTCCGGAA GGGATTGTAA TTATTGCAAC GGGTCCTTTA 900

ACCTCTGAAA GCTTATCTCA AAAAATTCAA GAATTTAATG GTTCTGAAGG CTTTTATTTT 960

TACGATGCAG CAGCACCAAT TATTGATAAA TCAACAATTG ACATGGATAA AGTTTATTTA 1020

AAATCTCGTT ATAACAAGGG GGAAGCGGCT TACTTAAACT GTCCGATGAC TGAAGAAGAA 1080

TTCAATGCTT TTCATGAAGC ACTAGTCAAT GCAGAAGTGG TTCCTTTACG CACGTTTGAA 1140

AAAGAGAAGT TTTTCGAAGG CTGTATGCCC ATCGAAGTAA TGGCGCAACG TGGCATTAAA 1200

ACCATGTTGT tTGGCCCAAT GaAACCTGTC GGACTAGAAG ACCCTAAAAC AGGTAAACGC 1260

CCTTATGCTG TAATTCAATT ACGTCAAGAC AATGCTGCAG CTTCTTTATA CAATATTGTT 1320

GGTTTCCAAA CGCATTTAAA ATGGGGCGAA CAAAAACGAG TATTCCGTAT GATTCCGGGA 1380

TTAGAAAATG CTGAATTTGT TCGCTATGGC GTCATGCATC GTAATAGTTT TATGAATTCT 1440

CCAGAGTTGC TAAAACCAAC CTACCAATCA AAAAAACGTG ACGACCTGTT TTTTGCTGGA 1500

CAAATGACTG GTGTTGAAGG ATATGTTGAA AGCGCGGCAA GTGGTTTACT CGCAGGTATT 1560

AACGCAGCTC GTCTTGCCAA AGGGGAAGAG CCGATTGAAT TCCCTCGCGA AACGACTTTA 1620

GGAAGTATGG CTTATTATAT TACACATGCA GAAGGCAAAC ATTTCCAACC AATGAATGCA 1680

AACTTTGGgC TTTTCCCAGA GTTACCTGAA CGAATTCGTG ACAAAAAAGA ACGCTATGAA 1740

GCAATTGCTA ATCGGGCCTT AGATGTTCAA GCTCAAGTTA TCCAATCTTT AGATTAAATT 1800

AGAAAAAATA TGAGAAAAGA CCGAGCATCA TTGCTAAGAA ATGCGATGAT ACTCAGTCTT 1860

TTAATTTTAG AAAAAAATAA CAAAAGCTAA nGAATGGTAA AATTGTGACC ATATTTAGAA 1920

ATATTGGnTA ATTCCTGACC ACCTTAATTG TCTTTTGAAA AAnGGTTATG CCTAAAGTTA 1980

CTTTGTAATC GAATGGGGAG GGCTATCACC TGGGAGGAAA AGAAT 2025

(2) INFORMATION FOR SEQ ID NO: 440:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10594 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 440:

CATCATGGAA TGGACTGGGG TTCATTAGTA TTAGGAATTC TATTTGTTCT GACTGCATTA 60

ATTTCTTTCC AAAATCCTGC CGGTAATTTA ATTGCCATTG TGATGGTTTT TGCGATTTTT 120

GCTATTATTA AAGGGATTTT CGAAATTTTT GTACGTAATC GTATGAAAGA ATTGCTAGGC 180

TACAAAGCTT ATGCACCAAT CATCTTAGGG ATTATTGATA TCCTCATTGG TGTTTACTTA 240

TTGTTTAACT TAAATATTGG TGTCGCTGTG TTACCATTkG TTTTTGCAAT TTGGTTTTTA 300

TTCGACTCAA TTTTCGGCTT ATTTACTTTA GATTTTGCTA AACGAGTAAG CACAGGTTAC 360

TTCTGGTTCA CATTAATTGT TGACGTTTTA GGGATTATCT TAGGTGTAAT GTTATTATTT 420

AATCCATTGT CTTCAGCACT AACTTTAAGC TTCTTAGTTG GCTTCTACTT TATGATGTTT 480

GGTATTAGTA ATATTGTATA TGCTTTTAGA TAAAAGAAGT AAAAAAACAG GTTTTCCGAA 540

AAGGAAAACC TGTTTTTTTT TAATAACGAC GGTGGCAGTT GCCATTATTT CCCGAAATTG 600

TTCGAACAAG CCACACGATA ATCGCTACGG GAACTAAAAT TTTTACTGCT AGCCACAATA 660

AGCTTCCCAC GATACCGACA ACGGCACTAA AGAGTGAACC GACAACTAAA AATAAAAgGA 720

CCACTAAAAG AATATTTATA AACATTGGTC GTTGATTATT CATACGTACT ACCTCCAAAT 780

TTTTCCTGTT TTCTATATAC AGAGTATAGC AAGAACGCGC aACAAGCACC TCCATCGGAA 840

GGCTGATTTT TTTCTCTGGC GAATGATGGA GCCGCTGTTG GTTATTTTGG ACTTTCTTAT 900

TGArATAATC AGAATATTCT GGTGTATATT ATAAAAAATG ATTTAAAGTG AGGGATTGGA 960

TGTCGACTGA GGGATTAATC GCTTTgTTAG AAAAGAAACG AAATCATTTA AGTAAAATTG 1020

AAAATCAAGT ACTGGATTAT ATTTTGACGC ATATGACGGA AATTAGTTCA CAAACCATTT 1080

ATGAGGTGGC GCAGGCGTTG TTTGTCTCAA CAGCCACTAT TTCTCGGACA GCAAAACACT 1140

TAGGTTATCG AGGCTTCCAA GAATTAAAAT ACGCAATCGT CCAATCGGTG CAGAGCGAGA 1200

GCACAGAAAG CGATAGTCGT AGTTTTCaAG CTATTaCGCA ACAGTTGATT GCCaATGTGC 1260

aAGATAGTTT TCAACAAATG GATGAAGCCA AAGTCGCGGA AATGCTTAAA ATGATTGAGC 1320

GTGCGAATAC GATTGAAGTG TTCGGTGTTG GTGGCAGTTT TCCTATTTGT ACTGATTTTG 1380

CACGAAAGCT AACATTTTTA GGAAAAAAAG CCTATGCACG ATCTGATTGG GATGAACAAG 1440

CGGCAGCAGT TAAGAATCTG GATGGACAGG ATTTAGCCAT TTTTGTCAGT TATACAGGAG 1500

AAACTAAGGG GATTTTAGCG TATGCACGTG TAGCGCAGGA ACAGCAAGTG CCGATGATTA 1560

GTTTGATTTC GACTAAAGGA AGCACCTTGG AAAAATTGTC GACGACGACG TTATTTGCCA 1620

AAGGGACGAC ACGTTATCAT CAGCGCGTGG ATTTAAGTTC AAGAATTGCT GTCATTTGTT 1680

TGTTCGATAC GGTTTTATTA ATGTACGCAG AGGCAAAGCA ACCGCGGTAA TTTTATGCAA 1740

CACCATACTA AATGATGTAA GGGGATACAG AAAGAGTGTA GCCTCTTTTT TCTTTGTTTT 1800

GGGCTAGTTA TAATGAGGGC AATGAATGTA AACGCAATCA TTTTTAGTTT TGTAAAGAAT 1860

GAATGGAGGA GTTAAGAAAT GGCAGTAAAG AAAAAGAAAG CGACTTTCTG GGAGTTTTTC 1920

CAAGGATTAG GTAAAACGTT TATGTTGCCA GTGGCTTTAT TAGCTTTTAT GGGGATTTTG 1980

TTGGGATTAG GAAGTTCTTT TTCCAGTGAA TCCATGATTG AAACGATCCC GTTTTTAGGA 2040

AAGCCCGCAG TGAAAATTAT TTTCCAATTT ATGTCGACAA TTGGTGGTTT TGCGTTTGCC 2100

TATTTACCTG TCATGTTTGC AATGGCCATT CCGTTGGGGT TAGTTCGGAA AGAAAAAGGA 2160

ATTGCTGCGT TTTCAGGTTT CGTAGGCTAT ACGGTGATGA ATTTAGCGAT TAATTTCTAT 2220

TTAGTGCAGA CGAATCGCCT CGTGGATCCT GAGCAGTTGC GGGAAGCCGG CCAAGGAATG 2280

GTTTTTGGGA TTCAAACGAT TGAAATGGGT GTTCTTGGTG GAATTATCGC TGGGCTTATT 2340

GTCTATAAGT TGCATAATCG ATTTTATACG GTGCAACTAC CAGATAGTTT TGCCTTCTTT 2400

TCTGGCGCAC GGTTTGTGCC AATTATTACT TCATTAGTAA TGGCATTCGT GGGCTTAGTA 2460

ATTCCGTTAG TTTGGCCACT TTTCGCGCTA ATGATTATGG CGATTGGCCA ACTTATCCAA 2520

CGTTCTGGTA TTTTTGGGCC TTTTCTGTTT GGTTCAGGGG AACGTCTACT GTTACCGTTT 2580

GGGTTGCATC ATATTTTGGT TTCCATGATT CGTTTTACTG AAGCAGGGGG TTCAGCGGTG 2640

GTGGCTGGCA AAGAAGTGTT TGGCGCGTTA AATATTTTCT ATGCAGAATT GCAAAATAAT 2700

TTACCGATTT CACCAAGCGC AACAGCTTTC TTGTCACAAG GGAAAATGCC GACCTTTATT 2760

TTTGGCTTAC CAGCGGCGTC ACTTGCGATG TATCATACTG CAGCTCCGGC GAATCGACAT 2820

AAGATTAAAG GCTTACTCCT TTCGGGTGTA ATTGCTACGG CTATCACTGG CATTACAGAA 2880

CCAATTGAAT TTTTATTCTT ATTTATTAGT CCATTATTAT GGCTGTTTCA TGTCATTATG 2940

ACTGGTCTAG GTTTCATGGT AATGGCTTTA TTAGGTGTTG TGATTGGGAA CACTGATGGC 3000

GGGCTACTGG ACTTTGTTAT TTTTGGCTTG TTACAAGGAA CTTATACAAA ATGGTGGTGG 3060

GTTCTCATTG TCGGCGCTAT CTGGTTTGTT GTTTATTATT TCGTCTTTAA AACAGTCATT 3120

GTAACCTTTG ATTTAAAAAC ACCTGGACGA GACAAGGTCT TAGATGAGAC TGAATATACC 3180

GATCAAGAAG TTCAATATAA GAAAACGGGT GGCTATGATG CGCCTGGGAT TTTAGCGGCT 3240

TTGGGTGGGC AAGAGAATAT TCAAGCTATT GATAATTGCA TTACGCGTTT ACGTTTAGTG 3300

TTAGCAGATG CTAATAAAGT AGATGATGAC AAATTAAAGC AATTAGGGGC TTTAGGCGTT 3360

GTACATTTAG ATGCGCAGAA TGTACAAGTG ATTATCGGAA CCAAAGTGAC AACGGTCCGT 3420

AATCAATTAG AGATGATTTT AGGTTAGAAA AAAGACAGAA GGTGAAGAGT GTGTCGTTTG 3480

ATAAAGCAAT TGATCGACGA GGAACGTATT GTACACAGTG GGATTTTGTA GAAGACCGAT 3540

TTGGGGAAGC GGACTTGCTT CCCTTCACTA TTTCCGATAC GGATTTTGCT GTGCCAGAGG 3600

CAGTTCTGGC AACATTACAG GAGCGTTTGA ACCATCCAGT ATTTGGCTAT ACGAGGTGGA 3660

ACCATCCGCA ATTAAAAGAA GCAATTCAAA CGTGGTACCA GACAAGATTC CAGACGAAGA 3720

TTGAAGAACA CTGGATTATG TATACCCCGA CCGTTATCTA TGGTATTTCT GCTTTGATTC 3780

AATTATTGAC TAACGAAGGG GAGGGCATTA TTTTACAAAC ACCTGCCTAT GATGCATTTT 3840

TTAAAGTTAT TCAGGAAAAT AAGCGGCAAG TAGTGGCGAA TGAACTGCTT TATCAAGAGA 3900

AGCGCTACAG CATTGATTTT ATTGATTTAG AAAAAAAACT AGCGCAACCA GAAAATCGCT 3960

GTTTGCTTTT ATGTTCCCCG CATAATCCAA CAGGACGTGT TTGGGAGCAA TGGGAATTAG 4020

AGCGAATAGT TTCATTGTGC CAACAGTATG ATGTGTTTTT ACTTTCTGAT GAAATTCATA 4080

TGGATATCGT TAATAAAGGC CATGTTCATC GACCAATTAC GCAATTTGAT TATAAAAAAT 4140

CTGCCATTAT TACCAGTGGC ACGAAAACCT TTAATTTTCC AGGGCTGATT TTTGCTTACG 4200

CCTTAATTCC AGATAACGAA TTGCGTGACG CTTTTCAGCT CAAATTAAAA AACGCCGATG 4260

GTTTATCATC AACAAGTATT TTGGGAATGC TAGCAACCAT GACGGCCTAT CAAAAATGTG 4320

GTACTTGGGT TGACGAGTTG AATGATTATT TAGCGGAGAA CCAACGTTAC GTGAAAGATT 4380

TTTTACAGAC ATATTTGCCA AAAATTAAAG TGACGGAGCT GGAAGCCACT TACTTAATGT 4440

GGTTGGATGT ATCAGCGGCA GTTCCAGATG TGGCCCGCTT ACAAGCGGCT TTAGTTTCTG 4500

TGGGGAAAGT CGCTATTATG GACGGCTCGA TTTATGGCGG TAATGGGCAA CGATTTTTAC 4560

GGTTAAATAT TGGCTGTTCG CAAGCAAAAC TACACGAAGG ACTTGAAAGG ATGCGCCAAG 4620

GTTTTGAAGC TGTTTTACAA AAGGATGGCA CAGCATCAGC GCTTAGGAAT AATTGAGAAA 4680

AGATAGCATT CCGAAAATGG TTGCAGTACA ATGGAAAAAT AGAAACAAAT AAAATAACAA 4740

AAAGTAACAA CTTGTCATTT ACTGGGGGAA AAGAAATGAA AACAATTAAA ACAAAGCAAG 4800

CACCACAAGC AATTGGGCCA TATTCACAAG CAAAAATTGT GAATGGTTTA TTGATTACTT 4860

CTGGTCAGAT TCCACTGGAT CCTGAAACAG GCAAGGTGGT TGGGACAACG ATTAAAGAAC 4920

AAACCAATCA AGTACTGAAA AATATTCAAG GGATTTTAAC AGAAGCTGGT AGTGATTTTG 4980

CTCATGTAGT CAAAACGACA TGCTATTTAC AAAATATGGC GGATTTCACG GCATTCAATG 5040

AAGTCTATGC TGAGTATTTT AGTGCGGAAT ATCCTGCACG CACAACGATT GAAATCTCTA 5100

AATTACCGAT GGATGTGTTG GTGGAGATTG AGATTATCGC AGAAGTAGTG GCGTAAAAAA 5160

TACAGTTAAA AGAGTGAGAA AAAATCGTCT TAGATTTTTT TCTCACTCTT TTTTTCAATA 5220

AAACAATACG ATTCAGCCAA AATGTCTAGC CGTGCTGGAA AATTTTCAGC AAATAGGCTA 5280

CCATGGAAGA AAATCGATAA AACGAAAGTT AAATGAATAG AAAAGAGATA TGAAAGAGGC 5340

TATAAGATGA AAAAAAATAT TAGCCAACTA ACGCAtGGTT GAAAGAACAC AGTTTATTAT 5400

TGAAACTAAT TTTTCTTGGT TCTGTATTAG TTTTCGTGGC TAACCAAGTG ACACATATCG 5460

CTCAAGGCAT GAGCTGGGCG GATATTTTTT CCACGATGGA ACAACAAAGT ACGGGGCGTT 5520

TAATTGGTAT GGTCTTGGCA GGACTATTAG GGGTTATTCC gATGCTCCTT TATGACTATG 5580

TAGTCGTTAA ATTATTGGAA AAAGAAGGTA AGCCACCAAT GAAACGAATG GATTGGTTAA 5640

CTTCGGCTTG GGTCACCAAT ACCATTAATA ACTTAGCGGG TTTTGGCGGC GTGGTTGGGG 5700

CAACGCTACG GATTAATTTT TACGGAAAAG ATGTACCGCG GGGCAAGGTT GTAGCAACGG 5760

TCTCAAAAGT TGCTCTGTTC TTGATTTCAG GGTTATCTAT TTTATCTTTT GTGGCGTTTG 5820

TTGATTTATT TTTTATTCGA ACCCAAAATG TTTTTCGTGA ATATTGGGTT TGGTTATTAT 5880

TGGGCAGCCT GATTGCGCCA GCGCTCTGGT TTTTCACGTA TTTAAAACGG CGAACATTGT 5940

TTAAAACATT TTTTCCAAAA GCAGTGCTCC TATTATTCGG CGCTTCATTA GGTCAATGGC 6000

TAGGTGGGAT GTTTGCTTTC TTGATGATAG GTCGCTTGAT GCAGGTGCCA GTTTCGATGG 6060

TTTCTGTTTA CCCAATGTTT GTTATCGCTA CGTTGATTGG GATGTTAACT ATGGTTCCTG 6120

GTGGTATGGG AACGTTTGAT GTCTTAATGA TTTTGGGCTT ATCACAATTA GGGATTGATC 6180

GATCACAAGC GATCGTCTGG TTGTTGTATT ATCGTTTGTT TTATTATGTG ACGCCTTTTA 6240

TGACAGGAGT TATTTTGTTC TTGCAACAAG CAGGCATGAA AGTCAATCAA TTTTTTGATA 6300

ATTTACCTCG GTTATTTTCA CAAAAAGTGG CGCATTTTAT TTTGGTGGCG GCGTTATATT 6360

TTGCAGGGAT TATGATGGTT CTATTGTCGA CTGTGACGAA TTTGTCGAAT GTCAGTCGTT 6420

TATTCCAAGT GTTACTTCCC TTTTCTTTTA ATTTTTTAGA TCAAACATTG AATTTATTTG 6480

TAGGATTCTT GTTATTAGGT TTAGCTCGAG GCATTTCTAT GAAAGTGAAA AAAGCTTATT 6540

GGCCTACAAT TATTTTACTA GGATTTTGTA TTGTAAATAC GGTGGCACGA ACGACTTCGT 6600

GGCAGTTGAT TGCTGTTTAT GCTGTTATTT TATTGGCGGT TATTTTGGCT CGTAAAGAAT 6660

TTTATCGTGA AAAGTTTGTT TATTCATGGG GCGCTTTAAC GGTAGATAGT ATTTTGTTTG 6720

GCTGCTTGTT TATCGGTTAT GCGGTTGCTG GCTATTACGC CGCACGACCA GCTGGTGGCA 6780

ATCAAGTTAT CAATCACTTT TTGTTGTTTC CGTCAGATGA TGTTTGGTTT AACGGATTGA 6840

TTGGATTAAG TATTTCCTTG ATTGGTTTAT TCTTTTTGTA TCAATATTTA GCCGAAACAA 6900

CTGTTACTTT AGGTGAGGGA TTTGAGAAAG CGCGGTTGAC ACGCTTTTTG GAAAAATTTG 6960

GTGGCAACGA AGGCAGTCAA TTTTTGTATT TAAAGGATTA TGGTCATTTT TATTACCAAG 7020

AAGAAGGCGA AGATCAAGTA CTTTTTGGGT TTCAAATGAA GTTTAATAAA TGCTTTGTCT 7080

TGGCGGATCC GATTGGGCAA CGAGAGAAAT GGACAGCGGC GACACTGGCG TTTATGGATC 7140

AAGCTGATTT GTTAGGTTAT CAATTAGTTT TTTATCGTAT TTCGGAAGAA TATGTTATGA 7200

ATCTTCACGA TTGCGGCTTT GAATTTATGA AAGTTGGGGA AGAAGGCTTG ATTCAATTTG 7260

ATGAGCCGTC AACAGTGAAT CAAACAGCCT GGACGGAAAC AGTGACTGAA AAAATAGCAG 7320

CGGAAGCAGC GGATTTTCAG TTTGAATTTT ATCCCGAAAC GATTAGTGAT GCCTTATATC 7380

AAGAGTTAGA ACGAGTTTCT GCGGACTGGT CACGGAATCA AAAAGAGCGC TACTTTATTG 7440

GTGGACGGTT GGATCCCGAG TACTTGAAAT GCAGTAGCGT GGGCTTGGTG AGACAGAAAC 7500

AAACGGTCAT CGGTTTCATT ACTGGAAAAG AAATGGAAAA AGGCAAAAGT ATTTCGTATG 7560

ATTTGTTACG TATTCGTTCA GACGCACCTG CTTTTACTAG AGAATATTTG TTCACTCATT 7620

TTATTGAAAC CTATCAACAG CAGGGCTATC AATTAATTGA TATTGGGATG GCGCCGTTGG 7680

CTAACGTAGG GGAGAGCAAG TACTCTTTCT TAAAAGAACG CTTTGTTAAT ATTTTTTACA 7740

AGTATAGTTA TCAAATTTAT GCGTTTCAAG ATACACGCAA GCGCAAAGAA CAATATGTCA 7800

CTAGTTGGCA ACCACGTTAT TTTGCTTATC CTAAGCGGAC CAGCGTACTC TTTGCTTTTG 7860

TGCAACTTTC CTTGTTGATT ACGAAAGGAA GACATCAAAG TGTTTCGCTT GTGGAAGAAG 7920

CAATGACTGA AATTTAAAAA AATCACGCGA TGACTTAGCC ATCGCGTGAT TTTTTATTTG 7980

TCTTCTTTTT TACGGTTTCG TCTAGCAGGA ATCAAGGTTA CGTAGGCAAC GAAACCACCA 8040

GTACATAGAA CGAATAACAC GATGCCAGGA ATCAATCCTT CTGGTAAATA AGAGAACTGA 8100

ATAGTATGGG TTCCAGCGCT GACAGGAACG CTTAGAAAGG CATCTTTAAA GGCTTTTGGG 8160

GTAACTTTTT TGCCATCAAT TTTGACACGC CAACCTTTGT CATAAGGAAT TGTAGTGACG 8220

AGTACTTGAT CTTTGTCGGC GGTAAAGGTA CCTGAAGCAG AGCGTTTGCC AGTTGTTAAA 8280

TCAACGCCTT TTTCTTGCAC GGCGGAAATG GCTTTTTCAA ATGCATTGGT ATTTAAGCCA 8340

ACCACTTGCG GCTGAACAAA ACTTACTGCT TTGGTTCCAT AAAAACTTAC TTTGAAGTTA 8400

ACGGTTGTAT CTTTGGGATA GTAACCTAAG TTATAATATT GGCCAGTAAT GCCAATTTGT 8460

GATTGCTGAC TTGAGCCGTT CACGGTGACC GTCGCAGTGG AACTTTCCAA TTGAGCAAAA 8520

TCCGTTGGGA ATAAACTCAA ATAAGCTTGC GTATTGGCGG GCACATTGAC TGTATAAGAA 8580

ATCTCTTTGG CTACGTTATG TTGTTTCTCC GTAAAAGTTA CGCCAGCTGT ATTTTGCGTA 8640

ATTGTCACGT TATTTTGAAG CGTCATCGTT GGTTGATAAA AGGTAAAATA CCGTTCATTA 8700

GTATTGGCTA AAGCATTTAG TAAATTTGTT TGACTACCTA AATTATCATT GAGAGGTTGC 8760

CGGACATTGT AAATCTCTTT GTCTGCTAAA AATCCTAACG GCAAAGCATT TTCGTTACGA 8820

TAAAGCTGAT ATTTACCAGC AGCAGCTTGT CGCTCAAAGC CGAATTTCAT TGGATTGTTT 8880

TCAGCAATAT TGTATTTAAT GCCCATCAAA GCATCCATTA ACAAGGTATT ATTTTGGTAA 8940

CGAATATTTA AATTCGTGCC TCGTGAGCGG AAACCTAAGG CGTTTAGGTA TGTTGAAGAA 9000

TGTCGGTTCC GTACGGAAGA AAACATGCTA ATGCCACTGT AGCCATAGTT AATGCCGTCA 9060

TTGGCCGAAA CACCATTTAA ATTTTCCAAG CGGTAAAAGG TATCATTTTC TTTTTTTGTT 9120

TGATCGACTA GATTTTTTAA GTCAGGATAA GGCTCAGAAT ACAAGCTGCG TGATGCGTAG 9180

TTCCAATCAT CTAAAATACC AGTGACCATC GCCCGTGTGT TAATGATTGC TTCACCAGAA 9240

ACGAGCAGCA ATAACAAAAT CACTAAGTAT TGCATAGGCG CTTTTTTCAA TTGATAAAAA 9300

GCAATCCCGA AAAAGTATAA TAGTAAAAAA GTTGCCGTTA ATACGAAAGA GGTTAAAGTC 9360

ACATACGTAT AACTAGTAGC ACCTTTTGTG CCCATTGCAA GAGCAAAAAT TGCTAAAAGA 9420

ATCAAAATCA AGCCGCTTAA GACACCGAGA TCATCCTTTT CAAATTTCTC CCAACCATAA 9480

CCTGCAAGCA AAATAACTAA GAATGAAAAG AGGAAACTAT AGCGGAAAAG AAACATGTTC 9540

GGCGCATGCA TACCGTGCCA GAATAAATTT AAGGGCGTGA TGTAAAAGCT AGCAATCAAT 9600

AGGACAAACA AACTGCCAAA AAGTAGTTTG TTTTTCAATG GAACTTCTTT CGTAACGAAA 9660

TAGAAGAGAC AAAAAATCAA AGGCAACAAC CCAATGTAAA TAAAAGGAAT GGACCCATAT 9720

TTGGTTGTAT CGTAAACGCC AATCATATTT TTCATAATAA TATCTAAAAA GGCAGTCGCT 9780

TCTGTTTTAA AGGTTGTAAT TTCAGAAAGA GTTTCACCAT TGGTTCGTAA ATCTAAGACA 9840

GCAGGTAAGA CCATAATCAT GGATGCTCCA CCAGCTAAAA GAGAAGTAGT AAAATACGCA 9900

ACGATTCGAG ATTTATATCG TTGCCAGTCT GTAAACGTTC GTGCAAAATA GTACAGGAAG 9960

GAAAACAACC CAATCATAAA GCCAAAATAA TAATTTGTAA TGAAAAGCAA GAAATAACTG 10020

ACAAACAATA ATGTCGGTTT TCGTTGGTCC ATTAAACGAT GAATTCCTAA AATTACGAGT 10080

GGCAGATAAA TAAAGGCATC CAGCCACATA ATTAATTCTG AATGGGCCAC AATGAAGGAC 10140

ATTAGCGCAT AAGACACACT TAAAGTCACA TGGGACCATT TAGGAATCTT AAAGGTTTGT 10200

TTTGCCAAGA ACCAAAAACT CAAGCCAGCA CTCCCAATTT TCAACAACGT TAGGAAATAG 10260

AGCGCATCAG GCATGTTTTG ATTGTTAAAG AAAAAGACGA GCGGCGTAAA TAAGCCACCA 10320

AGATAATAGG AAATTAAGGA CAAGTAGTTT AGTCCCAAAG AAGCATTCCA CGTATAGAAA 10380

AGACTTTGTT TTCCATGAAG AACATTGTTA AAGCTTGCGT GAAAATTTGA AAATTGAGAA 10440

AAGGCATCGC TGGCTAAAAC ACTGCGACTA CTTCCTGGAT AAATCCCAAT ACTCAAATAG 10500

ATAATTGCCA TAATTAAAAA TGGAaTGAAG rAACTgscTA ACATATAAGG CCAGTTGTtT 10560

TTTAAAAACT GCcTACTtTT AtTTTTCATA AATA 10594

(2) INFORMATION FOR SEQ ID NO: 441:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1087 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 441:

ATGAATTAGA GCCTTTAATC GAGTTGGGCG TAACTGGCCT CCGGnTnCnn TAGCGGCATT 60

ACGATAGAAC AAATGGCTCA TGCTTCACAT AAGATTGATA TAGGATTGAA TGCTAGTACC 120

ATTACACTGG AAGAAGTGGC TGAGTTAAAG GCTCACCAAG CGGACTTTTC TCGCTTAGAA 180

GCTTGGCATA ATTATTATCC AAGACCTGAA ACCGGCATTG GAACAACATT TTTTAATGAA 240

AAGAACCGCT GGCTTAAAGA ATTAGGCTTA CAAGTTTTTA CTTTTGTACC AGGAGATGGG 300

CAAACAAGAG GCCCAATTTT TGCTGGTTTA CCAACATTAG AAAAACATCG GGGGCAAAAT 360

CCTTTCGCCG CTGCCGTTGG TTTGATGGCT GATCCTTATG TCGATGCTGT TTATATTGGT 420

GATCCTACAA TTAGTGAGCG AACCATGGCT CAATTTGGTT ATTATCACCA AACTAATCAA 480

TTTCTGTTAG AAGTAGCACC TAGCGAGAGC CGCTATTTGA AGCGAATACT TGGGACTCAT 540

ACCAATCGTT TAGATGCTGC AAGAGATGTT TTGCGAAGTG AATTATCAAG AACGAGTGAA 600

ATGTTCAGAA AAGATGAAAT TGCaACGATT GAGTCAGAGC AGACTGAAGC ACGTCCAGTG 660

GGaACCGTAA CGATTGATAA TGaAAAATAT GGTCGCTATA TGGGaGAAAT TCaAGTCACA 720

CTTGTnGGAC TTGCCGAAAG ATGaAAAAGT CaATACGATT ACACGAATTA TTGaAAAAGA 780

TCmAACGATA TTGCCGTTAA TCcAAGGCAG GCnATCaATT TACGCTAGTC ACGGAAGGnA 840

CGATAGAGAA TGAATTTAGA AAACTTAACA ACTGAACGTA GAAATGAAAA TACAATGGGC 900

TTAGATGAGA TGAGTGTCAA AGAAGCTTTA CAAAAAATGA ATCAAGAAGA CCAAAAAGTT 960

GCAATGGCAG TGGGTCAAGA ATTAGCAGCC ATTGAACCAG TGGTAGAAGC CATCCATCAA 1020

AAGCTTCAAT CAAGGGGGCC GCTTGATTCT CCnTGGGCCC TGTACCAATT GGCCGGTAAG 1080

GGGCCTA 1087

(2) INFORMATION FOR SEQ ID NO: 442:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 762 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 442:

ATTGTGATTT TTTCAATTCA GATTCTTCAT CGTTAAGATG TCATTAAAGA CACTGACTAA 60

GTAATCGkTG ACTGTTTCTA AATTAGGTTC CATTCTGTGC ACCATCCATA TACTTTGATT 120

ATCAAAACAT TTGATGTTCA AACTATATCG TAAATAAGCA AAGAATGCAA GAGAATTTTC 180

TATATTTTTA ATGAAATATT CTTCAAATAT TTCAAAAAAG GCGCATTTTC AAGGGACGAG 240

GCTTTTAAAA AAAGGATCAA TTACTTTGAT ACCCAAACTA TTTTCACAGA AAAAGGAAAG 300

AAGTGCAACT AAAAAGCATC ACTGGCACTC CCCTGATTGA GAAGTGCCAG TGATTTTTGT 360

TATATAATTC TTAAATTCCC ACTAAGCTAT TGCTACAACT TGACTGGATA ATACGAGAAG 420

AGAGTGAACC AGAAGCTTTG GTGCTTCTAC TCCACTCCCT TATTAAAATG TAGGACGATT 480

AATTATTGAA TTAAGTCGTA AATTTCCATT GCTACAATAT CGATGTTATC GAATTGATAG 540

CTTTGGTGCG TATCAGATTC TGTTAATTCA AATGTTTCTG TTGATTTATC AAAACTTACA 600

ATACATTTCT CTACGCCTTC ACGTTCAAAG CGGCGTACTT GTACTTCATT ATCTGCTGCT 660

TCGATCATTG CTTCCAGTCT CTTAATAATC GCCACAAGTT GTGAATGTTT CATCAGTGGG 720

cCTCCTTTTG CAAACTAGTT CTGTACCATT GTAACAAAAC AT 762

(2) INFORMATION FOR SEQ ID NO: 443:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2012 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 443:

```
CAnnACACCT GCTTCTGTTG nATAAGGAAT GATTATCTTG GACAAAATCC AAAGTCTAAC     60
CCATGTTCTG CAAATTTAGC TTTGGAATAA CTnTCTGTAA AAAAGCCACG ATGATCCACC    120
GAAAACATCC ATTTCGATAA TTTTGACATC TTGTAATTTT GTATCAATAA CTTTCACGTT    180
TCTTCGCTCC TATTCTGCTG CTAAACGTAG CAAGTATTGG CCGTAGCCAT TTTTCTTCAA    240
TGGTTGTGCC AATTCTACCA ATTGTTCTTT TGTAATATAG CCCATTCGAT AaGsAATTtC    300
TTCcAAACAa GsGACTTTTA AATTTTGACG TTTTtCAATk GTTTCAATAA ATGkTGACGC    360
TTCTAATAAG GATTCGTGTG TGCCTGkATC TAACCAAGcA AaGCCACGGC CCATTACCTC    420
AACAGAAAGT TTATTTTTTT cAAGGTAGAC TTTATTAACG TCTGTTATTT CTAATTCCCC    480
ACGTTCTGAC GGTTTGATCC CTTTTGCAAT TTCAACCACT TCATTGTCGT AAAAATACAA    540
ACCTGtCAcT GcGtAwTTkG ACTTCGGTTG TGCTGGTTTT TCTTCAATAG AAAGCGCCCG    600
CATCTCTTCA TCAAATTCTA CCACACCGAA ACGCTCTGGA TCATTTACAT GGTAACCAAA    660
GACTGTTGCC CCTGATTCTT TAGAAGCCGC TCGTTGCAAC ATTTTTGATA AGCCGCCACC    720
ATAATAAATA TTGTCCCCTA AAACTAAGCA GACACTGTCA TCACCAATAA ATTCTTCACC    780
AATAATAAAT GCTTGCGCCA AACCATCTGG GCTTTCTTGT ACCGCGTATT CGATATGAAT    840
GCCTAAATCA TGGCCATCGC CGAATAAGCT TTCAAAACGT GGTGTATCTT CTGGTGTAGA    900
GATAATCAAA ATTTCATTAA TTCCCGCCAA CATTAACGTT GACATTGGGT AATAAATCAT    960
TGGTTTGTCG TAAATTGGCA TTAATTGTTT TGATGTTGCT TTTGTTAACG GATATAAGCG   1020
TGTCCCGCTT CCGCCTGCTA AAATAATTCC TTTCATGAAC AATCTCCTTA AATACTTCTA   1080
TTTTTAGTTT CTATAAAACC AAATATCCTT CCCTATTATC GCATTAAGCA TAGGGtTTGT   1140
CATGTTTTTT ACGTAAATTT TACAAATCTG TAAAAGTTTA GCGTTGTTCC TTTAAACGCT   1200
GATGTAAAGC TTTTAATAAA TTCACACGCC AATTTAATTT TTGTTGGAAA CTTGTTTTGG   1260
TAATAATAGG GCTGACATTC GCTCCATGTC TTCGGTAAAG CACTAATGGT TCTTTAATGA   1320
GACCCGTTTG CTTTTTCCGT GCAGCTAATA AGCCAATCCA CATATCATGC ATAGGAACTT   1380
CTGGCGGAAT GGGTAAAATG ACGTTTTTCA TCTCTTGACG AAAGGCCATA CCTGCCCCAA   1440
TATAGCCACT TTTTATCGCA TTTCGCCAAA ACCCTGGTTT GACTTTTCGA AACTTAAAAT   1500
AAGAGGGATT GGTAACTTGT AAATCCGCAT CAACAATTTT CAAGTCACTA ATAACCACTT   1560
GGATGTCAGG GTGCGTTTCA AAATATTCTG TCACTGTTGT TACTTTATTT GGCAACCAAA   1620
CGTCATCTTG ATCTGCTAAA AATATCACTT CGCCTTTCGT ATGCGTAAGc GCAAATGcAA   1680
AATTAGCAAT CACTCCTTGC CCTGGaCCTT GTAACAATTG AATTTGGGGA TAATTCGCTG   1740
cATACGTTCT CAAAATTTCC AACGTATGaT CAGTAGaAcC ATCaTCTGAA ATAATTAGtT   1800
CATCyTCtTC ACTGACyTGT AAAAGAATAC TATCTAATTG TTCCGCGAGA TATTTTTCTC   1860
CATTATATGT CGCAATACAA ACTGAGATCA TAGATTATCC TTTCATGAAC TTGAGAAATT   1920
CTGCCAATGT TTGCCGCCAA ATTTGGCGAT TTTTGACAGT TAnAAATnGC TTACTACTGC   1980
```

GTAAAAATAG ATGGCGTCGA TGnGGGAAAA CT 2012

(2) INFORMATION FOR SEQ ID NO: 444:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5308 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 444:

GACATTGGTG TTATATTACT TACAAGTTGT TTTTTCATTG GTTTTTATCA GGGTCCTGTT 60

CTGATAAAAG TTAGTGAAGA GACCTTCACT AACGTAACGA CTCGCCGTGC TAGACACCGA 120

GACGGTACGT TTTTTTTATT GGTTAAAATA GCTGCCGCTT GACATCATCG CGTCATGCGG 180

TTTTTTTTGA AAATAAGTAT TGTGGGCGCT ATTTATTTTG AATTTATGCT AAAATGAACA 240

CGATTGATAT TTGAGGAGCT GAAGATAGTT GACGAACCAT ATTGTTTTAT TTGAACCACA 300

AATTCCAGCA AATACTGGAA ATATTGCACG GACATGTGCA GCGACGAATA CACCGTTGCA 360

TTTAATTGAA CCTTTAGGAT TTTCAACAGA TGATAAACAC TTAAAACGTG CCGGATTAGA 420

TTATTGGAAT GATGTAAATA TTACGTATCA TAAAGATTCG GCTGCCTTTT TAGCCCATGT 480

GGCAGGCAAA CACTTGCATT TAGTAACCAA ATTTGCTCAC CAAACATACA GTGAGGTCAA 540

TTATAATGAT GGCGAAGACC ATTATTTCCT TTTTGGAAAA GaAACAACGG GCTTACCAGA 600

AACGTTTATG CGAGAAAACG AAGAAAAGTG TATCCGAATT CCTATGAACG ATGAACATGT 660

GCGCTCATTG AATCTTTCTA ACACTGCCGC ACTAATTGTT TATGAAGCCT TACGTCAGCA 720

AAATTTTCCG AACTTAGAAC TGACGCATCA TTATGAAAAT GATAAATTGG ATTAAAAAAA 780

GTTAAAACAG CAGCGCTTTT TTCAAAAATG TCAGAAGAAA GTGGTTTTAT TTTTACTGTT 840

GCCCAATTTC CAAGCCTAGA ACAAAAAACA ATTTAGATTT TTGTTCTAGG CTTATTTTGC 900

TTGGATTACT AGAAACTTCC ATGGATATTT TCGCAGGGAA GTGATAAACT AGAAAGACAT 960

ATGAAGAGAT AAAGAAGTTT GGAGAAGTTA CATGAAATTA TATTTTACAC GACATGGAAA 1020

AACTGAATGG AATCAACAGA AACGATTTCA AGGAATGACA GGAGATTCGC CGTTATTACC 1080

AACTAGTTAT GACGAAATTA AACAATTAGG CCAATATTTA CAAGATATTC CGTTTGAAAA 1140

AATTTATTCT AGTCCGCTAT TACGTGCCAA AAATACTGCG CGGGGAATTC AACAAGAATT 1200

GACTCATCCA GTTGAAATTG TTTATACAGA TACCTTAAAA GAATTAGGTT TAGGTCGTCT 1260

AGAAGGACAA TATATTGAAG AAATGCGTAA TTTTTATGGA GAAGAATTAG ATCATTTACG 1320

GCATCGTTTA GATTTGTATG ATCCAACAAT TTTTGATGGT GAACCAATCG AGCAAGCGAT 1380

TCAACGGATT TCAGAAACAG TTGCTGAAGC GGCGAAACAG CACGAAGGTC CAGTTCTATT 1440

TGTCGGACAC GGAGCAGCGC TGACGGCGGC AATCCAAGCG ATGGCTGGTA AACCATTATC 1500

AGAATTACGG ACAATGGGCG GGCTGTTAAA CAATAGCCTC TCAATCCTTG AAACCAAAGA 1560

GGCATCAAGG AACATGCCGT ACGACTTAAC GTTATGGAAT GATACGTCTT TTTTAGCTAA 1620

AGAAAAAGCA CAATAAGCAC ACAACAGAAA GGAGGGCTTG TAGTGGTAAC ACTTTTTGGT 1680

GAGGACGAAG AAAAAGCCTT TATTGTTGGA ACTGTACAAG CAATCTTTTT TGAAAATCCC 1740

AGTAATTTTT ATAAAGTCGT TCTTGTAAAT GTGACCGATA CAAATACAGA TTATTTAGAA 1800

AAAGAAATTG TAGTAACTGG AAGTTTTGGC CAAGTTCAAG AAGAGGAACC CTATCGTTTC 1860

TTTGGTCATT TTGTGGACCA TCCTCGTTAT GGCAGACAGT TTCAGGTGGA TAGTTATCAG 1920

CAAGAAAGAC CGACTTCGGC TAGTGGGGTG GTTAACTATT TATCAAGTGA CAAATTTCCT 1980

GGCATTGGTA AACGAACCGC TGAAAAAATT GTTGAAGTAT TAGGGGAAAG TGCGATTGAC 2040

CGAATTATTG ATGATCCTAG TGTTTTGGAA GAAGTTACTG TTTTAAATGA AAAAAAACGT 2100

CAAGTAATCG TCGAAACAAT TCGCTTGAAT CACGGGATGG AGCAAGTGAT TGTGGGACTA 2160

AATCGATACG GCTTTGGTAG TCAGTTAGCT TTTTCAATTT ATCAAACCTA TCAAGAAGAA 2220

ACCCTTTCTG TGATTcAAGA AAATCCCTAT CaGTTAGTTG AAGATATTGA GGGAGTCGGA 2280

TTTAAACGAG CGGACAATAT TGCTGAACAA ATCGGCATTC AGGCGGATTC TGCTGTTCGA 2340

ATTCGAGCGG CGATTTTACA TGAAGTGTTT GAACACTCTA TTCGTTCAGG CAATACGTAT 2400

GTTCAAGCAG ATGTTTTGTT AGAAGAAGCG ATTCGGACCT TGGAAGCAAG TCGCCCAGTG 2460

GAAATCTCAC CTGATCAAGT GGCCAATGAA ATCATTACGT TAgTCGAACA TGGAAAAATC 2520

CAACAAGAAG AGACTAAACT TTTTGAGAAC AGTCTCTATT TTTCTGAATG GGGCATTGGT 2580

ACTTCGATTC AACGTTTACT TTCTCGTAAA AAAGAAATTC ATTATGAGGA AGAAGAAGTC 2640

CAAAAAAATA TCCGAATGAT TGAGAAACGT TTAAATATTC AGTATGGCGA CTCGCAACAA 2700

GCAGCGATTG AAGAAGCGAT TAAATCGCCG TTATTTATTT TAACTGGAGG GCCCGGAACA 2760

GGGAAAACAA CTGTTATCAA TGGGATTGTC TCGTTGTTTG CGGAACTAAA TGGCCTTTCA 2820

TTAGATCTCA AAGACTATAC GCAGGAGATG TTTCCAATCT TGCTAGCGGC GCCAACTGGG 2880

CGAGCAGCGA AACGTATGAA CGAAACGACA GGTTTACCTG CAAGTACCAT CCACCGTTTG 2940

TTAGGCTTGA CAGGTCGTGA AAAAAATCCG AGTTTAACCG CCAAAGAATT AGAAGGTGGT 3000

TTGCTAATTG TAGATGAAAT GTCCATGGTG GATACTTGGT TAGCCAATAC GTTATTAAAA 3060

GCTATTCCGA CAAACATGCA AGTGATTTTT GTTGGTGATA AGGATCAGCT ACCGTCTGTC 3120

GGTCCAGGTC AAGTCTTACA TGACTTATTG CAAATTAACG AAATTCCCAA ATGTGAGCTA 3180

AACGAAATTT ATCGTCAAGG AGACGGTTCA AGTATTATTC CTCTTGCCCA TGAAATTAAA 3240

GAAGGAAAGT TGCCTGCTGA TTTTCAAAAA AATCAAAAGG ATCGCTCGTT CTTTGCAAGT 3300

GATATTGGTC ATATTGAAGA ATACATTCGA CAAATCGTCA CAAAGGCAAA AGCAAAAGGG 3360

TTTACGCCAC AAGATATCCA AGTTTTAGCG CCAATGTATC GTGGCGCAGC CGGCATTGAT 3420

GCCTTGAATA AGATGATGCA AGAAATCTTC AATCCAAATG ATGGTAAGAA AAAAGAAGTC 3480

AAATGGAATG ACACCGTCTA TCGGATCGGG GACAAAGTTT TGCAACTAGT CAATACGCCA 3540

GAATTAAATG TGTTCAATGG AGATATGGGA GAAATCGTCG GAATTACTTT AGCCAAAGAC 3600

TCAGAGGACA AAGTGGACGA GTTGGTTTTA CAGTTTGATA ATAATGAAGT AACGTATAAA 3660

AGAAATGAAT GGAATAAAAT TACGCTATCG TATTGTTGTT CGATTCATAA AGCGCAAGGT 3720

AGTGAATTCC GCATGGTGTT ACTGCCAATG GTTCATCAGT ATAGCCGAAT GCTGCAACGA 3780

AATTTGcTGT ATACAGCAGT GACACGTAGT AAAGAATTAT TAATTCTACT AGGTGAGGTT 3840

TCTGCTTTTG AAACATGTGT CAAAAACGAA TCGGCTAGTC GCATGACGAT GCTAAAAGAA 3900

CGAATCGTGA ACGCAGAACA AATGACCTTA ACGACTAGAA CTCaGTTAGA AGCCTATGAA 3960

GAAGGTTTAA CGGCAGACCA TCCATTTACT GAAACGGAAA CAAAAGCTGT CTTTTATGAA 4020

ACTGAACAGC AATCAACGAA AGCCGATCAG ATCAAAGAmA CAGACGAGCA ATTAGTTGAC 4080

ACAACGGTAC AAGAGGTTTC ACTATTTGCT GATGAAGGGG AAGAGTCAAC CTCACCAGCT 4140

GAACCACAGA AAGAAGAACA ATTACCTGAA AATCCTGTAT TGTCAATTCA AGAAATTCAA 4200

GAAAATAAAA TTAATCCAAT GATTGGTATG GCTGGAACCA CTCCATATCA GTTTATGTGA 4260

AAATAAATAA TTAATTATGT TTGTCTGTTT AAGAAGTCAA TTTATTTTAT TTGATTCAAT 4320

ATTATTAAAT AGTTTTGAAT GTTTGTTTAA GCTTTGCAAC AGTTTTACAA AAAAATAAAA 4380

AAGTGAGACG ATGAATAAAT ATTTTTTACA GATATTTATT CATCACTTAA ACTGATAGAG 4440

AGGAATTAAT TGTAAAAATG AAACGAAAAA GTCTAAAATC ATTTATATAT GTAAGCGCAT 4500

TACTTGTTTT TTTGAGTGGT TTTGTTGGTG TTAGTTTATT CGTTAAGTCA GAAGTAGTAG 4560

AGGCTACAAC GATAGGTGAT GACTATCCTG TTAAATGGAA AAATCTTCCT TTAGGCGGGG 4620

CGATTGATGA TTGGAGAATG TATACTAGAT ATTGTACCTC ATTTGTCGCT TATAGATTAA 4680

GCACAGCTAA TAACTTTGAG CTTCCAAGTG GCTTTGGAAA TGCTGATCGT TGGGGAACAG 4740

AAGCAATGGC GAGAGGTTAC AAAGTAGATA AGAATCCTAA AGTAGGATCG GTGGCTTGGT 4800

GGACATCTAC GCATGTTGCG TGGGTAGCAG AAGTATCTGG CGACAATGTT AAAGTTGAAG 4860

AGTACAATTA TGGATTCGAT GGTAAATACA ATACTCGTTG GATTAACAAG AATTCCGCCA 4920

ATGGATACAT TCATTTTAAA GATATGCCAC AAACACCTGT AGGTTGGTAT AACAACGGTC 4980

ACTATGATTA TTACTATTAT GCAGACGGAA CAAAGGCGAT TGGTCTAACT TGGGTAGGTA 5040

CGAAACGCTA TAATTTCGAT AAAAATGGTG CCATGTACAA GAATGCTTGG ACAAACAGCG 5100

wTAAATATTC ATATTATAGT ACATCAGATG GCTCATTAGC GGTTGGTTTA ACTTGGGTAG 5160

GTACGAAACG CTATAATTTT GATAAAnATG GTGCCATGTA CAAGAATGCT TGGACAAACA 5220

GCGTAAATAT TCATATTATA GTACATCAGA TGGCTCATTA GCGGTTGGTT TAACTTGGGT 5280

AGGTACGAAA CGCTATAATT TTGATAAA 5308

(2) INFORMATION FOR SEQ ID NO: 445:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 745 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 445:

GCTAAGCTTC CTGTTAAAAC CACGTTATTC GGAGATTTTG TTAGTTTTCT TTTCATATGT 60

TTTCACCTCA TCTACTAAAT TTTACACTAG AAAGGCTGAT TCGTCAGCCC ATCTAGCTTA 120

AAATTGTTAA TTATTTATCC GTATCTTTTA GGAAAATACT TCCCGTTGTT GTAGAAGCAT 180

TAATTTGAGC CATTGATTCT TCCAGTACAC GTCTAAAATG TAATTGTTGA TTGCCGCGAT 240

CTTTCTTTTC ACGAACAACT TCAATATCTG TTAAACGGCT GTTAATACTA CCTAAATTAG 300

TTTTCACTTG CCCTTCAACA CCTAAGTCAT TTGGCAAGGC CAATTTGATA TTCCCATTTG 360

CAGAGCTTGC TTCAACACGG CGAAcGTTTT TTCTTTAGCT GTAATGCGGA TATCTCCATT 420

AATCAGAGAG ACACTTAGAC TTTCTGGTGC CGCGGAAATA GACACATCAC CATTGACTGT 480

TTCAATAATG TTATCTAAAA TCGTTCCTTC TAGGACTTTA ATTTCACCAT TCACACCTTC 540

AATTTCTAAC ATAGTTGCAT CAATTTTnTT AAACGTAATC GTCCCATTCG TTGATTTTGT 600

GTAAACATCT TTCGCTGTTA ACTCTTCTAC TAAGACATTT CCGTTTAATA ATTTAACAGA 660

TACATGATCG TAAGTGCGTT TTGGTAAATA GAACGTTAAA TCTGCTTTCA CCCGTTTGTT 720

TGGCACTGGG AAAGAAATCG TTCAC 745

(2) INFORMATION FOR SEQ ID NO: 446:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 445 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 446:

TGGGTACggG CCCCCCCTCG AGTTTTTTTT TTTTTTTTTT TCTTTTAAAG CATTTATTTA 60

GAAAAATAAC CGTTCCCTGC CCGCCTTGCT CTGTGCCATG GGACTTGAGC ATCTTTCTTA 120

ATACTCTTCT GCATTGGGCC TGCAACCCAC AGTGGTGGGG TCCAGCACAC GCCCCTCGAA 180

ACGCTGCGGC ACTGCTGGGA GGCCTCACGC CCTTCCAGTC GTGACAGTTC CCGTGTTCAC 240

TCCTTGTCCA GGCCTCTGTC TCCGGGTCTC CAGGTGGACA GGGGCCTCTG CGGGATGGGC 300

GCAgcTCTGC CACAGTCTGT GTGCCCCTCC CGArGCTCTG CTCTCCCCGC ATCsGAAGCT 360

CArGGCTGGA ArCGCGGGGA rCTGACATGA rGTTGGTGGA ATGCsTkGGC gCTGTaCACC 420

ACCTCsGGGG GCGAgGCTTC TTGCn 445

(2) INFORMATION FOR SEQ ID NO: 447:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1421 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 447:

TGATGTATGT TTCGTATATT CCGCAAATTT ATGGAAATTT ACATGGGGAA AAAGGCAATC 60

CGACGCAACC ATTAGTGGCA ATGATTAATT GTATTTTTTG GACGATTCAT GGGTTATATG 120

GAGACGATGG AGAAACACGA GATAAGTCCA TTATTTTTGC TAATGTGCCC GGAATTATTT 180

TTGGyTTTyT TGCGTTTATT ACAGCATTAT AGTTTCCGAC GAATTGGAAA GTATTCCAAT 240

TCTGTACATA AAAAAGAGCG GGGGAAATTT CCCACGCTTG AAGAATTAAA TCTTTTTCAA 300

TTGTTTGTAA TTAATAACAA TCACACCATT GCTGTGCACC ATTAAGTGAT TGTTTTTTTg 360

CGtTTCGTCA ATTTCGACGA CAGCGGCATT TTTTAATTGT TTCGTAACAc GCCTGTTTGT 420

GGCTGTCCGT GAACAAGAAA AGTAACAkGG GkGyCAaCGG CGAAATCCTT ACCAGCATCC 480

GATGGAACAA CAGGAGTAAC CATACCATTA AAATTTGACA TACAAAAAAC CATCCTTTCT 540

TGATATCTCT ATTATACCAC ATTTTCAGAA AAAAGTTTGT AAAGTCGCTA TTTGGCTTAT 600

TTTAAAGAGA TTTTGCATAA ATAAATAACA AACAGACTTT ACAAATAAAG TAGAAAGTCC 660

TATAATAGAT TTGTTTCAGG CCAGCTTAGC TCAGTTGGTA GAGCAACGCA CTCGTAACGC 720

GTAGGTCGTA GGTTCGAACC CTATAGCTGG CATAGATTGA AAAGACAGAT ACACCAAGCC 780

GTAAAGGTAA TAGGTATATC TGTCTTTTTT GTGTTTTGAA AGTTTTCGCT GCCTAAAATT 840

GCCTACTTAT TAAAAATTAG TATTTAAAAC ATGTTATTTT ATAGCTCGAA AATATCATAA 900

TTATCAAATT TATCGCTCAT TGATTTTTGA GAATTAGGGT ATAAATGACC ATAAGTATCA 960

ATTGTTGTAG TAATTGAAGA ATGACCTAAT CGTTCTTTTA TAGTAGTATT CTTTTCGTTA 1020

TTGTGAATAA GAAGAGCTAC GTGTGAATGT CTAAAGTCGT GTAATCTTAT AGGTTTAAGG 1080

TTCAAGTTGT TTTTTACAAT TATTTTTTTT ATTTGTTTAG AGTAATAATC GCTCGAAGGG 1140

TGTTTTTCAC TGTATTGAAA AACAAGTAAT TTGTCACTAT CAATATTTTG AAAATTATCT 1200

ATTAGTAATT TTGGTTGTAT TTCTTGCCAT TTTTTTAACA TGGCGGTAAT TTTTGAGTTG 1260

ATAGTAATAT ATCTGTTGCT GCTTTCCGTT TTTGGTGTGG TAGTTACGTA ATTACCATTA 1320

ATTAGTTTAG CACTTTTATT TATACGGACT GTTTGATTAT AAAAATCAAT GTCGCnCCAC 1380

GTTAAAGCAA TCATGnCTCC TGGCACGCAG TCCTGGAAAA A 1421

(2) INFORMATION FOR SEQ ID NO: 448:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 787 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 448:

AACATGTAAA CAAATCGTAC AGATAACACG CAAGTTGCGT GTCGAGAGAG AGGCGGTTTC 60

TATTCTTATA GAAACTGCTT CTATTTTTTT ACGAAAAAAT GAAAACTATA GAGGTGATTA 120

AATGGAACAA AAnAACGTAC GATTATTCCC TGCCGTTTTA GCGACAGGAA TTATGTCTTT 180

TGCTGGCGTA TTAATTGAAA CAGCAATGAA TGTAACCTTT CCGACATTAA CAAAGGAATT 240

TGGTGTTTCC ACCGGTACCG TGCAGTGGGT GACAACGATT TATTTATTAG TTATTTCTAT 300

CATGGTGCCC TTATCCAATT ATTTATTAAA AACCTATTCG TTAAGACGCT TATTTATTGT 360

TGCGAATCTT TTCTTTTTGA TTGGTTTAGC GATTGATGTC TATTCGCCAT CTTTCAGTAT 420

CTTGTTATTA GGCCGACTCT TCCAAGGGGC TAGCACTGGG ATTGCTTTGC CACTGATGTT 480

TCATATTATT TTGAACTTTA CCCCGCTGGA AAAACGAGGA ACGATGATGG GGGTAGGCAC 540

ATTGACCACT TCAATTGCTC CAGCCATTGG TCCAACGTAT GGTGGAATCT TAACATCTTC 600

CCTGTCTTGG CACGCCATCT TTTTATyCTn AATnCCtATT TaCTGCTTyC TTkGTTwAtG 660

GGcTtAtCTG CTATcCAGAA aTACCTGTnA AAAAAACCAC CACTCTgrAT TgGtaGTTaA 720

TgGgAtgCCC yGTaATaGtG GtTGtgAgGT TTAAGnAAAT GGCCATGTTG GCGGCTTCAC 780

TCTGCAG 787

(2) INFORMATION FOR SEQ ID NO: 449:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 901 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 449:

TTTTACCTGC ACCATTCGGA CCAACTAATC CATAAATTTT CCCTACTTCT AAATCTAATT 60

GATCAACCTT TAAATTGAAT GTATCACTTA CTTTTTTATA TAAATTATTT ATATTTATTG 120

TTTGCATTTT GCAATTCAGT CCCTTCTTGA CAATGAAATA AGTGAAAGTA ATGTAGTTAA 180

CACGAACCAA AAACAAACTA CAATAATATT AGAAACATTA ACAGTATTCT GTTCATAAAA 240

CAGTGATCTT ATCAAATTAC TCGATGGAAC AAATGGTAAA AATTCAAAAA AGTTCACTAA 300

AGCAGAAGGC AATCTATCTG TTGGATAAGT AATAGGCGAG AATAAAAGTC CACCTATCAT 360

AATTATCTGA GTTGgCTAAT CCTACTACAT TAGGTGTTAA CCAATATGCC AAAGcAAAAC 420

CTATAGAGAT CmTACATAAT TGAATTAGAC TAAGAATAGC yAGGGCCTTC ATtGAGAATC 480

yAATTTGGAC ATCGAATCTT AAATATGCCG CTAAAAATGA CATAAACACC CCAGGTAGTG 540

AGGCCATTCC CCAAATTATA AAGTCGGAAA TTAATATGTT CAACCGAGAA ACTGGCAATG 600

TCCTTTGATA GTCCACAATG CCCTCCTGTT TTGCTGTACT TACTATTTGA GCAGCAAGAA 660

CGCACCCTAC CGCAATAATC CCAAGAGTAG TTGCTCCAGA TGACAGATAG ATAGCTGTCT 720

GATTATTCAC ATCTGGTATT AGAAGTGCTA ACCCGTAGAC AATTGCTAAT GATAGAAATA 780

CTTGAGTTAA TGTAAAAACA GGCAATAAGT ATTTAATGTC TTAATAAACT CCACCCTACT 840

AAAGGCAATA ATCCTTAnTC CATGATnGCT CTATAATTTG nATCATTTCC ACnATTAACT 900

C 901

(2) INFORMATION FOR SEQ ID NO: 450:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4239 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 450:

GAATACGAAT GATACTGACA CTTnATGAGA TCAAAGAACG TGTGAAACCT GTAGCTGAAA 60

AACACAATGT TCCAGTAATC TACATTTTTG GTTCCTATGC TAGAAGCGAA GCAAAAGAAA 120

CAAGCGATAT TGATTTTGCT GTAGACATAA CTGGAGAAGC TGATGAAGAA GAaGCTTACT 180

GGAATyTATA TGGAGATTTT TTTGaAGAGT TGAAAGAAGC GGTAGGACAT GATATTGATT 240

TAGTCGAATT AGAGACATTA TTTGATTCAC AAGATAAACC AGAAAGAAAA GCTTTGCTTG 300

AAAATATTTT GAAGGAGCGT GTAAAAGTAT TTGAAAGATA GTACAGAAAA AGATTTGCGT 360

GTTCTAAAAG GGACGTTAAT CTATATAGAA GATATTGAGG AAACACTCAA TCGTTGTGGA 420

AGACAGTATG AGGTCTTTTA TCACGATAGG GTGTTTTTTA ATGCCATAGG AATGGCTTTA 480

ATGCAAGTTG GAGAACTAGG AACAGGAAGC AATGGATTAT CGGTAGAATT TAAAGAGAAA 540

AATTCAGATA AAGTAGACTG GAATAAATTA AAGAAATTAA GAAATATGTT TGCACATTCG 600

TATGATTCAA TGAATGAGCG ATTTATTTTC CGCATGGCAA CAATGTTTAT TCCGGTGTAC 660

AAAGAATTTT GTATACAAGC AATAGCAGAA TTGCAAAAGA GACGAGGTAA ATAAATGTTA 720

AAAAAATTAT TTAGATATAG AGGAAGGCGT ATTCGTTATT CTTCAAGAAA CCTGTTGGCA 780

CTCTATCGAG TGTTATTTTT TATGCCAATT TTAGCTTGTT TAGGGTATTT TGTTGGCTAT 840

AAATGGATTT ATCCTTTATA TCTTTCAAAT CCGCCTGATT GGAAAATATA CATTGTGCCA 900

GCGTTGATAA TTGTAGGTAT TTCTATAGGT GCAATTGTGC TAATTACATT GCTTATTAAA 960

GCTTCAATTA TTAATTCGGG GTATTTTTCT AAAGTGGaAC AACGTCAAGT ATTAGCaCAT 1020

ATGATTATTG ACAACGGGTA TTACACaAAA AAACAAGTGA AAAGTwsTGA yGGyAAAACA 1080

ArAGAAAAAA TCAAGTTTCC AAAAATTTAT TATAAGTCTG CAAAAAATAG TATTTTyGTT 1140

TCATTTGAAA CAGCrGGTAA TAAATTTCAG GAAAAATTTG AAACCATTGG TGGCTTTTTG 1200

GAAACAGCGT TTCATGCGGA CAATTTAAAT AAGATTGATG AAAAAGGCTT TATAACCTAT 1260

GAATTAGCTT CTGACGTTTA CAATAAGCGT ATTTGGATAA AGGATATGCA AGCGGACAAA 1320

GGAAAGGTTC AATTAATGAA AGGGTTATAT TGGTATTTTG ATAAAGACCC ACATTTATTA 1380

TTAGGTGGCG GAACTGGTGG AGGTAAAACC TTTACCATTT TATCTCTAAT TTATGCGCTT 1440

TGCCGTGTGG GTGAAATGGA AATTTGCGAT CCTAAAAATT CTGACTTGAT GGCACTAGGT 1500

AAGTTACCAA TGTTTGCAGG AAAAGTGCAT ACAGGAAAAA CAGATATTGT TAATTGCTTG 1560

AAAAATGCAG TTGAATTAAT GAATGCACGT TTTGAAATGA TGAATAATAG CCCTGATTAC 1620

AAAATGGGGA AAAATTATGC GTACTATGGT TTAAAGCCTA AGTTTATTGT TATTGATGAA 1680

```
TTTGCAGCAT TTAAAGCAGA ATTAGCCAAT GACTATTCAA CAGATGGAGA AGTTGATGAG     1740

TATTTAACGC AATTGATTTT AAAAGCTCGT CAGTGCGGTA TTTTTCTAAT TGTCGCTATG     1800

CAAAGACCAG ATGGTGAATT TATTAAAACG GCTTTACGTG ATAATTTTAT GTTTCGAATG     1860

AGTGTTGGCC GTCTATCTGA AACAGGrATC CTTATGATTT TTGGTGATGA AAACAAAAAT     1920

AAAAATTTCA AATATGTAGA GAAAATCGAT GGACAAAAAG TATACGGTCG TGGGTACGTT     1980

GCACAAGGCG GTTCTGTCGC AAGGGAATTT TACAGTCCAC AAGTACCAAC AGATTTTGAT     2040

TTCATTGAAG AATTTATCAA AATTTCTAAA GAATTAGGTT ATGAAGATGT GTCAAAAGAA     2100

GTGCAAGAAG AAGTATCTCA AAAAATAGCT AAGCATGTAG ATAAAGAAGC CCTTGCGGAA     2160

ATTGATGAAG AGTTAGAAAA AGGTCAAGCA AAGTTAGCTA AGTTAAGTGA AAAATTTGAG     2220

TAATGAGGAA GGAGATTTTT ATTGGATTCT AGCGAAATAA TCAAACGTGT TAGAGAACGA     2280

GTGTATGTTG AAGTCAAGAA ACGATATAAA AAGCCTGATT TAGATATACG TATTCGTGAC     2340

ATTCTGTATG AACAATCAGA AAGCTATGCA AAATTAGTGA GGTTTTCTAA CGGAAAGCGA     2400

ATAAAGAAAT TAGCTGACCC TGTAGAGTTT GAAAAATTTA TGGAAAATAG GGGGGCTAAA     2460

ATTGTTGATG AAGTGATTGT TGGTTTAAAT GAAAAAGGAG AAGTATAAAT GGAATTTAAA     2520

AATAGTAATC AACTTATGGA ATATTGGGAA ACAAAAGTAA AGGAATTTGA AAAAGAACTT     2580

AATAGTGGAA TTTCCCAATT GTCAAAAGAA GAGAGACAAG AaTCATGTGT TGTTGaAGCa     2640

AAAAwATATT TAGAAGaAAA aAGAAAATAT GCGAATAATC TTCATGCGTT ATTTTATGTT     2700

GAAGGTAGTT ATCAGAATTC TAATGAAGTA GATAGGAAAT TTGATAGAGA ATTCTATCCA     2760

GCGTTTTATT CGTTTTTAAG TGAAATTGAT ATTTTAGCTC mTTAATAGrA AGGaATAGTG     2820

ATTGGTATGA ACGGtAAATT GACTCTTGAA GAGTTTTATA AAAAAATGTC TTCGGAAATT     2880

TATAGAAAAG TGAAATTAAA ATACAAGAAA AAAGATTTAG ATGATAGGTT TAGTCAGGTA     2940

TTGCATAATT CGTCTTTTCG TTTTATTTAT AGAAAATATC AGAATCGTCC TGATAGTTTA     3000

CTTACTTATC AGGAAAGTGA AATGGAATTG GATAAAAATT TAGATGGTCT TGTCGATGAA     3060

GTATTAAAAG GATTAACAAA TGTTCGTCAA ATTGATTTTT CGGAATATCT GGAAACTGTC     3120

AAACGTGCAA CATTTAAACG CTGTTCTGAA AAAACAACAA AGTATTTTTC TAGTCAAGAT     3180

TTTAATTCTA TTTTTAGAGA AGAATGTTTT GACTTTGTGA AGAGTGCTTT TAAAAGAGAT     3240

TCAGACGGAG AAAGTGTTAT TTGCTGTGAT GATTTAGATA TTTTAATGGA AATAGTAGTG     3300

AAGGATTGTG TTGAAAAAGT TATGCGTGTA ATTAATAAGT AATGCAGGGG GAGCAGGTGA     3360

GTGAATATGT TCTCATTTGA AGAAGTGCTT TTTGAATTAG ATAATGCTAT TTCTTTGAAA     3420

ACGTTGAAAA ATTGGGCAAA TAAAATTGAA AAATTAACAG ATGTTCGATT TGTTCGCCAA     3480

TACGCTAAAA ATAAACAGGG TAGAAACTAC TCGTATAAAG TTTTTAGTGT AGAACAGGTT     3540

GAACAGTTAA AGCAATTGGT TCAATTAAGG AAGCAAAATG TCCCGCTAGA TCAGGCAATG     3600

ATTGAAGTGT TTATGTCTGC AGAAGAAAAA GAAAGGCAGC AAGTAATTAC TATTGCAAAA     3660
```

ATAGATTTTG AAGAAAATAA AGCAACGGTG AAAGAATTAA TTGACTTAGT AAAAGATGTC 3720

TTGTCGGATA ATGCAGAGAT AAAAAAGCGA CTGAAAGTTC TAGAAATGAA GCAAGGAATG 3780

ACAAGAGAGT AGACTAGTAG TTGCTAGGGA TTGTTGCGTG ATTTGTAGGT GGTGTAGGGG 3840

TGGCGCCCTA AGGCGCCCGC CCCTACACCA CCgTACGCAA AATTAGCAAC GTATATCCCA 3900

AATACTTCTA GTCGTGAGGT GGGCTTGTCC CACCTCAAAA GTTTCTACCC CCGTTTTATA 3960

ACATGGGGGT AGAATTTAAC ATACCATATA AAAAAGTTAC TTAAACGCTA TGTCCCACAA 4020

GGGATTGAGC GATTTTGTTT TTGTCACATT TTTTGACTAA AGTTGTCACA TTTTTTTGGA 4080

AGTAGGTTTA AAGCAGTGAG AGGTGTTGAT TTAAAGCGAT ATAGAAAAGA GCTAAAATTA 4140

AAACAGCAAG AATTGGCAAG TAAATTAGGA ATTGAGCGTT CGTTAATTTC AAAAATAGAA 4200

TCTGGAAAAC GTGTTATTTC TAAAGAGTTA GAACAAAAG 4239

(2) INFORMATION FOR SEQ ID NO: 451:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1105 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 451:

GCATATTAAT GTAGTCCATT TCTCGCTTCC TTTCATGGTG TATGCATCTA TAGTATCCGT 60

TTAAACAAAT TAAATCAAGA TAATGATAAT TAACACCTTC TTAATGTTTA GGACGTTTCG 120

ATTAAAGTGG TTATCAGAGC TTTCCGATAG TAGCGGTTAT AGGAACGTTC CTATAACGTT 180

GGCAAGCAAT GAnCCGGAAT ATCCTTTTGG GATATTCCGG GTnGCTTGTT GGGGTTTCCC 240

CAAGCCCCAG GCCGTCCGCC TGCGCGTCCG GCnGAACACG CCTTTCGGCG TGGCTTTGCG 300

CTCCCTGCGG TCGCTTCCTC CTGCACATCC GCGCAGGAGA GGGCGGCGCA AAATGCGCTG 360

TCAGGAGGTT TACGAGATTG ACAGAAAACT TTATTTCTCT GAATCGTTCC CTGTTTCGTT 420

CTTTGTGATG GTAAGCTGCC CTTGCTGGCA TTTCACAGTG ATTTTGTCGC CGACAGAAAA 480

TCCGGCCTGT TCCAGCCATT TTCCTTGTAA AACAATCTGC GTCGGAGCCT GCTTGTAATT 540

GCCATGCCCG TAACATACGG TCAGTTTGCG GTTATCCATC CGTCCCCCTT TCCCGCAACC 600

TGCCGGCATA CGATTTCAAA TACATTCCCA TCCTCAGTCC TGACAGTGAG CAGCCCGTTC 660

CGCTCGTCTG TGTCCAGGTC GGCAATCCCC ATCCCCTCGC TGTCATTCAG TAAATCAAAA 720

AGCCTGTCCT TGAAATAGTT CAGTTCCATT TTTCAGTTCC CCTTTCGTAT TTTATTGTTG 780

CTTATTAGCG AATACTTAAA AACAAATCCT GCCGCCAGTC TGCCGGTTAA GCAAAACGGA 840

TGCTGCAAAA AATCAGTGCC GTCCCGGTTG GAAAGGGGCG GCGCTTTTGC GTCACCGTTC 900

CGGTTTGGTC TGTTCTTTTT CCTGCGCCTG CTCCGGCGGT TCCAATCCTA AAATCTTATC 960

CATGTTGCGG AGTGCGGTTt GATAAGTAAT CATATCCTGC CGCAAgTCTT GTACTGTtCA 1020

TACTGTTCCC GTTTTcTTCa GCAGCGCGGC ATATTCCGCT GAAAGTGCGC TACTTTGGGA 1080

ATAGCCTTGC CATTCAGGGC ATCAA 1105

(2) INFORMATION FOR SEQ ID NO: 452:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1095 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 452:

CGCCAATTGA AGCCCCACGA ACAAAAGGTT CATTTTGTAA GGCTTTGAAA TTTTTGGTTT 60

GTCCGGTGAG CATGACTCCC CAAATAGGTA ATTTTTCGAC TTTTTTATTT yTaTyTkGTT 120

GGATAAmCTT TTGGCCTyCT TGATTGTCAG CaAACAACGG GGATTTTTCT AAGGTTTCAA 180

AAAAGTGTTG CTTGTCGTTT TCAGGATCAA ATATTTCGTT TTGATCGTCC ATGCTTAATT 240

GAAAACCATA AGGCAGCGTA CCAGAAGGAC CTTCTGCTTC ATTAACCGTT TCGGAGTAAA 300

CATAGAGCCA AACTAAGTTG ACATTGCTCG GTAACATTTt GCGGACTTCT TCATATGGAT 360

AAGGTTTATC AAAAGAAATC GCCACTTCCG CTACTTGGTT TTCCATTTTA TCTAATTCAC 420

TTAATTCGTT TTTGATACCA CCATAGTAAT TGGTAATCGC TGGGTTATAA AAAGATGCTA 480

CTTTCTGCTT GGTTTGTCGG TTGTATTCAT AAAAATTCTT AGAAGAATAA TAGCTACCAG 540

GATAGACTTC ATTCCAGTCA ATCGAATTTC TGAAAAAGCT ATATCTGCTA CGTAAATTAC 600

TCCaCTCCAC CGTGTAACCA TCAATATTTT TTGAACGATT GGTCACAACT TCGCCCCCAA 660

AAGGCGAACT ATTAGCAATT ACTTGGGAAT CTAATTGAGT GTTTGGTGAA GCAATCATTT 720

CTTGATCATT AATTACTTCT AGTAGGGTAT GATAGTTTTT CGTCGAAAGT TTATCCATTC 780

CCACAAAAAA ACCGACAAAC AATACGGCAA AACTAACAAC GGTTATGGCA CCAATCACGA 840

TAAATTGTTT TCTTTTCGCT TTACGAATAC TTTTTTTTAA ATCCACTTCT ATTCCTCCAG 900

TTCTTTTTTT AATTTTTGAC GTGTGCGATA CAATTTTTTG CGAACATTTy CAaCmGATAA 960

TTCyAAAATC mACGcAaTCT CGGcATACGA TAATTGaTAA TAATATTTCA AAAGTACTAG 1020

TTCTTGTTCC TCATCTTTTA AGGaATCGAT CAATAATAAT AACTCTTGCC CTTCCATTTT 1080

TTCTATTGAT TGCAC 1095

(2) INFORMATION FOR SEQ ID NO: 453:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5828 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 453:

TGTCGGGGAG ATGGAGTTTC CTTCCTCAGA GCTTTTCTAT TACCAGCAAC AACAATTTGC 60

TTTTCCGATT GATACTTCAA TGAATGTAGA AATTGTCACC AATAAAAAGG CATTAGCTAT 120

CGTTCGCAAC AAGAAAAAGG AATTAAAAGA TTTAGACAAC CACGCTTACC AATCTGACAA 180

TGAAACGAAT TCCAATGTTT TGGATGCTTT GGATTCCGTT GATGAATTGG AGACCACCTT 240

GGATCAGTCA AAAGAATGAT GAGGAAAAAT AGACTAGTCA AAACAATGCC TACAATCCGT 300

AAAATTTTCC TTTTCATCTC CTTCACTCCA TTCTCTcTTG aACCGGTGGG CGAGTATCAA 360

AGGCATGGAA GAGTTCTTCA AAGACCGGGT GAACCTGAAC CACACCTACC CGACCATACA 420

AATCCTGCAT CAAGCATTGG CCGTTTTCCA AATCTCGTAG GCGTTTTTGA TTACTTTCAT 480

CTTCTTGATC CAACCCAAAA AATTCCAAAG TCTTCTTGAT TTCATGAATA TCGGTACTAC 540

GAAAAGCAAA TTTCAATCCG ATATTATTCT TCAAGCGTTC ATCCAGTAAA TCGTCACTAT 600

TTTGCGTGAC AAAGTAAACA CCGGCATTCA TGGCCCGACC AGCACGAACG AGCTTATTTG 660

AAAGGGTCTT tCCTTGAGCC ACTTGAAGAA AACTCCAAGC CTcATCCAGG TcCACAATcT 720

TGAAGATTtC TCGATTACTA TGGATAAAAT CCAAGGCAAA CGTGGaAATG ACGATTAGCA 780

TAGAGACACT TAGCAACTCC ATCGTGGTAT ATTCTTCAAA AGTGGTTTCT GCATCTGGTA 840

AAACCAAATC CGCTACTTGA ATGATATTCA ACTGCTTATC TAAACTAATT GATTGGGTAA 900

TCGAACCATC TGAGAAAAGC AAATGGGCAA AATCATAATC GGTAAAGGAC TCGATATGAT 960

CGGCGATATT TTCCGCGACT GGAGAGTTCT CCTTTCGAAG CTCATTAATG ACTTGCAGCA 1020

ATCCTCGTTG ATCACTTTGG GTAACCGTCC GAATGGCTTT GCGCAATACT GGAAACTTCT 1080

CCCCATCCCG ACTGGAAATT CCTGTTAAAA AAGTCAGGAT ATCAATTGCT AAACTTTCTG 1140

AATCTTTCTT TTTCTTCATA ATCACATAAG GATCAAGTAA GCCTTGATTT TCTGGTTTAC 1200

TAGTCAAGTT GAGAATATTG ATTTCATCCG CAATTTCTGG CAAGTTTTCT TTCCAACCAC 1260

CTCGTTCTCC TTTTGGATCG ACAATCACAG CTTGGCCACC AAAAAGAACA GCATAGTAGA 1320

CCAAGAGATT GTTACTAAAG GATTTTCCGC CACCAAGAGA ACCAAGAAAA GCCGCAGCTA 1380

ATGCATTGGT TACTGAGCCC TTTACTCCTT GTGAGGCTAA TGCGGGTTTC AGGTAAACAT 1440

TACGCCCCGT ATCCGCATTG TAGCCAAAGT AAATCCCTTC CAACTCTCCT AACATTTGAG 1500

TGGCGCCAAA GCCAAGTCCT GCTAAAAAGT CAGATGTTAC ATACTGGATA TAGTCATTCA 1560

TATACCGTTT ACTAGCTGGT AGGAATTCTT CATGTAGCCC CATCGTATCG CCAAAGGGCC 1620

GGACTAGTTT CACATTGGTA TCATCATAAA AATCGAACAC TTCATCACAA CGGCGTTTTA 1680

GTTCATCCAC TGATTCTGCT CTCACCCGGA CCACATAGCT CAATTTGTAC ATGGATTCTT 1740

TTGACTGATC CAAGGTGGTC TCCAATTCAT CAACGGAATC CAAAGCATCC AAAACATTGG 1800

AATTCGTTTC ATTGTCAGAT TGGTAAGCGT GGTTGTCTAA ATCTTTTAAT TCCTTTTTCT 1860

TGTTGCGAAC GATAGCTAAT GCCTTTTTAT TGGTGACAAT TTCTACATTC ATTGAAGTAT 1920

CAATCGGAAA AGCAAATTGT TGTTGCTGGT AATAGAAAAG CTCTGAGGAA GGAAACTCCA 1980

TCTCCCCGAC AATGGTATTA ATTGTCAAAT AGGTTACATA AGACTCATGA TTCTCATGTT 2040

CAATCCTTAA GTAACGAGGT TTCTCTTCAA TCAAACAACG ACTGGGACGT AATAAGTCAT 2100

AGCGTTTCAC CAAaGTTTCT GATTTCAATT TTTTCTTTGG TAACTGAAAG GCGTATTCTT 2160

CAAACGGTGT ACCTTTTTCG CCATAAATAT GTTCAATCAA GTAAGTAAGA TCACTGGGAT 2220

TCACTCGTCT GACTTTGAAT CTCCGCGCTA ATTTACTTTC CATCAGTTTT TCAAGCTTGG 2280

TATATCGGCG AATTTCTTCA TTTGATAATG AAACAAAATC GCCCATCAAA TGATGATTGA 2340

CCTCATAGAT AAACTCTTGA AAGCCGGAAA AGAAGGATTT TTTGAGACTT TTCAAATTCA 2400

CTTCTTCATC TGTAGCAATC AGTTTGAAAC CAATAAAAAA GCGATAATCA ATCTGACTAT 2460

CGCCAATCAT GGATACTAAT GCGTCTGTCT GTAAATCAAT CCGTTGTAAT GCTATTTCTT 2520

TGAGTCTACC TATAACTTCC TTTTTAGATC GTTCCTGAGT GGCACGAATA CTACTTTCTG 2580

TCGCAATTTG TAGCGCATGA ATCTTACCTT CTCGATTTTG GGCAATCAAT TGGCGGAAAT 2640

TATCATGCAA TTGATATTTT TGTTCAGGAG ATAAGAAGGA ATAGTTATAA GGTATTAACT 2700

CATAATAAGC AAAACACTCT CCTTCTTGAT TAAAAACAAG ATTATTTTCG ATATACTTAA 2760

TCGGGTACAT AGTTCACACT CCTCACTAAG GTAATGGGTT CATCCCATCG TTTTTTCTGT 2820

TCTTTAACTT TTTTTCCTGC AAAAGTGACC TTAAATCGTA AATGATAACT GATAAAGGAT 2880

CGTAAAAATC CCAATGGTTT CTTGCCATCA AAAGTCTTTT GCGACATAAA CCAAGTAATC 2940

GCCACCGGGA TGCCAAAGTA TTTTAAAAAT GCCCCCTCTA TCACAGATAG TGGAGGCACA 3000

TGGGCAAATA CAATGACAAA CAGCAAGGAT AAAACAAACC AAGCCATTTG GTTAAAAGTT 3060

ACAGGAAAAG GCAGCTGAAA GTCATTGATT GCGTAAATGa CTTTTTCAAC TGCCCAGATA 3120

CTGGTGTAGC TTTTTATTTT TTTCATTAAC TCATCTCCAA ATATATTGTT TTTTAAATTA 3180

AATATTCTGT TATAATTAAA ATTAATTAGG GGGTAGATAA GTTTGAAACA AGAAGAGCTT 3240

AATCTTAAAT TATATGAGTT GCTATTGAAA ATCCAAAATG AATCTGACAC ACTAACAAAT 3300

GCTGACATTT CAGAATCAGA CGCAGATTTT TATGGATATC TTATCTCTAA AGAATTAGTT 3360

TATAATTTAA AAATAGAACA CTATTACAGC GGTCCAAATA TAGATGCTAG CCAAGCAATA 3420

ATCACAGATA AAGGATATGC ATTTATTAAT CAAATTGGTG AAGCAGATAA ACCTATTAAA 3480

ATGAATCGTG AGAATCGATA CAAACAACTT CAGATATTTC TCAAACGATT AGATGACCAT 3540

GACGCTGATT TAAGAACTCC AAATTATCGT GTTCAAGAAA TTGACTACTA CGACTTAATT 3600

AAATATGCTA TTGATTCAAA TCTCGTTAAG GGAATTGCTA TTCGATATGC TAGCAATAAA 3660

CCCCATCTTT TTTTATCTCA ATTTGCAAGG GTTACTACAG AGGGATATGA TGTATTAGAT 3720

ACTCCCTATC CAAAAAAATAA TACACCCGGA GCAACCATCA GTAATACATA CAATATCTAT 3780

GGTGGAGATC AAAAGGGTGC GAGTTTTGGC AGTAACAACA CAACCAATAA TAATTAAAAG 3840

GAGAGATTAG TATTGGACTA TCGAACCAAA GAAAAGAAGT TAAAGCAGCT CTTAAACTTC 3900

ATTTATGATG GAATTACAGT TACGTCAGAC TTATTGGAGT TATCTCAACA AGAATTCGCA 3960

TCATTACTGG AAAGCGCAAC AAAATCTGGc TACATCCgAG GTGCATCCAT AACAAAAACC 4020

AAAATGTACC CCATAATTTG GACTGACGAT GATATGTCAA ATGGACGTAA AGATAAGTGA 4080

TTTTTTTCGG CACCAGTAAG AGGTGCCGTT TTTGATTTAT TGGATAGTCC CGACAAATTT 4140

GTAGTAGATG GTAATGCGCT TTTCCTTACC CTTGCGGCAG CGCCCCATAC CCGGCGTTTC 4200

GTAAATCGTG ATCCGGTCAA TCAGTTCGTG CAGCATGAAG CGGTCGAGTT TTTCAAAGCT 4260

GGTGTATTTT TGGATAAGGG CGGCGAAGCG GTCAATATCA GCCTTTGCAT CCCGAACCTT 4320

TTCCAGTGCT TTTTCCAGTT CTGCAACTTT TGCTTTCAGT GkTTCCCGTT CCGTATCATA 4380

GTCCGCGATA AACATCGCAA AGATATGGTC GGGCAGTTTG CCGCTCATGT TGTCCTCGTA 4440

GGTGCGTtTC AGTTtCACAT CCAGGTCAGC CAGCCGTTTT TTTGCTGATG CAAGGTCACG 4500

CTTCATTGCT GCGGTCTGTT TCTGCTGGTC GGCGCATTTC AGTTCCATTA GCTTTTGCGC 4560

GGCTTTCTCG GTGTCCTCGC TGAACAGCCT GGCGTTGCGC CGAATGTCCT CAAGCACAGC 4620

CTGATGAAAG GTGTCACGGT TGATGTAGTG GGCGGAGCAG GCTAAATTCC CAATCGCATG 4680

cTGCGCCCGC AGGTATAATA GAAGTATTTC TGCTGGCTGA CTTTCGGAAC CATTGAAGTG 4740

CCGCAGkTTC GCAGAAGAAC AGTCCGGCAA ATAAATCGGG TTCGTCTTTG GCTTTGATGG 4800

TCTTTTTCCG CGCCCCCATA CGTTTCTGCG CTTCGTCAAA CAGTTCCCGT GATACAATAG 4860

GTTCGTGCAT ATTGGGAACA ACGATCCAGT CCTTTTTGGG TATCGGTTCC CGCTTTTTGC 4920

TGCGGTAAGA GGGTGTGTAC TGCCTGCCCT GCACCATGCT TCCTACATAG ATTTCGTTTT 4980

TCAGCATGAA CTGCACATAA GTCCGTGTCC AGAAACTGCG GCGCTCAAAT TCGCCGTCAC 5040

GTTCCGGGTT GTGCTTGCGG AAACGGGTGT ACTCTGTTGG AGACAAAATC CCCTCCCTGT 5100

TCAATGTCGT TGCAATGCTC TTTGACCCTA TCCCGGCCGC ACACATTTCA AACATCCGCC 5160

GGACAACCGG CGCGGTTTCT TCGTTGATAA TCAGTTTGTG CTTATCGTCA GGATGCCGCA 5220

GgTAGCCGTA GGGgTCGAGT GCGCCAAGAT ACTCGCCGCG CTTTGCTTTG GTATGAAGCG 5280

TAGACTTGAT TTTCGTGGAA ATATCTTTGG CGTACATATC GTTCAGGATG TGCTTGAACG 5340

GCGTAATATC CATGCTGCTG GTCTGGTTCA GGGTGTCAAC GCCATCGTTC AGGGCAATAT 5400

AGCGCACATG ACGCTCAGGG AAGTAAATCT CGGTGTACTG CCCGCACATG ATGTAGTTAC 5460

GGCCCAGACG TGAGAGGTCT TTGGTAATGA CCAGATTGAT GCGCCCTTTT TCAATATCGG 5520

CAATCATCCG CTGAAAGCCG GGACGGTCGA AATTTGTACC TGAGAATCCG TCGTCGATAT 5580

AGGTGTCAAC AATCTCCCAG CCTTGTTTCT GCACATACTC GGCCAGCAAT GCCCTCTGGT 5640

TCCCAATGCT CATGGATTCA TGGTCGGCTC CATCTTCGGA AGAAAGCCGG CAATAGATAG 5700

CGGCTCGGTA ATTCTGTTGT yCyCTCATAA AACTGCtCCT wATCAGTCCG aGaAACAACA 5760

GGGTTTACGC TCGTCTGTCA TCTATAnTAT ACCACATTAA AGCGTTAAGA GTCAAGCTGA 5820

CAGCCnTn 5828

(2) INFORMATION FOR SEQ ID NO: 454:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1246 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 454:

GCACTTGACT TGCTGGGCTG CCACTTACAG TGAACAAGCA CTTAATAAGG tACTTAGAcG 60

AGTAAGCtCA TTAAAATGTA GGGTTTCCTC ATATAAAGTC CTCACTTTGG TTAATTGACT 120

ATGTAAAAAA CAAATTGCTT GTATCTATAT TAAAAAGAGT GCTACAAGCA ATATTTTAAT 180

CGTAGTTCTA GTAAATAGCA ATACTTTTAT CAAAATTCAT GTTTGGAAAT ATAATTTGAA 240

AAATATTTTT AACTAGCAGA AAAACATTTC TAACAAAAAG TATTGGCTCA AAGTAAAGAG 300

ATCGTTTTGT TAGAAATGAG AAATGGCAAC ATAAAAGTTT TCTATTTTTC TGCAATTAAG 360

ACAGTTTCAA AAGTGAATTC ACTAGGAATA TGATTCGCAA TTGAGTATTC AAAAATATGC 420

CCATTGCCTA GGTAAGCAAT TTGTTCTGCA CGCATCAGGA AATCATTATC GGTTAAGTTT 480

AAATATTTTT TTTCCAACGG AGAAGGACGA ATCCCAGAAA TGCTGACCCG CGCACTTTGA 540

ATTTTTTGAT GCAATTGTTC CGAAATATAT TGATAAATGG AGACTTTCAA TTCTTTCATT 600

GTCAAGTTCG GAATTAAAGC AACGGGCATC CAAGTGTATT CAATAATTGA GGGCACATCT 660

TCGATAATCC GTAAGCGAAT AATTTTGTAG ACTAACTCTT CTTCTGTAAT ACTTAATTTT 720

TCAGTCACCC AAGGCTCAGG TACTTCTGTT GCAAAGCCTA AAACCACGGA TTCAACTTTT 780

CCATCGTAGG TTGAGTAAGT GCCGGTTAAC GGAGAAAATT TAGTAATGGT TCGATTTTGA 840

AAGTCTTGAA CAAAGCTGCC CGAGCCACGT CGTCTGACAA CATAACCTTC TTGCACAAGC 900

AAATGTAAGG CTTTTGTAAT GGTTAGTTCG CTGACCTGAT ATTTTTCTGC AAGAAGACGA 960

CCTTGGGGTA ACTTTTGGTT AATTGGATAA ACACCTGCTT CAATATCCTG TTTTATAGCT 1020

GCATAAATTT CTTTATAACG TGTATTTGTC CGCATCAGAG TCcTCCTTTA TCACCAGTAT 1080

TTCTCTCATT ATACCATGAA TTTAACGTGT TAAAACTATA TATTTATTTT TAACGTAACC 1140

ACTTAATAAT CAAGTACTTA AATAAAAGTA GTTGACTTAT TCTTTTTAAA ACTATATAGT 1200

TTTAAAAGTA AACAGAAGAA CTTATTGATC TGATTGGTAC TTTTTG 1246

(2) INFORMATION FOR SEQ ID NO: 455:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2211 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 455:

CAGTTTAGGT TGGTCAAGAG TnATTGGACA CCAGTTATTT GAAAATTGGT GGAATAAATT 60

TCTTCCCTGC TTCGTCATTA TCTACAGCCA AAACAAGtTG GACTTTTTCC GtTGTTAGTT 120

TCGACTTTTC TAATTGTTCA AGAACGGTTG GCTTTTCTTC CTCTTTGACA TAAGAGCCAA 180

```
TTTCATTGGC AATAAGCGTA GATACTGCGC CCTTTTTAAG TCCATTCGCA CAAAATAATA     240
CCGCATCTCC AATCGTCTCT TTTTTTAATT CGTAATAACT CATGAGATCT ATAGGCGCTT     300
CAAAAACAAT GATTTTGATT GGCTTTTCTG ACGTAGCTTC TGCAATTTTT GGTGGATACC     360
CTACACGAAC AGTCAACCCA TAATAGCCAT TTCCCCAGAC ACGCTTTAAC CGTCCTCGTT     420
CGCCATGTAG CTTCTTATTT TCCCAAATAC CTTGTAAGGC AACGCCTTTA ATTTTTTCTT     480
CAAGACCAAC GTGCTTGAAG ACTATTACTG GTTCACTTTG TCCGGTCTCT TTATCTGTAT     540
ACGTGGATTG TGCCATGAGG TTTTGTTCAA AAAAGAAATC AATCGTTTCT CTTGATAACT     600
TACGTTCCTG AAGCAAGTAA TCAATGGTGG CATTTTGTTG TGTATGTTCT TTCATATAAT     660
AATGAAAGTT CGTCGGAGTA GGTTCTTTCA ACGTTTCAAT AGCCCCTGCT TCTACGGTCT     720
GTAAATACTG CAATGCTTCA GCGTGTGTAC ACTCTTTTAT CACTTGAACA AGCTTAATAC     780
TCCCACCACC AACTTGGCGA GAGTTCCAAT AAAAAGAATT CTTCTTAGGG GTTAAAACAA     840
AACTATCGTG CTCTGACCAG GTATAACTTT GTCCCTGTCG TACAAGTGTT ATTCCTAAAC     900
TTTCTGCGAC ATCCACAATA CTTTTTCGTT CAATTcTcTA ACACTTTTTG TCGTtGTGCA     960
TGTGTCATtG GCATTTtCTG TtCCTCCTTT TTCTATTTTC TGAAATAAAA AACGCACGTC    1020
TTTTTAAACG TGTGTTTCTT CTTtCAGTTC ACTTGTAATT TCTCCATGGT CAATAATATC    1080
kGTTGCCGAG ATTTTATCCA AAATAGCTTC TTCATCAGTC ATATATCGGC GCCACGTATA    1140
CCACTTATCA TCAGTCGGTC CTGTTGCTAG TTCATCTGCT CGGTCATGCT GAAAAGCATA    1200
GAATTTTTTA TCTTTAAAAA CATGCTCTTT TGATATAAAA ACCAGACACT CGTCTCCTCC    1260
AATAAGCGCA ACTTCACTCC TGTCAATCAA ATCGCGCCCG TATTTATCTC GATTTTCGGA    1320
TCCTCCACCT TGATTTCCTT TATTCATGCT ATGTTTCCGA ATAGATAATG TTTGTTTCCC    1380
TGCACGCGCA GACAGATACT TGGTGGTTTC TTCTTCGTCT CCACCAAGGT ATAGTAAACA    1440
CGCACAGCCA TTTAAAATGT TTTGCCACCC ATTCCGATAA AGCGTTTGTA ATTGGTTCAA    1500
GGACTGTAAA ACAATGGTAA ACGACATTTC TCGCTTACGA AACGTGGTCA ACATTTTCTC    1560
AAAGTGAGGA ATTCGGCCAA TATTGGCAAA TTCGTCAATT AAGAAGCGGA CATGTAGGAG    1620
TTCTTTTCCT TCTTCTAGTA CTCGTTTTCC TAAACGTACT TGGTCTGATT TATTTCGCAA    1680
CACTTCTCCG ACTGTCGCAA ACATCAACGC TGCAATAAAA TTAAACGACG TATTAGTTTC    1740
AGGTATCGCA ATAAACACCG CTGTTTTTTC TTCATTCCAG GAGTCAATAT CCATCGTGTC    1800
TGTACGAATC ATATTGACCA CTTCTTCATG GTCAAAAACA GAGTAACGTG CAGCAGCTAT    1860
CCCTAGAACG CTTGCTCTCG TTTCACTCTC ATACAAATTG TTAAACAAGG TCCATTGTCG    1920
GTAAGCGTAG TTGTCAGGAA TGGCCTCATT TAACTCCTCA AACCATTCTT CCACTGGGcT    1980
AGGTACATTT TyATCTTTTC GTTTTAGGTG ACGTAACATG TCCGCAATCA TTCCTAAATG    2040
TGGCGTATAA TCATTTCTGC GACCGTCAAA CCATAAATAC GCAATAAGAG AGCGAACCAA    2100
TAAGCCTTCG GCTTTTATCC AAAAATCTTC CCCGGTCGTT TTTTCCCCTT CTTTTGTTGG    2160
```

CTTCGGTGAC AGCTTCTAAA AACTCGGTCC ACGTCTAAAT CnGAGTGCAT A 2211

(2) INFORMATION FOR SEQ ID NO: 456:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 826 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 456:

TACCGATTGA ATTCATTGCT TCCGTTTGCG GAGAATGTGT CAACACATAT GATAAAACCG 60

TTTGTAAAAT AACTGATGCT ACAGCGAAAA CTTTTAAAAA ATCGCTGATA TTCAGATTTA 120

TTGGTTGTTC TTTACTCATG TACGTCCTCT TTCTTTCAAT AATTTAGCGT GTAAGTAAGC 180

GCTCGGATTA AGAACTTACT CACACGCTAT TCACGCTTCT TTGATTTACA TTGCTAGCGA 240

TTCATAATTG GCTAACAAGA TGCCGCGTTC TTCAATCGCT TCTTTGACTT CTTGAGACGT 300

TAATATTTCC ACTTCACGAA TTCGAGGCAT ACAATACCCT GATTGATTTT GTAAAATAGT 360

ATCGATAAAC GCAGGATGAC AATTGATTTC AAAGACCTCT CCTTCACTAC AAACAACCAT 420

ATCAAGCAAT TGCAAGATGG TTTCAGTTGA AATCGCCTTG TCATAAAATT GGTACAACAT 480

TTCATCTGGC GTGCGAACAT CCTGATATAA CTCTAAATAG TCTTkTGTTT CAATACTTCT 540

TGAGGCGTTA CGTAAAGGTA ATTGATATTT GCGAGCAAGT GCCAAAGcTA CGCCCAATAA 600

CTTTTTATTT TTTCCGTGAA CATTATGATG AGAATCAATA TGATCAGGGC GTCTGCCACT 660

TTTCATAAAT GAAATAATTT GAGCATCCCA TTCGTTATAA ACCTCTTCTA AATTGACTTT 720

TTCTTCaAAA ATCGATTGAT GCCAArAATA ACCAGCkTCA TCTACTAAGG AAGGGACCAT 780

TTCTCGAGGT AAAATAGGTT TTGCTTGGTn CAATGTCAAT GTTAAA 826

(2) INFORMATION FOR SEQ ID NO: 457:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1332 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 457:

TAAGCTAGAT GCGATAGAGT TATATTTAAC AACAGAGTTG TCCTATCAAG AAGTAGCTAA 60

TCTTCTAAAG ATTCGCAATC CATCATTGAT TGCGAATTGG CTTAGAACGT ATCAGAAGTT 120

TGGTATAGAG GGACTCTCTA AACAGAAAGG ACGTCCACCT ACCATGTCTA AAAAGAAACG 180

AAATGAACCC CAACCACTAC CAAGTGAACG ATCTCAAGTA GAGAAAyTAG AAAAAGAAAA 240

TAGAATGTTA AAAAkTGAGA ATGCCTATTT AAAAGAGTTG AGCAGACyGC GTTTAGAGGA 300

CGAATAAAAA ATGGACGAAT CAAACGAATT ATTCATAGTT TCCAAGGACA ATTTCGATTA 360

AAAGTTATTC TGGAAACCTT AAGGTTTCCA AATATCACTT ACATGTACTG ACAAAAACGT 420

CTCGACAGAA CAGCTCTAAC ACAAATAATC GAAGAAGAAA TTCAAGTGAT TCGTAAAAAA 480

AACATAGGCA TTGTACAAGA ATTTGGATAA ATCTACCTTC CAACAAATTC TTAACACTAA 540

CTTAGTCACG AGTTAGTGAA AAGAGTATCA ATCTTAACAA AGTTATTCTT GAAAAGAACG 600

GGTATCTTAG TTGGTACCTG TTCTTTTTAT ATGATAGATA TTGAAATTTA AGGATAAAAT 660

CTTTAAAATA TGGATATCTT TTAAAGATTT TATAATTGAC CTTCTTATGA GAGATATCTT 720

GTTCAAAAGA AGGGAGCTCC TATCTTTTTC CAAGTTTTAA GTAGGAGTTT TTATATGTAA 780

TACTTTGCTT AGTTTAAAGC TATCGTATTA TATGAAGAAA ATACTTATTA AGGTATCaTC 840

aATTATGAAT GTAAAAAAGA AAACTATTTT TATAAATnAT TTATATAAAA AGTATTGCAT 900

TAAAAAATAG GTTGTTTTAA AATAAAGTTG TCATCATAGT ATGACTCTTT TCGATAGAAT 960

CTCCTATTGG AAACGATATG GATTCTTTAT TGAAAAAGAA TGTTGGAGTA ATGATTAGTC 1020

AAGTCCAGAC TCCTGTGTAA AATGCTATAC AATGTTTTTA CCATTTCTAC TTATCAAAAT 1080

TGATGTATTT TCTTGAAGAA TAAATCCATT CATCATGTAG GTCCATAAGA ACGGCTCCAA 1140

TTAAGCGATT GGCTGATGTT TGATTGGGGA AGATGCGAAT AATCTTTTCT CTTCTGCGTA 1200

CTTCTTGATT CAGTCGTTCA ATTAGATTGG TACTCTTTAG TCGATTGTGG GAATTTCCTT 1260

GTACGGTATA TTGAAAGGCG TCTTCGAATC CATCATCCAA TGATGCGCAA nTTTTGAATA 1320

TTTTGGTTGA TC 1332

(2) INFORMATION FOR SEQ ID NO: 458:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6021 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 458:

CTATATAATT AAAGTATTTT AGATCTTAAC TAGGATCTAA AAACTGTATA TAAAAAACCA 60

TCTATACTTT ACAAAAATAA TATTTTTGGA GTTATAAATA ATGAATGAAT TTTAATAGTA 120

AATTAAATTT TGTTATACTT CAAGTTCTGG GTTGCATGTG CAGAATGTGC AGGTGGCTCC 180

AGCAAATTAT wATwCmCCmC AmCAAaTCGT GATTGGTGGT GAGGTTGCTG CTGTTGATCA 240

AGCGATGACA CTTCTCAAAG AAGCTGGTGT GAAGCGAATG ATTCCGTTAA ATGTGAGTGG 300

CCCTTTCCaT ACGGCGCTGT TACAACCaGC ATCaAAAAAA TTGGCTCAGG ATTTAGCAAA 360

ATTGAACTTT CAAACGATGC AAATTCCTGT CATTAGTAAT ACGACTGCCG AAATTATGCC 420

CCAAGAGGCA ATTCAAGCGT TATTGGAAAA GCAAGTCATG TCTGCGGTAC GTTTTGAAGA 480

CAGTATCGAA ACGATGAAGG CTATGAACGT AGGAACGATG ATTGAAGTTG GTCCAGGGAA 540

AACATTAACT GGTTTTGTTA AAAAAATTGA CAAAACAATT GAAATGCACC GTGTGGAAGA 600

TGTTGCAACA TTAACAGAAA CGTTAACAGC GCTTACTGGG AGGTAAAAAA TGGAATTAAC 660

AGGAAAAAAC GTATTTATTA CTGGAAGCAC TCGTGGGATT GGTAAAGCAG TTGCGTTAgC 720

```
TTTTGCGAAA GAGGGCGCCA ATATTGTCTT AAACGGTCGG AGTGAGATTA CGCCAGAACA     780
ACGACAAGAA ATTGAAGCCT TTGGGGTAAA ATGTATTGGC CTTTCTGGAG ATATTTCCGA     840
TTTTGATGCG GCAGGTGAGA TGATTCAAGC AACAGTTGAC CAATTAGGCT CGATTGATAT     900
TCTGGTGAAT AACGCTGGGA TTACGAATGA CAAATTATTA TTACGAATGA CAAAAGAGGA     960
TTTTAATGCC TGTTTAGATA TTAACTTGGT AGGAACTTTT AACATGACCC AGCAAGCGGT    1020
TAAACGTATG ATGAAACAAC GAAGTGGTCG GATTATTAAT ATGGCTAGCG TTTCTGGTTT    1080
AATGGGAAAT GTTGGTCAAG CAAACTACGC TGCGAGTAAA GCCGGTGTCG TTGGTTTTAC    1140
TAAATCGGTG GCACGAGAAG TTGCGCCGCG CGGTATTACC TGCAATGCGA TTGCACCAGG    1200
GTTTATCCAA ACAGAAATGA CGGATGTTTT ATCGGAAAAA GTTAAAACAC AAATGAATGC    1260
GCAAATTCCC TTACAAACGT TTGGGCAAGT CGAAGATGTC GCAGCTACAG CGATTTTCTT    1320
AGCTAAAAGT CCCTATATTA CTGGGCAAGT CGTGAATGTC GATGGCGGCT TAGTCATGCA    1380
CGGATAAACG AAAGGAGTCA AAAGGATGAA TCGAGTAGTT ATTACCGGTT ACGGTGTCAC    1440
CTCGCCGATT GGCAATGAAC CTGAGACTTT TTTAGAAAGT TTGAAAACAG GGAAAAATGG    1500
TATCGGACCG ATTAGTAAAT TTGATGTTAG TGAAACAGGT GTCACCTTAG CAGCTGAAGT    1560
GAAAGACTTT CCAATGGAAA AATATTTTGT TAAAAAAGAT GGCAAACGCA TGGATAAGTT    1620
TTCGTTATTT GGCATTTATG CGGCCTTAGA AGCAATGGCT ATGAGTGGCT TGGATACCAG    1680
TCAACTAGAT GTGGATCGGT TTGGCGTTAT GGTGGGTTCC GGAATTGGTG GTTTGGAAAC    1740
AATTCAAAAT CAAGTAATTC GGATGCACGA CAAAGGACCT GAACGTGTGG CACCGTTATT    1800
TATTCCAATG GCGATTGGAA ACATGGTGGC GGGAAATATT GCTTTACGTG TTGGTGCCAA    1860
AGGAATTTGT ACTTCAACAG TGACAGCTTG TGCTAGTGCA ACTCATTCAA TTGGGGAAgC    1920
GTTCCGCAAT ATCAAACATG GCTACTCCGA TGTCATTATT GCTGGTGGAG CAGAAGCGCC    1980
TATTACAGAA ATCGGTATTT CAGGATTTGC TTCCTTGACC GCTTTAACAA AGGCGACTGA    2040
TCCaGAAAAA GCCTCAATTC CCTTTGATAA AGAACGTAGT GGATTTGTGA TGGGAGAAGG    2100
AGCGGGTGTC TTTATTTTAG AATCATTAGA TCATGCTCTA GAACGAGGCG CAACTATTTT    2160
AGGTGAAGTG GTTGGCTATG GTGCTAATTG CGATGCCTAT CATATGACAT CACCAACACC    2220
AGATGGGAGC GGCGCAGCGA AAGCCATGGT ACTAGCAATG GAAGAAGCGG GGATTTCTCC    2280
AGAAAAAATT GGCTATATTA ACGCCCATGG AACAAGCACA CAAGCCAATG ACAGTGCAGA    2340
ATCTAAAGCA ATTGAATTGG CTTTAGGTGA TGCCGCAAAA ACGGCGTATG TAAGTAGTAC    2400
AAAGAGTATG ACGGGACATC TCTTAGGCGC CGCTGGTGGA ATTGAAGGAA TTGCTACCCT    2460
AAATGCTTTA CAACATCAAT TTATTCCGCC AACCATTAAT GTGGAAAATC AAGATGAAGC    2520
AATTACGGTG AACGTTGTTT TAAACGAAAG CAAAGAGCAT AAATTTGaTT ATGCTTTAAG    2580
TAACTCGCTA GGCTTTGGTG GCCaCAATGC AGTTATTTGC TTAAAACGCT GGGAGGATTA    2640
ACATGCAGTT AGAAGAAGTA AAAGCACTAT TAACACAATT TGATCAATCA ACTTTAACCG    2700
```

AATTTGATTT ACGAGAAGGC TCGTTTGAAT TATATATGAA TAAAAATACA GTTTCTGGGC 2760

GAAGTGCCGT AGAACCAGTG GCGCAACCAC AAGAAACACC AGTAGCCGCT AGTGGTGTGT 2820

CAGTGCCTGT TGAAACAGTT TCTGTCGTTG AAGAAACGCC AACTAATACA CCGACAGCCA 2880

ATGAAAAGAC AGAAGAAATT ACGTCACCTA TTGTAGGAAT CGTTTATTTA CAACCAGCAC 2940

CAGATAAAGA GAACTTTGTC AAAGTAGGAG ACACTGTTAA AACGGGTGAC GTTGTGTGTA 3000

TCGTTGAAGC GATGAAaTTA ATGAATGAAA TTACGGCGAC AGTTGACGGT GTGATTACTG 3060

AAATATTAGT CAACAATGAA GATGTGGTTG AATTTGGTCA GCCGCTATTC CGAGTTGCTA 3120

AAGGAGAATA AAACATGAAA TTAACaATTA CAGAAATTCA AGAAATTTTA CCACATCGCT 3180

ATCCCTTTTT ATTATTAGAT AGTGTTGAAG AAGTCATTCC CGGGGAACGC GTTGTAGCGA 3240

AAAAGAATGT AACGGTTAAT GAGCAAGTTT TTCAAGGTCA TTTTCCTGGG AATCCTGTTT 3300

TGCCAGGAGT TTTAATTATT GAATCGTTGG CTCAAGCAGG CGCCGTGGCA TTACTTTCAA 3360

TGCCTGAATT CAAAGGGAAA ACGGCATACT TTGGTGGATT AGATAAAGCA AAATTTAGAC 3420

ArAAGGTAAC ACCAGGCGAC ACATTAATTT TAGAGGTTGA ACTTTTAAAA GTCCGCGCTT 3480

CTGCTGGAAT GGGTAAAGGT GTGGCAAAAG TTAATGGAAA AAAAGTTGCC GAGGCTGAAT 3540

TAACCTTTAT GATTGGATAG GTGAAAAATG TTTTCGAAAG TATTAATCGC AAACCGCGGA 3600

GAAATCGCTG TTCGTATTAT TCGTGCCTGT CGCGAATTAG GCGTTCAAAC AGTTGCTGTT 3660

TACTCTGAAG CGGATCAAGA AGCGTTACAC ACACAGTTAG CAGATGAAGC CATTTGTATT 3720

GGACCGGCGA AAGCAACTGA TTCTTATTTA AATGTTCAAG CTGTTTTAAG TGCAGCGATT 3780

GTTACCAATG CTGAAGCGAT TCATCCAGGG TTTGGTTTCT TATCTGAAAA TAGTCAATTT 3840

GCTTCAATGT GCGAAGAATG TAACATTACT TTTATCGGTC CTAAAGCAGA AACGATTGAT 3900

GCAATGGGCA ATAAAATAAA CGCACGTCAA CTGATGCAAA AAGCCAAGGT ACCTGTTATT 3960

CCTGGGAGTG ATGGCGTCAT TGATTCTGTC GAAGAAGCGT TGACGATTGC TGAAGAAATT 4020

GGTTACCCAG TGATGTTAAA GGCAGCTGCT GGTGGTGGCG GAAAAGGAAT CCGAAAAGTT 4080

CTTTCAAAAG AAGAGTTACC GAAACACTTT ACTTCAGCGC AgcAAGAAGC CAAAGCCGCA 4140

TTTGGTAATG ATGATATGTA TTTAGAAAAA ATTATTTATC CAGCACGTCA TATCGAAGTG 4200

CAAATTTTAG GTGATCAATA TGGGCATGTG ATTCACTTAG GGGAACGTGA TTGCTCACTA 4260

CAAAGAAATA ATCAAAAAGT GTTAGAAGAA TCACCATCGA TTGCTATTTC TGAAGAAAAG 4320

CGTCAGATGT TAGGCGAAAC GGCCGTTCGA GCGGCACAAG CTGTGCATTA TGAAAATGCC 4380

GGAACCATTG AATTTTTGAT GGATCCAGCT GGCGACTTTT ATTTCATGGA AATGAATACC 4440

CGCATCCAAG TCGAACATCC AGTTACGGAA ATGGTAACAG GGATTGATTT AGTCAAAGCA 4500

CAATTAGAAA TTGCTTCAGG TGAACCTTTA GGCTATACCC AAGAGGATGT TATCATGAGT 4560

GGGCATGCGA TTGAGTGCCG AATTAACGCT GAAAATCCTG CCTTTAATTT TGCTCCTTCA 4620

CCAGGTAAAA TTCAAAATCT CTTGTTGCCT AGTGGTGGAA TGGGCTTGCG TGTAGACAGC 4680

GCTATGTATT CAGGGTATTC AATTCCGCCT TATTATGATT CAATGATTGC AAAAGTCATC 4740

GTACATGGTG AAAATCGTTT TGATGCCTTG ATGAAAATGC AAAGGGCTTT AAATGAAATT 4800

GTAACAGAAG GTATTATTAC GAATGCCGAA TTCCAATTAG ATTTAATTAC ACATGACAAT 4860

GTTTTGACAG GTGACTATGA CACAAGCTTT TTACAAGAAA CATTTCTACC AAATTGGGAG 4920

CCAGAAAGCA ATCATTAAAG AAAGAACGCC GCAAGAGCTG TGCTCTTGCC GCTAAAATAC 4980

GAGGAGTAAG CAAATGGCAT TATTTAAAAA GAAAAATTAC ATTCGCATTA ATCCAAATCG 5040

TGCTCATGCA AATGATGCAT CAAAGAAGCC TTCTGTTCCA GATAATATGT GGGCAAAATG 5100

CCCTTCTTGT AAGCGAACAC TGTATACAAA AGAAATGGGT GCAGAAAAAA TCTGTCCGCA 5160

TTGTGGCTAT AGTTTTCGGA TTGGTGCATG GGAACGATTA GCGATAACCG TTGATGAAAA 5220

AAGTTTCCAC AATTGGGACA GTGAATTAGT GACAAAAGAT CCACTAAATT TTCCTGGTTA 5280

TCTTGAAAAA ATTGAAAAAA TGCAAGAAAA AACTGGGCTA GATGAAGCTG TCTTAACGGG 5340

AGAAGCAACA ATTGAAGGAC AAGCTGTTGC AATTGGAATC ATGGACGCTA ATTTTATCAT 5400

GGGCAGTATG GGAACGATTG TTGGTGAAAA AATCACACGC TTGTTTGAGC GGGCGACAGA 5460

AAAgCATTTA CCAGTAGTGA TTTTCACTGC ATCTGGTGGT GCCCGTATGC AAGAAGGAAT 5520

TTTTTCATTG ATGCAAATGG CGAAAATTTC GGCCGCTTTG CAACGGCATA ACAAAGCAGG 5580

CTTGCTGTAT CTTACGGTAT TGACTGATCC AACGACTGGC GGTGTTACCG CAAGTTTTGC 5640

GATGGATGGC GATATTATTT TGGCAGAGCC TCAGAGTTTA ATCGGTTTTG CTGGCCGCCG 5700

TGTAATTGAA CAAACGATTC GTCAAGAGTT GCCAGATGAT TTTCAAAAGG CCGAGTTTCT 5760

TTTAGAACAT GGTTTTGTAG ATCAGATTGT GCCAAGAAAT CTTTTGCGTC AACGATTGAG 5820

CGACTTGTTA CGTTTACATT CTTTGGAAGG GTGGCGCTAA CATGGAAAAG AAAACAGCCA 5880

ATGATGTTGT TACCTTGGCA AGAGCCCAAG ATCGTTTGAC AACTTTAGAA TATATAGAAG 5940

CAATTTTTGA GGACTTTTTA GAATTTCATG GGGATCGTTA CTTTGCCGAT GATTTAGCTG 6000

TAGTCGGTGG TGTTGCTACT T 6021

(2) INFORMATION FOR SEQ ID NO: 459:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2813 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 459:

TTTTTTAAAC CGAACAATTT TGGCAAAAAA CGAAAAAAGA ATTAACAGAG GGGACACGAT 60

TTATCGCAAT TGGTTATGGG CCCAATATTG GTACGGCGAA AGAGTTTGAG ACGAAGTTTA 120

CAGAAACAGT TCGCTTACCT TCCCAAGGAT TtGAAGTCGA AGCCTATATG CATGGGCCTT 180

ATCTTGAAGC GGATAATACA CATCTTTTGT TTTTTATTGA AAATCAAAGT CCTATTCAAG 240

CACGTAGCCA AGCCTTGAAA AATTATCTAG CCAGTTATGT TGGCCAATGT TACACAATTA 300

CAACGCGTTC AACGAGAGCA TCAAATACGT TAGATTTAGC AAGTGATGTC GAAGAAAAAA 360

TTAGTCCTTT ATTGTTAGTT ATTCCATTTC AATACTTGGC GCATCAAGTA GCTACAAGTA 420

AAGGAATTGA TTTATCTCAA CGGATTTTTG ATGATTTTGA TTGTGTCCTA AAAAGTAAAA 480

TAGCAAAGAA AGAGGGAAAT CAGCCATGTT AAAATTTAAT GAAGAACAAC AAATCAAAGA 540

CATACGAGGT GCCTTAGCCT TGCGTTCAGA TGTTGaAAAA GTTGTTGaTC AAGTTTGGAC 600

GAAAGGATTT GATGCGGTTT ATTATATAGG AATCGGCGGT ACTTATGCTT CaGCCcATGc 660

AAGCaGTTaC TTATGCGaAT GGAAAAAGTA ATTTACCAAT CTATGTTCAA CATGCTGCTG 720

AGTATTACAC AACAGGCAAT AAACGTTTAA CAGAAAAATC ATTCGTCATC TTGTCTTCCG 780

TTACTGGGAC AACACAAGAA GTGGTGCAAG CCGTTCAAGA AATCAAAAAG GTCGGCGCAA 840

CGCTATTTGG TTTTATTGAT GCAAAAGATA GTCTTTTAGC AGATTTATGT GATTATGTGG 900

TGACTTATCC AGCTCCAGGA ACTGAGCAAA TCAAGTTCTT TATGGTAGCT GATCGTTTAA 960

TGCAAAAAAA CAACGAATTT ACAGATTACG AAAATTATTA CCAAGAATTG GAAGAATATT 1020

TGCCTACAGG TTTAGCTGAA GCCGAAAAAA GTGCCGATGC GTTTGGCTTA GCTTTTGCGG 1080

AAAAACATCG TCATGATAGC ATTCACTATT TTATTGGCGC AGGGAACCAA TGGGGAGCGG 1140

TTTATTCTTA TGCAATGTGC TATTGGGAAG AGCAAAGTTG GCTTCGTTCA AAATCAATTC 1200

ATGcAGcAGA GTTCTTACAT GGaACGCTAG AAATTGTAGA TGAAACTACG CCAGTAACGT 1260

TGTTTATtGG AGAAGACGAG CAGCGTCCGT TGGcAGAACG AGTAGCaAAA TTATTGCCGA 1320

AAATCTGTGC CAATTATACA ATTATTGATT CGAAAGcGTA TGAACTTCCA GGrAtTAGTG 1380

AAAAATATCG TGGCCGAGTT GTTTCCATTT CTCCCTAATG CATGCGGTAA CGCAACGCAT 1440

TGATGCGCAT GTAGAAAAAT TAAATTGTCA CCCATTAGAG ATTCGTCGTT ATTATCGTCA 1500

ATTTGATTAC TAAGACAAAA TAGCCTCTTT TTCCAGCAAG CCTGTGGTAG ACTGAAGAGA 1560

GAATACTTCA GAAAGTAGGA AAACCAATGA CAAAACCGAA AATCTATACA ATGACCTTTG 1620

CCAGTGTTTA CCCGCTATAC ATCCAAAAAG CAGAACGTAA AAATCGAACG CAAGCGGAGG 1680

TTGACGAAAT TATTTACTGG TTAACTGGTT ATGATGAAAA AGGCTTGCAA CAACAATTGG 1740

CTCAAAAAGT AACGATGGAA ACATTTTTTG AAGAAGCACC CAATTTTAAT GAAAACGCCC 1800

AATTGATCAA AGGCGTAATT TGTGGcTATC GCGTGGAAGA AATTGAAGAT CCTTTAATGC 1860

AArAAATCCG TTACCTAGAT AAATTAATCG ACGAATTGGC CAAGGGAAAA GCGATGGAAA 1920

AAATTTTAAG AACAkGAAAA cAGGGTCtCa AkGAGACCCt GTTTTCaTGc yCcCtTGTAC 1980

ACGGAATTTA ACTATTTCCT GTATTAATGA AAGAGGTAAC TCTTGCGTGT AGGGAAATTG 2040

AATGGCACCT TTTGAGGTTT TAAAAGGGCG TAATTCTTTG CTGAAACGTT GGATGGCTGA 2100

TGGCGTAGGA TAGAAACCTA AATGTTGCTT AGCCGCCGCA AAATGCACTA AATTGCCATT 2160

TAAGTAAAAG GTCGGCATTC CATAAGCGAT TTTTTCAGTA GTCTCCTGAG GAACAACTGC 2220

TAAAATTGCT TGGTAGATTT CTGTTAACTG TGTTTGATGT TGGGGTTGCT GGCGAATATA 2280

GTTTGCAATG ATGGTCATAT AAAAACACCT CCGTTTGTTT CAGTATACTA TGTAAAAAAT 2340

AAAAGGCGAA GGAGATACAA AAAAACCGTC TAGTATkTTT TCTTTACTAA ACGGCTTTTA 2400

GCTTAAATAA CTACCTAAAG GGTCTTGCAT GGAAATGTTT GTCACkCCTT AGTTTAATTG 2460

TTTTAAGTGT GCAATTGCTT CTTCAATTGT GTGACCAGTT GCTGTTTTAC TTTCATTTTT 2520

kGCATCAAAA ACAATGAAAA TATTTTTGTC TGTGTCTAAT TGAATACGGT ATTGAATGTT 2580

TTTTACTTTG TAAGTCATAG TTATTCTCCT TTCATAAAGT ACTCTAAAAA GGCGACGATA 2640

ACTGATAACG AAATTCCATT GTTACTATAC CACTGGAAAA TTATTTTTGC AACAAATAAC 2700

ATTTTTTCTT TACCATTCCT TAGACTTCTT TGCTAGCTAC TTCTTTACAT TCTTTTTCAA 2760

GAAGCATAAA ATAAATACAT TACTAGTTTT TGAATAGGAG ATCGATGTGA AAA 2813

(2) INFORMATION FOR SEQ ID NO: 460:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2818 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 460:

CACTTCTTCC TGTTGAATGA GTCGCTTAAG CGCCAATGTa GGTCACGCCm CTACTATAaG 60

CAtTACATTA GGTAAGACAC GATACGAAAT CGTTTGATCa TTGATGACAT TGCTTTCAGA 120

AATAATGATT CCATCAGAAG TGACCGCTTC AACAAAAGCA ACATGTCCAT ATTGATTATC 180

TGCGCCAGCA ACTCCTGGTT GGAAACTAAT CGCACGACCA GCTTTAGGCA ATGAACTAAC 240

TTGATAGCCT AATGCACGTC CTTTTCTTCC CCAGTCACTG CCATTTCCCA TAAATTCATC 300

CACAGAAGTG CCCAACTGTT TGAAACGATT ATAGACATAC CACGTACATT GGCCCaCGGG 360

ATAGCTACCA CTTCGATTAT AATTGATCCC ATCATAGACT GGAAAATGCa TTTGTTGcTG 420

tTCCGCTTCC GACAGyTTAT TTTTAGATTG GATAATCAAA CCGTCCACTG TTTTTGGTAA 480

ATCAAATTGG GTTAAGTTAT AAACCGCAAT TAACGAATTT AATTTATTAT CGTATTGCTT 540

gTCTGTTGCA TATTTACCTG TTAAAAAACG CGTTGCTTGA AGGTAGTTTT TGGCTTCCGA 600

ACGCCACGCT GGTTTATAAA AATCTTGATT TCCGTCAATC CCTTGTCGGA GTAATTGAAC 660

ATAATCTTGC AGTGAGTCAT TATAGCCACC GTAATCGCGA AACCCAGCGG AAATTGTATA 720

AAGTTGGCCC CTCTGgATCA GCCTCTTTTG TTGAAAAGCT AACACTATTC CTGGAACGAA 780

CCTTTCACGC CAAAGAGATa TAATAAGGTT CTTTGGCTAG TTGACTTTCG CCAGAGCCAC 840

TTTCTAAAAT AGCCTGAGCT ATCATTACTG ATGCAAAAAC ATCGTATTTT TCGCCAAGTT 900

CTCTGGCTGT CTCACCAATT TTTTTAATAA ACGCTGTGGT TTGTTCTGAC TTCTGCATTT 960

CAAATGACGC ACCATATGTT GCTAAAACTT CTTTGCCATA CTCCGTTAAT ATGAACTCCT 1020

TAAAAGGATT AGTCACTACC AAGCGTTTTT CCTCTTTTGA GTCTGGTTGT TGTTTGGCAG 1080

TATAGATATT CTCTAACTGA GCGATTACTT CTGGTTTCTC TTTAGAATCC GATTCCGCTG 1140

CGGCAATTAA ATATTTTCCT TGGCCArGAT AAACACCATT TTGAAGCGTT TTTCCTTCAG 1200

CTGTTTGCCA AAAAATCAGA TCACCAGGTT CTGCCTCCTC TATAGAACGA ACTTTACCAA 1260

ATGTCATTTG TTCTTGCGGC GTACTGCCAA TGTCTATTTT AAATAATTGT TGGTAAATTT 1320

TTTGGGCAAG GAAACTACTG GTCAATGCTT GCTCCGTCAA ATTAAATTCT TTTTTTGTTT 1380

TCCCAACGTG TCTTAATGCT TCAGCGACCA CGACTGCCCG TTTCTTTTCG CCAAAAGACG 1440

ACAACAGTGG AAGTTCGTAG CCTTTTAAAT TTGAAGAAGC TATTTCATCT ACTTTGAgGG 1500

ACGGATTTAA CGATACAGCC TGATTTTCTT CAACAAAGTT ATTGTTGGCT TCAGGAACTT 1560

GgAATAGGTT CTTCGGGTTC TTGCTTGGtC ACAGGAACTT GTGGtGGtTG CGATTGTTcA 1620

ACAGGCTGCG TGaTTTCTGa TGATACCGTT GGcTCTGTAG GTTCTTTTTT TGTTTCAGAA 1680

CTAGTAwTTG TGGTTGTCGT TTCTGTCGTT gACTGACTCG TTGATTCGGT AGTCGTATCG 1740

GTAGAACTAC TTGATTCAGA TGACGAACTA CTAGTGGACG TTGTCGATGT AGCAACAGTT 1800

GTTTCGGTTG TTTCGGTTAC TTTTTCTTCC GAAGTAATTG TCGTTTCCGA GGACAGAGCT 1860

GTTTGTAGCG AGCTACTTGG TAATTCTGCA GCCAACGTAA CAACTGGAGC GACTTCCAAA 1920

CTCAGAAGAA CGAGAATACT TGTTAAAACA AGTAGTTGAT TCTTGATTTT CTTTCTTTTC 1980

ATTCGTAAAA ATCCTCTCCT ATTACTTACT TGCGCATTTG CACACTGTTT TTGTACAAAT 2040

GCGCCTTGTA AGTTAATGAT ACGCTAATTT TGCTCTTTTT TATTTTTGAA GAATAATTTT 2100

CCTAAACCAA TACTTACTAT CAATAGAAAA CCACTGATTG ACCAGATTGG CGAAACAACT 2160

TCATTTGCTT TTGGCAAAGT ACCAGAAGTG GTGGTTTTCG TTGTCGTTCC ACCTGTTGCT 2220

GTGCCCTTTC CTTGGCTGGT TACACCACCA GTACCGCTAA CCGTGCTATG ATTTCCATTG 2280

TTGCCTGTAG AGTTGCTACT GATTGGCTCT TCTGCCATTG CTTCAATGGC TTGTGATAAA 2340

ATACTTAAAC ATTTGTCCAC AGCTTCTTGC GTCGCATTTG GGTCCTCAAC AAGTGCTTTT 2400

GCTTGTTTAA TTGCTTCAGC TAAAGCTTCT TGCGATGCTT TTGTAAATTG TTTTCCTGCT 2460

TTAGGTTTCA AGGTTTCTGC TTTTTGAAGC AGTTCTTTCA ACTTTGTTTT ATCCGTTTGA 2520

AGGACTTGTT CCTCTAAACT ATCCAAAGCA GTTTTTACAT TTTTTTCAGC AATATCAATT 2580

TGTTCTTGCC GAGCATTTTT ATCAGCTAAA ATCCCTTCCG CTTGGTCAAT TGCTTCAGTT 2640

AAGCGCGCTT TAGACTCGCT GGTAAATTCT TTTCCGGCAG ATGGTTTGAT TGCTTTTGCT 2700

TTGGCTATCA TTGCCTCTAA GGCTTTTTTA TCTGTAACTA ACACTtCTTC TTGTAACCCA 2760

TCGATGGCAC TAGTTAATTC AGCAACCGCT TCATCAATCA TTTTTgATTA GCTGCTGG 2818

(2) INFORMATION FOR SEQ ID NO: 461:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1326 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 461:

ATATTCTTAT GGTCACTTCT GCCACAGCGG AATATTGTAA TAGTTAGAGC GCATTGCACT 60

CACCAAATTA GTTTACCACA GACAAAATAA GCTGTAAAGA GAATTTTCTT GAAAGTTCCC 120

TTATTTTTGC TCTTCTTGTT TAATTTCCTC TAATAACTTC AACATTTCAG GCGTCATTTC 180

TACTGATTCC AACATGTCAG GTCGACGTAA AAAGGTCCGT CGTAATGATT CTTTTAGTTG 240

CCATTCTGCA ATTAATTTAT GGTTACCAnT CGTTAGCACG GCAGGGACTT CCATATCGTT 300

GAAAATAGCA GGTCTGGtAT ATtGTGGATG CTCTAATAGG CCAGTGGaAk GkGAATCCgT 360

TTGTGcAGAA AgATkGTtAC CTAAgrCaTC aGGTAATAAA cgGACTGTCG CATCAATCAT 420

GACCATTGCG CCTAATTCGC CCCCTGTTAA AACGTAATCC CCCAATGAAA CTTCATCTGT 480

GACCATCGTG CGAATTCGCT CATCATAGCC TTCATAATGT CCACAAATAA AAACTAAATG 540

TTCTTCTTGT GAAAATTCCT CGGCCATTTT CTGATTAAAC GGTTTTCCTG CTGGATCTAA 600

TAAGATGACC CGTTTTGGTT GCTGGTTTGT CTCTTCTTCA ATTGCCTTTA GGTTATCATA 660

GATCGGTTGT ACTTTTAAAA GCATCCCCGC GCCACCGCCA TACGGATAAT CATCGACCGT 720

TTGGTGTTTG TTATCTGAAA AATCGCGAAA ATTAGAGACA TTCAATTCTA ATAACTGCTT 780

TTCCCGAGCT TTCCCAATAA TTGATTCCCC CATTGGtCCT TCAAACATTC TAGGAAAAAG 840

AGTTAATACA TCAATTCTCA TCGTCAATCA ATCCTTCTGG AATGTCCACA TCAATCCGTT 900

GATTTTCGAC GTCGACTTTT GtAACAACGC CTTCAATATA CGGkAAAAGT AAATCtTTTT 960

TCyTTGGyCG TTGGACGACC CACACATCAT TCGCACCAGG AGATAAAATT TCTTTAATTT 1020

TTCCCAGTTC ACTGCCGGTT TCATCGTAAA CAGTAGCACC AATGATTTGG TGATAATAAA 1080

ATTCGTCCTC TGCCAATTCA GCGAGATTGT CTTTTGCAAC TTTTAAAATA CCGTCACGAT 1140

ATTTTTCAAC ATCATTGATA TTTGGATGTC CTTCAAAGCT TAAAATGTCA AAATTCTTAT 1200

GCTTGCGGTG GGACTTCACG ACCAATTCAA TTGGTGCTTT TTTCTCTTGA AATAGTAGTA 1260

ACATAGCCCC TTTTTGATAA CGTTCTTCTG GAAAATCCGT TTTCGAAATG ACTCGAACTT 1320

CTCCTG 1326

(2) INFORMATION FOR SEQ ID NO: 462:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2845 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 462:

CAGAGAAATT ATTTTTCTAA TTACTGGAAA AATAGTTTTT CTTTTTTTGT GTAATCAAAG 60

AATCAACCAA TTAATTCATT GACACAAATT TTAAATGGAA CGATTTCTTG TACATCTAAA 120

GTAGTTTTAG AAAGCATTGT AAAAAGCAGT TAAAAGCAAA CTTATGAGCT CACTCTTACC 180

TGCTTCAATT TATTTCTATA GATTAGTAGT TGCTCTTACA ATTCTTCATC AATGAGTTTA 240

GTCTTGTTAT TCTCCATCAC CTTAAAAAAT GGATAGTAAA TAGCAGCACT AATTAGGATT 300

AATACAAAGC AAAAAACAAT ATTTCGCCAG TCCATACTTG AGAGATATGC TTGTGCAAAA 360

AATGGTGTGA AAGATGGGTC GATGATAAAT CCAACGCTAA TCAGTCCAAC ACTCTGCGCA 420

ATATACGTCA ACAACAATGA AACAATTGGT GTAAGCAAAA ATGGAATAGC TAAAATGGGA 480

TTGAAAACAA TCGGAAGCCC AAATATGACC GGTTCATTGA TTGAGCATAT CCCGGGAATA 540

AGGGAAAGCT TTCCGATTGT TCGATATTCT GAAACTTTGC TACGTATCAT TAAGAGTACT 600

AAACCAAGAG TTCCACCCGC ACCACCTAAA ACCGATATAC GGAACATTTG TAGATTCATC 660

AGATGAGTCA TTGCTTCTTT ATTTGCCATC TGTTCTGCAT TCAAACCTGT TTGTGCCATT 720

CCTAAAGTGA AGACGATTGG GaAAATGATA GATGAACCAT TGATACCGAA TAACCATAGA 780

ATGTTTCCCA AAGTGACAAT CAATAAGAAT CCACCAAGGC TACCAGCAAT ATTTGTAGCA 840

GGCGTTAAAA CCTGCATCAC TGTTTCTGGG AAAATCTTTT TCGTTGTCGC TAAGAAAATC 900

AGATTGATGC CATAAAAAAG AATGATGTTT GTCAACAGTG GAACTAAAGT ATTGATGAAA 960

GTAGCTACCA TAGGAGGGAC AGTATCTGGT AATTTTAACT TGATGTTTTT CACTTCAAAC 1020

ACTCGATTGA CTTCAACAAC TAATAACCCG ATGATGATAG CTACGAAAAG GCCATTAGTT 1080

CCCAAATAAT TTAAGTTtAT TGTTCCTTTT GTCACCGGGG TACATACCAT CAGGTAAATG 1140

ACAGAAGCGA CCATACCATT CATAGATGGG TTGGTTTTGT ATTCATTTGC AAGAGAGAAC 1200

GCAATACCGA AGACACTGAT TAATCCGATG ATCCCCATTG TCAAGTTGTA TGGCGCAGTA 1260

ATCAAGTCAT AATTAGCGAC TGCAAAATCC TTCCATCCAG CCATAATTTT CATAAAGAAA 1320

TTAGCGGTCT CTGGATTATA ATGATCAATA TTTATTGGTG GATTTGCAAA AATCAGAAAA 1380

AAAGAACCAA TAACAATAAA CGGTAATCCA AACATCATTC CACTAGAAAT GGCTTTAAGA 1440

TGTCGCTGAT TTCCGATTTT TTCAGCAATC GGGCTAAGTA TCTCATTCAA TTTTGTGATA 1500

ATCTTAGAAT TTCCCATAAG TATTCCCCCT TATACTTGCC AAGGATTAAA AACAAACGTA 1560

CTTGGTACGA CATGATTCGG ATCATATTTC TTAATAATTT CTTCATATTG TTCTTTGTAA 1620

TCATTCTCTA CTTTTGCTTG CGGGATAACT TTACTTGCTT CATGTAAGAA ATGCTCTGAT 1680

CCTCGTCCCA TCTCCTGACC TCGTGTCGTG TGTTTATCAA GAGCGATATC TGGAATTTCA 1740

GGGACATATC CCATAGCAAA GCTTTTAATT ACAATATTTT TTAAAAGATC TGAGGAACGA 1800

TCTTTCTCTG ATCCGCATAA GTAGCGAATG GCATGAAAGA AAAACATCGC ACGATCTGAT 1860

TCATTATATT GAAATTCCTT ACGCATCATA TTCAAGTTAT TGATCATCAA TGCTGCTATA 1920

GGATTTCCCA TACCGATATC CTCAACACTG ATTGCCAACA AACGTCTCCA AAGCTTTTCT 1980

TCAAACTGTG GTGAAGTAAT ATACATCTCG TAAGCAAATT CGCAAGCGTC TTTTTCTTTT 2040

CCTCTACGGA TAGATTTTTG TAATGCAGAA ATGACTTCGT CACCAGCTAA ACCGTTTCGC 2100

GTAGTCATCT TTGCCCAAGG ATCTTGAATA AATTTTTTTT CTGTCATGTT ATAATACTCC 2160

TAACTAAAGA ATTTGAGGTG GACCTTATGG ATACTGAAAT GCTAATCGAT AAATATCAGT 2220

TGAATGATTC AGAAGCTGCT GTTTTACGGT TTATGGATAA AAACCGTGAC TGCTTGAAAA 2280

GTCTTGGAAT TCGAGAAGTA GCAAAGGCTT CTTATGTCTC AACTACTGCA ATCATCAATA 2340

TGGCAAAAAA AATCGGTTAT TCTGGTTATA GTGAGCTGGT CTTTGCCTAT GGAAACAAAC 2400

ATTCATCAAT TAGCTTTCCT GAGAATTTCT CAAACGAAGA AAAAGATGAA TTTATCCGAT 2460

TATTATCTCG CTATCGAGAA AAACGGATTA TGGTTCTAGG CTCGGGTTTT TCACAAAATA 2520

TTGCAAACCA TTTTTCTGAA ATGCTTAACC TTTATGGATT TCGTGCAACT GCCAACAGTC 2580

ATTTAGAATT CCTACGAGAA AATGTTGAAA AAGATGTCCT GATTTTTATT GTTTCCAACT 2640

CTGGTGATAC TCTTCGGTTA GCAGAACTTG TTAAATTAGC AAAAAATCaT CATATTGATC 2700

TCaTkGCTTT CgTTGGtGaA AAGAAATCCA AAATTGGACA ACTTGCTACA CTAACTATAA 2760

GCaCgGAAAC mtwTaAtCCy CaGGkAGtTT CAGAATATCA ACCAAACCTG TTCTTTGGGA 2820

CTGCnTTGAA TCAATTCGnA CTATT 2845

(2) INFORMATION FOR SEQ ID NO: 463:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2373 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 463:

ACGACCACAC CAATTAATGA AATGGCAATC ACTGTAGCTA CACCTGCACT TGTTCCAAAt 60

CCaTATTAGT ATTtGGCAAT ACTTTTGGTT TTTCTTGTGG AGTTGTAACA TTCACTTTGT 120

TAGAATCTGT CGGTTCATCA TTAATCACCA ATTGTGCCAC GTTTGGAATC ATAATATTCC 180

CTTTATCATC GACATATTGC GAAAAGTCTG CGTCTTTCAA TTTCCCTTTA ATGTACACAA 240

ACAATTGTTT TTCTTTTAAT TTGTTTGGTT CATTAGCATT CCAAGAAACT TTTTGTGTTT 300

TTTCATCAAC ACTAATTTTT CCTTGATCCG TAATGTCTGT CCCATCTGAA AGAGTGACTT 360

TTACTTCTTT ATTCGCTTCC AATACATCTT CTAAATCATC TTGaATGACT AAAGACGTCA 420

TATCTTTGGT ATTTGGTACT TGCACATCTA ATTTGAACGT CACTTCATCG CCCACTTTGG 480

CACTGGTCAA ATCCACTTCT TTATCGCCtT CGAGGATTTT CTTAGCGATA GACGGGTCTT 540

TTGGTGTGAC AGTGGTTGTT GTTTCATCAG TTGGGACTTC TTTGTTGTTT TGGATAATTT 600

TTGCTTTATT TGGAATTTGT AACTTGCCAT CAATAAGGTA CTTATCTAAG CTGGCACCTT 660

CACGAATTTT TGTTTGGAAA TTGAATTTGT ATTCGTGATT ATAGAAATCC GCTGTTTTTA 720

ATACGTCAGC CTTTGCAACC ATACGTACTA AATTGCCTTC ATGctTCGCT TCAAATTTAT 780

TGGTAACGTC CTTTCCTGCT TCGTCCGTGA CTTTCACATC AGACACACGT TCCAATTCAG 840

GAATCAGTTC GTCTTGCATT TCATAGTTGC TATAATAAAA TTCTTTCATT TCATCAGGAA 900

CTTTATGTGT GATCGTATAG CTATATACTt CGCCTAATCC TTCTAACGTA TTTTTATCAA 960

CATTTGTTTC ATCACTATCT GTCACCGTTT TAGTAGGAAT CATCGGCGCT GTCCGAACTG 1020

GTTTTTTGGC ACTATAGCCA AAATAACTAT ACTGGTTATT ATTTTTTGAC GGTAATGGAT 1080

TCCCTGAATT TTTTTCGCGT GTAGCGAAAT CTTCACTATA CGTGAAATCG AATCGAGAAC 1140

GATTAGAATA CAAGAACGTA AAATCACCAC GATTATCACC ATCATCTAAA GGAATATCAT 1200

CAGAATCGTA AAAAGTTTTT CCACCATCTT TTTCTTCTTT GTATTTAATG ATTGTATCTT 1260

TCGTTGGAAC GTACAATTTA TCGATTGCTT TATACGAATC TTTGTCAATC GTCATTCCTT 1320

GTAAATTATC AATATCAATC CATGAATAGT AGCCTGAAAC TGGCTTTTCT TTGTCTGTGC 1380

CATGTTCTAG GTACGTCCAT GTTTGTTTCA CCCATGAGAA ACCACCTTGG TAGGTTGCAA 1440

TTGCGGATAC ATTGTTAAAA GCAACAATAT TATTTGACAC CATATTTTGT GCATTTTTAT 1500

ATTGTTCCCA ATCTTGAATG GTAATTTTTA AATCAATTGG TGTCCCATCA TCAGCTGTAC 1560

CTACATTGTT GTAACGAATC CACGATTGCC CTTTCTTCGT TTTCGTTACA TCAAGACCGA 1620

TTTTCCCGTC CCATTTTTGT GGAGGCATAT TAACGTAAAC ATCATTCGTC CACCCATCAC 1680

CACCTGTTTC AACCGTAGTG ACACCTTTAA TAAATTTTGC AATAAAGCTA TACTTTGAAC 1740

CAGCTTCATC ATCTGGTGGT CACTACAGAA GCATTCGGCG CATTAATGAT tGAATTAGCG 1800

CTTGCACGTT GGAAAAAATC TGCTTTCACT GTTTGTGGTA CTTGTCCGCT TTCTtGTGCT 1860

TGAACCATCA TTGGCATTTC CATGCCTGCT AAAATTGTTA CAGTTGCTAA TCCTTGAATA 1920

ATCGTTGTTT TCTTCAAAAT ATTTCCTCTT TTCATTATTA ATTTTTTGTA GTTCTATTTC 1980

ATTTCTCCTA GATTTTGCAT AAAAAATAGG AGAGCCACGA ATCTTCTTCA TTGGTTCACC 2040

TTTCCACAGT TATTCACTTA AAAAGTTATT TGTATAAACT ATATAGCGTA ATACAGGAGC 2100

TTCTTGATTC TTTTTATTCT GGAATTCTTC TTTAAAATTT ACATTTTCAT ATACATACTC 2160

CTCAACAACT ATACAACTTT TCTTTGCTAT ACTAGCATTA AAAATCAACT CATTTTCTAA 2220

AGAAATTTCT GAAAGAACAA AACCTTCTTT CATTGAATCT GCATGCtTCA AACTATCTCC 2280

CTTTAAATAG AGTTTCTTTT CATAAAAAAG TAATCTATCC CTTACTCTAT CAACAACTAC 2340

AGGAnACAAT TCTTCTAACT CTCGAnCATC TCT 2373

(2) INFORMATION FOR SEQ ID NO: 464:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11036 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 464:

CGATGATGAC CTTAAGCGAC CGAAATGAAG CGATTGATGT GGTAACAATC AAAGATCGCC 60

TAGAACAAGA GAATTTGTTA GAAGATGTTG GTGGACTAAG TTATCTGTCA GATTTAGCTT 120

TAGCTGTGCC GACAGCTGCC AATGTCATTT ATTATGCAAA AATTGTTGAA CAAAAATCGT 180

TGCTGCGCAA CTTAATCCAA ACCGCTACAG AGATTGTAAC TAAAGGCTTT GAGCAAGGTG 240
AAGATGTTGA ATCTATTTTG GATGATGCAG AACGAAGCAT TTTAGAAGTC TCCGAAAAGC 300
GGaATCGTAG TGGCTTTTTA TCAATTTCAG ATGTtCTAAA TACaACTATT GCGAATATTG 360
ACCAATTATA CCAAAATGAT GAAGAGATTA CTGGTTTACC TACAGGCTAT CAAGCATTGG 420
ATAAAATGAC TGCTGGCTTG CaAGCGGAaG AGCTAATTAT tCTAGCGGCC CGTCCAGCGG 480
TAGGGAAAAC CGCTTTTGCA TTAAATATTG CCCAAAATGT CGGCACCAAG ACAGATAAAg 540
CTGTCGCTAT TTTTAGTTTA GAAATGGGGG CAGAATCATT AGTGAACCGG ATGCTTTGTG 600
CAGAAGGATC CATTGAGGCC AGCCATTTAA GAACAGGGCA ATTATCTGAA GAAGAATGGC 660
AAAACTTAAT TATTGCGATG GGTAGTTTGT CCCGTGCGAA TATCTATATT GATGACACAC 720
CAGGGATTAA AATTACCGAA ATTCGTGCGA AATGTCGTAA ATTAGCACAA GAAAAAGGCA 780
ACTTAGGTCT AATTTTGATC GACTACCTAC AATTGATTGA AGGAACTGGA AAAGAGAACC 840
GCCAACAGGA GGTTTCTGAT ATTTCACGTC AATTAAAAAA ATTAGCGAAA GAATTAAAAG 900
TCCCGGTTAT CGCTTTGTCA CAGCTTTCTC GTGGCGTGGA ACAACGTCAA GATAAACGTC 960
CTGTGTTAAG TGATATTCGT GAATCAGGTT CCATTGAGCA AGATGCCGAT ATCGTTGCCT 1020
TTCTTTATCG TGATGATTAC TACGATCGTG GTGAAGGAGA AGACGGTGAT CATGAGCCAC 1080
CAGAAGTTGA CAATGTTGTA GAAGTCATCA TTGAGAAAAA CCGTAGTGGT GCGCGTGGAA 1140
CGGTCGAATT ACTTTTCATT AAAGAATATA ATAAATTTTC TTCGCTTTCG CCGAGGACAG 1200
AATTTTAAAA AACGAACTCA TTTTTTGAGT TCGTTTTTTT GCTTATTAAG GGATTAAAAT 1260
CAGAGAGTAA AACACGAACA ATATTCGTAT ATTTTATTAT TTTACAAAAA CATAACTAAA 1320
AATGTTCGTA TTTTTGACAG AAACGCTTTT GATTGTGAAA GAATATTCGA TTTTTACCCT 1380
TGAAAGGGAC TGGTTTTACT GGTACAATGA TTCAGGACAA AAAAATTATG AACGAGGTGT 1440
TCGAATGTCA TCAGTAGTAG TTGTTGGAAC GCAATGGGGC GATGAAGGTA AAGGAAAAAT 1500
TACAGACTTT CTAAGTGAAA ACGCAGAAGT GATCGCTCGT TACCAAGGCG GAGATAACGC 1560
CGGACATACC ATTAAATTCG ACGGCGTAAC GTATAAATTA CATTTAATTC CTTCAGGGAT 1620
TTTCTATAAA GAAAAAATTA GTGTAATTGG TAACGGCGTA GTTGTGAATC CAAAATCTTT 1680
AGTAAAAGAA TTAGCTTATT TAAAAGAAAA TAATGTTGCC ACTGATAATT TACGTATTTC 1740
AGATCGTGCC CATGTCATCT TGCCGTATCA TATTAAATTG GATCAATTGC AAGAAGATGC 1800
GAAAGGTGAA AACAAGATTG GGACAACGAT TAAAGGAATT GGCCCAGCCT ACATGGATAA 1860
AGCAGCACGG GTGGGAATCC GTATCGCTGA TTTATTGGAT AAAGAAATTT TTGCGGAACG 1920
TTTGCAAATT AACTTAGAAG AAAAAAATCG TCAATTCGTT AAAATGTTTG ATAGCGAaCC 1980
ATTGAATTTG ATGATATTTT TGAAGAATAC TATGAATACG GACAACAAAT TAAACAATAT 2040
GTTACAGATA CTTCGGTTAT TTTAAACGAT GCATTAGATG CTGGAAAACG GGTTTTATTT 2100
GAAGGGGCAC AAGGTGTCAT GTTGGATATC GATCAAGGAA CCTATCCATT TGTTACTTCC 2160

TCTAATCCAG TAGCTGGTGG CGTAACTATC GGTAGTGGCG TTGGTCCATC AAAAATTAAT 2220

AAAGTGGTTG GTGTCTGCAA AGCGTACACT TCACGTGTCG GTGACGGCCC ATTCCCAACA 2280

GAATTATTTG ATGAAACAGG AGAAACCATT CGTCGTGTCG GTAAAGAATA CGGAACAACA 2340

ACAGGACGTC CGCGTCGTGT CGGTTGGTTT GATTCAGTAG TCATGCGTCA TTCAAAACGT 2400

GTATCAGGGA TTACAAACTT GTCATTAAAC TCGATTGACG TGTTAAGTGG TTTAGAAACG 2460

GTGAAAATTT GTACAGCTTA TGAACTTGAT GGTGAATTAA TTTATCATTA TCCAGCAAGC 2520

TTGAAAGAAT TAAGCCGCTG TAAACCAGTT TATGAAGAAT TACCAGGTTG GTCTGAAGAT 2580

ATCACTGGTT GCAAAACTTT AGCCGATTTA CCAGCTAATG CTCGTAACTA TGTGCATCGG 2640

ATTTCAGAAT TAGTTGGTGT GCGCATTTCA ACATTCTCTG TAGGTCCAGA CCGTAACCAA 2700

ACGAACGTTT TAGAAAGCGT TTGGGCACAA ATTTAAGCAG ATTTTTAAGG GAACGTAAAG 2760

GATGGCTTAT GCCAGTTCAG CCTTTACGTT CTTTTTTAAT TTTATATAAA AATGAAAAAC 2820

TATGGTAAAA TAAAAGAAAA CATGAAGGGG GCAATTGTCC GTGAATTATC AATTAGTGAC 2880

AGATTCTTGT TGTGACTTGC CGTATACATA CTTAAAAGAA AACCAAGTAC CTTTTATTAG 2940

CATGAATATT CAAATTGACG GGAAAGAATA TCGCGATGAT TTAGGTGAAA CGTTTGATTA 3000

TCAGAACTTC TTAACGGCTA TTAAAAATGG AAGTATGCCC ACTACCTCTC AAGTTAATGT 3060

AGGACGCTAC AATGAATTTT TCCGTCCTTT TGTGGAGCAA GGGTTACCGG TGATTTATTT 3120

AGCTTTTTCA TCTGGGTTAA GTGGCTCTTA TCAAAGTGCG TTGCAATCAG TAGAGATGTT 3180

AAAAGAAGAG TATGATAATG TAGAAATTCA TATTATTGAC ACAAAAGCAG CTAGCTTAGG 3240

TCAAGGGATG TTAGTCCGAG AAGCCATTCG ATTACAAACA GATGGTCATT CATTAGGAGA 3300

AGTTGTTGCC TATCTTGAAG AGCAAAAAAT GAAACTTCAC TCGTGGGTAA CGGTAGATGA 3360

TTTAAAACAT TTAGAACGTG GCGGACGGAT TTCTAAAACA GCGGCTGCAC TTGGTGGCTT 3420

AATGAACATT AAACCAATTA TTCGCGTGGA TGCGGCAGGC AAACTAGCAT CCGTTGGTAA 3480

AACTCGTGGA CGAAACAAGT CCTTGCAAAA GATTGCGCAA GAAACGATTC AAGGCATTGT 3540

AGAGCCAATG AAACAAACAC TATTGATTGC ATATGCGGGA ACGAAAGATG ACGCGGAAAA 3600

AGTCAAAGAA TTAATTGAAA AAGAAATAGA GGTTAATGAG ATACTGATTT ATCCTCTAGG 3660

ACCCACCATT ACTAGTCATA CCGGAATAGG TTGTATTGCT GTGTTTTCTT TTGGTGAAAA 3720

AAGAAAATAA AAAATAGAGA AGTCATTTTT TAATGACTTC TCTATTTTTT TAATAATCTG 3780

AACTGAATAA AACGTCCATT TGTTTTTGAA TTTCTGGGTT TTCCAAGAAT TCATCATAGG 3840

TTGTTTCTGC CCGATCAATG ACCCCTTTAT CAGAAACAGC AATAATTCGG TTCGCTAATG 3900

TTTGGATAAA TTGGTGGTCA TGTGAAGCAA AAAGAATTGA ACCAGTGAAA GCCATCAACC 3960

CATCATTTAA TGCAGTGATT GATTCTAAGT CTAAGTGGTT CGTTGGATCA TCTAAAACTA 4020

AGACATTGGC TTTGGAAAGC ATTAATTTTG AAAGCATGAC ACGCACTTTT TCGCCTCCGG 4080

AAAGAACATT GACTGGTTTT AGTACCTCTT CACCAGAGAA TAACATCCGA CCTAAGAAAC 4140

TACGTAAGAA CGTATTGTCA TCTTCTTCTT TACCAGCAAA TTGACGTAAC CAATCTAAAA 4200

TGGTTAATGG CTCTTCAAAG TCTTTGCTGT TGTCTTTTGG TAAATAAGCT TGGCTAGTTG 4260

TAACGCCCCA ACGAACAGAA CCTGTATCGG GAGTAATTTC GCCCATAATC ACTTTGAATA 4320

ACGTAGTCGT TGTAATATCT GAGTCAGCGA TGAATGCCAC TTTATCATCT TTTGTTAAGT 4380

TGAACGAGAT ATTATCTAAG ATTTTTTTGC CATCAATCGT TACAGAGACA TTTTCAACTT 4440

GTAATAAGTC ATTGCCGATT TCACGTTCTG GTGTAAAACC AACGAATGGA TAACGACGAG 4500

AGGAAGGTTG AATGTCATCT AAGGTAATTT TATCTAACAT TTTTTTACGA GACGTTGCTT 4560

GTTTTGATTT TGAGGCATTG GCGCTAAAAC GAGCGATAAA GTCTTGTAAT TCTTTGATTT 4620

GTTCTTCTTT TTTGGCATTA GATTGTGCTT GCAATTTTGT CGCTAGTTGG CTTGATTCCA 4680

ACCAGAAATC ATAGTTACCA ACGTAAAGTT TGATTTTACT AAAGTCTAAA TCTGCCATGT 4740

GAGTGCAGAC TTTGTTTAGG AAATGACGGT CATGGGAAAC CACGATAACG GTATTTTCAA 4800

AGTTGATTAA AAATTCTTCT AACCAATTGA TTGAGCGGGT GTCTAAACCA TTTGTCGGCT 4860

CGTCTAATAG TAAGACATCT GGTTTACCAA AAAGAGATTG AGCTAATAAT ACTTTGACTT 4920

TTTGACCAGC TGTTAATTCG CTCATTTTTT GATCATGTAA TTCTTCTGGA ATGTTTAGCC 4980

CTTGGAGTAA AACAGCTGCT TCAGGTTCTG CTTCCCAACC GTCAAGTTCA GCAAATTCGC 5040

CTTCTAGTTC TGCGGCACGA ATCCCATCTT CATCTGAAAA ATCTTCTTTC ATATAGATAG 5100

CATCTTTTTC TTTCATTACT TCGTAAAGAC GTTTATGTCC CATAATTACA GTTTCTAAAA 5160

CAGTGTAGTC TTCGTAGTCA AAGTGATTTT GTTTCAGCGT TGCTAGTCGC TCATTAGGAC 5220

CCATTGAAAC CACGCCAGTT GTCGGTTGAA TCTCACCTGA TAAAATTTTA AGAAACGTTG 5280

ATTTTCCTGC ACCGTTTGCG CCGATTAAGC CGTAACAGTT GCCAGGAGTA AATTTAATAT 5340

TCACTTCTTC AAAAAGTTTA CGATCTGGAA ATTGTAAACT TACATCATTA ACAGTAATCA 5400

ATTGTTTTCC TCGCTTTTAA TATATTCTTG AACCTCTCTG CATTATATCG GCTACAGAAC 5460

CTTTTTTCAA GGAGTTTCAA TCTTTTTACA GAAACGAATA AAAAGGTGGG TTTTTTATAG 5520

AAAAAAGGCG CTTTTTTAAA ATGTTTCATG TGAAACAGAA ATTTTTTTCA TGAAAGGGTT 5580

TTTAGTGTAT TTTTTTAATG ATACACTGTA TCATGTAAAT TAATTAGTGT ATCGATTAAA 5640

GGAGCCATCA AAAATGAAGC GGATTGAAAA AATCTATCAA TATGTTAAAG AGCAAACGAA 5700

GAATTTAACG CCAATGACCT TAACAAGTGA TGCGGGAGTC ACGACACAAG AAATTTCAGA 5760

ACGTTTAAAC ATTCAGCGGA CCAATGCTAG TAAAGATTTG AATCAACTGG TTCGTGAAGG 5820

GCGCTTGAAA AAATTATCTG GTCGACCAGT AAAATATGTC GCGCAGGTGG CGTTTCAGCA 5880

TCGACCATTA ACCAAGCCAG TTAAAAGTTA TCGGGAAAAA TCAATGGATC ACTCCACAAA 5940

AACTTTTATC GAACCATTGG TTGAACCAGT AGTCACGACA AGTAAAACCC CAGAAGTTGA 6000

AGATATTTTC AAAAAAATTA TCGGTTCAGC TGGAAGCATG AAAACGCCGG TGGAACAAGC 6060

CAAAGCAGCG ATTTTATATC CGCCAAAAGG CTTGAATTGT TTAATTACTG GACCAACAGG 6120

```
TTCTGGGAAA ACTTATTTTG CTCATGCCAT GTTTCAATTT GCGAAgCTGa ACCAAATTGT   6180
GGCCAAAGAA AAAGAGTTTG TGGTGTTTAA CTGTGCAGAC TATGCGCATA ATCCAGAACT   6240
TTTGATGTCG CATTTATTTG GGTATGTAGA AGGGGCCTTC ACAGGCGCTA CGAAAGCGAA   6300
AGAAGGGATT ATTGATGAAG CAGATGGCAG CATTCTGTTT TTAGATGAAG TACATCGGTT   6360
ACCACCCGAA GGACAAGAAA TGATTTTTTA TTTTATGGAT CACGGCGTGT ACGCGCGTTT   6420
AGGTGAAACA GTGAAATCAC ATCATGCGGA TGTTCGGATT ATTTGTGCTA CGACGGAGAA   6480
TCCTACTTCT TCTTTATTGA ATACTTTTGT GCGACGCATT CCGATTATCA TTCAATTACC   6540
GAACTTTAGT GACCGTCCTG CGAAAGAAAA AATTGACCTC CTAAAAGTGA TGGTTTCAAT   6600
GGAAGCAGCG CGGATTCAAC GGCGAATCTC TTTGTCAGAA GATGTGGTCA AAGCGTTAAT   6660
TGGAAGTGTT TCTTATGGAA ATGTCGGTCA GCTAAAATCG AATGTGCAAC TTGTGACTGC   6720
GCGAGGGTTC CTTAATCAGA TGGAACAAGA GGAATTAATG ATTACGATGG ACGAATTAAC   6780
GGACAATATC AAAGAAGGCA TTATGCAACT TGCCAGTAAT CGAGAAGTAT TGTCGGAGTT   6840
ATCGAAATAT CTGGAACCGC AATTAGTGGT ATCACCAAAT GAATCTTTAG TAGCTATTCA   6900
GTCAGATGCC TATGAATTGC CTTACAATCT ATATGAGATT ATTGGAGATA AAGCGGCTCT   6960
CTTAAAAGCA GACGGCTTGG ACCAAGAACT AATTAATAAT TTTATTACAA CGGATATCAA   7020
TGTTCATTTA AAATCTTTCT ACAAAGATTA TGGGTTCACC TTTGATACAG AAAATAAATT   7080
GGCTGAAATT GTTGACCAAC GAATTATTGA TGTCACTAAG AAAATTTATA ATTTTGCAGC   7140
GAAgCGGTTA TCGTATTCGT TTCAACCCAA CTTTATTTAT GCAATGAGTT TACACATTAG   7200
TTCTTTCTTG AAACGCATCC AATTAGGCAA AGATCCTAAG CATCCTTTAA ATGAAAGTAT   7260
TCGAAATATG GTTTTAGATT ATCCAACAGA GTTTGAAACA GCAAAAGAAA TTAAAAATAT   7320
TATTGAAAGT AGTTATCAGA TGGAGATTCC TGAATCTGAA AGTTACTATT TAGCTGTTTT   7380
GCTCATTTCT TTAAGAGAAA ATCCTGAGGC TGGCCGGATT GGCATTGTCG TGGcTGCGCA   7440
CGGAAATAGT ACCGCTAGCA GTATGGTACA AGTGGTTTCT CAATTGTTAA ATGTTGATAA   7500
TTTGAAGGCT GTGGATATGC CGTTAGATAT GCCACCGAAA GAAGCTTTGC GTAAAATCGT   7560
TGAAGCAGTT GGTGAAGTGA ATGAAGAGAA TGGCGTACTA TTATTGGTGG ATATGGGTTC   7620
ATTAAGTACT TTTTCAGAAG AAATTGTACG CCAGACAGGC ATTGATGTGC GGACAGTAGA   7680
CATGGTGACC ACGCCAATTG TTCTTGAAGC GGCCCGCAAA ACCGCATTAA TTGATACGCA   7740
ATTGGAAACA CTTCACGAGT CTTTAAAAAA CTTTCATGGC TATGCCGATA TTCGCCAAAG   7800
TGAAACGAAA CAAATCATTG AAAATTGGAA AACAAGAGCA ATTATTGCTA TCTGCGCTTC   7860
TGGAGAAGGA ACTGCGCGCC GAATGAAAGA GCTTATTGAG GAAGCTGTCT TGCCACAAAT   7920
AGACTGGCAT TTAGAAGTGA TTCCTTTGTC CATTGTAAAT ATGAAAGAAG TCTTGCCTAA   7980
AATTCAAGAA GACTATGAAA TTATTGCGAC AACTGGGATT ACCAATCCCA AAATTGGTGT   8040
TCCTTATATT TCCATGGAAA ATTTCTTTTC CGGTGAAGCA GAAAAAATTA TTCAGCAATT   8100
```

ACTAGCAGAG CGGAAAGAAG AAGTAATGAA TAAGCCGTTG GACGAAGCGG CAGCAAAAGA 8160

AATTTGTCTA AAATATATGG AGCAAAGTTT CACTTTCATT AATGGAAAAA AAATCATTGA 8220

TCTGTTGTGG TCATTCGCTA AAAATATCCA AACCCAATTA GCGATGCCAG AAGAGTACAC 8280

ATTTTACATC AATTTGATTA TGCACACATC GGGCATGTTG GAGCGTATTT TACGGAATGA 8340

TACACTGACG GTTTCAGAAA AGGAACTTGG TCGATTGGTT CAAGAACCAA TTTATCCAGT 8400

GATTGTGTCC TCAATTGAGA CTATGGAAGA AGCTCTTAAT ATGGATGTTC CAGCAGAAGA 8460

AGTTTATTTT ATCGCACAGT TAGTCAAAAA TGCTCAGTGT ATCGAAGATA AAATCACCGA 8520

AATTGATACA CTAGATTAAT ACAGAAAAAC GATACACTAT TTATTAGTGT ATCGTTTTTT 8580

TTATTTTCTT AAAAAAACCT ATAATTTTCT AGGTTTCGAA CCGATACACT AAAGTTGGCA 8640

CGAAACTTGC TTATAATAAG GTGTCCCTTT TTACCTAAAA TTTTTAAGTA AAAAGACAAA 8700

GGTCCTAACC TTTAAAAATG AATTTAAAGG AGCATTCCCG ATGGTAATGG ATATTCGTTT 8760

AGCTCGAATT GATGATCGAT TAATTCATGG CCAAGTAGCA ACTGTTTGGG CAAAATTGAC 8820

GAACATTAAT CGCATTTTAG TGGTTAGTGA TTCCGTTGCT CATGACAATT TGAGAAAAGC 8880

CTTGTTGGTT CAAGCTGCTC CACCTGGTGT CAAAGTAAAT GTCATCACTG TTCAAAAAAT 8940

GATTGATGTT TACCCTGATC CTCGTTTTGA TTCGTTTCGA GCCATGTTGT TATTCACTAA 9000

TCCTTTGGAC GTGAAACGAG TGGTCGAAGG CGGTGTACAA CTAACTTCTG TGAATATTGG 9060

CGGAATGAGT TTCAGCACAG GCAAGCGAAT GATTACGAAT GCAGTCGCAG TGGATGCCAA 9120

CGATTTAAAA GCATTTCTGT TTTTAAATGA ACAAGGTATT GAATTGGAAA TTCGAAAAGT 9180

CGCAGCAGAC AGCCAAGTGA ATTTGATGGA TTTACTGAAA AAAGAGCGTA gCAAAGAAAA 9240

TAACGCCAGC TGACGTTAAT TTTTACAGAC GAAAGACATT GAAAGGATGT GGTGGCCTAG 9300

GACCAACATC ATTGCTTCAT GAACTGTATT GATAAAGGCA AGTGCTGATG TTGAAAAAAA 9360

TCAAGAACTA TTGAAATTTT TTCATTTATT TTTGTCAGAA TCTGATTTAT CTGATTTTTT 9420

TCAAAAAGTG AAAGCGTTTA GCAGTACAAC CATGAAAAAA AACATCAATT ATACAGGAGG 9480

TATAAAAATG GTAGGAATTA TCCTAGCAAG TCACGGTGAA TTTGCTGAAG GGATCTTGCA 9540

ATCTGGTGCA ATGATCTTTG GCGAACAAGA AAATGTAAAA GCAGTTACGT TAATGCCTAG 9600

CGAAGGTCCC GATGATGTAA AAGCAAAAAT GCAAGAAGCC ATTGCATCGT TTGACAATCA 9660

AGACGAAGTA CTATTTTTAG TTGACCTATG GGAGGAACA CCATTTAACC AAGCGAACTC 9720

ATTGTTAGAA GATCATAAAG ACAAATGGGC AATCGTTGCA GGTATGAACT TACCAATGGT 9780

TATTGAAGCA TATGCTTCTC GTTTCTCTAT GGAATCAGCA CAAGAAATTG CGACACACAT 9840

TCTTGAAACA GCAAAAGATG GTGTCAAAGT TAAACCAGAA GAGTTACAAC CAGCAGAAGC 9900

ACCAAAAGCT GCTGCGACAG AAGATGCACA GCCAAAAGGT TCGTTACCAC CAGGTACAGT 9960

TGTTGGCGAT GGCAAAATTA AATTTGTGCT AGCACGTATT GATTCACGTT TATTACATGG 10020

TCAAGTAGCA ACAGCTTGGA CAAAAGCAAC ACAACCAAAC AGAATTATTG TTGTTTCTGA 10080

TGCCGTTGCA AAAGaTGACT TACGTAAGAA ATTAATCGAa CAAGCAGCTC CTCCAGGAGT 10140

AAAAGCCAAT GTTATTCCAA TCAGTAAAAT GATTGAAGTA GCGAAAGATC CACGTTTTGG 10200

CAATACAAAA GCATTATTAT TATTTGAAAA TCCTGAAGAC GTCTTAAAAG TGGTTGAAGG 10260

CGGCGTAGAA ATCCCAGAAG TAAACGTTGG TTCAATGGCT CACTCTGTTG GAAAAGTCGT 10320

TGTAAGCAAA GTATTATCAA TGGGTCAAGA AGACGTTGAT ACGTTTGATG AATTAAAAGC 10380

CAAAGGCATC AAATTTGATG TGCGTAAAGT GCCAAATGAT TCAAAAGCGA ATATGGACGA 10440

AATTTTGAAA AAAGCAAAAA ATGAGTTAGC AAACGCGTAA CTCGTCAAAT ACAAATCAAA 10500

ATAGGAGGCT TATCATGTCT GTTATATCAA TTATTTTAGT TTTACTCGTT GCCTTTCTAG 10560

CTGGTATGGA AGGGATTTTG GATGAGTTCC AATTCCACCA ACCGTTAGTA GCATGTACAT 10620

TAATCGGATT AGTTACTGGT AATTTAGAAG CCGGAATTGT TTTAGGTGGT ACTTTACAAA 10680

TGATCGCGTT AGGTTGGGCG AACATCGGGG CTGCCGTTGC ACCGGATGCC GCTTTAGCAT 10740

CTGTTGCCTC AGCAATTATT TTAGTTTTAG GTGGCCAAGG TGTTAAAGGT GTTCCTTCAG 10800

CGATTGCCAT CGCAGTTCCT TTAGCGGTTG CAGGTTTGTT CTTAACAATG ATCGTTCGGA 10860

CAGTTGCGGT TCCAATCGTT CACATGATGG ATGCGGCTGC TGAAAAAGGC AACATTAAAC 10920

AAGTTGAAtG TaCACATTTT CGCTGTATGT TTACAAGGTA TTCGTATTGC GATTCCAGCA 10980

GGTGCGTTAC TATTCATTCC TGCTGACACT GTTCGTAATT TCTTAGAATC AATGCC 11036

(2) INFORMATION FOR SEQ ID NO: 465:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1440 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 465:

AACGGTGATG CAGGAAAAAG AATCCTCTTC GGCAATCGCC GAAAGCAGTA GCGGAAATGC 60

GGTTGCAGTA ACTACTGAAA CCACAAATGA GATACAAAAT AGTGATACAG ATGGaAAAGC 120

TGTTTCTGCr GAGAGTGTTT TTTCAGAAGC AGaTTATAAA CAAGCAACTG CACTGGAGTT 180

AGCAACTTTA GTAAGAGAAA AAAAGGTGAC GAGCGAAGAA TTAGTAAAAA TTGCGTTGGC 240

CATCACTAAA CGTGAAAATC CTACATTAAA TGCGGTCATT ACTTTACGAG AAGAAGCGGC 300

TTTGACAGAA GCAAAAGCTT TACAAGATAC AGGTCAACCA TTTCTAGGTG TGCCGCTCTT 360

ACTAAAAGGG TTAGGACAAT CCTTGAAAGG TGAAAGCAAC ACGAACGGTT TTGGTTTTCT 420

CCAAGATCAA GTAGCTGGTG GGACATCGAC CTTTGTCAAA GCACTACAAA ATGCCGGATT 480

CATAATCATT GGTCAAACGA ATTATCCAGA ATTAGGTTGG AAGAATATTT CAGATTCAAA 540

GTTATATGGC GTTTCCGTCA ATCCATGGAA TCCTAACCAT TATTCAGGTG GTTCTTCAGG 600

CGGACnGGGT GCTAGTGTCG CCGCGGCATT TGTTCCCATT GCTTCTGGAA GTGATGCTGG 660

TGGCTCTATC CGCATCCCTG CTTcTTGGAC AGGCACCGTA GCTTGAAACC TTCTAGAGGA 720

GTAATCATTG GTAATTCTAA TAGTGCAAAA GGTCAAACTG TTCACTTTGG TTTAAGTCGA 780

ACCGTGGCGG ATACAAATGC ATTATTTGAA ACCTTATTAA CCAAAAAAGA TCTTCCTGCG 840

GGACATTTAA GTCAAGCGCA ACCCATTGCT TATACAACAG AATCCCCTGC CGGAACGCCT 900

ATAAGTGCCG AAGCAAAAGA GGCGGTCGCT GAAGCTGTTG CTTTTTTAAA AGACCAAGGG 960

TACACATTGG TTGAAGTGAA GCACCCTGTT GATGGGGAAC GTCTAATGAA AAATTATTAT 1020

ACTGTAGCTG CTGGCTCAGC AGGAATTGCC GATTTTATGG CGCGGCAGAA ATTGAAAAGA 1080

CCGCTTGAAC GAAATGATGT AGAACTGTTA ACATGGGCGC TTTTTCAAAC AGGAAAAAAT 1140

ATAACGAGTG AAGAGACAAC TGCGGCTTGG ACAGATATTG CTTTACAAGC ACAAGCGATG 1200

GACGAATTTT ATCmACAGTA TCtATCyTAt TAmCACCAAC GACGGCTGCa CGGCACCTAg 1260

TAaTTGATAA TCCGTTACTA AAACCAGAAC ATGCcAGCAC AGATtGGAAA AAAaTTGATC 1320

mATTGTCACC cAGCcAGAaC aAAaACmATT GgtTTaTGaC CaTgGCTGAC GGCGTTCACa 1380

TTACmcCgtT TACCCCACCA AGCCAAATTT TAnTTGGGGC CATCCCGCCA TTAAGGGGGT 1440

(2) INFORMATION FOR SEQ ID NO: 466:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 821 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 466:

TGCACTGAAT CATCCTAACC TAGAAGTTCG CTTAATAACG ACGGTTGCAG GGAATGTAGA 60

TGTAGAAAAA ACAACAAATA ATGCCTTGAA ACTTGTCGAT TTTTTGGGA AAAAGGTTCC 120

TGTAGCCAAG GGGTGTAATT GTCCGTTACT TATTCAATTA GAAGATTCTG CTGAGATTCA 180

TGGGGAGACT GGTATGGATG GCTTTGAGTT TCCGCAACCT ATCTCTACGT GTTTAGATAT 240

TCATGCAGTG GAAGCTATGC GAAAAGAAAT TCTGTCTAGT GATGTTCCTT TGACCATTGT 300

TCCAATTGCT GCATTAACGA ATATTGCTTT GTTACTGACC CTCTATCCAG AAGTCAAAGA 360

AAATATTGCA GAAATCGTCA TGATGGGTGG CTCATTAGCA AGGGGAAATA CAAATACAAG 420

TGCCGAATTT AATACGTATG TCGATCCACA CGCTGCACAA ATTGTTTTCC AATCAGGTGT 480

CCCGCTAACA ATGGTAGGAT TAGATGTTAC AAGTCAAGCT GTGTTAACGA ATCATGAAGT 540

AACAGCCATC AGAGCGCTTG GTAGAGTTGG GGAAATGTTT TATGGTCTTT TCCGTCATTA 600

TAGAGGCGGC AGTCTAACTA CAGGTTTGAA AATGCATGAT GTCTGTGCGA TTGCCTATTT 660

AACTTCACCA GAATTATTTG AGACAACTGA AACGTTCATT GAAGTTGcTT TGGAAGGTCC 720

AGCTGCGGGA GcTACTGTTG CGGATTTAAA AATGAAGTAT CATAAAAACA CGAATGCCGT 780

TGCCTGTAtA GATGTAAATG TTGAAGCTTT TCAAAAATGG G 821

(2) INFORMATION FOR SEQ ID NO: 467:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1420 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 467:

CTTAGGGATG GTGACTACTA GTGACGGTCG AAGCTTCGCT GCAGCAGATT TGCCAGGCTT 60

AATTGAAGGA GCTTCTCAAG GTGTCGGCTT AGGAACACAG TTCTTACGCC ACATTGAACG 120

GACGCGTGTT ATTTTACATG TCATTGATAT GAGTGGCATG GAArGACGCG ATCCTTATGA 180

AGATTATTTG GCAATCAATA AAGAATTGGC TTCACACAAT TTACGTTTGA TGGAGCGACC 240

TCAAATTATT GTTGCCAATA AAATGGACAT GCCTGAAGCA GAAGAGAATT TAGCGAAATT 300

TAAAGAGCAA CTTGCGAAGG AACGCACCGA TGAGTATGCT GATGAGCTAC CTATTTTCCC 360

AATCTCTGGG GTTACTCGCA AAGGTATTGA ACCTTTATTA AATGCAACAG CTGACTTGTT 420

GGAAGTAACG CCTGAATTCC CATTATATGA AGATGAAGTG GTAGAGGAAG AAACCGTTCG 480

TTATGGCTTC CAACCAGAAG GTCCAGAATT TACGATTGAT CGTGAACCAG ATGCTTCATG 540

GGTTTTATCT GGTGAAAAAT TAGAAAAATT ATTTGAAATG ACGAATTTCG ATCATGATGA 600

AACAGTAATG CGTTTTGCAC GTCAATTACG TGGTATGGGT GTCGATGAGG CATTACGAGC 660

ACGTGGCGCG AAAGATGGCG ACATTGTTCG GATTGGTAAT TTTGAATTTG AATTTGTCGA 720

ATAAGCATAA AACCAGTTCA CTCAGTGAAC TGGTTTTTTA TTTTAAAATT AGTATTAAAT 780

TAGCGAATAG ATTGAAATAA TAtTTTTAAT AGCtTATAAT TTTaAATGtw ACaAAATAAC 840

AAACGAAAGG AGCGAAATTG AAATGGAAAA CAATCAAGCA ATGGTTCCTA AGTTTTCAGT 900

TAAAAGTGCC GGTTTTCTTT TTAGTATGGG GCTCTTATGT GGTGTCCTGT TGCCGAGTGT 960

TTTATTATGG ACTTTCAGcT GCCGTTTAAT ATTGGTGCCA TGATTTTTTT GCCTGTTGGG 1020

ATTGCTTTTG CGTTAGCGTA TAGTCATTAT TTATTGAAAC GAAAACAGGT TTTTGTAAGC 1080

GTTTTATnTG GTTGTTTTTA ATTTCTCTTA TTTTATAGAG CTTATCTCTT ATTTATGGCT 1140

GATTAAAGGA GTTATTTTTT AAGAGAAAAA CAGAGGAGAG TGTTATTTAT GTTAGATAAA 1200

CTGACTGTTT GGATTGAAAA TCATCTAGCT GGACCAATGG CTAAAATTGC CAATCAGCGC 1260

CATTTACGTG CGGTGCGGGA TGGGATTATT GCGACATTAC CTTTAATTAT TGTCGGTTCT 1320

TTTTTTCTAA TTATTGCGTT TCCGCCGTTA CCAGAATCTT GGGGCATTAC TCAATTTTTA 1380

ACAAGTAATG CTGCTACAAT CTTATTGCCT TACCAATGAC 1420

(2) INFORMATION FOR SEQ ID NO: 468:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2179 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 468:

GGTATATCCA tGTATwACAG tATAgCCmTT aCAtAAtCAT GTAACTaTCA ACAATCaTCA 60

AAGgTTTCCw AAtCTTTATG GATTGCTGgT ATCCTATCGA TGGTGTGTCA CCATATCCCA 120

TTGTATGCTT ATTTCCCAAG CATGGATGGT ATTGGCATTT GGCAATTCAC ATCCGCTTAT 180

ATTGCAGGTG GTTTAGATGG TAACGTAGAT TTAACAGGAA TTACGGATAG TGGTTATACA 240

GATACAaATA AACCAGAAAC GGATACGCCA GCAACAGATG CAGGCGAAGA AATTGAAAAA 300

ATACCTAATT CTGATGTTAA AGTTGGCGAT ACCGTCAAAG TGAAATTTAG TGTTGATGCT 360

TGGGCAACTG GCGAAGCTAT TCCAGATTGG GTTAAAGGAA ACAGTTATAA AGTGCAAGAA 420

GCAACTGGAA GCAGAGTATT GCTAGAAGGC ATTTTGTCAT GGATTAGCAA AGGTGATATT 480

GAACTATTAC CAGATGCAAC AGTTGTCCCT GATAAGCAAC CAGAAACAAC ACATATAGTT 540

CAACATGGCG AAACATTATC AAGCATTGCT TACCAATACG GTACCGATTA TCAAACTTTA 600

GCTTCGTTAA ATGGATTGGC TAATCCAAAC CTTATTTACC CTGGACAAGT TTTGAAAGTA 660

AATGGATCAG CAACAAGTAA TGTTTACACA GTCCAGTTGG GTGATAATTT ATCAAGTATA 720

GCGGCTAAAC TTGGCACGAC TTATCAAACA TTGGCAGCTT TAAACGAATT AGCAAATCCT 780

AACTTGATTT ATCCGGGTCA AACGTTGAAC TATTAAGAAA TAAGTAAAAG GCCTACTTCT 840

CTTTGTGAGA GGTAGGTCTT TTTTCATTTA TACTAATATA TACATATCAT TGAATGCTAA 900

AAAATTAAAA TTTGTCTTAA GTAAATTTTC AAGTTATGGT ATACTTTGCT GGTGAATATG 960

GGAAGGAGCA TAAAATTTGA AGAAACATTC TAGGAAAATA AAAAATAATA AAGAACTAAC 1020

TTTACTGATT ATAGGAATTT GGGTTGTTTT TTTTCTTTTT AATATGTATA TGCCGACAGT 1080

GAGAGCTGAT GACTTAGTTT ACGTAAATAG ATTAGATAAA TTAGGATATT TGGGTGCTTC 1140

AATAGAACAT TATAAAACAT GGAGTTCAAG AGTTATTATT GAATTATTCT TGATGTTTTT 1200

TTCTAAACAC CTTATGTTAT GGAAATTATT AAATTCAACG ATAATGACAG GTTCTATAGT 1260

ATTGCTTTGC AAATATGTTT TCAACAAGTT ATATTCTAAA AATTTGCTAC TAGTTTTTTC 1320

TATATTTTGT CTAATTCCTT TAACAATAAT GGGGGAAACT GGTTGGATCG CTACTACGTT 1380

AAATTACCAA TGGCCAGTAG CTTTTGGTTT GCTAGCTTTT TACCCTTTTT TTCAAGTGTT 1440

AAGCGGTAAA GAAATAAGCA AAAAAATATA TTGTATAAGT ATTCCTTTAT TAATATTCTC 1500

TGCTAACCAA GAGCAAGTTA ACGTATGCTT TTTTGTTTTA ACTAGTCTAG TAAGTTTGTA 1560

TTTATTATTT AGAAGAAATT ATAATTATAA GTTATCGGCG CTTTCCATAA TAAGTTTAGT 1620

GGAATTATTT TTTTCTTTGA CGGCCCCAGG GAACACATTG CGGGCTGCTC ATGAAATTAA 1680

TAAATGGTTT CCTGAATACA AAAAATTTAA CTTTGTGAAT AAATTAGACC TTGGTATATC 1740

ATCAGAGAGT GTAAAATATT TTGTGTAAAT GAAAAAATCC ATACAAAAAA GGAAGTCGCT 1800

TCTGTAGAAT AAAGTTAACG ACAACCAATT CACAGAAAAG AGGACTTCCC TATGAATGAT 1860

TTTACTACAG AAATTGTGCA AACTCTAGTC ACTAAAGGCG ATTTAAATGA ATTATTCCGT 1920

TCGCACTTAG AAAAAGCGAT AAACACACTC CTACGGACTG AATTAACGGC TTTTTAGAT 1980

TACGAAAAAT ATGATCGCAC TGGTTTTAAT TCAGGTAATT CGAGAAACGG TTCTTACTTT 2040

CGATCAATCA AAACCGAATA TGGTGaATTA ACATTGGgAA tACCTAGAGA TCGTAATGGT 2100

GAGTTTAAAC mACmAACTTT ACCAGCCTAC mAAAgGAmCA cGrTACaTTG GaaCCmCTAT 2160

TATCCnTTTA TCCGAAAAG 2179

(2) INFORMATION FOR SEQ ID NO: 469:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2005 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 469:

TTTATATAAA TAAAATAGTA ATTGGAATGG CAATCAGCAC AaTCtGtGTA AAgCCaTGAa 60

TAGTGAaTTG GCTGTgATTT GGTCATCATC GAAAaGAGnC CAaTCGCAAC GGTGTAATTT 120

TCAACTTTAT CACCCATAAT AACTTTTGCG AAAATAAAAT CCATCCAAGG GCCCATGAAG 180

GCGGTTAAAG AAGTATAGAC AATAATTGGT TTCGACAATG GCAACGTAAT TTTTGTGAAA 240

ATTTGCCACT TGTTGGCACC GTCAATCATA GCCGATTCGT CTAAAGCCAT CGGAATGGTG 300

TCGAAGAAAC CTTTGGCGAT GTAAAAACCT AGCGCCGAGC CTGCAGAGTA AACCAGAATT 360

AAAGCGAATA AACTTtCAGT TAAATTTAAG GCTTTTAAAA TATAGTAAAC GGCAATCATA 420

CTCATAAAAC CAGGGAACAT ATTTAAAACG AGCGCCACTT TTAAAAATGG TTTTCTTAAG 480

CGGAAACGCA ACCGTGACAG CGCATAAGCC ATTGCGATTG TAATGAAAGT TGATAACAGA 540

CAGCTTGCGA TGGAAACGAA CAAGGTGTTC AAAAACCAAC GGACAAAGGG GAAGTTGCCA 600

GTCTGATTTG TTAATAATAC TTTATAATTT TCTAAAGTGA ATGTTTTGGG AATGATATAG 660

GGAACAAATG CACCGCCTTC GCCACGAAAA CTTGTTAAGA TAATCCACAC AATGGGGAAC 720

ATCCAAAAAA CTGCTAAAAT GGTTAGAATG ATGTAAACGC TGGTAAGTGC AGTGCGTTGT 780

TTTTTCTGGT AACTTTTTGC TTTAGCCATA AGTTACGCCT CCTTAAATGA ATTCGTACGT 840

GTATAGGCAA TTAAACTGAA GACTGCGGAG AGAATGAAGA TTAAAATACC AATCACAGAG 900

GCTAAATTGT AATCCATTGT ATCGACAGTC AATTTGTATA GCCAGGTAAC TAAAAGGTCA 960

GTTGAGCCTG CACCGTAGAA GTTAGAATTG GCTGGCCCAC CACCTGTTAA TAGATAAATG 1020

ACATTGAAGT TATTGATGTT GCCAATAAAC TGTTGAATCA GCGCTGGGGT CATCACAAAT 1080

AGAATTTGTG GAAACGTAAT ACTTTTAAAA ATTTGCCATT TATTGGCACC GTCAATTCTA 1140

GCTGCTTCGA TTTGATCTTC AGGAAGGTTT TGGATAATCC CTGTAGAGAT TAACATCGTA 1200

GCAGGAATAC CAATCCACAT ATTGACAACA ATAATGGTTA ATTTTGCCCA ATTCGGATCT 1260

GTTAAGAAGG GGATAGAAGA ATCGATTAAG CCCCAATTTA ACAACAGAGC ATTGATAGGT 1320

CCAGAACCAT TTAAGAGGTT TCGCATAATT AAAAGTGAAA CAAACGGAGG TACGGCCATC 1380

GTGATAACAA AAATAGTCCG CCAAAATGCT TTATATTTTA AGCCTTTTGT ATTGATAAGC 1440

AAGGCTAAAA TGATTCCAAA AAAGAAACAA GTTGCGGTTG CGAAAACAGC CCAAATCAAG 1500

GTCCAGCCAA GTACTGGAAA GAACGTTGAA GCCATATCGC CTGTGATTAC ATTACCTAGA 1560

TTTGCCAATC CAACCCAAGT AAAAAGGTTT TTCGGTGGCA GATGGTTATG GTCGTAGTTG 1620

GTAAAGGCAA TTGAAATCaT ATACAACAAC GGCAAGATGG TAAAGAAAAG AACGCCTAGT 1680

AAGGGGATCG TCATTAAAGT TGCATGGAAA CGTTCGTTTA ACAAGCTACG TAAATCATCC 1740

ATAGTTGTTG GAATTTTTTT GCCGGCTGTC TTTAATTCAT ATAGGTGACG GGCGCTTTTT 1800

AAATTGATTA CATATAAAAT GATTAGCGAT AAACAGACTA AAATTGCTGC AATTCCGAAT 1860

AATAATAAGA GCATTGAATT ATCGCCTTCT TTTAGAACTT CAATCCCTAA AGAATCATCA 1920

TAtACTAATC CTTGCTGTTG GGTCCCTAGT GTTCCTAGCA TCATTAATGC GTGGAAACCA 1980

TTACGGATTA ACCAATAGAT AAAGC 2005

(2) INFORMATION FOR SEQ ID NO: 470:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 529 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 470:

GGTCCACCAA TATATTTAGG ACGAGAATAT AATGCTGTAA CATGGGCCGC ACCTAACGTT 60

CGAATGCTCA AGCGAAGAGG AACTTGCACA TGTTTTACAT GCATACCGAT AGATGTGTCG 120

CCAATGTCAA TCCCTGCTTG AGCAACAATA TGTTCAACTT CCACAGGTTC ATCAAATTTT 180

TCAAAGGCAG CCACTGAACA AGAACCGCCA GCATGTAATT GTGGTACTAC CGAGACAATT 240

TCAAAGCCTT TTTCTTCAGC CACACGACGT TCCATCGmTA AAGCACGATT AATATGTTCG 300

CATCCTTGGA CAGCCAAAGA GATGCCTTTT TTATCTAATT GTTCAATTAA GGTTGAGACA 360

ATCCATTGTC CTACTTCGGC ACTGGAATTT TTACCAATGA CGCCACCTAC TACCTCACTG 420

GTGGTACAAC CCAAAACGAA TAGGTCCCCT TTTTTTAAAT TTCCTGCTGT AATGACTTCT 480

TCTGTAATTG TCACTAACTG TTCTTnTAAC ATTTTnTCAT CGATAGTCA 529

(2) INFORMATION FOR SEQ ID NO: 471:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 451 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 471:

AAGTTTTAGT CGTATCCACT TGATTATCTG GAATAATACT TTCTACTGTA AaCCCCATTT 60

TTTCTTCTGT TGAAGTAyTC TCTGCTTGAA CAACTGAAGG AAAGAAAATA GATAATAGTA 120

AAAACGCTAT AAACCCCATA AAAAATTTTT CTCGTTGTTT CACCTACATC CTCCTAAAAA 180

AGGAGAGGAT GCATGATCCT CTCCCAAAAA ATTTACCGAT TAGATCGTAT CATTCAATGA 240

CCAAGTCAAT GTCCCACTGT ATTGTTGATT AATCTTCGCT GCATTCGCAG GAACTTCTAG 300

TTTAACATTT GACATCTCCA TAGCTGTTGA TCCTawGAAG cTAGCGCCGC CAGTTCCAGC 360

ATCTTTCATA ACAGGCACTG CAGCAgCTCC AGCAGTCACT GATTGTGAAG AAGCATTAAC 420

GGTTGCCGTA cGTGCTCCTT TATCAACAAT T 451

(2) INFORMATION FOR SEQ ID NO: 472:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1551 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 472:

TTCCTTGATG CAATACGCTT TCTGCATCAA TACGTCCTTT AyAGTTyTCG ACTAATTGTT 60

TyGCAATCGA TAGACCTmAG CCATTACCAC CTTTATTTCT TGCACGGGCT TTATCTACTC 120

GATAGAATCG ATCAAATATT TTCTCCAAGT CTTCTTCAGT GATTCCTTCA CCAAAATCTT 180

GCACAGCTAT TTCAAATTCA TTCATCGTTC GAGAAATGGA AATATGCACT TCTTTTCGAT 240

CGGTTGAATA TTTAATCGCA TTATCTAATA GAATAATTAG CAATTGTTCA AAGTGATTAC 300

GATAGATTTT CAGCTCAACT TCGGTTGGTA AATCGTCATC TAATGTGATA TGAAACTCAG 360

GATAAACCAA TTGGAAGTTA TTAAATACTT GGTAGACTAC TTGTTTAGCA TCTGTTCGTT 420

CATTTGCATA TTGGGTGTCC ACTTGTTCAG CGCGTGAAAG GTCAAGCATT TCTTGGACCA 480

AACTCTTCAT ACGACTAATT TCTTGTAAAC TGGCCTTTAA TGATTcATCT AAAATTTCAG 540

GATCGTCTTT CCCCCAACGA TTTAAAAGGT TTAAATGGCC TTCCATAATC GCAACGGGCG 600

TTCTTAATTC ATGGGAAACA TCTTCAACAA ACTGCTCTTG TTGTTCGATA TAGCGTCTCA 660

TACGATCTAA CATTTCATTA AAGATTTCCG AGATATCTGC TAACTCATCT CTCGTATTAA 720

TCTCCGGCAT GTGGACATCT GATTGCGGGT CTTTGCGGAT TGTATCCATC GTGTCTCTCA 780

ATACTTTTAA TGGTTTTAAG AAATAAGAAG AGAGGATAAA GCCTAAGACA CTACTTACAA 840

TCAGAGAAAT CACTTCTAAA ACTACAAGTG TTAATAATAA ATGATTACGA ATTTCATAAA 900

AAGAAGAAAG TTCATAAAAC GCTTGAATAT AGCCAATCTT TTCTCGTGTT TCTTTTGAAA 960

AAATAGGTTC CACGGAATAA AAACCGGTTT TATCAAAAAC GGTTCGAACA ACAGGTAATT 1020

GTCGATCCAA TTGTAATAAC TTATCATATT CATTTTGTGT TTTAAAAACT AGCTTTTGGT 1080

TTGTGTCATA CACAGATAAA TAAAGTTCTG GTTGTCCTAG TTCAGATATG AAACTGTCCA 1140

TTTCCATAAA CGAACCTTCT ACTGCTGTAT GCTTATTATA GTAGTTTTCA TCTCTTTCAC 1200

TCGGCGTTTT TAAATAGTCA AAAACGTCTG TTACCGTTAA ATTTTCATTG GCATTAGCTA 1260

ATCGATTTGT TACTTCTGCA ATCGTTGCTT CGACATTTTC CTTTTCTTTG GcAACAATAA 1320

GACTGACAGA AGACTTATAG GtAATCACCG CAAAAATGGt AAAAACTACA AATATAAAGA 1380

AAGAACTTGC GAAAGCCCAC TTTATAGTTA AAgATGGGcC TTCCAGTTCT TTTTTAATCG 1440

TtCTTTTCAT nAATTAGTTC TCACGAAcGC ATAAcGTAGC CAGTTCCAcG GACAGTTTGG 1500

ATGTAGCTTT CTtCTCCAGG TACGTCAATT TTATTTCGTA gTAGCGGATA G 1551

(2) INFORMATION FOR SEQ ID NO: 473:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4120 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 473:

TTCGTGAATT TTTGTAGCTy TCAATTTTTC CtAAACGTCG TAGCCtATAT TGCTACCTTA 60

AGTATCTCCC TGACCATCGG ATTGAATCTA CTGGTACCTT TTTACTTAAT TTATACAAAA 120

TCATCGCAAT TCAACCCCAA AGGCTACTTT ATCAATTTAA TGACTTTTAC CTTTTGCCAA 180

TTTCTCCCAA TTAAAATGAC TTTTATCCAA CAATCTAAAG CATTTTTAGC CGGATTACTT 240

CTTTTTTTAA TCGCAGTGAA ATTCTATAGC TTTGTTTCTC AATATTATGC AAAAAGAACA 300

TCTGAAAAAA ATGAGTTTAG CTTTTTCGCT GATTTATTGA ATTCATATGG TAACGAACAA 360

GAATATCAAG CTAACTTAAA CAAATTAGCC AAATTGACTC AACAATTGTA TTTCCAAGcA 420

TATCTTTCTT CTAGTTCTAA AAACAAAATT ACTCCTAAAA CTCGTATTGC ATTCAACTAT 480

GCCTTTTTAT CCCAACGTGC TCTATATTTT TTTACAAGTA TTGATAAATT GGTAAAGCCA 540

GATTCAATTA TTGAAAAATG GATGACAAGA ATTTCTGATT TTTTTGAAAA GTCCAGTCAG 600

TTTACTACGA TGTCAGTTAT yGAAAAAAAT GAGTTGAGAA TACAAGGGAA AATACTTATA 660

ACTTCTTTAG AACACTCCGA TGAAGTAGCA CTGATACTTT TAAGAAACCT GCTTATTCAC 720

AGTTTATTTA TATTACGTTT ATTGGATAAT GATCCAGAAA TAAGCAATAA TCATTGGTCA 780

ACGCCTCTCC mTCaCCrGAC CAAATATCTT TyCCAACACA AaTTTAGaAT tGATTCaTTT 840

GaATTAcGTG TTcGCaCtAC GTTTCaGTAT CGTTCTGGCT TTCTGTTTTT CTTTATGCAC 900

ATTGACTCAT ATTACTCGTG GAACTTGGTT ACCGATGAAT GCTTTCTTAC TGCTTCAACC 960

AATGAATGAA GAGACTACTT ATCGCTTAAA GACTCGTTTT TTAGGAACCA TCTTTGGATG 1020

TTTGATAAGT ATCTACACTG TGCAGCTCTT TAGCAACAAC TTACTCCATC TAATTTTGGC 1080

TAGCTTATTA GGGATCTGTG CCTTTACGAT AATTCCAGGA ACTACATTTC AAGCAACTTT 1140

ATCAACTATT TTCGCATTGT TCCTATCATC ATTAGTTATG TCTAATTTAA TGGCCGCTGG 1200

ATTGCGCTTT GTTTATGTTT TATTAGCTTG TTCATTAGTT TTTCTCGTCA ATCAATTCAT 1260

TCTTCCAACA GGCATAAAAA ATCAATTTCG ATTTAATTTA CAAGAACTTT TTCATATTCA 1320

TCAGACTAGT TTAATATTCC TTCAAAATTC ATTTCATCAC AAGATTGATT ATGGcACAAT 1380
CGCTGACTTA CAAATGaATT ATCACTTACT CCATGGTGAA ATTAAGAATT ATTTCGGCAA 1440
AAGCGATAAA TTGAAACAGA TATCCTACAA AGAATTTTTA TCTATATCTT GGCAAATGAT 1500
TACTATAGTT GAACAACTTT TACTACTTAT TGACAGTAAA AAAATTACAG AAAATGATAT 1560
TACAGTTATT GAAAATACAA TTGACTCTCT ATGTTACGTA TTGTCTCTAA TCCAAAACTT 1620
GTTAGACTGC AATTTAAAAA CATTAAAAAT ACCTGAAGTA CTTTTTAAAC GTGaTTTAGC 1680
AATTGaTCCT GAAATAGCGT ATCTGckATT TCaATACTCT aAAAAATTaT CAGCmATGtA 1740
TCGCCTCGTA CTTTGnATTG TCATTCAATT AAATTTCGAA AATrTTCGAT AATCTACTAT 1800
CTAAGGGAAT TATAAAGGGT ATTCTTATTA AAGAATGAAG AATACCCTTT AATTTACCTT 1860
TTCAAAATCT ATCCTTTACT CTTTTATCTG TAGTGCTAAT ATAAATGATT GTAGAGATAG 1920
TTTGTATATA AGGATATTTT TGTTCAGCTT AAAGGAGTTT TTACTCTATT TAATGATTAC 1980
TACATACAAT TATAGAAGTA CAAGTGATAA CaATAAGTCT ATACGCATAG GrAATAGyCA 2040
TyCATTAGCa AkGmCTTCCC ATTTTTTTGC TTTTTGTAAT ATTAAAACGT ATTGAAATTT 2100
TGATGTAATC TTAGATGTAG GATCAAkaTA TTTAACTAAA ATATTTGTCT TTATTGCTTG 2160
TCCTTTTTTC TGATAATTGA TACTGCTAAT TTCAACAAAC TTTAGATTCT TATAAATTGg 2220
CTCCAACACA TCTTTTTTTA CATAAtATCC tAGTTCTTCT TTAGAAGCkG TTGGATATAA 2280
TTTAAATAGA ACCTCTAAAA ATTCTTCAAT TTCTTCTTwG TkCTTTTCTT TAATCATCGA 2340
ATCATCTGTA AACTGCTTAG TTTTATAATT TGCTTTAACA GGTTTACTAA TCATTGTTGG 2400
CAAAGAAGTC ACAATACAAT TTCCAAGTCT ATCTTCAAAA ACAGAAATTC GGTATGCTGT 2460
TTTTTCTTGT AAAGCTTTTT TTTCCTCTAC TATTTCTTGA TCCACCAAAA ATGATACGAC 2520
GAACCATTTT GACTTTTTCT CAACATTTAG AATTTTCACT GAGTTCACCT TAGAACTTGT 2580
TGGTATATCT GCTCTAACCA TATCTTGACT TAAATCTAGT ACATCTTTTG TAAGATAATG 2640
CCCTAGAGAA CGTATTCTTT TTTCAATTGC TACTCTATCA TTTTTCCAAG AGAAATATTC 2700
TTTGGgCAAA tCTACAGTAA AATTTTCTAC ACCTGAGGgT ATTCACAATT TtTTCCTTCA 2760
CAACCACTTT CTCATGAGTT GTATGTTGAT CAATAGCGGT GAAATTTTTG TAGATTCCAA 2820
AGGATAAACT ACTAAGAAGT AATAACCAAA GGACAAGCAC CATTTTTCGA TGCTTCCCCA 2880
TTGAAACAAC TTGTTCTTTT TTCTTTCCCG AAATCTTCTT TTTATTTCGT TCTATTTTTA 2940
TCTTCATCAT CTGTCCCCTT TTCTATTTTT ATACCGACCA GCACCAATGA TATGTTGCTG 3000
CCAATAAGTT TCGGTAAGAT TAGTCCATCC AACGGGATCA CCTGCATGGA ACATCCGATT 3060
GTCTCCAGCA TAGATACCTA CATGTGTCAC ATAAGTTCCT GCATCATACG TTGAGTGAAA 3120
GAAAATTAAA TCTCCTGGCT TAGCATCTTT CAACTCGATA TGAGACATAG CATCATACTG 3180
ATCCTGCGCG ACTCTGGGAA GTTGCAAACC AACTTTGGCA TAACACCACT GCGTTAATCC 3240
GCTACAATCA AAACCTGTTG CTGGACTTGA ACCTCCGAAC ACATATGCCG TCCCTTCATA 3300

CTTAAAAGCT TCATCAAAAA TTCCTTGAAC GGTTTTATCA TCAAAtTCGT TGTAGTAAGA 3360

TATTGTTTCA CTAACATTAC ATAAAACATA TTTCCATAGT TATAACGCCA ACCACCATTC 3420

TCCTTAATAG CGATTGGATT CGAATAGGTC ACTTTTTCCC CATCTGATTT TTCTTTTGAA 3480

AAAGATTCCG CTAAGGCAAA GCTGTACTTC TTTCCATTTT TTCCAACATA GTCTAAATAT 3540

CCACCACCAT AGTTATACGA TTGAATGATG CTCTCTTGaT CACAACCGAT ATCCTTTGCT 3600

CGATTGACTA AGGAAGCAAA ATAAATGCAC CCTTGCTTGa TTGAAGCTTC GGTGCTTAAC 3660

GAGTTAgGTG GGgAATCCCG CTGGATTCnG AACTTTGGCA yCACATCTTC CaTCGTACCG 3720

CCaGACTCTA CCTGaATAaT gcTAGAAGaA TCGGaACATA CTCTAAAtAT CCGaTkCTTT 3780

AGCaTACTTC TcTACAAGTG GaCGATGCTT TAAAACTTCT AAGrAAACAT TACTCCCTCC 3840

AgAGGACAAT CCAGTACTGC TATTTCCTTC ACTATCGTCA CCAAATAATA GACCGACACA 3900

AATCAAAATT CCCATAAACA AAACCGTTAA AAGAACTAAT AGAACGGTAC TTAACTTTTT 3960

GATTGTCATG CCATATCACC TGTCTTTACC TTCTCCCGGT TTATAGAATT CTTTTGATAT 4020

TGCTTCTTTC CATTTTTCGT TAGAGTGGAT TGATTTTTCA TTTTTTTGAT ACTCATCATC 4080

TTCGGTGCAA TCTTAGGATA AATGGTTTGT AGGGAATCCT 4120

(2) INFORMATION FOR SEQ ID NO: 474:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 830 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 474:

AGAAAGAAAA CACAATTGTC TGAAnCGTTT TACGCACAGA CGACACAGTT GTTTTGGAGa 60

AAATGAaCAC AACAACTGAT GGATTAAGTG CACAGGAGGC TCAAAAGCGT TTAGAAACGT 120

ACGGGGAAAA TGTGCTAGAT GAAGGGAAGA AAAAATCACT TGCCGTGAAA TTCTTTGAGC 180

AATTCAAAGA TTTTATGATT ATTGTTTTAT TAGCGGCAGC GGTCATTTCA GCCGTCTTTT 240

CCCATGATGT GGTAGACTCG ATTATCATCT TAGTGGTTGT TATCTTGAAT GCCATCTTTG 300

GGGTCATCCA AGAAGCCAAA GCGGAACAAG CCATTGAAGC ATTAAAAGAA ATGTCTTCAC 360

CAAATGCTAA TGTTCGCCGT GATGGACACG TCATTACTGT TAAAAGCGAT GAATTAGTGC 420

CAGGTGACAT TGTTTTATTA GAAGCAGGGG ATGTCGTTCC AGCCGATTTA CGTTTATTGG 480

AAGCGAATAG CTTAAAAATT GAAGAAGCAG CGTTAACCGG TGAATCTGTT CCAGTTGAAA 540

AAGAAGTAAC CATCTTAGAG GGAACCGATA TTGGTATCGG TGACCGAATC AATATGGCTT 600

ATTCCAATAG TAATGTGACG TATGGACGTG GTCTTGGCGT TGTGGTTGGC ACAGGAATGA 660

ATACAGAAGT CGGTAAAATT GCTGGCATGC TAGCAAGCGA ACAAGAAACG GAAACGCCTT 720

TAAAACAAAA CTTAAACCAA TTGGGTAAAA TGTTAACAAT TGCCATTTTA GTCATTGCAG 780

CGGTGATGTT TGTGGTAGGA ATGATGArTG GTCGTCATGG ATTGACATGT 830

(2) INFORMATION FOR SEQ ID NO: 475:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1219 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 475:

TTGTTCATTT TTTTTGAAAG TTTTGAGAAT ACTATAGATA GGAGTTTTTT TATGGAGGAA 60

GATATTGTAA GGnTGTTTTA TAAAGGATCT GAAATTATTT TAATTGGGAC CAGTCATATC 120

TCAGCTGAAA GTGCTGATTT GGTCAGAAAA ACAATTCAGG AAGAAAATCC TGACACTATT 180

TGTATTGAAT GGGATCAAAA AAGATACAAA AAAAACATAC ATCCAGACGA ATGGGATGAT 240

ACTGATATAG TTAAAATAAT AAAGAATAAG CAATTTCCAG TTTTTATTTT TGGTGTAATT 300

TACAAATTAT TTCaAAAAAA AGTTTCTCAA GATATGAATA GCCTTGTAGG AAAAGAATTT 360

GTAGTGGCTG TTGATGAGTC AAAAAAATTA AACATAAAGT TTTATCTAAT TGATAGAGAT 420

TCTTCTCTTA CTTTTAAAAG AGCATGGAGG ATGTTAAATT TTCGTGAAAA AGTGAAATTA 480

CCATATGCTT TTGGAAAAAT ATTTGAGGGT GCGGAAGAAA CTGAAGAAGr GGTTCAAAAT 540

CTGCTAGAAT CkGAAAATTT TGAACCTGTT TTTGAAGAAT TGAAGGAGTC CTATCCTAAT 600

TTGTGGGAAG TTTTCGTAAC TGAAAGAGAC GATTATTTAG CAACAAAAAT TCAAAATACA 660

GCGAATGGAA AGACTGTTGC TGTTTTAGGG AAAGCTCATT TAAAAGGGGT TTCTGACAGA 720

TTAAAGAATA ATCAAAAATC CGATTTACAA AAATTAGAGA TTATTCCACC CAAAAAATTT 780

GGAAGTAAGT TGCTTGAATG GATTATACCA GGTATATTAC TTATTCTTCT AGGAGTATCC 840

TTTTACCAAG GGACACAAGT GGGAATAGAG CAATTGTTAC GTTGGCTTTT ATGGAATGGT 900

GGATTAGCTG CATTGTTTAC AGTATTTGCT TTAGGACACC CTTTAACTGT TCTAACGTCC 960

TTAATATTTG CTCCTTTAGC TACTTTACTT CCAATGGTTT CTGTTGGAGT ATTTTCAGCA 1020

ATAGTGGAAG CTACTGTTAG AAAACCAAAA GTAAAAGATT TTCAAACGAT GGATTTAGAT 1080

CTTCAATCTA TTAAAACAAT ATACAAAAAT CGTGTTTTAA GAGTGTTTTT AGTATTTTTC 1140

TTATCAAGTC TTGGAGGGCT TTTAGGTAAT ATTATTGGTG GGTTAGGAAT AGTTAAAAAT 1200

TTAnTTTAAA GGAGTGAGT 1219

(2) INFORMATION FOR SEQ ID NO: 476:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 943 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 476:

ATATTTTTGA AACGTTTGTT TAGTTACAAC ACGTGTTGCT ACCACCAAGC CTTGGTAAAT 60

TAAATAGTAT TTGATTTTTT TCTGACAAGG TAACCACATC AGTTGCCAAT CTCGTTGGTT 120

AGCTTGACTA ATTCGTTGAT TTTGGTTTGC AAAGCGCGTC AGAAAGAGCC AATCTTCCTT 180

GAGTTTTGCC GCTTTTTCGA AGGCTTGTTT TTCAGCAGCC TCTTGCATTT GTTGAGCAAT 240

TCGTTTGATT GGTTCTCGTC CTTGATTATG GAAAAAGTCG AACAATTCTT GTTGAAATAG 300

TTCGGTCGGA ATGGCACTCT CTTTTTGAAA GGTTTGATGC CAGTAGTTGG TTGGGACGAG 360

ATCGTAGATA CTATCCAAAA TTATTTTTAA ACGGTTTAAT TTTCGGCGAA TCGAATAAGG 420

ACCAAAAACAG TTCTCTTTAG TGGGAATATT TAAGAGTTTA ATCTCTAATT GCCGCTCATT 480

AACGGCTACC GATACATAAG AATATTGTTC ATAAGCATTC ATTTGGCGAT TGTAGCGCGG 540

CCGATACTGT TGAATCAGTT GACATTCTAG GAGGAGTGCA TCCAATTCTG TGTCAGTATG 600

GAGAACATCG AAGTCTATAA TTTGTTTGAC TAAACGCATC GTCTTTCGAG AATGTTGATT 660

ACTATTAATA AAATAGCTCG ACACGCGATT TTTTAATTTT TTGGCTTTAC CCACATAAAT 720

AATTGTGTCG GTTGCGTCTT TCATTAAATA AACACCCGGT GTTAACGGTA AtTGATGGGc 780

TTTTTCCTTT AAATATTCTT TCATAAGTtT CCTCTTTtCt TAAAAGGATC GGCTTCAAGT 840

aTACAAAGTT TTCTGAAATA GTCAAGTAAc AGAAATAGnC ATAGGTAATT TCACTGAAAA 900

ATAGTTTTCA TTTTTTTTAA AATATGGATA ATTCnAGGGA ATT 943

(2) INFORMATION FOR SEQ ID NO: 477:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 908 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 477:

ATAGATGTCT TTTCTGTCTC TTTTTTTGTA TAATAAACTT ATCAAATGAT GTGTATCTTA 60

CAACTTTGTT CGTTTGCAGA AAGAACAAAC GATGTTTTTT ATTGATAAAA AGTTATATAC 120

TAAAAGTACG ATGAGTATAA CTAAAGGAGA TGAGAAATGT GAGTGTTACC CTATCTAGAA 180

AAATTGTTAA TATTATTTCC ATTTTGGGCA TTGTTGCAAC CATAATTGCA ACCATCTATT 240

TTATCCGATT AGGTGTCTTC AAGGACATCA ATGCACTAAG AGGACTAGTA GGGCAATCTG 300

TCATTCTCGG TCCGATTATT TTCATGTTAA TTCAGATTAT TCAAGTGGTT ATTCCGATTA 360

TTCCTGGTGG TGTTAGCTGT GCTGCAGGGg TcTTGATTTT CGGGCCAACA ATGGGCTTCG 420

TCtAtAATTA CGTTGGTATt GCACTTGGCT CTATTATCAT TTTCTTATTA GGTCGTAATT 480

ATGGGAAACC ATTCATTATG AGCCTTGTCA GTGATAAAAC ATATAACAAG TATATCGGTT 540

GGTTAGACAA TGAAAAACGT TTTGAACGAT TGTTTGCTTT AnCAATTTTC TTCCCGATTG 600

CACCGGATGA TGCGCTTTGT TTAATGGCCG GTTTAACAAA AATGTCTGTG AAAAAATTTA 660

CTCTGATTAT TTTATTAGCA AAACCTGCTT CTATCTATTT ATATAGTCTT GCGTTAATTT 720

ATGGCGGTAC ATTTTTAACT AGCTTACTGG GAATGTAAAA TAAATAAAAA AGCGAACGGA 780

CTAAATCAAA ACTCTGCCTG TGtCAGTTTg ATTTAGTCCy TTTAwTAtGG AATTTTGGtA 840

GAAcTGGGGG rctTTTGawG kTTAgtTAtt TTAAAAAtAG tATTGkGaAa CgCCCAGATT 900

TTCCTTAA 908

(2) INFORMATION FOR SEQ ID NO: 478:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 605 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 478:

CGAACGAGAT TTAAAAAAAT nACGTAGCCT TTTTACAAGA ACAAGGCCTG CATTCTTGGG 60

ACGAAGTTGA CCGTTACATG ATTACAGAAT TTTTGCAATC CCTTCACGAA GAACAGCAAG 120

CTTCAGCaAG TGTGATTCGG ATGATTTCTA GTTTGAGAGG ATTTCATCAA TTTTTACGTC 180

AAGAACgCTT AAGCGAACAT AATCCGATGC AACATATTGA CTCCCCCAAG AAAGTTCAAA 240

AATTGCCAAG TACATTATCT GTTGACGAAt GACTCGATTA ATTGAAACGC CTGATACAAG 300

TAAACCACTA GGTATGCGGA ATCGTGCTAT TTTAGAAGTG ATGTATGCAA CGGGCCTGCG 360

AGTTAGTGAA CTAATTGAAA TAAAATTAGG GGATCTGCAT TTATCCATTG GTTTATTGCA 420

AACAATCGGA AAAGGGGACA AAGAACGAAT TATTCCTTTA GGCGATTATG CGATTCAATG 480

GATTGAGCGC TATATGAACG AaGCCAGACC TCAATtGATT AAAAACCATC CGAATGAAAC 540

ACACCTGTTC GTTAATCATC AkGGCGAACC GTTATCTaGA CaGGGCATTT GGAAAAATTT 600

AAAAC 605

(2) INFORMATION FOR SEQ ID NO: 479:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 955 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 479:

TGGCGTCGTT CTTACGATAC TCTTCCTCCA TTAATGGAAG CAACAGATGA AGGTTCTGCA 60

GCAAACGACC GTCGTTATGC AATGTTAGAC CAACGCGATA TTCCTGGTGG CGAAAACTTG 120

AAAGTTACTT TGGAACGTGC ATTACCATTC TGGCAAGATG AAATTGCGCC AGCTTTAAAA 180

GACAACAAAA CTGTTTTAGT AGCGGCTCAT GGTAATTCAT TACGTGCTTT AGCTAAACAC 240

ATCGAAGGCA TTTCTGATGA AGATATTATG GATCTTGAAA TCCCAACAGG TAAACCATTA 300

GTTTATGAAT TAAACGATGA TTTAACTGTG AAAGAAAAAT ACTACTTATA AGAACTTGTT 360

TCTTATTAAC AAAAGCGGTT TGCCAAGGCA AACCGCTTTT TTATGGgATT AAATAAATtG 420

ACTTTtCGGkT AAACmATCTT TATAAAyCGC TACTAwAtCC GCTGCATCCA ATGGTACATA 480

AGCCTTTTCG GCAATTGCAC TATGGTCCAC AGCTTGTTGT GCCATTTCTT CAAATTTATC 540

TGCTTCGATG CCAACTTCTG GTAATGTCAT TGGAATGCCA CAAGCCACAA AGTAATCATA 600

CAATGCTTGA ATACCTTTTT TCGCTGCTAC TTCCTCCTCT TGTTCGACTA TTCCCCAAAC 660

ATTTCGAGCA AACTGTGCAA ATTTGCCTAC GGTTTGTTCA GACAAAACGT AATTCATCCA 720

ACGAGGCGTT AAAATTGCTA AGCCAATCCC ATGCGTAATG TCATAAAACG CACTTAATTC 780

ATGTTCCATT GGGTGACAAG ACCAAACACC TTGTTTGcCG CGCCCTGTTA AACCATTtAA 840

GGsTAAACTA mTTGsCCACA TTAAATTAGC ACGTGCATCA TAATCnTCyG GGATTTCyaA 900

AGgCAaTAGG rCAATTTTga TAACsGTcGs ATtAAGCCTt CTGACACnAA ATCTT 955

(2) INFORMATION FOR SEQ ID NO: 480:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3604 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 480:

TTGGCTTTTA TCTGTGTTAT ATTTTTTTTA GTTACTTTTA CTATTGTGTC ACTACCATTA 60

ATCACTGCTC TTAAGACACA AGGATTTTCT TCTGAAGAAT TAATTGAATT AGCTAATTTT 120

GATTTAAATG TTGATATCCC TTTTTCTGAA CTAAACGTCT CAATTATTTy AATTAGTTTT 180

TCTGGTGCCG TTATTGACGG AAGTATGGCT ATTTGTAGTy CTACTTACGA AATATATACG 240

AAGAACCCAG ATTTAACTTT TAAAGAACTA TTTAATTCTA GTTTTAACGT AGTAATTGAG 300

GTATTAAATT CGACTATATA CACTCTCTTA TTTGCTTTTA TAGCATCTAA TTTTGCATTA 360

GTAATTTACT TACAGGATTT AAACTATTCA TTTTTAGAAT TAATTAATTC TAAAATTTTT 420

GTAGGGGAAC TATTAGTAAG TATTTTAACT GGTTGTGCAG GAATTTTAAT TTTACCTTTG 480

TCCTGTTTAC TAGGTAGCTG GATATTAAAA AAAAGATTCC ACACTTTGTA TTGAAACCCa 540

AAAGCTAGGC CGAATACCTT TTATAGAAAT ATATTTGGTA TTTTTATACA ATTAACCGTT 600

CACTTATATT GAACAAAATT GCAGCCAAGT TTCGTTTTGA TTTATTGGTG ATTGAAATAA 660

TTAAATCCAT TTGTTATGTA AGTCTTGATT AGTGTGAAAA AACACCCTCT AATCCAGTCC 720

GCTCTTTTCC AAATACTTAA ATAAAAGAAT CCACAACTAA TTTATCTGTT ATTCTATGTA 780

ATATACTCAA CAAAATTCAC AGTAACTTAA CTATGACTTT TCAGTAAAAA TCAATTGTGC 840

ATACTGATAC AGTAAGTTTA AACTTTTATT GTAATAGTAT AATTCTTAAT TTATATTTTT 900

TACTTCTATT CCTAATTGAT TTATTCAAAA TTATCACAAA ATCGTAAAAG TTAAAATTGC 960

ATATTCTAAC ACAAGAAAAT CAAAAAAAGA TTTTCTTGTG TTAATATTTT TGAATATACT 1020

AAAGTTCGGA TAAGTAGAAA ACACCCTACT GATTTATATA ATAGTAGGGT GTTTTGTTAA 1080

TATAATTTAC TTAAATTCTA TAACTCACAG TCTATAATCA CCTTTTATTC ATTTAAATTT 1140

AATCCGATAA CATACTTAAC TCTACGTCAA AGTATAATTT TACCACACGA ATCACCATAT 1200

ATAGTTGTAT AACTATTTAA GATGACTTCA TTAACTTGTT CAGGTTTAGT AAAAACAACT 1260

ATGACAGATG GATCTTTTCC TGCTTGTTTA ATTTGATCAA AATCAACTTC GGCTAATAAT 1320

GTTTCGGTAG TAACTTTCTC TCCTGCATTA ACAAAAATTG AAAAAGGTTT ACCCTCCAAT 1380

TCAACTGTGT CTATACCAAT ATGAATTAAA TACTCAATAT CTGCATCGCT TTTTATACTA 1440

ATTGCATGCT TTGTTGGAAA AACAGCTTCA ATTCGTCCAG TCAAAGGAGC ATATATTTTG 1500

TTACCTTCGG GAACAACTGC AAATCCTTCT CCAAGTACTT TTTTTGAAAA TATTGGATCA 1560

TTTACTTCTG AAATAGGAAT AATCCTCCCT GAACAAGGAG CAGTAATGGA AATTCCTTCT 1620

TGTTTTTTAT TTTTAAATAG ATTAAACATA AGATATCTCC TAATTCTCAA TTTCTAAAAT 1680

TTGATTAATT TCTTGACTAA GTAAATCAGC TTTACCGCCA AAAACTGCCT GAATACCACC 1740

ATTCACTTCA AAAACAGCAG TAGCACCTAG TTGTTTAATA GTTTCTTTGT CAACCATTCG 1800

GCCATCTACA ACAGCTAAAC GTAAACGAGT AGCACAAGCT TCAACGGgAT TCAATATTGG 1860

TTTTACTACC AAGAGCTTTA ATGATTATTT GAGCTTCATC ATTTAATGAT GTTTTTTTAT 1920

TTAATTTATT TTCAGTAGTC GTTGGCAACT CTTCTCCCAT TCCAGGAATA GCTACTTTAA 1980

ATTTCTTAAT GAAGAAAGTA AATACGCAGA AATAAACAAT TGCCCAAGCA ATTCCGAAAG 2040

GAATTAATTT CAGCCAATTT GTCTTATCGT TTCCTTGGAG TACACCAAAT AATAAGAAAT 2100

CAATTAATCC ACCAGAAAAT GAATTTCCTA CTCGGATATT TAAAATATCA GCAAAATAAA 2160

AAGATAGACC ATCTAAGAAT GCATGAATTA TATATAACCA TGGGGCTACA AATAAAAACA 2220

TATATTCAAC TGGTTCAGTA ATCCCTGTTA AAAATGAAGT CAAGCCACCA CTAAAATAAA 2280

GTCCACCATT CTTTTTACGA TTTTGTTTaG GTATAGCACG ATACATAGCA TAACAAGCTG 2340

CTGGTAAACC AAACATCATT GTCGCGAATC TACCTGCAAA GAAACGAGTA CCATAAGTAA 2400

ATAGTCCTGT ATGGTTTGGA TCAGCTAGTT GCGCAAAGAA AATATTTTGT GcACCAGAAA 2460

CAGTAACTCC AGCTACACTT TCTGtaCCTC CTAAAGACGT wTACCAAAAA AGTGGATAAA 2520

TAGTATGATG TAATCCTACT GCGCCAGTCA AACGAAGTaA AAATCCATAC AAAAAAGTAC 2580

CTAGACTTCC CATTTGTGCA ATATGTTCTC CAACTACTGC TAATCCATTT TGAATTGGTG 2640

GCCAAATTAA ATAAAAGAAA GAACCGATAA TAATTGCCGC TAAACTAGAA ATGATTGGAA 2700

TAAAACGCGA TCCGCCAAAA AAGCCCAAAA ATTGCGGTAG TTCAATTTTC CGATATCGAT 2760

TGTGTAGAAA TGCAACAGTT CCGCCAATTA CAATAGATCC AACAACACCA GTATCAAGAG 2820

TTGCATCTTT AGCAGAAAAT AACGCAAGAA AGCCGCTGAT CGTTGCTGTA TACACTAAGT 2880

AAGATACTCC ACCGGCTAGA CCTGCTGTTC CTTTATCCCC ATTTGCTAAC CCTACAGCAA 2940

TACCTATAGC AAAAATCAAT GCTAAATTAG TAAAGACAGC ATTTCCTGCA TAACTCATAA 3000

TACTTAATAC TGTATGGATC CATGGCTGTT CTAAAATTGG ATACGCATTC ATTGCACTTT 3060

CATTTGTTAA AGCACCACCA AGGCCTAAAA GTAAACCTGC TACTGGTAAG ATTGCAATCG 3120

GTAACATAAA TGCGCGACCT AATTGTGAAA ACTTCTTAAA CATCTTTTTC CTCCTTATTC 3180

AATGCATCGA CAAATCTTTT AGCAATTTCT TGTGGTCGTG TAATTGCACC ACCAACGACA 3240

ATTCCTGCAA CACCCAGCTT TTTAATTTGT CTTGCTTGTT CAGGTGTATG AATTTTCCCT 3300

TCAGCAATAA CACAAATATT TTCTTGTACT AATTGTTTAA TTAATTTCAT ATCTGGGCCA 3360

TTTTGTTTTT TGCTCTCCTT TGTATAGCCA CTTAAGGTTG TACCAACAAA ATCCACACCA 3420

GCTTGATAAG CATTCTTTCC TTCCTCAAAA GTAGAAATAT CAGCCATTAA TAATTGTTGA 3480

GGATACTTTT cTTTACTTTT TCTATAAArG TATTAATAGA TAATCCGTCA kGACGTTaCG 3540

TAAAGTACAG tCCAAAGCaA TTACCTCCaC AcCAATTGcA cTAATTCAcT ACTCCCCCAT 3600

AGTG 3604

(2) INFORMATION FOR SEQ ID NO: 481:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 646 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 481:

GCGCATGGTT ATATGGGTTC GGCAGAAACG ATGTCTGTCT TTGCGAAAAT GTATCATGAT 60

TGGGGCTACA ATGTGTTAGC GCCAGATGCG CGTGGTCACG GGAAAAGTCA AGGAGATTAT 120

ATTGGGTTCG GCTGGCCCgA TCGTAAAGAT TaTGTCCAAT GGATTGAGAA AGTTTTAACC 180

GAAAATGGTC AGCAAGAACA AATCACTTTA TACGGTGTCA GTATGGGTGC TGCTACAGTC 240

ATGATGACGA GCGGGGAAAA ATTACCAGAT AACGTGAAAG CCATTGTGGA AGATTGTGGC 300

TATTCTACAG TCAATCAGGA GTTGCAATAT CAATTGAAAG AGCTTTTCAA TTTACCAAGT 360

TTCCCATTAG TTAATGTTAC TAGCGGGATT ACGAAACTTA GAGCTGGCTA TTTCTTTGGG 420

GAAGCAAGCG CTGTTAAACA ATTACAAAAG AATCATTTAC CAATGCTCTT TATTCATGGT 480

GAAAATGATA CGTTTGTGCC GTTTAGTATG TTAGACGAAk TTTATAATGC AACCCAAGGa 540

CCTAAAGAAA AATATGTCGT GcCTGGaGGC AGAGCATGGC GAAAGCTTTA TAATAAAAAC 600

CCAGAAAAAT ATAAGGGAAC AGTAnGCGGC TTTTCnAGAT AAATAT 646

(2) INFORMATION FOR SEQ ID NO: 482:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1322 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 482:

AAAGAAGAAT TAATATTATT ACTGAATGAT AAAGATTGTG CCTATGATGT ACGGATTCAA 60

ATTCAACAAT TTCTTGAAGA AAGGTGAGTA AAGACAAGAT GTTCAGAAAC AAATGGAAAA 120

ACAGTTCACT AATTAAAAAT GTGTAAAAAG GGTCTGATGA AAAAAGGAGA CATAATAGTT 180

ATTATCTTTT TAATAGCTAT CTCTTTTTCT CCATATTTTA TTTTTTTTCA CAATAATCCA 240

TTTAACTCCA AAAGTTTTGA CGACACTAAA TATGCTGTGG TCAAGATAGA TGGGAAAGAG 300

ATTGAGCGTA TAAATTTAGA TGATTCAAAA GAATTTATCA AAACATATTA TCCATCAAAA 360

GGGCAATATA ATACTATAGA AGTTAAAAAT GGGCACGTTC GTGTAAAAAA AGATAATAGT 420

CCAGATCAAA TTGCGGTGAA AACAGGATGG ATATCAGAAC CAGGGCmAAC TAGTATCTGT 480

ATTCCTCACA GATTCATTTT AGAAATTGTT CAACAATATT CTAAGGATtA TTaTATTtAC 540

TAAAATCATT TTTTGGCATA CATGCTAAAA AATGATTTTT TGTAAATATT TTATTGCATT 600

AATATCAATA TAGTTTAGTT TTATTTGGAA GAATGTAATA ACAAATATGA GATGGGATTG 660

TCCTTATTAA ATAACATTTA TAAATGTTAT ATTTTAAGTG AAAGTAATGA TTTACAGTTA 720

ATTAAATGGA GGTTTTATGA ATAACCTATT TATTTTtGTT TGCTTTTGTT TTATGTGGTT 780

AATGATTTTT GGAGTAATTT TATACTATGT TGGTCTAGTT AATCACAGAT ATATACATCA 840

TACTTTAATT TTGGGTTTAG TAACTATTAT TTCTGGTACT CTATGTTGGT TATTTGTAGG 900

ATACTCATTA AGTTTTTTTG GAAATATTCA ATATAGTATC kTTTATAGCC CCTTAGCTTC 960

TAGTGAAATT GTTTCTATAC TGATACAATT ACTTTTTTGC TTATATTCTG TAATTATGAT 1020

AATTGGTTCA GTATTAGAAA GAGGTAATTG GAAATATATT GTGTTGTTTG TTCCTTTGTG 1080

GATTGTATTT GTTTATGCAC CTATTTGCwT TTCGTTATGG GGACATGGAA ATTGGCTAGG 1140

GAAAATGGGA GTACTGGATT ATTCAGGCGG ACTAGTAGTC CATACTACCG CTGGAATAGG 1200

AAGTCTTGTT TTAGCAATAA CATCACCAAT ACGTTTAAAG AATTCATTGA TATTTAAATC 1260

ACAAGAAATG aTAGCTTTTG TAGGAATGCT GTTTATTACC TTAGGaTGGt TCGGaTTTAA 1320

TA 1322

(2) INFORMATION FOR SEQ ID NO: 483:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1268 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 483:

GACCTGGTTT GGATCGGATT AATGCGTTAT TAGAAAAAAT TGATCATCCA GAAAACAAAG 60

TGCCAACGAT TCATATCGCA GGAACCAATG GTAAAGGCTC AACCGTGACA TATTTACGCT 120

GTCTATTAGA AGAAATGGGC TTGAAAGTCG GAACCTTTAC CTCTCCTTAT ATTGAATCAT 180

TTAATGAAAG AATTGCGATA AATGGTCAAC CAATTTCAGA CGAACAATTG ATTACCTATG 240

TGGAAAAGTA TCAACCAATC ATTAAAGAAT TAGACCAGAT CACTGAAGTG GCTGGAATTA 300

CAGAATTCGA AACATTGACA GGAATGGCCT TGGACTATTT CGTGAATGAA CAAGTGGATA 360

TCGCCGTTGT TGAAGTTGGC TTAGGTGGAT TACTTGATAG CACCAATGTC GTGAAGCCAT 420

TATTGACGGG AATTACGACA ATTGGGAAGG ATCACACAGA AATTTTAGGT GAAACAATTG 480

CTGAAATTGC CTATCAAAAA GCGGGTATTA TTAAAGAAAA AGTCCCAGTA GTAACAGGCA 540

ACATTTGTGC TGATGCGCTA gcAGTCATTG AAAAAGTCGC ACAAGAAAAA CAGAGTCCAA 600

TTTTTCGTTT TGGCAAAGAG TATCAAGTGG AATATCTACA TCCTGATACA CAATGGGGGG 660

AAGTATTTAA TTTCTATGGA GAAATGGGCA AACTGACAAA AATTAAAGTT CCCTTGTTGG 720

GGCGTCACCA AGTAGAAAAT GCGGCGGTAG CTATTCAATT ATTTGACAAG TATTGCCAAC 780

TACAACATTT GCCATTTAAA GAACGAGACA TTACTCAAGG GTTGGCTAAA GCGCAATGGC 840

CAGCTCGAAT GGAACGTCTT AGTGATGAAC CGTTGATTGT TTTAGATGGT GCCCACAATG 900

ATCATGCAGT AAAACGCTTA GTTGAAAACT TGCGCAAAGA GTTTCCACAA CACACGATTC 960

ATATTTTATT TTCAGCTTTA GCAACTAAAG ATGTGGATGA GATGATTCAG GACTTAAAGC 1020

AAGTACCGAA TGCGCATCTT TACTTAACTA GCTTCGATTA TCCTAAAGCT ATTGCACTTA 1080

CAGAAATGGA AAAATATGAA GACGATTTAA CAGAAATTGT GTCTTTATGG CAATTTGGCT 1140

TAGGCGAAAT TTTAGAAAAA ATGTCGACAG ATGATTTATT ATTAGTAACA GGTTCATTAT 1200

ATTTtGTTTC TCAAGTAAGA GAATTACTTT TAACTATAGG AGGAAATGAT GAAGAAATTT 1260

GATGGCGT 1268

(2) INFORMATION FOR SEQ ID NO: 484:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1217 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 484:

nAAGCAATAA TTGAACCGTA AATTGCaCCA ACCCCtTTGT TCGTACCaAT CATGATTAAA 60

GGCGCAACTG CTTGAATGGC ACAACCTAAA ACGACTGGCA AGCCAATCCC AAAAAACTTA 120

TTCACCGTTA ATTGTAAAAG TGTCGCAACG CCGCACATAA AAATATCGAT TGAAATCAAA 180

TACGTCATTT GTTGCGAATC AAAATTCAAG CCGGTTCCGA TTAATAAAGG TACGGCAACT 240

GCTCCAGCAT ACATCGCTAA TAAATGTTGC AATCCTAAAA CGGCGGCTTT CCCATTTGGT 300

GTTTCCGTCT GTTGTGGTGT ATTCAATTGT TTTGTTTCCA AGATTATGCA TCCTCCTCTA 360

AAAACGTCAC CGTTCCTTCA CTTAATGAAG CAATACGTGC TAAAGAGACT ACATTTaAGC 420

CCATGTCTTC TAATAACTGa CGACCATCTT GGaATGaTTT TTCaATTACG aTACCAATTC 480

CTTCAACCTT TGCACCAGCC TGTTGGCATA ATTCTACCAA TCCTTTGGCA GCTTGTCCGT 540

TAGCTAAAAA GTCATCAATA ATCAACACTT TATCAGCATC AGATAAAAAT TTTCTTGAAA 600

TGGAAATTTG ACTTGTCACT TGTTTTGTAA ATGAATAAAC TGAAGCCGTT AAAAGCTCTT 660

CATCCATGGT TAAACTTTTT GCCTTACGAG CGAAGATCAT TGGCACGCCT AATTTTTGCG 720

CTGCGTATAA GGCAGGAGCG ATTCCTGATG CTTCAATTGT AATTACTTTG GTAATGCCTG 780

CTTCTGCGAA AACCTCAGCA AAACGATTAC CCATAGCCTC CATTAATTCT GGATCCACTT 840

GGTGCGTAAT AAAACTATCT ACTTTTAAAA CACCTTCACC TAATACACGG CCATCATTTT 900

TAATACGTTC GACTAATTCC TTCACAATAT GAACTCCTTT ACTTATTTTT CGATAGATGC 960

TTTGTTCTCA TCCTGAGGAA AGGACGGTTC GTTAAAAACG AACGAAATAA AACAAAGTGC 1020

TATTTATATT CGTGTTCTAA CAAAAGAArA AAGCCCATTT TGCCAAAGCA AAAAAGGGCT 1080

ATCCGCTCTT CTTCACTTAT AGACCAGCAT TTACGGTGCT GGGTAGAGAn GGTCAACCGT 1140

ATTTTGACCT TATACAAGCT TCTATGTGTC TCTAnGGATA GTACAAAAAT CTCTAACTnT 1200

CCAATGGGTA ACAACTT 1217

(2) INFORMATION FOR SEQ ID NO: 485:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1929 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 485:

AACCATCCCC TAAGGATTGA TTCAACTAAC GCTTTACGGT CGATTGAGTA AGAAATTGCT 60

TTACGTAAGT TCGCATTTCT AAATGGTGAT TTTTCATCAC GTTGATTTAG TTCCATATAT 120

TGTGTTGATG CTTCTTTTTG ACTAACAAAA GCTGGGTCAT TGGCCATTTG TTGGGCTAAT 180

TCACCAGAAA GAACGACATC GTCTGTTTGT CCATCTTGGA ACAAGTTCAA CGCGGTTGGT 240

GATTCTTTCA CGACATTCAC ATCTACTGAG TCCAGTTTCA CAGTATCTTT ATCCCAATAT 300

TGATCGTTTT TCTTGAATGA CCATTTTGTA TCTGTACCAG GACCATCAAA GCCGTCTAAG 360

ACGAATGGAC CATTGTAGAC AGCACTTTCG CTGTTTGATG CATAATTTTT ACCATATTTT 420

TCCACAATGT CTTGACGTTG CGGGAAGAAT GATGGGAAAG CTAATAAGTA ATCAAAGTAT 480

GGTGTTGCTT TTTCTAAAGT GATTTCTAAT TCTGTATCAC TGACTGCTTT AATTCCTAAT 540

TCTGATTTAT CTTTTTCCCC TTTAGCAATG GCATCACCAT TTTTTACAGA GGCATACAGA 600

TAAGCATATT CAGAAGCTGT CGCTGGATCA ACTGTTCGTT GCCATCCGTA AACATAGTCA 660

TTAGCAGTCA CTGGTTTACC GTCTGACCAT TTTGCATCTT TATTTAATTT AATTTTGTAT 720

GTTAGTCCAT CTTCAGAAAC TTCTGCTTTT TCCGCTGCAC CTGCAGGTTG GACTTTGTTA 780

TCTTTGTCTA AACGATAAAT TCCTTCATAT ACATTATTTA ATGCAATAAA ACTAATTCTG 840

kCTGTTGCTA GTGATAAATC AGCTGTTGGC ATTTCTTGTT GTTCTGTGAC ACGTAAAACT 900

TGTTCGCCAC TTGCCTTACC GCCACCAGAC GCGCTTGAGC TATcTGACGT TCCGCCGCCT 960

CCACAAGCCG CTAAAGTTGT TACGCTAAAT AATGTAATCA CACCGAATGT TAATGACTTT 1020

TTCAATTTCA TAGTGTTGCC TCCTAAAATA ATGCAATCTG TATTTTCTGA AAAATCaGaA 1080 kTTAATAATT ATTaTkGCTA TGGaATATaT CTtACATTTt TCATATTAAA AtgCAAGTCA 1140

AAtGkTAACT TTTtTTAATT tTccTCcCTA TCAGAAACCA CTCTTTTTCT TCCTGTCTTT 1200

TATTGCAATT CTTTTGTCAG TTTTCTAAAC ATTATGGTAA ATTAGGAAGT AAAGGAGTGA 1260

CAAAATATGA AAAATTCTAC GTTTTTAAAG GATTTGTTTG CTATTGATAA AGCAAATGAC 1320

GAaTTATTTC GGTTaGTTGG TGTCGCAATT TGTATGGCAA TTCCATTATT AATTGGTTAT 1380

TTTTCCAATA ATTTACTCAT TGGGACCTTT GGCtCGATGG GCATTTATAC GTTTATTTAT 1440

TATCAGCCGC TTCCTTTGCC TCAATTATTG CGCAGACTGA ATATTGTTGG CTTTTTTATT 1500

GTTCTGGGCA ACAGCTTAGG AATGCTAAGT CACCATGTAC CTTGGTTGAT TCCCATtACA 1560

ATTGCCATCG TCGCTTTTCT TGCACGCTTA CTTTTTAGAT TGTATGGCAT TGAGAAGCCT 1620

GGTGCTTTGT TAGTTATCAT GTCAACCGCA ATGGGAACAA GCAACAATTT TCCTTTACAT 1680

AAGATTCCCA TCATGGCCAG TTTTGTTCTT TTAGGAGTAG TGACCGGCAT TATCATGGGC 1740

ATTGTGCTTC ATTTCATCGA CAAGCGCCCC TACGTTTTTC AAAAACGGAT GTCTCTTCAA 1800

GAACGACTTT ATATCGATCC CGCTTCTTTA TTAGATGCTT TGCATTATGC GGCCATTTTA 1860

TTTTTAGCGG CTTATTTAAG CCAATCGCTT CATTTAGTCA ACGCCTATTG GATGACGTTT 1920

ACTTGCGCA 1929

(2) INFORMATION FOR SEQ ID NO: 486:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3893 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 486:

TTGGTGATTA CGACTGCATT CAGTATTTCT ACGATGTTTT TCAATATTTC GGCTACCTAT 60

CCCTTCTTTA TGGTAGCTTT TCTACAAATT GTGACCTTTG CCGGTATCTA TTTCAAATTG 120

GGTGACATTA TGAGTATGTT CAACCTTCAG AGTAATGACA GTCAGTCAAT GGGCAGACGG 180

GTCATGCGAA GACCGCAAAT GTTGATGAAC CGTAAATTGC GGCAACTCAA TCGAAATGTG 240

GGCCGAACTC TGGCTTTTGG CGGTGCTGCA GCAGTTGGGA ATAAGTTGGC AAAGGAACAA 300

TCAAAACCCA AATTTAAGCC AGCAGGTTCT TCTCTACGAA AGAATTCACG ATTACCTAAC 360

GATCGCGAGG TATCCTCAGA CTCGGCTAAG GAGAATCCAA TTTCAAATAA TAAGAAACAA 420

TCCCGAATGA ATTTGACTGG GAGAAAAATA GGCAAAGTTC TGGATACCCA GGCTTTGGTG 480

AAAGATAAGG CAAAACAAGT AAAGGATCaG GTACGGAATA CGCCGACCAA CTTGAAATAC 540

ATCCTCCATA AAGGTCTGGA AAAGACGAAA AAAGCACCGA AAGAATTTAA GCGAGGGCTT 600

GTTCAAGAAA AAGCAGATCG AGAGAAGCTG CGAGACAAAC AACGACAACG CAGAGATGAG 660

AAAATGGATG AAAAGCGAAA AAACTTGGGC GAAGTAACAG ACCGTCATGG AAAAAGAAGA 720

AGTAATGTTT CGATAAAAGA GGATCCAAAA CCTCAGAAGG ATAGGATACT GAAAAATGAA 780

CTTCCAAAAC GGATTGTGAT GGTAAAACCG AATTCTGAGG TAAGGAGAAG ACTTGTTAAA 840

CAGAATTTGG TTGCGAGAGA GAATAACCAA CTATTAAGTA AGAAAAATTC TTTCCAGCAA 900

GTTAAGCAAC GATCTACTCT TCCGAAGAGA GCCAATCAAA AAGTACAAAG AATAATGAAG 960

GCTCGTCCaA AACCAAAGTC GGGTGATAAA AAATGAATTT AAAAGAATTG AGTGTCATTT 1020

CGTTAACCAT GTTGATGCTC TCTTTTATGG GCATCTTGAT GTGTGTAAGC TTATTGTTTG 1080

GGGAAGATAG TGAGGGGAGT GGTACTTCTG GTATTTCTCC TGGTGGCAAT AGTGTCTCAG 1140

AAGAAGTCTT GAAACATCGC TCAATGGTTG AAAAATATGC TAAAGAATCG GACATATTAG 1200

ACTATGTACC AATTCTTTTG GCAATCATCC AAGTTGAATC TGGTGGCACA ATGGAGGACG 1260

TCATGCAAAG TTCTGAGTCT GCGGGTTTAC CACCAAATTC ATTAAGTACC GAAGCATCTA 1320

TCAAGCAGGG ATGTATTTAT TTTGCTTCCT TGGTTAAGAG CTCCAAAGAA CTAGGGTGTG 1380

ATCAAGATAG TATTATTCAA GCCTATAACT ATGGTGGCGG CTATTTGAAT TATGTCGCAA 1440

AAAACGGGAA GAAGCATAGT TTTACTTTAG CCGAATCATT TTCAAAGGAA AAATCGGGTG 1500

GTCAAAAAGT AGACTACCCC AACCCAATAG CGATTAAAGA AAATGGCGGG TGGCGCTATA 1560

ACTATGGCAA TATGTTTTAT TGCCTATTAG TGAAGCAATA CCTGACCACG ACACAGTTTG 1620

ATGATAAAAC AGTCCAAGGG ATTTTTGACG AGGCTTTTAA GTATGAAGGT ACTGCCTACG 1680

TATTTGGCGG TTCGAGTCCC GCAACGGGTT TTGATTGTAG TGGATTGACC CAGTGGTGTT 1740

ATGCCAAAGT TGGTCTGAAA CTTCCCCGGG TGGCTCAAGA TCAGTATGAT GTCATGACTC 1800

ATATTGACTT AAAAGATGCC AAACCTGGTG ATTTAATTTT CTTTCACTCA ACCTATGATG 1860

CCGGTACTTA TGTGACTCAT GTAGGTATTT ATGCAGGAGA GAATCGCATG TTTCACGCAG 1920

GAAATCCAGT GGGTTGGACG AATCTTACAG AAAGCTATTG GCAGCAGCAT CTGATTGGAG 1980

CCGGCCGATA CAATTAAGAG AAAAGAGGAA AAATAATGAA GATAAAAATT GAACGGATTC 2040

AAAAGGAAAA GATGTCAAAA AAGAAAAAAG AGCGGACAAT TTCAGTGGGC AAGCATCGAA 2100

AAATGGTGCT TGCCCTTTGG CTTTTACTGT CaTGTAGTCT TGCtTTTGGg TATCTACAAA 2160

TATTTCACCG CCATTGaTCA GCATACaCTC ACGTAAAAAT GGTTGTAAAn GAAAArGTAA 2220

TAAaTACTTC TGGAGTTGAA AGTTTTACAA AGAATTTTGT GAAAGAATAC TTTTCATGGA 2280

AGAATAACAA GGAAGTAATT GAAAAGAGAA TGAGTAATCT TGGACAGTAC CTTACAGAAG 2340

AAGGCCTTGC TTTGAGTCAA GATATGGTTC GAGTAGATAT TCCCACTAGT TCAGAAGTAA 2400

ACTCAGTAAA AATTCTAGAT GTTGAAAAAC GTTCGAAAGA TTTTGTCGTT TCATTTTTAG 2460

TAGACCAAAA GATTACAGAA GGGAAAAAGT CAGAACAAAT TTCTTCTGCG TATCGAGTGA 2520

AGATATTTGA AGATGAAAAA GAAAATCATA TAGTGACGAG TTTACCAACT ATGATTAGTA 2580

AGCCAGGTAA GGCAAAATAT GAAGTAAAAC AAGTAGAAAG TGATTCTGAC ATTGATGCAA 2640

AAACAACGCA AGAAATTACT GAGTTTCTAG AGACATTTTT CAAACTTTAT CCAACTGCTT 2700

CTGAAAAAGA GCTAGAGTAT TATAGTGAAA ATGAATCTAT GAGCCAAGTT AGTAGCGATC 2760

TTAGATTTGT AGAAATAACC AATCCTATTT ATCTTGAAAA GGACGGTTCG GTACAAGTAA 2820

AAATCGTAGT AAAATACCTA AATGAAATTG AAAAGGTGAC AAATAGTTTT CAATACAATC 2880

TATTGTTAAA CAAAGAAGAA AATTGGAAAA TTATTAATGG AGAATAATTT TTTAAGGAGA 2940

ATATAAAAAA CAAGGAAAGA ATGAATCATT GTAAACTTGT GGGAGCGTCA ATAATTTTGT 3000

GTAAATGATT CTTCCCTACT GCAGATTGTT TCTAATTCTC AATCAGTTGG CTTAGCATTT 3060

GTTCAAGTTC GCCTTCGGCC TGTTGAAAGC CTTTATGACT GCGACCTAGG AATTTTTGGT 3120

TGTAAGTATC AAAAGACGAG ACTAAGAAAC GTTCCATCGA CTCTTCGTTT TGGAACTGTT 3180

CCTTGCGGTG GCTGTATTTC TTGATTTGCT TATTAAAGGA TTCAATCAAG TTCGTAGAAT 3240

AAATACTCTT GCGTATGGCC AAGGGGAAAT CATAGAAAGT GAGCAAGTGA TCGTTCGAAA 3300

GAATCGATTC AACCACTTTT GGATAGCTGG TTTTCCATTT TTCCGCAAAA GCACCACGTG 3360

CTTCCAACGC CACCTTTTTA GATGAGGTTT GATACACCAT TTTAAAATCA TCACAGACTT 3420

CCTTACGATC ATCGACACGC ACTTTGTGAC TGATATTACG GGAAACGTGT ACACAACAAT 3480

GTTGAAAACG AGCTTTGGGA TAGAACCGAC TAATCGCTCC TACCATCCCC TTTAAGCCAT 3540

CCGTGATGAA CAGGAGGACA TTTTTCAAAC CGCGCTCTTG AAGGTCCAAT AAAATTTCCT 3600

CCCAAATCGT GATGGATTCA GTCGGTGCAA yCGCATAGCT CAATACTTCC TTTGATCCGT 3660

CCGGGCGAAT GCCAACTGCG ATATGAATGG CTTCCTTGGC GACGGTTTTC CGCTTTAACG 3720

GAATATACGT TGCGTCCATA TAAATAGCGG CATAACGGTC ATGAAGCTCC CGCCCTTTAA 3780

ACGCCGTTAC CTCTTCTGTA AATGATTTkG TTATATTGGA CATGGTTTGG GgCGTGTAGT 3840

GaTGCCCATA CATTTTCTCA ATCAAGCTGC GATTTCGACA TGGTAATACC TTG 3893

(2) INFORMATION FOR SEQ ID NO: 487:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 601 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 487:

CTCGTTCATC TTATCCAGCA CCGCCCGCTG TGCAGGGGTA AGGACACGGG GCGGGGTGAT 60

GCGTACCCAT TTTTTCGGCA GTTCAAAGGA AAGCCCGCCA TATCCATTAT CCCCGGTCTG 120

CCGTACCTGT TCGGGGTAAG TGCCGCACAG TTCGGAAAGC TGCCTGCAAA GACTGGTATT 180

GTGGGTGTAG ATGCTGGCGG CGGTTTCTGC TTCGTTATAC AGCACGATAG TTTCCCTCTC 240

GGTTCCGGAC AGCTTCATCG CTCCACCTCC TGTCCGCGCG GGGGGTTCTC CGGCGCGTCA 300

CGGCCTTTGG CGGCATCTTC CCGGCACTGG TTCAGGGCAT CCAGCACCGA GGGTTTGGGC 360

GGCTCGTTAT TGATGATCCC GTCAATCTGG TTATAGTTGC CCTCCACGCT CATTTCAGCG 420

GCGGCAAGGT AATTTTCGGG GATtACTTCC TGAAAgGCGC GGACGTTCCC GTCCCCCGCC 480

AGTACCGGCA CATAGAGCCG CCCGGTCTGt AAATACATCC CGTCCAGGCC AGCCCCGCcT 540

TTGTAGCGTC cTGCGtGTCA CGCTCACACT CCAAAAGGGC AATCCCGTTT GCCAGAAGAT 600

A 601

(2) INFORMATION FOR SEQ ID NO: 488:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2109 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 488:

ACTCTTCTGC TAATTCTAAA AATACTTGAT CAACCACTTC ATCACCAATC GAATGGCAAT 60

CAAAATGTTC AGGATAATGT CCAGTTAGTT CCTTAAAGCG CGTCAATTGC GCAATCGTTT 120

CAAGTCGTAC TTCTTCATAT ACGAAGGTTA CTTGCTTTTG TTGACGATAA TAAGAAGAGC 180

GATGAAAATT GCCTTGTTCA TCTACCAGTG AAGGAATCAT ACATGGGTCC GCACAAGGTT 240

TCCCTAATAA AAAATTCGTA TGCTGACCTA CAAATAATCC TGGATAATTT TTTGCTAACG 300

CAACGGCGTG TtCTGCTGTT GGTAAATTGG GCATCAAAAG TGTCGACGTT GTTAAACCAT 360

CGACATAACT TTTGATAATG CCATGATTAA TCGCTTCTGA CATCCCAAAA TCATCAGCAT 420

CAACAATTAC TTTAACCAAC AGCTCCCCCT CTTTTCTTTA CGTATTTTTG aTGACAATTC 480

CGACAATAAT ATTAACAGCT GATTCACAAT CATTAATCGT GCGCTCAAGC TGaTCATAAA 540

CATCTTTCCA CTTGATAACT TTCAATAAAT CTTTTTCGTT AGTCATCAAA TCTTTCATGG 600

CGCTACTGTA TAATTTATCG CCTTGAGATT CAATCGTATT CACTTCATCA ATCAATGAAA 660

CTAACGTTTT GGAATTCTTA AATTTTGCAA ATTCCTTCGT TGCTGATTTA ACGCCTTCTG 720

TCGCTTTAAC AATATAAGCA GTTAATTCTA GCGCGGGCGC AATCATTTCT TCGACAACTA 780

AATGATCAAG TAAATAAGAC AAACTATTAA TGCTATCAAT AATGTTATCT AAGTGATCCG 840

TAATATCGAC AATATCTTCT CGATCAATCG GCGTAATAAA GGAAATATAG AGTTCATTTA 900

AAATTTTGCG AACAATTTCA TCATTCTCTT TTTCCAATTG ATGAATCGCT TCTGATTTTT 960

TTACTAAGTT TTCTAATGAA TAATTGTCTA CAATTTCTTG CAAAACCTTT GCCGCTTCAT 1020

AAGCATTTGT TGCTAACTGG TTTAATTCTC CAAAATAATC AAATTGTTTT TTACGTGCCA 1080

TGTTTTTCTC TCCTTAAAAA ATTGCAATAA ATAATTTTGA CATCACATAA GCGATCrGCC 1140

CACATCCTGG GAAAGTTAAA ATCCATGmwA TCAGCATTTC TTTCACGATA CGCCAGTCAA 1200

CACTTGAGAT TCGCTTAGAA GCACCTACTC CCATAATCGC TGTTGTTTTT GTATGAGTAG 1260

TTGAAACTGG AATCCCAAAA GcCGaTGsTw AAAATAAACA AaTAGCCGTG CTTAAATCTG 1320

CAGTGAAGCC TTGATAGCGT TCAAGTTTTA CCATATCCAT CCCGACTGAT TTAATGATTC 1380

TCATGCCGCC CACAGAAGTT CCAATGCCCA TTGTAaTGGA aCACrGAaCC ATAcCCATAA 1440

CGGAtCACAA AACCGCCACt GTTTTTTCGA CTAAATTATT ATAAAAAAGC ACTAACATGA 1500

ACACGCCCaT AAATTTTTGC CCATCTTGCG CACCATGTAA AAAGCATTAG CTCCTGCACC 1560

AAgGCTTGTC CAACTGTAAA AAATTTGTTG GCTTTTCTTC TAGGCACAGC CGAGAAAGCT 1620

TTCACAACAA GTTTGGCAAT TAAATAACCG CCGCCAAAGC CCATAACGGT AGAAACAACT 1680

AGACCAACAA GCACTTTGAT CCATTCCGAG CCATTGACTG CACCCAAACC ACCTAATGCC 1740

ATTGCCGAAC CAGTCAAGCC GGCAATTAAC GCATGGCTTT CACTTGTGGG AATACCAAAA 1800

TACCACGCAG CAACTGCCCA AACCACAATT GAAAAAAGCG AAGCAGCCAA AGCCACCTGT 1860

GAAGCGTTTC CATTTGCGTC AAAACTAACA ATGTTACTAA TCGTTTCAGC TACTtGGGCA 1920

TTAAAAAAGG TCATGACTAA TGCACCTAAA AAGTTCATCA CAACAGCCAT CCAAATTGCG 1980

ACATTCGGTT TTAACACCCG AGTCGAGACA GCGGTTGCAA TTGCGTTCGG CGCATCCGTC 2040

CAACCGTTAA CAAAAATTAC ACCCATTACC AAAGTAATCG TTACCAATAA TGCAATACTC 2100

ACAAAGACA 2109

(2) INFORMATION FOR SEQ ID NO: 489:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7712 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 489:

GTAAATTCAC TTCAATTGCC CCTTCCTTCT CCTGTTCTAT TTTACGTTAT CTTGCTGGCA 60

GTGTCACGAA AAGAAGAAAA GAATCTGGAA TAATTGCCGT TTGACAGTTT ATCCAGATTC 120

TTCTTACACC ATTTCTAACG AtACGTAATC CCTAATGTAA CATCCTTACG AATTAAAACT 180

GGCGTTTTTT CCAGTGTTAA ATCGGCGCCA TCGACTTGAA TGGCGACGTT ATTTGAGGCG 240

CCTAGAATCA GTGTTGCATT GGCGGCATTC GCAGGTAAAG CTGTTGTTTG GGTCTCTCCT 300

GCTTGCAGGG TGTGTTGGTA CGTATAGTTG CCATCGACCA TCACGCCAAT CCAGCAAGAG 360

CCTTCAGTCC CTTTAAAGGT AATGTTGATT GGACTATTCG CATCTTTAAC CGTTGCAGTC 420

ACTTGATTGT TAGCATCACT CACCACAGCC ACAGCTGTTT TCTTCTCTGG TTCGCTTGAA 480

GAAGACGGCG GTGTACTACT TTGTGTGCTA CTTTCTGGGG CTTGACTTGA TTCTTTGGGC 540

TTTTCCACAC TGGCACTAGT TGCCGGTGTT TGAATAATCG GTTCGGCTTT TTGATCTTGC 600

CACATAATGT ACAGCACCAC GACCGCAATG GCTAAACCAA TTAACGAAAA GGCAATGGCC 660

GGTAAGTTCC GCAGTAACCA ACTGGATTGC TTTTCTTCGT CATGCAATTG CTTACGTGAT 720

CCTTCAAGCT CTTCATACAC TGGCTTGGGT TCTTCGACAA TTGCTGGTTC TTTGCCATCA 780

TAGATATCTA CCAATTGATT GCCATCTAAG CCGACCGCAC TTGCATATTG GCGGATAAAC 840

GCCCGCACAT AAAAGGTGCC TGGTAAGGAC TCAAAATCGT TTTGTTCAAT AATTTCTAAA 900

TATCTTTTTT GAATTTTTGT CATTTGTTGC AGTTCATCAA TTGAGATGTT CTTCTGCAAT 960

CTTGCTTCTT TTAGTGTTTC TCCAATATTT ACGCTGGCCA CTCGTCCATC TTCTCCAGTC 1020

TATTTTGTTA ATTTATCGAT TGTCTCATCA GTCTTTTTCA ATTGCCAAGG AAAACATTCC 1080

CGTCCTTTAA GAATACCACA AAAGAAGCCA ACTGAGGATA ACAAAATTGA AAGATTTCAA 1140

AGATTTTTCT TTTTCTTAAA CAGTkGGTTT ATTGCTCGGT TTCTCCTTTA GGATACATGT 1200

AGAAACGGCT CAATCCTTCT TCACGAATAA AGGTCTCTGC TACTTGGCGG ACGTCGGCTA 1260

ATTGAATACT TTCAATGACT TCAATCTTAT CAAACAGCGT GGTTTCCCCA AACAATGATT 1320

GTGAAAATTG GTTGGCGATA TACTCTAAAG AGTTTAAGGA TTGGAAATAT TTACCAATCA 1380

TTTTTTTCTT CAATAAGGAC AAATTTTCTT CCGTTAATTC TGGACTTGTT GTGGCAGATA 1440

ATAAAATAGC TTCAATTCGT TCTGCTAAGA CTTCTGGTTG CTCACTATCG CCGCCAATAT 1500

CCGCAAAATG GAAACTGCGA TCTAAATTAA AGTCATAACT AAAACTATCA TCTAACAAGC 1560

CTTCATCATA AAGAGCCAAA TAGTTTTGAG AGGTATTACC AAACAATAAT TGTAACAACA 1620

AGTTCACGGT GACTTGATAT TTCATTAAGT CACGACCCGT TGTTGGCACT TCATCTAAGC 1680

CTTTCAAACC AACAAGCACT TTTGCCCGTG TGACTGTCAT TTCAATGGCG CTCTCTTTGA 1740

CAATGTCAGC AGCTGTTTCT TCAGGAAAAT GCCGACGAAT AGGTTGGGCG GCAGGAAACT 1800

CTTTCGCCGC TTGGTTTTCA CGAATGAAGG CCATTAACGC TTCAGGGTCC ATTTTCCCTA 1860

CCACGAATAA GGTCATATTA CTTGGATGAT AAAACGTATT GTAACATGTA TACAAATCTT 1920

CTGCGGTAAT CTCCGCAATA CTTGCCACTG TGCCGGCAAT ATCAATATGC AATGGATGCT 1980

TCGGATAAAG ATTGCCTAAA ATCCCAAAGA ACAAGCGCCA GTTAGGATCA TCTAAATACA 2040

TTTGAATTTC TTGTCCGATG ATGCCTTTTT CTTTTTCGAC CGTTTCCTTC GTGAAATACG 2100

GCGATTGCAC AAAATCAAGC AACGTTTCTA GATTTTGTGT CACTTGATCC GTCGTTGAAA 2160

ACAAATAACT CGTTTTCGTA AAACTGGTAA AGGCATTCGC TGAAGCCCCT TGTTGCCCAA 2220

ATTTTTGGAA GACATCGCCA TCTTCTTTTT CAAACATTTT ATGTTCTAAA AAGTGGGCGA 2280

TTCCGTCAGG CACTTCAATC ATTTCCTCTT GACCAATTGG CACGAACGTA TTGTCAATGG 2340

AACCATAATC GGTCGTAAAC AAGCCATACG TTTTTTGGAA ATCATTTTTC GGTAATAGGT 2400

AAACCTTTAA ACCATTTTCC AGTGTTTCTG TATACAATGT TTCTTTAATT TGTGCATAGT 2460

CTTTTTTAAT CATTTTCCGT TTCTCCTTCC AAAAAGAAGA TCGCTTGAAG TTCCAAGCGT 2520

TTGGCTACTT CTTGGATTTC TGGAATAGTC ACGGCATTGA TCCGCGCAAT CCATTCCTCC 2580

GCCGTCAGCA TCGTCTGCGG CATTAATTCG TTCAAGTATT CTTTTTCCAG CCACGCGCCT 2640

GCATTATCTA ACGCTAAGAT ATACTGATTT TTTAACATTG CTTTCGTTTG TTCAATTTCT 2700

AATTCCCTAA TCTTACCTAA GCGAATATTT TCCAATTCTG TCGAAATCAA ACGCAAGACC 2760

TGATTCCGAT TTTTGCCATC GATCCCTGTT TGAACCGTCA TGAAGCCaCG GAAGGTGTCG 2820

ATACTACTAG AAGCATAATA CGCTAAATGC TCTTTTTCAC GGACATTCAT AAATAATTTA 2880

GAATGCGGGA AACCGCCAAA AATCCCATTA AACACTTGCA ACGCAAAATA ATAGCTATCG 2940

CCATAATAAA TATCTGTGTT GTAAGCCAAG TTTAATTTCG ATTGGGCTAA CACCTCACGC 3000

TCGGTCCGCT CTTCAATCAC GTTGCGAATT GGTTGATTGT AGAAAATCGC GGCTTTGCCT 3060

TCTTCACGAG GTGTAAACGG TAATTGCTTG AACAATGGCA CTAACTCCGC CTCATTGACA 3120

TCACCTAAAA CAAAAATATC CACTTGATCT TCCGCCAGCA TTTTTTGATA ATACGCTGCT 3180

AATGACGCCG CTGTTTCCTC AGctAACGCC GCAACAGTCC CGAAACTAGG GATTTTTTGA 3240

TCCTCCGATT GATTAAAATA GACACTTTGA AGCGCCAAAG AGGCATAGGT TTGCTTATCC 3300

TCCACAATAC TTTCCAAGTA CGCTTTCAAG TTCTCTTTTT CCCGTTGGAA GGTTTCCGCC 3360

TCAAATTGGC CCGCTTGAAT ATTGGGGGCA AAAATAATTT CCTTCAAAAA GTCCACCGCT 3420

TCGGCGAGAA CCTGACTATC CTGTAAATAA TGATCATTGA CAATGTTCAT AGAAATATTA 3480

AACCAATGTT GATTGCCTTT TTTGGAAACG CCAATCCCAA AGcTGGCGCC GTAAAGcTCC 3540

GCCAAACGCT CACTTAACTT GACTTGATTC GGGTAATTCA AGCTGTTGGT TTCCATCAAA 3600

CTGGACAACA GCGTTCTTTT TGTAATAGTT TCATGATTTA ATCGTGTGTT AAAGCGGACT 3660

AAAAGACGCA CCGTTTTATA TTTTTCTGTA GGAATGACGT GTAAATTCAC GCCTTTAACT 3720

AATTGAACAG ACATTGCCGC ATCGCTCCTT CATTTGCTTC ATCTTACCTA AAGAATGCTG 3780

TGAAATATAT GTATCGCGTA AATTCTAAAG CATTTTTTAA TGTTTTTTCC AATAGCTCCA 3840

TTATAACACA GCTTCAGGTG CTTTCTGAAT GATAAGAGTT TTTCAGTTCA AATTTTCAAG 3900

AAGATTCGGT ATACTGTTTT TAAAGTGAGG AGGGAAAAAA GAATGATTAA AGAATTTAAA 3960

GAGTTTATCA TGCGCGGGAG TGTCTTAGAC CTAGCCGTTG GTGTTGTCAT TGGTAGTGCG 4020

TTTACAGCGA TTGTGACACA AGTGGTGGAA GGCTTGATTA CCCCGTTAAT TAGTCTGATT 4080

TTTGTCTTGA CGACTGGGAA AAAGAGCGCC GACGATGCAT TGGGTGCCCT TGTCTATAAA 4140

GTCGAAGGCG TTGAATTCAA TATTGGTAGC GTGATTAGTG CACTTATTAC ATTTTTAATT 4200

ACGGCCTTCG TCCTCTTCTT GATTGTCAAA GCGGCCAACA AAATGAAAAA CCGCGGCAAA 4260

AAAGAAGAAG CGGCAGAAGA AGAAGTGGTA CCAACGAGCG AAGATTACTT AAAAGAAATT 4320

CGCGACTTAT TAGCGGCGCA AACACCGCCT GCTGAAACTG TCAAAACAGA CAGCACTTTC 4380

ACAGAGAAAT AAGCAAAAAG TCATCCTAAG GGGATGACTT TTTTGTTTGG ACTTACACAG 4440

ATGGTCCTGC GACGAGGTTC CACGTGATAC TGCCTGTGTA GGCCCCTGCT TTAATATTGT 4500

TGGCTGGTGT TGTGATGTTT AAAGCTGTAT TAAACAGTGA AAATTGCCAT AATCCTTGGC 4560

CTTGACCAGC TGCTGCGTTA gcCATTAATT GGGCTTCCCC GCCTAAATCA AGGGTCCAAT 4620

TTTGCGGAgC AGCCAGTTGT GGTCTGACTT CTGGTGTAGA GACTGCTGTC GTCAGAGCTA 4680

CATAACTATT GCTTGTTGCC TCGTATTCTA AATCAAGTTG AGTAATCAGC GCCACTTTAA 4740

TGTTGCCTTG TAAGCTTTCA GTGCCTAATT TTAGGCCATT AGATTGGGCC GTTAAATGCC 4800

AGCCTTCTTT GCCGCCGCGT ACGTCACCGA CACCAATCAT TTCTGTATGG CCGTTGGTGG 4860

CTTCGGGGTC TCTTTGGTTA AGAATTGGTT TCGTTAAATC ATCTGAAATT GATAATTTCC 4920

CAAAATTAAA TTCTCGAGGA ATAAACAGTA AATCAAAGTT ACTGTTTGAG CCACTCGGGT 4980

CTGGATCAAC GGTATTGGCA TCGCCGTTGC CATCTGGCAA CGTGCCGCCT TCAAATGCTG 5040

CTGTGACTGG CGTATTCCCT GTCGTAGTCG CCTGTGCCGA AAGCGGGGCT AAACTTGCGC 5100

CAACCACTGC GCTACCAACT AATAAACTTG CCATCATTTT CTTTTTCATT TGTGTTCTTC 5160

TCCTTTACGA TTGAATGATT ATTAGAGGGC AGACCAATAG CAGTTATCCG CTGATGGTCC 5220

GTCCTCTGTT CTTTTGTTGC TTGTGCTTAA ATTGATGGTC CTGCGACTAG GTTCCAAGTG 5280

ATATTGCCTG TGTAGGCACC TGCTTTGATA TTGTAAGCTG GTGTCGTAAT ATTTAAATAT 5340

GGTCTAAATG AAAATTGCCA TAATCCTTGC CCTTTTCCTG CTGTAGCATT CGCAAGAAGC 5400

ACTGCATCTC CGCCTAAAGC AATTGAAAAA GCTGTTGTAT CAACAGTTTC CGGTTTGGTA 5460

ACTGGGTCGG CAAGATCAGC ACTTTGATAA TCATACGTTT TAGTTTCCTC ATTATAAATT 5520

AATGGATACA AAAACATTGC TCCTGTTGTA ATACTGCCTT CCAAACTTTC CTCGCCGAGC 5580

TTCATACCAG TCGATTGAGC AGTCACATGC CAGCCTTCTT TGCTCCCTCG TATATCTCCT 5640

ACACCAACAG ATTCCACATT TCCAGCACGT CCTTCATCTG TTTTATTAGG AATTGGTTTA 5700

GTTAAATCAT CAGAAATAGA TAATTTCCCA AAATCATATT CTCTTGGAAT AAAAAGTAAA 5760

TCAAAGTTAC TATTTGGCTT CGTCGGGTCT GGATCGACTG TATTTGGAAC GCCATTACCA 5820

TCTGGCAACG TGCCGCCTGC AAATTCTGCT GTGACTGGTG TATTCCCAGT GGTCACCGCT 5880

TGTGCTGAAA GAGGTGCTAA ACTTGCGCCT ACCACTGCGC TACCAACTAA TAAACTTGCC 5940

ATCATTTTCT TTTTCATTTG TGTTCTTCTC CTTTACGATT GAATGATTTA TTAGAGGACA 6000

GAATAGCTGA GGTTCTGCCC TCTGTTTTCG TTGCTTGTGC TTAAATTGAT GGTCCTGCGA 6060

CGAGGTTCCA TGTGATGTTG CCTGTGTAGG CACCTGCTTT GATTTGTTGG TATGGAGTAG 6120

TAACTTCAAG ATTCATGGTA CGAATACCTG AATACCATAT GCCTTGTCCT TTTCCTGAAC 6180

TTGCTGTCGC AATTGCAATA GCATCACCGC TTAACGGAAT ATTCGTTCCC GAAAAATCAG 6240

GTGTAGTTGG ATCAGCTTCA TAATTAGCTA AACTCTCAGC ATCAGTTAAA TAAACATATA 6300

TTTGTTTTGT TGCATCAAAT TTTGAATAAT ATCCGATTTG ACTAAAATTA AGGTTTCCAG 6360

CTAAGGCTTC ACTCCCATTA GCCAAAGTTG CTGAAGCTGT GACATGCCAC CCTTCTTTCG 6420

TTCCTCTTAT ATCTCCGATT GCATAGCCAA GATATTGTTG TGTAGAACTA ATATTTTGTA 6480

GACTATCGCT AATTTTTGTG GTGTTAAATG TAACTGCTTT TGGTATCTTT ACTAAGTCGA 6540

AATTAGTGTT AGGTACTGTC GGGTCTGGTG CTGTGTTATC TCCAGCATTT CCTGTCTCGT 6600

TATCTAATGT TCCTCCACCA AACTCCACCT GCACTGGCGT ATTCCCTGTC GTTACAGCCT 6660

GTGCCGAAAG TGGTGCTAAA CTAGCGGCAA CGACTGCGCT ACCAACTAAT AAACTTGCCA 6720

TCATTTTCTT TTTCATCATA ATATGTACGT CTCCTTTAAC TGTTCTGAAT TTCCAGAGCG 6780

CGCGGTTCTT TTTTCTGCGC TTCCGAAGCT ATAGCCTTAT TATACAAGAA TCTGCCCTCT 6840

TCGTAATTTT AAATATTTAA CAGCCATTCA TGAAAATTGA TGATTCCTTT AAAACCAAGC 6900

ATTTCTCTTG ATTTCAAGCT TTAATTTCAG CACTTTTTGA TATTCACAAA TGTTATTTAA 6960

TGATTTTTTC GTCAAATTCG TTTTCTCCTA AAATCTTATA CATTTACCGG CGGTTCTTGC 7020

GAAAAAAACCG TCAATTTTCA TTTAATGACA GCCGTTTTTG TATTTGTTCG TCACTAAGAC 7080

TTACAATTTA CGGCGACGGG CGAATTCCAC AAACTTAAAT TTGTTTAAGC GGTGACGGGA 7140

TTCCGTATAG CCAAAACAAC GGGTATCTTC CAAGTTCATG ATACTACGAA CCACGACGAC 7200

ATGATTGCCT TCGACTTCAC CAATCAGCTG CTTGTCCATC TCTGTGAGCG GTTCCACGGT 7260

AATTTCTTTC TGGGCATAAC TAATCGCCAA GCCAAGGTCG TTTTCCAGAT AATCGAAAAT 7320

CGAATCTTCG GCACGGTGCT TCGGCAGTTC TGGGACAATC TCTTTAATCA AGTAATCAAT 7380

ATCTAAAATA TCGACTTCCC CTTCAATCTC CCGTTGCCGA ACTAAGCGCC ACACAGGGGC 7440

CCCTTCCGTC CAACCAGTCA ACTGAGCCAA GTTGCGCTCC ACAGGAATTT CTTCTAAGGA 7500

AACCACATGA GTATGACTTT CAATATGTTG CGTTTGTTGT AACTCCTTGT AACTGGTCAC 7560

CCCTGAAATA GGAAAATTAA ACTTTTTAAA GTTGAGCACA ATCGAGCCTT TGCCTTGTTT 7620

TTTCTGAATG TAGCCAGCGT TAATCAGTAA ATTCAAGGCT TTCCGAATTG TTtCaCGGGA 7680

AACATTGTAA GTTTCCACTA ATTTGATTTT CA 7712

(2) INFORMATION FOR SEQ ID NO: 490:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 569 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 490:

CGAACTTTGT GTTTTTCGCC AGCCATGTCA AAATAACTGA CGGTAATTTT GTTATGGCGA 60

ACGACATACC CAGGTACATC CCGATCGACA GATAAACAGC CTTCCCCTTC ACCTAAACAT 120

ACATCTTGCA CAGAATGGCT CAAAATTTTA GGGTTGTACA TCACCGTACT TAATGAAGGC 180

GTTTCGTTTT CTGGGTCGTT ACTAGGAACA TGGACAGCAA TAATTCGTTT AGAAATATCT 240

AATTGTGGTG CAGCTAAACC AACGCCACCA CGTAATTGTA ATTCTTCGGC TTTGACTGGA 300

TCTTGACTAT TTTTTAAGAA AGTCAACATG TCCTCACCTA ATTGTCGGTC TTCCTCAGTG 360

ATAGGGACAG GCACTTCTTC TGCCACAGCA CGAAtGTGGG GATTCCCTTC ACGAATAATA 420

TCTTTCATTG TAATCATTCA ATGGTTCCTC CTTTGAAACA TAAAAAAATT CTTCTTTTAT 480

TTATTCaTGC GCTGrAATAA TCAGTTGtTT ACCACCyTTT AGtGkAnCmC aAAAyCmCCA 540

AAAAAGGGGs tAgAwTGgTT TAAAAAACn 569

(2) INFORMATION FOR SEQ ID NO: 491:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1125 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 491:

GnTTTACAAC TTATAATCGT ATCTGGCGTT TACCCACGTC TCTGCATCGC CATTAAGTCT 60

GGAATAAATC GCCTGCATTT TGCGAACGAA AAAGTTGCAG TAAAAAGAGA AATTCATCAT 120

CTTTCAACCA ATTTTTTGGT AATGATCAAA TAGTAAGTTT GAGATTGTTG TTTCACCAGC 180

TTCTAAATAT TCACGAATTG ATAACATTGT TGTTGTCCAC TTCCCCTTGC TTGTTTCaCT 240

TATTGCACTT GATAAACATC AAAATAAATA GCTTTCATAT AATCAATTAA ATAACTAGGA 300

TTCAATTCTT CTCCCgTTGC ATCATAAATT AATTGATTTG GTTTACGTGA AGCGCCGTAC 360

TGATGAATAT GTTGTGTTAA CCaTTTACGG ATATCAGAAT AATCCTCCGA AGCTAAAATT 420

TCATCAACAG ATAATTCTTG TTTCATTGCG TGAAATAGTT GCGCGGCGTA CATATAACCC 480

AAGGCATAAG ATGGGAAATA ACCAAAGCTA CCTCCAGACC AATGAACATC TTGAAGGGCA 540

CCTTCTAAAT CATTTTCAGG TGATACGCCT AAGTACTCTT GGTATTTTTC GTTCCATACT 600

TTTGGCAAAT CAGCGACTTC CAAAGAACCA TTAAACAACA TTTTTTCAAT TTCATAACGA 660

ATAATGATAT GCAAAGGATA GGTGAGACTA TCGGAATCAA TCCGTATCAA ACTTGCTTTC 720

GTTTCTTTTA AAGAAGCATA GAAATCTTCA AAAGCAATAT CATCAAAAGT GCCTTCTGCG 780

CATTCTTGGA AAAACGGGTA TTGTTTTTGC CAAAAAGCTC GGTTGCTACC AATGATAATT 840

TCATTAAAGA GCGATTGTGA TTCGTGAATC CCCATAGAGG CGCCTTCGCT TAGTGGTGTA 900

AAGTCAAATT TAGGATCAAA ATTTTGTTCA TACATCCCAT GACCCGCTTC ATGAATCACA 960

CCAAATGTTG CCATTGAGAA ATTATTTTCT TCCCAGCGnG TAGTAATCCG CGCATCATTT 1020

CGATTTAGAG CTGTCATAAA TGGATGAACT GTATCATCGA ATCGGnCTTT TGAAAAAnCA 1080

TAGCCTAATT GTTCAACCAC ACCAATCACA AAACGGnGGT GGTGG 1125

(2) INFORMATION FOR SEQ ID NO: 492:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 651 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 492:

GAAATtAcaA CAACATTGTC TAGTGCTGCT CCAATTCCTC CTAaCATTAA GCTGGCAAGA 60

AAaACAGTTC GATAAGGTCT AGTTAAAAAT TGCATTTCTT CATAGCGCAG TCCTTGATCG 120

TTTAATAATT TCATAGTAAT TGTGGTAATA AAATAaGrTa GAAAaaCACT AaATATTGTT 180

GCTAAAACCT TCCTTAAATC AATTTTTTTT AGTCCGTAGA GAGTACCAGT TGATATAATA 240

ATaGCAATTA TTGTATAGAT TGATATTAGC AATAGTAAGT TTATGGATCT GTTTCGCTGA 300

TTTATCCAAA GCAAGAAAAA AAGAAAGATT ATATTGAATA AAATACCAAC TAAAGAAAAA 360

AATCCTTTTT TTCCAGCGAT ACCTAAGACT AATGAAACAA AACTAACTAC TAAAATAAAA 420

GTATAGCCAT CTCTTTTTAG AGAAATAAAT TCCCACTTTT TAGTGGAATT ATCTAATAAT 480

ATTTGATTAT TTTTATTAAA TACTGGaTTG GTACTTTCAT TTCCATTAAA TTTTACTTTT 540

ACAACAACAG TTTTTTTTGt ATTATTTACT AGTTTAACTG TTAGTAATTG TTCGCTGkCa 600

ATTATTTTTA CtGaTTTTAC aATACCmACC nGGGGAnGTA TAAAAGGTAA C 651

(2) INFORMATION FOR SEQ ID NO: 493:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3501 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 493:

GATGGTTCAA CTAATTTCTA TTATTTTTGC GTTTTTGATT ATACACACTT TATTTAAAAC 60

AAGAGAAACA AGTAATtCTT CAAGAAAATT AAGAATGTTA GTGTACACTT TTTTGGGGAT 120

AATTTTTCTA TTAGGCTTAG TAAGTGTCGT ATTAGGAACC TATGGGCTGT TGAATTCTTA 180

TAACATTAGT GGAAATTGGC GTATAAATAT GTCCTCTATT TTAGTTGTGT TGGCTGGAAT 240

TTTACTAGTT GTACTTTCTT ATGTTGGTAT TTCAAAATTT AAAATGAATT GTTAGGTGAA 300

TTAAATGAAG AAAGTTATCT CTACCATTTT ACTAGTAATC CTTTTTGTGG GATCAGGTTA 360

TGTGGTAAAC GCTGCTCCTG AAATGAAGAG TTTGACTAAT TTCGAAACGA ATGATCAACA 420

TTATGATTTA CAAGGGATAG AACATTCCTT TTGCTAAAGT GTTAAACTTT AAAACAAGTA 480

GAAAAGATAT TACAGCAAAA ACGATATTAA AATTTGTATA TCTAAAAAAG ACAGACAATG 540

AATTGAATAA TGCTTCTGCA TCTATTAAAG ATTGTATTCG CTATGCCACT TCTATTGGGG 600

TAAATATGGA ATAGCTTGAT GCAATGAAGA ATGAACTTGA TTTCAACCAA GTGAAGAGTG 660

AGATTGATAA GGATCATCCA GTATTACTAT GTTATAAACC AAACAAAACC GATTTAATGG 720

AACCTTATTT ATTAGCTGTA GCGCATGGAT ACGTATACGT TGCTCCAGAT GGTCAAAATA 780

CTAAAGAACC ATTTATGAGT ACAAGTATCT GGATGTCGAA CTATATATAT CCTTCTATCC 840

TGACTACGTT GGCGAATAAA AATAAGTAAC AGTTACATAT TTTTCAGTGA ATAATGCGCC 900

ACAAAAAGAT TATACATTTA TTGGTATGAT TATTTTTAAG TAAGAAAGGT TGTAAGGAAA 960

TTGAAAAAAT ACATTTTGTG TTTGAACACA TTTGATTGAA GAACATCAAC GCGACTATCT 1020

CTATGGTAAT TAGGGGCACG TACTAAGTGC CTTTTTGAAT GTAGAGTCAT TTATTATGCT 1080

TAAGAAAAAA ACTTACACAA AACTCTATAC AGATCCGATT TCACTGTAAC TGCTTTTAAA 1140

GAAAAGACGA TATAAAGATA GTATAAAGAA ATAAACTTAA GTTTCACACT ATCTTTAGGT 1200

CTTTTTTTTT ATGCTTACAG AGTAAAGAAA ATTAGTTAGT TTATTTAACT GGTTGTTTTG 1260

AAAAGGAGTG AAGTGAAATG ACGATTATTT TAGGAATTAT TGGACTCTTA TTAATGATCT 1320

ATCTTTTCTG GTTTTTATTT AAGGGGGAAG AATTATGAAT AGCTTGATTT ATCAAGGATT 1380

```
ATTTTTCCTA TTCATATTAT TGTTATTGGC AGTGCCGTTA GGcTGGTACA TTAAAGAAAT    1440
CATGCAGGGG GAAATCCCTA AAGGAGTATC CTTCTTACAG CCAGTAGAAC GTTTATTTTA    1500
TAAAGCGATT GGACCTATCA GCAAAAAAGA AATGTCAGCG AAACGCTATG CACTGAATGT    1560
GCTGCTGTTA AGTTTCTTCT CAATCGTACT TTTGATTGCT ATTTTAATGA CACAGCACTT    1620
TTTACCGGGA GGTTCCTCAG TAAAAAATCT AACGCTGCCG TTAGCGATAA ATACGGCCGT    1680
CAGTTATGTT ACGAATACCA ACTGGCAAGC ATATGTAGGT GAAACGACTT TATCAAATAC    1740
ATCGCAAATG TTTGGTTTGA CAGCTCAAAA TTTTGTTTCA GCAGGTGTGG GACTTTCTGT    1800
TTTGGTAGCA TTGCTTCGAG GACTTTCACA AGTCAAAAAG AAAGCTTTAG GAAATTTTTG    1860
GCAAGATTTA ACTCGTAGTT TAGTTTACAT TCTGTTACCG ATATCAATCA TTCTTGCTGT    1920
TTTATTAATT TCACAAGGAA CCGTTCAGTC TTTCCAATCC GGAGTCGCTT ATCAAGGTTT    1980
GGAAGGGAAA AGTCTTTGGT TGCATTTAGG TCCGGTAGCT AGTCAGGTGG CAATCAAACA    2040
ACTTGGTACA AATGGCGGTG GTTTCTTTGG TGCTAACTCA GCATACCCAT TCGAAAACCC    2100
AACACTATTC TCAAACTTTT TAGAGAATAT TGCGATCTTG TTGCTCCCAG CGGCATTGAT    2160
CTTTGCTTTT GGCTTTTGGG TCAAGGATTG GAGACAAGGA CGAACGATAA TGATAGTATC    2220
CTTGTTTTTC TTACTATTAG CTTTCGTCGG CGTTGCTTTT AGCGAGTATT ACGGGCCTGA    2280
ATTTACGCAT GTTCTTGGAT CTACCAATTT GGAAGGAAAA GAAATTCGCT TCGGCATTGG    2340
CTGGTCTAGC TTATGGTCCG TCGGTACGAC CGCCGCTTCG AATGGTTCGG TGAACGCGGT    2400
GTTGGATAGT TTTACACCGC TTGGTGGCGC CATTCCAATG TTCTTAATGC AGCTTGGCGA    2460
GATAATTTTT GGTGGTGTTG GGAGCGGACT GTATGGAATG TTGGCATTTC TATTGCTAGC    2520
AGTTTTTATA GCAGGTCTTT TAGTTGGGCG GACACCTGAA TATTTAGGTA AAAAAATCGA    2580
AGCTTTTGAT ATCAAAATGG CTAGCCTAGT TATCTTAACA CCATTGATGT TGACGCTTTT    2640
CGGGGCAATG GCTCTAGTCC TACATCCAGA GGTAATGTCT TGGTTAACGA ACCAGGGTCC    2700
CCATGCTTTC ACTGAATTGC TTTATGGGGC AACTTCATTA GCCAACAACA ACGGAAGTGC    2760
TTTTGGTGGG TTGCTAGCTA ATACGACATT CTTAAACCTT TTAGCTAGTT TAATGATGGC    2820
TGTTTCAAGA TACTTGCCGA TCATCGCCAT CTTACTTTTA GCGGAAAATA TGGGAAGTAA    2880
AAAGAAAGCA GCTTTAAGTT CAGGAACATT ATCCACGGTT AGTCCAACAT TTGTTTCTAT    2940
GTTGATTATT GTTATTTTGG TAATTGGAGC TTTAAGTTTC TTACCATCAT TGGCACTAGG    3000
TCCGATTGCG GAATTCTTTA CGATGAAATA GATGGAGTGA GTAAAAATGA AAAAAATCTA    3060
TCAATGGGCA GTGGGACAAT CCTTTAAAAA ATTGGATCCA CGCCAACAAG TTAAAAATCC    3120
CGTGATGTTT GTCGTTTATT TAGGTGCGCT CATCACTACA ATATTATGCT TCTACCCAAT    3180
GGGTATACCG TTATGGTTTA ACATTTCAAT AACCATTTTT CTTTGGTTGA CTTTGCTTTT    3240
TGCGAATTTT GCTGAAGCTG TTGCAGAAGG ACGAGGAAAA GCACAGGCCG ATAGTTTGAA    3300
ACAAGCAAAA AAGGAAGTCA TGACTTATAA AATTAACAGT TTAGAAGATA TCAAAGAAGA    3360
```

GAATTTCATT GAATTACAGT CTTCTGATTT AAAGCGAAAC GATTTGGTCT ATGTACGTGC 3420

AGGAGAGCAA ATCCCAGCTG ATGGTGATGT AATCGArGGG GCTGCaTCAG TAGATGaAAg 3480

TGCCATTACT GGcGAATCCA G 3501

(2) INFORMATION FOR SEQ ID NO: 494:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1856 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 494:

ATTCATCAAC GACAGCGTTG ACAATCGCTC CTAAAATAAT AATCGTGGCA GCAAAATTCA 60

ACCATAACAT CAAGATGATA AAACTCCCGA TGATTTGATA ACTTGCGATT CTAGAACTAA 120

AATATTTTAC GTACAAaCCA AAAaTTTGTG ATAATAGCAT CCAACCGACT GTCGAAAAAA 180

TAGCCCCTGG CAAAATAGAA CGTAACGATA ATTTCCGATT AGGAACGACC GCATAAATCA 240

AACACATAAT CACTAATAAA ACAACGGTCG TTAACGGCCA TTTTAACGCT TGGAAAGTAT 300

CGATCACAGA TGTTGAGTAA TGGAAAATTG GTTGCAGTAA TTCAATAATG TACTGACCCA 360

GACCTAAAAT TACTACTACA CCAACAATCG CTACCATAAA TAACAAAATA ACTAAAAATG 420

AGACCACTCT TACGAGAATA AAATTTTTCC TTTGTTCGAC ACCAAAAGCT TTATTCATCG 480

CATTTTGTAG TGCATTAATA CTTTGACTGG CACTCCAAAA TGCTGCTAAA GCGGAAACAG 540

AAAGCAACCC GCCTGAACGT TGAGTCAGTA GTGAACGAAT CGCTGGTTCT AAGTTCTTAT 600

AAACATCTTT CGGAATCGCC TCTGCAATAT AAGGCAACAC AGAATTAGGG TCGATTCTTA 660

AATAAGGTAA GACATTACCT ACCGCAATCA GTAATGGAAA CAGCGAGAGT AATAAATAAT 720

AGGCAACAAC CACTGAGCTG TTGCCTATTT CGGCTGTGAC CATGTGACTT TGCGTTGTCT 780

cAATAAAACG CATCAAAGAT TGATTCTTTT TGATGTTGTC GAGAAACTTC ATCACACACT 840

CTCCTTATTA ATAAGTCAAT TCTTCGCCTT GTGACGTTAA GATTCTTGGA CCTTCTTTCG 900

TAATCGCTAA TGAATGTTCA TATTGACAAC TTAGACCGCC ATCTTCAGTA TAAGCAGTCC 960

AACCATTTGG GTCCATTTTC ATGCGCCAAG TTCCTGTGTT AACCATTGGC TCAATGGTAA 1020

TAACCATTCC CTCTTTTAAA CGTAGTCCTT TACCTGCTTC ACCATAGTGA GGAATCATTG 1080

GACTTTCATG GATAGTTGGG CCAATGCCAT GACCTACAAA GTCGCGAACT ACCCCGTAGC 1140

CTTCTCCTTC TACATACGTT TGAATTGCAT GACCAATGTC GCCAATTCTA TTTCCTACTT 1200

GTGCTTGTTC AATCCCAAGA TATAGCGCTT TTTTGGTTAC TTCCATCAAG CGATCAATTT 1260

CAGGTGTTGA TTCACCCACG ACATAAGACC AGCATGAATC AGAAATTGCG CCCTTTAAGT 1320

CGACACACAT ATCTACTTTG ATTAGGTCGC CATCTTTTAA TACTTTTTTA CGAGGAAATC 1380

CGTGGCAAAT TTCGTCGTTG ATACTACAAC AGGTAGCATA TTTGTAACCT TCATAGCCAA 1440

TTTGAGCCGC CACGCCACCA TGACTTTCAA TAAAGTCTCT CACAAATACT TCAATATCCC 1500

AGcTTGTAAT GCCAGGTTTA ATAAACGTCC GTAAATGACG GTGAACATCC GCTAATAATT 1560

CCCCTGATTC ATCCATCATT TCAATTTCTC GTGGTGATTT TAATGTAATC ATACTATCTC 1620

TCCTAAACTT GTTTCATTCA AATTATTTTA TCACATTTTT TGCAAAAAAG ATTGTTCAAA 1680

CACCAATTTT CAAAAAGTCC CTTAATAAAA GCACTCGCTA GTCTTTTATA AGAAATGCGC 1740

ATCTGCTATA ATAAGTGACA GAATTGTATG AGGGGGAAGA AAGTTTATGA CGTTAGCTAA 1800

AATTGTGTAT GCAAGCATGA CTGGGAATAC AGAAGAAATT GCCGACATTG TTGCAG 1856

(2) INFORMATION FOR SEQ ID NO: 495:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6645 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 495:

ATATGGTTAA GGAATTTGGA AATCGTGTAG TAATGCCTTT TGTAGATCAA GTAAATGTTT 60

ACTTATCTGA TATTGCAACA GATATGGGAT ATGATGAGGA GAGAAAATAT ATGATACATG 120

TAAGTGGTGG ACAAGCTCAG GTAAATATTT CCAACGATAA CAGCACCATA AATGCTAACC 180

AAAATGTTCA AATCAATCAA TCTAAAGTTG ATGATTTAAT TTCTGATTTG AAAAGTAATA 240

TTGAAAAGGA ATTAGTGGAT AATGAACCTA TTAAAGATGT TTTGTTATCT CAATTGGCTC 300

TGATTGAGTC ACAACAGACT GAGAGTGAAC CTCAAAAAAA TGTTTTGAAG ACGGCATTTG 360

AGACAATGCA ATCTTTACTT AAGACTGCCC CATTAGCAGT AACTGCTGTT GAAAGTTGTA 420

ATAGATTGTA TGAACTGATG TCACCTTTAT TTAATTAGGA GCAAAAATTA AAATAGTTTT 480

AGAAGCCAAG AGCAAAAAaT GTtCTTGGCT TTttACTTTG CAaTGAaTGA GGTtATTTAT 540

GACATTTTtA GATTtATTtG CCGGGaTtGG TGGTTTcGTT tAGGTATGGA ACAAGCCGGC 600

CACCACTGTA TCGGCTTTTG TGAAATTGAT GATTTTGCCA GACGGAGTTA TAAGGCAATT 660

TATGATACAA GTGAGGAGGT GGAAATGTAT GACATCACAA GTGTATCAGA TGAGTTTATT 720

CAATCCCTCG GACCAGTGGA TATCCTTTGC GGCGGATTTC CGTGCCAAGC TTTTTCAATT 780

GCGGGAAAAC GGCAAGGATT TTCCGATACT CGAGGTACTT TATTCTTTGA AATCGCTCGG 840

TTCGCCGCTC TTCTCCAACC TAAGTTTTTA TTCCTTGAGA ACGTCCGGGG ATTACTCAAT 900

CACGAAGGAG GGGCTACGTT CGAGACAATC CTCCGAACGT TGGATGGATT GGGGTATGAT 960

GTGGAATGGC AAGTGCTTAA CTCAAAGGCC TACGTTCCCC AAAGCCGCGA GCGGGTTTTC 1020

CTTATCGGAC ATTCTCGAGA CATATGTACC GAACAAGTAT TTCCTATCAT TGAATCGTCT 1080

GCAACATCTG ATCAAAACCT CAGAAACTTG TTAAATACCA ATCCTTCCAA CCGAGGAATG 1140

AGTGAACAAG TCTATGGTTC AGATGAGTTT GCCCCAACCG TAACAGGCGA TGAAGGAATT 1200

AAAATCGCAC TACCAGTAAA TGATGGAATT TCTGTAGTGG GTATGTTGCC GGGAAATTTT 1260

GAACAAGGTA ATCGGATTTA CGAGGTAACT GGAACAGCTC CCACCTTGTC AACGAAGCAA 1320

GGGGGAACGA AGATTATGAT TCGAAAAAAT AAAATAGAAT ATCAAGAGGT TAAGCCAGAG 1380

GGCAGTGTGA ATCTGGCTTT TCCTCATTCC AAAAGTCGCC GAGGTCGTCT TGAAAAGCAA 1440

AGCGTCCATA CGTTACTAAC AGGAGATCAA CAAGCAGTCA TTACAGACCA GTATCAAATT 1500

CGGAAGCTTA CACCAAGAGA ATGTTGGCGT CTTCAGGGTT TTCCCGATTG GACTTTTGAT 1560

CGAGCCTCTC AAGTCAATTC AGATAGTCAG TTGTATAAGC AGGCCGGAAA CTCTGTAACA 1620

GTTCCGGTCA TTTTTGATAT TGCCAAGAGA TTAAAGGAGG TGAAAACGGA TGATTTATGA 1680

TGAATTGATT GGAGAAATCT AATGGGTGAT TGGCAAGATT CACTCAGATC CGGAACTGAA 1740

AGAAGAATTA CGTCAGCTTA ATTTTGATAT TCGCAAAAAT GTTGTGAAAG TCACTGGTGA 1800

TTCGTATGTA ATGGACGAGG GAACAGATAC TCGAATTGAG CTGAATCAAG TCATCAAAGA 1860

ATTAGAACGG ATTGCGTACC TTGCGAAAGA ATCGGATATC CGTCAATACT TGTTTGAGAT 1920

AAAAGCAGAA CTGGAAGTAG GTGGCATTAC TGAAGGGTAG CAGCCGAATA AACTAACTGA 1980

TGGGCGCCGA CCGCGGGGGA TTTGCGAAGC AGAAATCCTC TGTGGTGCTG TGCCTGTTCA 2040

TGGACGGCGT CAGCCGGCCA TGATGTCTCA CCCCCCCGTA TCTAACAGGG GGGTACAAAA 2100

ACACAAAAAA TGAGCAAGTA ATCATTAAGA ATTATTGGTG TAGCAAGGAT ATTAAGTAAT 2160

TTGCATAGGT GTTAACTCAT TAAAAAAAGT TAACACCTTG AATTTGGAGG TGAGAACTTG 2220

GCACAACGGA ATTTAGATTA TCGATTATTA AAAGACCGAC GTAATGAATA TGGTGTTTCA 2280

CAAAATAAAT TAGCTATCGC TTGTGGGCTT AGTCGGACCT ATTTAAATCT GATAGAAAAC 2340

GGTGGCGTAA CTGCCTCTAC AAAAACCATG AGAAAAATTT TTGACCAACT GGAAAGTTTT 2400

AATCCCGATT TACCTTTAAC GTTACTGTTT GACTATGTAA GGATTCGTTT TTCCaCAACG 2460

GATGTGCGGA AGATTATTCA AGAGCTTCTC CATCTGAAAT TTGAGTATAT GCTTCATGAA 2520

GATTACGCCT TTTACTCCTA TCAGGAACAA TATGTCATGG GAGATATTGT AGTGATGTTG 2580

TCTCATGAAG AAGATAAAGG TGTTCTTTTG GAATTAAAGG GTCGTGGTTG CCGGCAGTTT 2640

GAAACTTTTT TACTCGCTCA AAAGCGTAgC TGGTACGACT TTTTCGAAGA TTGTCTAAAA 2700

ACCGGTGGTG TGATGAAACG CTTGGACTTG GCAATCAATG ATCGAGTGGG GCTACTGGAT 2760

ATTTCTGATT TAACGAAAAA ATGTCAGAAG GAAGAATGTA TCTCCTTGTT TCGTACCTTT 2820

AAAAGTTATC GTTCCGGGGA ACTTTTAAAA GCTGATGAAA AGGATGGCAT GGGAAATACC 2880

TTATATATCG GCAGTCTCAA GAGTGAGGTT TATTTCTGCT TGTACGAAAA AGATTATGAG 2940

CAATATATCA AGTTAGGGAT TCCATTAGAC CAGACTGAGA CGAAAAATCG TTTTGAGATT 3000

CGACTGAAAA ATGACCGTGC TTATCATGCG ATTCAAGATT TGTTAAAAGG TCGTAGCATT 3060

GAGAGCACCA CGTTTTCTAT TATCAATCGT TACTTGCGAT TTGCGGATAA GGTGGAAGGC 3120

AAACGGCGAA CCAACTGGCC CTTGAATGAA CAATGGGGTC GCTTTATTGG ACGGAATCGG 3180

AAGGAAATTC AATTAACTTC TGAACCGAAA CCTTATACCA TTGAACGAAC CTTGAATTGG 3240

CTTGGACGTC AAGTCGCCCC TACGTGGAAG ATGGCAAAGG AATTGGATCG GCTGAATCAG 3300

ACAACGTATA TACAGGATAT GGTACGGAAT GCACGGTTAT CAGACCAGCA TAAAAAAATT 3360

TTGGAACAAC AAAGTATGGC AATCGAAAAT TTGATTGTAT GAAAAGAGGA ATTTTGATGG 3420

GAATTGTAAA AGATATTTTA TTGGTATTTG GTGGCAGTAT GCTTGGTGTT ACTGTGATGT 3480

GCTTGATGCA TGCAAGTACA GCAGCAGACC GAGTAATGGA AAAGAAAGGA AAGTGACGAG 3540

ATGAATTTTG GACAAAATTT ATATCAATGG TTTTTGACCA ATGCTCAATC CTTAGTCTTA 3600

CTAGCGATTG TAGTCATTGG TTTATTTTTA GGATTTAAGC GGGAGTTTTC TAAATTAATT 3660

GGCTTTTTGG TGATTGCCTT GATTGCGGTA GGTTTGGTCT TTAATGCTTC TGGTGTGAAG 3720

GATGTCTTAC TCAATTTATT TAACCGAATT GTTGGTGCAT AAGGGGGAAT CACGATGGAG 3780

CAGATGCGTG TTTATATTGC GAACCTTGGA AAATACAATG AAGGAGAGTT GGTGGGGGCG 3840

TGGTTTACTC CACCTGTCGA TTTTGATGAA GTCAAAGAAC GAATTGGGTT AAATGATGAT 3900

TATGAAGAAT ATGCGATTCA TGATTATGAA CTACCCTTTG AGATTGATGA GTATACATCA 3960

ATTGAGGAAA TTAATCGATT GTGTGGCTTA GCGGAGGAAT TAGAAGGAAC GCCAATTGGT 4020

GAGGTTGCTT CAGAGATTCA GCACGCTTTC TTTAATTCTT TCGAGGAAAT GGTGGAGCAT 4080

GTGGATGATA TTGTGTATTA TCCAGATTGT AATGATATGG AAGATTTAGC TTATCAAATG 4140

GTAAATGAGG GGTATTTAGG AGATGCTCCA GAAAATTTTG TAAGGTATTT CAATTATTCC 4200

TCTTTTGCTC GCGACTTGGA GATTGAGGGA AACTATTTAG TTACTAATCG AGGTATTTTT 4260

GAATATCCAA TCTAAATATA AGCCATTCAA CTTGAATGGC TTATATCTCA ATTTATTAGT 4320

GGAAGTTTAT CAGTGTTAGA AAGCGCCCAA ACAATTGCAT TACCTGCTAA AGTAACAATT 4380

TTAGGATGTT TTTCGAGAAA ATCAGAAAAT CTTTCTAAAA CGTTAGGTTT AAAATCTTTA 4440

GTTTCTACGA TATTGATAAG CTCTTTGCCA ACTTGTTGCT CTTCAGGTGG TAAAGCATCA 4500

ATATATTCTT TTAATTCATC AAGAGAATTA TTCCAATTGT TTGTAACGGT ATTATTGTTT 4560

CCGAATGTCG CTCCTCTATA GTCACCTGAA GAGTTAAAAT TAAAGGTATT TGTCGTCTGA 4620

ATCGATTGAC TATTTTTTTC TGAAGGATGA ATTTTCTCTG CCCCCTCTTC TGTTAGTTCG 4680

ACATCATCTG ACGAAAAGAC GAATAGACGT AGAGGAAAAT GCCGAGAAAA ACAGGGGAAA 4740

TGCCCGATTA TCTGTCAAAA CTCCGACAAA AAATGCAGAA AAGATGGCCT ACCTCTGGCA 4800

CACACCACAA AAAAACAGAA TATCGAAACC ACAAGCCGTG CAGTAACTGA CAAACAATCA 4860

GCTACTGCAC GGCTTTTTTT GTTATAGATC GCTTTGCCCC GACGACACAG ATCGCCGGGG 4920

CTTTTTTCAT TTAGACAGGA GGCTTTGAAA ATGCCGGAAA CATTGAATCT GAATTACTAC 4980

TACGGGAATG AAGCCGACCA GTACAGCTTT TACCGTATCC CCAAAACCCT GCTGACCGAC 5040

CGCCGCTATA AGGGCGTGTC ACTGGAAGCA AAGGTACTGT ATGGCTTACT GCTTGACCGC 5100

ATGGGGCTGT CGGCCCGGAA CGGCTGGCTG GACAATAATG GGCGGGTATT CCTCTACTTC 5160

ACGCAGGAGG AAGCCATGAC CATGCTGGGC TGCGGCAAAG ACAAGGTGAC AAAACTGTTC 5220

CGCGAATTAG AGGGGATTGG CCTGATGGAA CGGAAAAAAC AGGGGCAGGG ACATCCCGCC 5280

CGGATATACG TTAAAAATTT TATCTTGGAA CCGGAACCCA CCGCGCCGGT TCAGACTGCG 5340

GAAAATCAGC AGTCAAGACC GCTTCCCAGT GCCGCCGTCA AGACTGCTGA AAAACCGCAG 5400

TCTGCACCAC GGAAAACCAG CGGTCAAGAG TGCGGAATTT CCGCCCCTAA TAATACTGAT 5460

ATAAATAAAA CTGAAAAAAG CGATACTGAT CTATCCATCC TACCCCCTAC CCCCTCGGCA 5520

GTTTCTTCTC CTGCTGGTAA AnrCGCTCCG CGCAGGATGG GATTGGATGA GATGGAAAGC 5580

TACCGTGAGC TGATCATGGA GAATATTGAC TACGACATTC TCTTAGAACG TTATCCCTAC 5640

GAAAAGGACT TGCTGGACGG CTATGTGGAG TTAATGCTGG AGGTCTGCTG CTCCACACGG 5700

GAATCTGTCC GTATCTGCGG GCAGGAAGTA CCCACCGAAG TTATCAAAAG CCGGTTCCTG 5760

AAGCTGAACA GCGAACATAT CGGCTATGTC ATGGACAGCC TGAAAAGCAA TACTGCCAAA 5820

ATCGGCAACA TCAAGGCATA TACGCTGGCG GCACTTTACA ATGCCCCTGT CACGATGGGG 5880

CAATATTACA CATCCCTTGT CAGCCACGAT ATGGCGCATG GGCTGCTGAG TGGTTAAACC 5940

ACCCAGGGCT TTGCGATACA TCAATCAAAC CAAAGGAGGA CATTTGCATG GAACATGAAA 6000

TCAACATTCT GGTGGTGGAA CCGGGCAGAC CGCCCAGGCC GGCACGGGTG GACTACTCGC 6060

TGGAAACTTT TTCACAGCTT GTCGGCGGGG ATCTGGCAAT GGGCTGTTTC CTTCCCCAGC 6120

GGGTCATGCT GCTTTATAAC GAGGAGGCCA ACCGCCAGGG ACTGCCGCCC AACCGCTGCA 6180

ATCCCTTTGT GAACGAGTGC ATCAGCGGTA CATTCCTGCT GTGCGGGCTT TCAGATGGGG 6240

AGGGCTTCTG CTCCCTTTCT CCTGCCCAGC AAGCCGAGTT TCAGCGCCTT TTCGCCAGCC 6300

CCGGCGAGTT TATGATGCTG GGGGCAAACC ACATCTGCAC TTCGCCTGCC GAGTTTGCCG 6360

AAACTGTCAG CAGGCTTTGG GGCGGTATGG CAAGCGGTGA GTCTGTGGTG CTGACCAAGT 6420

GGGGCGGCAC GGAAAGCGGG GCCTGAACAT GACAAAACAA AATAGCAATA CCGACCGCCG 6480

CAGTTTGCAG GACTGCGGCG TTTTTTCATC GGCAGAATAC CATTCAGTCG GGTTAAGCGG 6540

CGGCAAGGAT TCCTCGGCGC TACTGCTCCT GATGCTAGAA AAAGGGATGC CGATTGACAG 6600

CGTGATATTT GCCGATACCG GCATGGATTT CCCGGAAATG GCGGC 6645

(2) INFORMATION FOR SEQ ID NO: 496:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1075 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 496:

GGAACACAAG TnAGTACACA AGATACTGTT GAGGTTGATG GTGTACCAAT TTATCAAGAA 60

GaATATGGTT ATTaTTTATT CTATAAACCA AGAGGGGTTA TTTCTGCGGT GTCTGACGAT 120

AAAGGCCGTA AAGTAGTAAC GGATTATTTT ACAGATGTTA CTGAGCGAAT TTATCCAGTC 180

GGTCGTCTGG ATTATGATAC TTCTGGGTTG TTACTTTTAA CGAATGATGG AAGTTTTTCA 240

CAAAAATTAA CCCATCCTAA ACATGAAGTA GATAAAGTCT ACGTAGCAAA AATTAAAGGT 300

GTTCCAACTA AACGGGATCT ATTGCCATTA GCCAAAGGCA TTCGCATTGA TGGCAAACGC 360

ACAGCACCAG CCAATTTTCA AATTCTTTCA GCGGACATTA AAACAGGTAG CAGTGTCGTG 420

GAATTAACGA TTCATGAAGG ACGTAATCAT CAAGTGAAGA AAATGTTTCA AGCTGTTGGT 480

TTCCCTGTTC AAAAATTAAA ACGTGAACGT TACGGTGAGT TGACCTTACA AGGTTTGCGA 540

CCAGGGCAAT ATCGTGATTT AACTAAAAAA GAAATCAGTC AATTATTAAA TCAAGCTAAA 600

TAAGATAAAA CGACTCGATC TCTTCGCGCT TATGGCGAAG GGATTGTTTT TTTTGAGAAA 660

CGAAGCAATC AATGAAATTT AATAAAAAAG CTTAGTTAGT TGCATTCATT GTTCAAATCG 720

GTTACACTAA GTAAGTAAAA AAATATAATA CAAGGTTCGT CTTCAGGGGC AGGGTGTAAT 780

TCCCGACCGG TGGTTATAGT CCACGACTCG TTTTTAACGA TTGAATTGGT GTAATTCCAA 840

TACCGACAGT ATAGTCTGGA TAAAGAAGAT AGGGCTTATT TGAGACGCTT TTTCATCAGA 900

TAATCCTACT CTATTTTTCC CTGCAGAAAA ATAGGGTTTT TTTGTATGAC AAAGAAGCGA 960

ATCAAAAAGT TCGTTGAAGa TGGGTCCTTA ATTGGAGGAT TtCAGATGAA CAACAAGGTA 1020

CAAAAAATGG TCAGCATTGC AATGTTGGCC GCAATnGGTA CAGTATTACA ATTTG 1075

(2) INFORMATION FOR SEQ ID NO: 497:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1689 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 497:

TACCACATCC CACCACTTTC TACCACTATA AATCTAAAAA CTTCTATTTT TTAACTGTTA 60

TTTAAAATTT CTGTAATATT TAnAGGGTTT TATGTAACAA GCATCACCCA ACAAAAAATG 120

TTATTATAAC AAGGTTTTTA CTATGTTTTT TATAAAATAA CCCCACAAAT AACAAAAAAA 180

CCCCACGGAG gaAGtTGTGG GGGACATTTT CCCACAGCGT ATTTCCCCTA TTGTCTCCAC 240

GCTTTATTAC AGTAGATAAT GACTGATATT TAGCAAGATA AGCAACACAT ACAACACAAA 300

CGTTAGTAAA AACGTCAACC GCCAAAACAT CTTGAAATAT CGTCTATATA AAATCTCTCC 360

ATAATAATAA GCTTGAAACA GGGCAACAGC GACTCCTAAT AACAAAATAG AAATCACAAA 420

ATACGGCATA ATCGATTCAT CATAACTATC TTTGGATAAC TCATGGATCC CTATGAATAA 480

AAACGGAATG GCCAAATCCG GTGCTTTTAT TTTCCAACGT TCCGTTAGGG AAAAAGTGGA 540

CACGAGaAAA TTGCACGCAA AGAGCACAAT TACCGGAAAA ATATACCAAA ATAACATTAT 600

TGCCGAAAAC GACACCATTC TCTCAACAAA TCCTTTCATT CCACCTGCTA TAACTATACT 660

TCATCAACGT TACACAGATG TTTCATTTCG TTTAAATATA TGTGCTAAAA GCCCTTTTTG 720

CTTGCTTTTT CATATAAAAA ACGGATTACT TATCAAACCA TTTTCAAAAC GAGGCCTCAC 780

CTATCGCGGC CACTTATTTT TGAATGTTTT TTATAAGCTT TTCTTCTCGT ATTGTGTCAT 840

ACTTCTATTA TAATGGATTT AGAACAAAAA ATGTTGATTC GGGGTGAATC TTTTGATTAA 900

TTATTTATTA TTGAAAAATG ACTATTTTGA ACCTTGTTCA CCAGACGATG AAGCCCTAAG 960

TTGGATTTCG GTAGAATCTC CTACAGAAGA AGAAATCGAG CGTCTCGTCA ATCAATACCA 1020

CTTACCAACA GATTATCTTA CCGGCGTTTT AGATGATGAA GAAAATGCTC GAGTGGAAGG 1080

CTTCCGCCAC GAAAAATTAC AAACGCCTAC GTTACTTTTA TTTCGTTATC CCAAAGCTAG 1140

TATCAGTCCT AGTGGCTATC TGCAAGTAGA AACGGTGCCG ATTGCTTTAA TCGCAACCAT 1200

CGATAACAAA CTAATTACTG TTAGCAACGG GCCTAACGAC ATCGTTCATG GGATTCAAAA 1260

AGAGGCGTTC ACCCATCAAG ATCTTTCTAT TGAGAAGGCC TTAATTTTAG CCCTTTCATG 1320

GAAGATGGCC CTTTCCTTCA ATAAAAACTT ACAAGCACTT ATCCAACAAA CCAACAAACT 1380

TGAAGGAGAG CTCCAAGTAG CCACAGAAAA TAGTCAATTG TATCAGATTA TGGACATTCA 1440

AAAAAGTCTA GTTTATTTCG AAGCAGCTTT AACGGATAAT TTAAAAGTAT TAAAACGTTT 1500

ATATAGCGCT GAAATTTTTA ATCATCCAGA AAAACATTTG CCATTTTTAA GGGATATTCT 1560

GATTGAACTA GAmCAAGGAC TGAATACGAC TAAAATACAa CTAAAATTAG TTGATAcATT 1620

AGTAATCCTT TTTCAGCAAC CGTTCCAATA CTTAAATAAT GCATGAAATT TACAnCTTAC 1680

AATGGCCTG 1689

(2) INFORMATION FOR SEQ ID NO: 498:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1360 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 498:

ATATnATTTG ATTAnnAGTT CATCAAATTA CCCTGCATTA CAACAATCTA TATCAGGTAT 60

TTACTCAATT TCAAATAAAT TATTTTTTGG CTCATTAATA TTTGCAGGAG TTGTAGTTTC 120

ATTATTATTA TTCTTATGGA TGAATGCTAG AAAGAAAGAA ATAGCAGTAT TACTATCATT 180

AGGTATATCA AAATTAGAAA TTTTTGGTCA ATTCCTAATT GAGATGGTAT TTATATCAAT 240

CCCCGCATTC GTTGGTTCCT ATTTCTTAGC TCAATATACA GCTGACAAGC TTGGGAATAA 300

TATATTGAAT AAGGTTACTG GTGATATAGC TAAACAAATA GCTAAACAAT CTGCATCTAG 360

CCAATTAGGT GGTGGGGCAG AAGCTGAAGG ATTTAACAAG ACATTATCAA GTTTAGATAT 420

CAATGTATTA CCTAAATTCA TGATTTATGT AGTTATATTT ATGAGTTTAG TACTTCTTGT 480

ATCATTGATT ATTTCTTCAT TCAATATATT AAGAAGGAAT CCArAAGAAT TATTAATTGA 540

CAATAATTAA AATTTTCGGT GCTTTTTAGC ACCAAAAATG AGAGTGTAGT ATAAAGGCTC 600

TTtGTCAAAT GATGTTGGTA AAGAGAAATC TTAGAATAAT TTGGACAGAA CAGArAGAAG 660

AAACCCTCTT CCtGTTCTGk TTATyTGGTT AAAAATGtTT TGCCATGACG AATCATGATC 720

CGTAGTTTGA AATTGTAmAA GCTGCGAAAG CCGTACGCAT TGCGTTTTAA TACTTTGATG 780

TGGTTGTTTA AACATTCCAA TGGTCCATTT GAATAGCTGT AGTGAAGCGC ATTTTGGATC 840

TGTGGCAGAA ACTTTTTAAA CGTTCGTAAA GTGGTTGTAT AGGTTTCAGG AAGACCGCTG 900

ACTTCTCTGT TTAAACAGTC CTCAAATAAT GACTGATCTC TTTTCTTTAC CGCATAAAGA 960

ATGTCTTGGT ATAACTGATA GCCTTGGCGC AGGTTGGGAC AAGCAGATAA CAGTCGATCG 1020

ACAATCTCAC TTTCAGTCAA ATAATCTTTA AAACTGGCAC GCCAAACCCT TTTTTCGTAG 1080

TTTAACTTTG CTTCTTCTTT TTGTAGGAGT TTCCAATAGC GTTTCAATTT TTTCCCTAAC 1140

CCTCGTTGAG AGTGGGAAGT TCCTTTAAGC CGTGCCGTGG TTTCTTTCAC TCGATGATTC 1200

AAGAAGGTTC TGCCAATATG TTGGACGATg TGGAAACGGT CGATGATCAA CTGAGCATTC 1260

GGGAAAAGCT TTTTCACCAA TGAAACATAA GGCGCATACA TATCAATAAC AATCCATTGA 1320

ACGGCTTCAC GAGTCGCTAA TGAAAACCGA GAAAATAGCG 1360

(2) INFORMATION FOR SEQ ID NO: 499:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1656 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 499:

TTTTTTATCG GnCTCGCCAT GCCCGTTTCC AACTTGGACG GGAAGnTTGA AGCTTACTGC 60

GAAAAGGAAG TTGGCTGCTT nCAGGAAAAG TTGGAATTCG CAACCACTTT AAACCAATGA 120

GgGAAtTACm AGkTGrAGCT TCTTCmATGn ATAwTAArGT CTCTGAGtTA aAAGarAAAA 180

TGGCGGATaA TGaAGAATTA TTGGCTCAAA TTTCGACAGA CAAACAGAAA GAGCAACAAC 240

GCTTAGTCAG CGAAAAAGCA GCAGCTGAGG CACAGAAAAA ACGTGCGGCA GAAACTGCCA 300

AAGCAGAAGC AGCGAAGAAA CAAGCAGCTG CCGACAAACA AGCGCAAGCA GATGCGGCGG 360

CTAAACTAGA AGCCGAAGCT GCTGAAGAGT CAAGCAGTCA AGCATCAAGC TCTTCTACAA 420

GTATGGAAAG CAGTAGCACT GAATCTTCAA GTTCAGCTAG TTCATCCGAA TCGTCATCAA 480

CGCCAGAATC ATCGACAGGC GAATCATCAA CACCAGGAAC TGGyTCTGGT CGTGTGCTTC 540

AGATGGAATC AACTGCTTAT TCTTGGCGCG AAGCTGGTGC AAGTAATCTA AGTGCTaCTG 600

GTATTGACTT AAGTAAAGAA AGCAATGTAG TTGCAGTCGA TCCAAGTGTA ATTCCTCTAG 660

GTTCATTAGT GAAAGTCTCT GGTTATGGCT TTGCTATTGC TGGGGATAcA GGTGGCGCTA 720

TTCAAGGCAA TATTATTGAT GTTCATTTTG ATTCAGTAGA CCAATGCCGT CTTTGGGGAC 780

GACGTCAGGT TACAGTAGAA ATTCAATAAT AAATTTGTTA TACTAAAGAT AAGATAAAGG 840

GAATTGGCTA AGTCAGTAAT GATAGTAGCC AGTTCTTTTT TAGTTAAATG TAAAGGAGAA 900

TTACAATGAT TCGCTCAGCA ACAAAAGAAG ACGGTCAAGC GATTGCTCGG CTGGTCTTAG 960

TTATTTTAAA AGATATGGAG TTACCAATTT TAGAAGAAGT CTCAGAAGAA CAAATGATTG 1020

ACCTGTTAGC AGAGGCAACG GCGTATCCAA CCTATCGCTA TGGCTATCAA CGAATTTTAG 1080

TTTATGAACA TGCAGGTGAA GTTGCAGGGA TTGCTGTCGG CTATCCTGCA GAAGATGAAA 1140

AGATTATTGA TGAACCTTTA AGAGAAGTCT TTAAAAAACA TGGCTTAGCA GAAGATGTTC 1200

GGCTATTTAT TGAAGAAGAA ACCTTGCCTA ATGAATGGTA TTTAGACACG ATCTCTGTGG 1260

ATGAACGTTT CCGTGGAATG GGAATTGGTT CTAAGTTATT AGATGCTTTA CCTGAAGTAG 1320

CCAAAGCTAG CGGGAAGCAA GCCTTGGGTC TAAATGTTGA TTTTGACAAT CCAGGGGCGA 1380

GAAAGTTATA CGCAAGCAAA GGCTTTAAAG ATGTCACAAC AATGACTATT AGTGGTCACT 1440

TGTACAATCA TATGCAAAAA GAAGTGTAAA AATAACGAAG CTAAAGTCAA TCACTTTAGC 1500

TTCGTTATTT TTCATTAATC CGTTAAAGGA TCAAAATCTA ATCCAAAATC ATCGGAAGTT 1560

TTTTGTTTTT TCTTATTATA TAACCAGTAG TCAAGGACAG CTGCTATGCA GATGCACAAC 1620

GGTGCGTCCT CATCATTGGC AACAGTTAAA ACAAAA 1656

(2) INFORMATION FOR SEQ ID NO: 500:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3522 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 500:

TTACTTGGAC TGCTTCTGCT TCAGTACTTT CTTTCATTTT TTCAGTTCGT TGATCTGGTT 60

GGCACCCCAT TAAAACAATC ATTGAACAAA ATAACCCGAC AACTATCTTT TTAGTTTTCA 120

TTTTCCCTCC TTTTAAAACT CATTGAATTA CCATTTAATT TTAATATAAG TAGAATGAAA 180

AGTATAGCTA AGCAATCGTT TTTTCTTCAT ACGGCGATAA ACTTAACTGT CCAGATAAAT 240

ACTCACCTAA ATAGATTTCC TGTAATCTTT GAAATCCTTG TTCTTGGATT TGCGTTAATA 300

GCCAGTGCTC ATCTTTTTTG ATTAACTCTA GGACATCCAT GTTGGCTTGA CCGTCCACAA 360

TAATCGGATA ACGAATACTT TCATCGCCAT ATTCGACGAT TGTTAATTGA CCATTTTGTT 420

CTAAAACCGC CCGTTTAACA AGCTGGACTT CATAAATACC ATTTGCTCGT AATTTAAACA 480

TTAGATCATT TGCAGAAATA CCATTTTTCA AACATTCCTT CACCACAACC TGTCCATCCT 540

TGATTAAAGT AATGGGTTTA CCATCCACAA TTTGCTTCAC ATAGCGATTG TGTTCTTTTG 600

CAAATTTTAA AATTAAGACT AATAATGTCC ACAAAATTAA CACTAATACA AATTGGAGCA 660

CCGTAATTGA ATCATTATAA ATAACACCAC CAATAATGCC CCCCAATACA TAGTTTTGCA 720

CTTGGTCCAT AGCCGATGTG GGCGCCAAAT TCCCTTTTCC CATCAAATTA ATTTGGACAA 780

TTAAACAAAT AATACCTAAA GCAAATTTAA TAATAATTGG ACTATAAAAC TGCATTAATT 840

TTCCCTCACT TTACTTTATG ATTTCAATCG TTGGTGTTAA TAAATACGTT TCTTGTAACG 900

TATAAGACTG CTGATCCGTA CTTAAACTCA TTTTGTAAAA CACTTCGTCA ATTTTAACTA 960

```
TGGTTCCGTC AGCCAATTGT GTACCATTGA CATAAACTTT TTCTGGAGAA ACTTTTTTTT     1020

CCTGTGCCAC TTGTTCGACA AAATGTACCA TTTGTGAAGA TTGTGAATGC TGATTTTGGT     1080

TCAGTTGATA ATTGGAATAT TGAACCCCTA AAACAAATAA TAATGATAAA AAAATAATAA     1140

TACTTAATTC ACGATATTTC GTTTGAATTT GATGTCGCAC ATAAAAACTA rACACAATGA     1200

CTAGCCCGAT TAACAAAGcA AAAATAAAAA AATACTTTAA GTAATCATTT AAATTAGaTT     1260

GGCTTTTTAA ATACTCGATG CCATAAAAAT TCATGTTCTT GTCCACTCCT TGATAGGTTC     1320

GTTATAGTAA GTATTCTATC TCAAAAAAAA CAATCACCCC TACTTTAATA CCTTATrACA     1380

AGAAAAAGCT CAGCGCAATC ATCTTGAGAA TGACTGCGCC GAGCTTCAAT CTTCTAAAGA     1440

TTCAGCTACT GATTGATTAC AAATTAATTT TACATCAAAC TTTTCAGTTC AACATCTGGA     1500

TATTTATCCG CAAACCAGcG TTCCGCAAAT TGGTTTTCAA ATAAGAACAA AGGTTGATCG     1560

AAACGGTCCC GTGCCAAAAT GTTTCGGCTA GAGCTCATGC GTTCATCTAA ATCTTCTGGA     1620

TTAATCCAGC GGGCAATTTT AGAACCCATT GGCGTCATTA CCACTTCAGC ATTGTATTCA     1680

TTAGACATTC GGTATTGGAA AACTTCAAAC TGCAATTGAC cTACGGCGCC AATGATATAT     1740

TCTTCCGTTA AATATGTTTT GTATAATTGA ATGGCTCCTT CTTGAACCAG TTGGTAAATT     1800

CCTTTATGGA AAGATTTTTG TTTCATTACG TTTTTGGCAG TCACTTTCAT GAATAATTCA     1860

GGGGTAAATG ACGGCAACTC TTCGTATTGA ACATTCATTT TCCCTTCATA CAAAGTATCG     1920

CCAATTTGGT AATTTCCTGT GTCATACACG CCGATAATAT CGCCTGCAAC AGCTTCTGTG     1980

ACGTTTTCGC GGGCATCCGC CATAAATTGC GTCACGTTAC TTAATTTGAC TTTTTTGCCT     2040

GTTCTACCTA AGGTCACGTC CATGCCTCTT TCAAAAACAC CTGAACAAAT TCGAACAAAA     2100

GCAATTCGAT CACGATGGGC GGGGTTCATA TTCGCTTGAA TTTTAAAGAC AAAGCCTGAA     2160

AATTCTTTTT CATAAGGACT TACTTCGCCA CCCTCTTCTG TTTTATGCGC ATGTGGCGCA     2220

GGCGCAAATT GTAAGAAGGT TTCTAAAAAT GTTTGAACCC CAAAGTTTGT CAAAGCAGAA     2280

CCGAAGAAGA CTGGCGTTTG ATCACCACGA GCGATTTTTT CCTCGCTAAA TTCATCGCCT     2340

GCTTCTACCA ATAACTCAAC ATCTTCCAAC ACTTGTTGAT AAACGCTGTT ATTGTGCAAT     2400

GGTAAATCGC TTGGTATATC GCCATCTTTT AAmGGAATAA AACGCTCCCC GCCATTATTT     2460

TCTGGACGAT ATAATTCCAC GCGTTCATTA TAAATATCGT ATAGACCTTC TAGGCCTTTT     2520

CCCATACCAA TTGGCCAGTT CATTGGATAA GATTCAATAT CCAATAGTTC TTCCAATTCT     2580

TCTAAAAGTT CCAATGGTTC CCGTCCGTCA CGGTCCAATT TGTTAATAAA AGTAAAAATT     2640

GGAATACCAC GTTTTTTAAC AACTTGGAAA AGTTTTTTCG TTTGTGCTTC AATCCCTTTG     2700

GCACTGTCAA TAACCATCAC AGCACTATCC ACGGCCATTA AGGTCCGATA CGTATCCTCT     2760

GAGAAATCCT CATGCCCTGG GGTGTCCAAG ATATTAATCC GTTTGTCTTG ATAGTCAAAT     2820

TGCATCACAG AACTCGTTAC CGAAATTCCC CGTTGTTTTT CAATTTCCAT CCAGTCAGAT     2880

TTGGCAAAGT TCCCTGTTTT TTTCCCTTTG ACTGTTCCAG CTTGGCGGAT GGCGCCCCCA     2940
```

AATAATAATA ATTGTTCAGT AATCGTTGTT TTCCCGGCAT CCGGATGGGA AATAATCGCA 3000

AACGTACGGC GACTATCGAC TGCTTCTTTT TGaTTGATCG TCATAAAAAT CTCCTTATCA 3060

CTTnGTGATG TyaTyTATTT ATTTyCTTCA GCtAATTTTg ATTAACCTTT GCTAGTATAT 3120

CATGTCAAGC CTTGAGAATA AAGAGCTGCT GTTGGCTGTT TTTTAAAAGG AAAACCTGCG 3180

ACATTAGTCA AAACGACTGA TGTCACAGGT TATTTTTAAC GTCTTGGTTC GTCATCTTCC 3240

ATTTCTTTaC GGAAAAAACG ACGATTTTCT TCTGGTTCTT CATCGACAGT TTCTTCATCG 3300

TAGAAATGAA CGCGTAAGAC TAGCAAACGT GTGCCTTCCA TTTCTTCAGC AGTTAACTTA 3360

ATATTGCCCA CTTCAAAAGA AGGTTTTTCC CCTTCATCAG GAATGGTTCC TAATGCTGTA 3420

ATTAAGTAAC CCGCCATTGT ATCAACATCG CTCATGTGTA AATCTGTTTC AAATACTTCA 3480

TTAAATTCGT CAATTAACAT GCGTCCTTGA ACCAnnTATT CA 3522

(2) INFORMATION FOR SEQ ID NO: 501:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1693 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 501:

TTGTACTAGn AGATaACTTC CtGATcTAGT GCCtTTTTCA GAAGCAATCT CTTGGGCAAT 60

AATTGAAGAA TAACTGcTGa TGAAGACAAA TAACAAAATT GTTATCCCCA TTGCTAACGC 120

CATTTGAATC CCTGTGTTAT CTTCTCCTGT TTTTATTTTG CCATTTTCAT CAAAACTAAT 180

TTTGGCTTTT TTAAATGTTG CTTGTTGATT CAACATGGCA ACTTGCTCAG CTGTTAGGTT 240

TAAACGATTG GCATTCGCCA TTGTTTGTAA TTGGTTTAAC ATTTGAGGAA TTAGCATCTC 300

GGTTGCATTT CCTAAAGAGC GTTCAGAGTA AAGCTGACCA GAAATCTGAT TATCTGTAGT 360

GTCCAAAAGT AAATAGGCAT CAATTTTCTC GTCTTTTAAT TGTTCTTGGG CCTTTTTCTC 420

AGAAGAAACA GCTTTGAAAG AAAAGTCCTC TGTTTTATTC TTTGTCAACT CTTTAACGAG 480

CGCTTGATTT TCCGAAACCA GACCAATGGT AGTATCTCCT GAAAAACCGc TAGCTAGAGA 540

GCCCGCTAAG TAAACAATTC CTACAAGAAG AAATGGCGCT AAAATCATGA TTAAGAAAGA 600

CAGCGACTTC ACATTTTTCT TATAGACATC AGTCGCAATA ATCCAAAATT TACTCATCTT 660

GCCCACCTAC TTTCATTTTG AAAATTTCTT CCAATGTTGG TGGTTGTTGA TTAAACATTG 720

GAATGTAGCC GTTGGCTGTC GCTTTATCAA ACAACACTTT GCCCACTTCA GCGTTTTCtA 780

ATGTCGCTTC TAAGACGCCa TCTTGtCGCA AGTGTGCTTC TTTCACAmCT GcTGTAGCAG 840

CAACTTCCTC AACAGATAAC GGTGATTCTA AGAATAATTT TGTTCGACCA AATTGTTGTC 900

GAATTTCATG AACCTGACCA TCTAAAACCA TGCGTCCATT TTTTAACATG ACTAAATGGT 960

CACAGATTTT CTCCACATTG TCCATATTAT GACTGGAAAA AATGACACAA GAACCTTGAT 1020

TTTTCAATTC AATAATGCCG TCTTTTAATA ATTCTGCATT AACTGGATCC AAGCCACTGA 1080

ATGGTTCGTC TAAAATAATT AACTTCGGTT CATGAATAAG TGTCGCAATC AATTGAACTT 1140

TTTGTTGATT GCCTTTTGAT AAAGATTTAA CTTTATCTGT TTTTTTGCCT TTAACTTGAA 1200

ATTTTTCCAT CCAAAAGTCA ATTTTAGGTT CAATTTCTTT TTTAGTTTTT CCCCGCAACG 1260

AAGCGAAATA AATTAGTTGG TCTTCAATCG TAATTTTAGG ATATAAACCC CGTTCTTCTG 1320

GTAAATAACC AATAATATTG TACTCTTTTT CATTTAATGG TTGACCATTC CATAAAACCG 1380

ACCCGCTATC TTGTGTTAAA AAATTTAAAA TCAAGCGAAA TGTAGTTGTT TTCCCGGCAC 1440

CATTTTGGCC AATTaATCCT AAAATTTTCC CGTCCGGTAT CGTAAAGGAT ACATGATCAA 1500

CTGCTTTAAA ATCCCCAAAG GTTTTGACTA ACTCTTTTAC CTCTAACACA TAAACGCCCC 1560

CTTCACTTTT TcTAAACCGC TGTTTAGTTA CyCATATCGT ATCTTTCCTG TGTTGGGCTT 1620

GTCAATGACA ATTGTnATGG AAAGATGTTG CACTTTAATC AGACGCTTnA TTTTCCGCCn 1680

AACGAAAAAG GCC 1693

(2) INFORMATION FOR SEQ ID NO: 502:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1495 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 502:

nAAAnCTCGG TGGTGAAAAA ATGACAGACA AAAAGACAAC CGAAAAAAnA GAnACAGAAA 60

CTGCTGCAAC CTTTACAGAA GAACTAACAG CAGTTCTTGA AGAAATAAAA GAAGAACCTG 120

TAAAAGAAAA GAAAAAAGGG GAAATTGTCA CCATTGTGAA TGAAACAGAT ATTCAAATTG 180

GTGAAAGAGA GTACCGAATT GTTAAAAATC ATCGTGATGC CTTTGATCCT GAACGCTTAG 240

GTGAACGGTT TAGTGAAGTA TTAGCTCGTT ATGATTATAT CGTAGGAGAC TGGGGCTATG 300

AACAGTTACG CTTGCGTGGC TTTTTTGATA CTTCTAATCG TCGGGCAGCG CCGGATCAAC 360

GCATTGATAC ACTAGAAGAC TATTTGTATG AATATTGTAA TTTTGGCTGT GCCTATTTTG 420

TAATTGAACG AACAGGGGAA CGTAAAGAAA AAACAACGCA TCGTCGTCGT CGCCATAAAA 480

AAAATAATCG TTCTAACGCG TTTGTCGAGG AAAAGAGTGC GCCCGCTAAT AAAAATAGCA 540

AGCCAGTGAT TCGTACGAGA AAAGAAGAAA ATAAAAAAAC GACTACACCC AAACAGCCTA 600

GTCCCAAAAA ATCTGGCAAA GGCTTTACGA TTCGTCAACG TGAGGAGTAA ACAGAGTGAA 660

AAAAGACTAT CAAGGATATT TAATTGATTT AGATGGAACG ATTTATTTAG GAAAAGAACC 720

AATTCCAGCC GGAAAACGTT TTGTCGAGCG GTTACAGGAA AAAGATTTGC CTTTTTTATT 780

TGTAACGAAT AATACCACTA AATCTCCTGA AACAGTGGCA CAACGTTTAG CAAATGAATT 840

TGATATTCAT GTCCCAGCAT CCCTTGTTTA TACAGCGACT TTAGCGACGA TTGATTACAT 900

GAAAGAAGCC AATCGGGGCA AGAAAGTCTT TGTCATCGGC GAGGCTGGTT TAATTGACTT 960

AATTTTAGAA GCAGGTTTTG AATGGGATGA AACAAATCCT GATTATGTTG TCGTGGGACT 1020

TGATACAGAA TTATCTTACG AAAAAGTGGT TTTAGCTACC TTAGCCaTTC AAAAAGGGGC 1080

GTTGTTTATT GGGACCAATC CAGATAAGAA CATTCCGACA GAAAGAGGCT TATTACCTGG 1140

TGCAGGTTCA GTTGTGACGT TTGTGGAGAC AGCTACTCAA ACAAAACCAG TTTATATTGG 1200

CAAACCAAAA GCGATTATCA TGGAACGAGC CATTGCGCAT TTAGGCGTGG AAAAAGAGCA 1260

AGTGATTATG GTTGGAGATA ATTATGAAAC CGATATTCAA TCAGGGATTC AAAATGGTAT 1320

TGATAGCTTG TTAGTTACAT CAGGTTTTAC CCCTAAGTCA GCTGTTCCAA CTTTGCCAAC 1380

GCCACCAACC TATGTTGTTG ATTCCTTAGA TGAATGGACT TTTTGAGTAT GATGAGGTCG 1440

AAATTGCGTT TAAGGGAAAC ACTTGGACTC ATTTGTTTGT TTTTAACGAT TGTGG 1495

(2) INFORMATION FOR SEQ ID NO: 503:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 648 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 503:

GAGGAGTGGA AATGAGCATT GCATTACCAA AACAAAAACA TTTAAGAGGT CCGTTACTTG 60

ATTACCTGAG AATTGCAGAG GATAATGGCG ATTTGTTTGC ATGGAGAAAG GCGTGCGAAA 120

TTGGGCGAGA AATATTCTCA GGTGATTTTT CAGACAATGC TAAACCATTA ATTGTTATAT 180

ACAAAGATGG ATCTTCAGAA GTATTTAACA CAAGAGCTGA TGTAATTTCA GCGTGTCGAA 240

TCGGAAATGA AACTTTGCGA AAATGTTTGG AGACTGGCAA ACAAGATAGA TTGGGTCGCT 300

GTTATGACTA TGCCATTTTA GAGTAGCAAA TTGTAATAGC TAATCGTATC TTTTGGCTCA 360

AGAGGTTAGT CATTATGAAT TTTAAAATTA TGAGTTTGGA GGGGGAGCAA TGGGAAAAAA 420

TTTACTCAGA GAGAAGAAGC GATTGATACG ACAAAAAATT CTTTTTCTGA CGGGTGAAAA 480

CGAAtCTTGG gATGAAGAAT CCAGAAATCG TGAAAGAAGT CCAGAGGCTC TCTAAGCAAC 540

TAGAGTCaGA TCTTATAGCa GATaAGCgAC CATTGCCCAG TTTAGATCCT GATAAATTAA 600

CAAAAGAGAA GTATCAACAT TTCTTAGATT TAGGTTATCA AGTAGGAG 648

(2) INFORMATION FOR SEQ ID NO: 504:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1243 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 504:

CACTGTTTTC GGATCTACTT TGGCAAGTTC TGTCGTATCC CCTGTCACAA TGTATTGTAA 60

AGAATCGCCT CGTTTCACAT TTAAATCATT GATATTTTTC TGrCCATCAG TGACtTCTTT 120

ATTCGGmACA AGAACAGGAA CCACATTGGC TTTGTATCGA TTGAATTCAA TCGTTGCTTT 180

TTCTGGTTGC TTTGGACGTC CTTTTTCTTG ATAAGGCGTA ATTGATTTTG CTTTTAAATC 240

CGTATTAAAT GCGAACCACA TCGAATTTCC GTCATACCCT GGATTGTTAA CAATTTCATT 300

CCCGAAAGTG AATTTAATTC TATTTGGTGT TGAAATAACA GTTGCGCCAA TATATTCTTT 360

GCCTTGAACT GCTTCCCAGT CACTTATTGA AATCCCATTC GTCCCTACAT CAATATCCGT 420

TTTAGAATAA ACTTTATTTC CATGACGAGC TACTTTTGAT CCATTAATTT CATGAACAGT 480

ATCCGTGTTT CCTACTTTTA CGAACTCATA CGTAATGTTT TCACCACGTG AATTTAATGA 540

AGCACCTGAA AATACAAATG GATTCTCTTT CGTTGGCAAG ATTTCTTTGC CATCTTCTCC 600

AAAGAAGCGA ACGAGCATTT CAATCTCAGC ACCAGCCTTC TTTGTACGAG GGTCAATGTT 660

AATTCCATAG ACAATTGTGG CCACTGGATC ATCATGTAAG AAAGCATTCA ATCGACCTTT 720

TGAATCAGGT GTGGCCTTCA CTGTATATTG GAATTCAGCA CTAGCAATTT TTTTACCCAA 780

ATAACTAGCA TTCAAATTGC CATAGGTAAT AGTTACTGGT TGATTTTTGT GAAGAACAAC 840

CAATGTTCCT CCTGCACCGT TATTATAATA AAAATGTGCT TTTTTCGAGT CTTTAGTTGT 900

TaACCCGCTT GAATCTACAC TAAGCATACT GATAAGTAAT TTTTTATCTT TATGTTTATC 960

AGTTAGTTTT TTAGCGtCTA TATATTCTGC AGCTTTTACA GAAACAACTT TGGAATTTTT 1020

TGTTGCTTCA CGAkCGAAAA TTAGCGTTTT aTTTACTGGT TTTGCkACAT AGCCATTTTC 1080

ATTTTTATGT TTGTTGTATT CCGCGACTTC TTTCTCGTAG CGTTCTTTTT CCGCTTTGTt 1140

TTTTGgCCGC AATTTCTGGC ATTTTCTTTG nGCAATTTTT TCAGTTTTGG GCTTTAATTC 1200

AGCTGCTTGT TGCGCAAACT TTTTCGCTTG nTCTTTTTGT AAT 1243

(2) INFORMATION FOR SEQ ID NO: 505:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1010 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 505:

ACAAACTGGT ATTTAAATCT TGTCTTAGCG AAAAAACAGC TGCTAAACAG GCAATCATAG 60

GTGAGCCACG ATGGAAAAAT TGAAATAAAA GnATACACAG CATGACAGCA ATTGCTGTTT 120

TTAACGTTCG CATGCCTAAA CGAAATCGAC CCACTATCAT AAATATTCGC TCCTAAATAA 180

CTTTTCTCTT TAATGATAAC CGAAATAGAG GTAAAATTAC TAGAGAAAAG TGAAGAATTA 240

TTATTTTTTA AGGGGAGTAT GCATGAAAAT TGGTTTACGA ACAGTCAAGA CAGCCATTAG 300

CGCTGCATTA GCAATGATTA TTGCGGAAAA ACTGGGATTA CTTTATGCCC CTTCGGCAGG 360

GATTATTTCA GTTTTAAGTG TTACTAGCAC GAAAAAAACA TCTGTCATGA CAGGAATTTA 420

TCGTTTGTTA TCCTTAGCTT TAGCAACTAT TTTAGCGTAT ATTTGTTTTA CTTTTTTAGG 480

TTTTACAGCG ATTGCTTTTG GTATTTTTTT ATTATTATTT ATTCCAGCAG CCGTGTATTT 540

TCAATTATCT GATGGGATTG TTGTGAGTTC CGTTTTGGTA ACTCATTATT TAGTCGAAAA 600

AAATCTTTCT TGGGCGATTA TTGGAAATGA GTTTTTATTA ATGAGTATTG GTGTGGGACT 660

AGCGTTATTA GCTAATTCTT ATATGCCTGA TACAGAAAAA CGTTTAAGAG AAGACCAAGA 720

AGTCATTGAA ACAAgTTTCG TAAAATTTTA AGGGAAAtGG CGCTACATTT AAACAACGCT 780

GCAGGAGAGC GAAATTTGGT CATGCATTGT GCAGATTTAA AAACCTTTAT TCGTACAGGT 840

GAGACGTGGG CCAAAAATCA TGGCCGAGAA TCCAATTGCn GGTCTACCCC AATACGkACt 900

ATTTAGAGtA CTTTGCgATG AGAaAAATGC aAAGTAATAT TCtAAAAAmC ATGTTGGAAT 960

TATTAGaAGA TATTmCyGTT GATGCGCAAC AAGTTCGCTA TATTCAACAA 1010

(2) INFORMATION FOR SEQ ID NO: 506:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 789 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 506:

GATAAACTCT TCGATTCTTT CGCCGTTGAT CCATTAAAGA AAGAAGTACC TCTTGAAGAA 60

CTGGCAAAAC AACGACAAGA AGTAAAAAAA GAAGAAGCAA AAAAACAAGC AAAAGAGTAA 120

TTTTGCTTGT TTTTTTCTAT AGAAATTTTT CAATTGCTGT CCAAGCTGCT TCTTTTCCTT 180

CTTTTGTTTC AGAAGAGAAA ATAATAAAGT CATCTTGTTT ATCAAAATTC AAGGCTTTTT 240

TGATGACACT CTCGTGTTTG TTCCATTTGC CCCGCGGAAT TTTGTCTGCC TTAGTTGCTA 300

CAATAATAAC TGGCAAATCA TAGTATTTCA AAAATTCATA CATTTGGATA TCATCTTGGG 360

TCGGCGGATG ACGTAAATCA ATTAAAGAAA CGACTGCACG TAATTGCTCA CGTTGCGTGA 420

TATACGTTTC AATCATTTTT CCCCACTTCG CACGTTCTGT TTTAGAAACT TTGGCATAAC 480

CATAGCCAGG AACATCCACA AAGTGTAAGG CATCCTCAAT TAAATAAAAA TTTAGCGTTT 540

GCGTTTTACC AGGTTTGCCT GAAGTTCTTG CTAAATTTTT ACGATTAATC AAGGTATTAA 600

TAAAaGATGA TTTCCCAACG TTCGAACGTC CAGCTAAAGC AATCTCCGGT AATTGAGATT 660

CAGGaTATTG kTTaGGGGga TACCGcACTA ATCACGAwTT nCyGCATtAt GTACGkTCAt 720

kGTTTTcCyC CGkTCyTCyA tTCAgTCTTT CATGCAAAAG CTTAACAnAA ACAGCTCTnT 780

CGGAGATCC 789

(2) INFORMATION FOR SEQ ID NO: 507:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 643 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 507:

GGGAGTGAAA CGCCCTATTT GGTAATGGAT GAGCGAGGAA TGGCCATGTC GACAACAGAA 60

TTGGCTGAAA ATCTAGCAAC ACTCTATGAA TCGCCAAATA ATGTTTTATT GGTGGCATTA 120

GCGGGAGAAG CGATCATTGG GACCGCTTCT GTGTCTGCTT CTTCTAAAAA GCGGATGGAG 180

CATATTGGTG AAATTGGCAT TAGTATTTTA AAAGAGTACT GGGGCTATGG ACTGGGCAGT 240

ATCCTGATGG AAGAACTTAT TCGCTGGGCG CATGAAAGTC ATGTTATTCG TCGCTTGGAA 300

TTAACGGTAC AGGATCGTAA TCAACGAGCG ATTCATGTCT ATAAAAAATT AGGTTTTGAA 360

ACAGAGGCAA TTATGCCTCG GGGCGCCAAA ACAGACCAAG GTGAATTTTT AGATGTTCAT 420

TTAATGCGTT TGCTGATTGA TTAATTATTG GTTAATAATT TTAATTTCTT AAATAATCGA 480

ACGATAATTG CATGAAAAAA CAAAATTTTC GCAATTATCG TTTGTTTCAT ATTTATTTTT 540

CTGATATTTT TGTTATAATT ATAGAAGAAT TaACGAAAAA ATGAAAATTT ATTTTGGGAA 600

GGAATGTAAG AAATGGCGAg GCAACGAAAA TGAAGTTAAC TAT 643

(2) INFORMATION FOR SEQ ID NO: 508:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6298 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 508:

TCTTGCTCAT CTTCATATAC TCCGTTTCTG ATTTTTTCTT TATTTTCAAC GACAATTGTC 60

GACATATTCT GGTGCGAAAC ATTGCTATTA TACAGGGAAA TGCACGAAGA ATCAAAAAAG 120

GCATACTTAG AAACAAGAAA ACTGTCTGAA TCTGCTCGCT AGAATGAGTG ATTCAGACAG 180

TTTAATCTTT TTACGGCTTT AACTACGTAA GCCACGCCCT GTTTCTACAA GGTACCAACA 240

AATGCCATAT AAAACGACAA TAAATAAGAT TAAAATGGTC ATAGAAATAC CAATTGGCAC 300

ATCTTGAACC CCTAGGAAAC CGTAGCGGAA CCCAGAAATC ATGTAAACAA TCGGATTAAC 360

TTTTGAGACA GCTTGCCAAA ATGGCGGCAA CATGGAAATG GCATAGAAAA CCCCACCCAA 420

ATAGGTTAAA GGCTGTAACA CAAAGGTTGG CACAATCGAT ACATCATCAT AGGATTGCGC 480

AAAAATACCA TTTAACAAAC CAGCTAAAGA AAAAAGAATC GCTGTCATCA ATAAAGTCAT 540

GATGACCATC AGCCAAGAAA CAACGTGCAA TGGCACAAAA AATAGAGAAA TGATGGTGAC 600

TAACGCACCG ACCAAAATAC TACGACCAAC GCCCCCAATC ACAAATCCCC AAATAATCAC 660

ATGAGTTGGA ACTGGCGCCA CGAGAATTTC TTCGATATTT TTCTGGAATT TTTGTGAGAA 720

AAAGGAAGAT GAAACATTGG CGTACGAACT TGTAATGGCT GACATCATAA TTAAACCTGG 780

CACGATAAAT TCCATATAAG AAAAACCGCC CATGTCGCCA ATTCGCCCGC CAATCATTTT 840

CCCGAAAATC ACAAAATATA AGGACGTCGT AATTACTGGC GGTACCAGTG TTTGGACCCA 900

AATCCGTAAA TAACGGTTGG TTTCCTTCAC AGCTAAACTT TTTAAAGCAG TAAAATACAG 960

ACTAAACATG TTGATCCTCC CTTTGGTACG TATCCTCGGT AATCTTCAAG AAGAGCTCTT 1020

CCAAACGGTT GGACTTATTC CGCATTGACA ATACTTTGAT TCCTTGTTGA CTTAGCTGTT 1080

CAAACAAGTG ATTGACTCCT TGGTTTCGTT CCACTTCCAC CGCTAACGTT AACTCGTCCT 1140

CAAATACGCT TTGATAGCCT TCAATAACTG GCGCCTGCGT ATAAGGTGCT AAATCAAAAA 1200

TAAACGTTTC AAATTGTAAT TTTGCTAACA AGTGCTTCAT GCTGGTGTTT TCAATCAATT 1260

CACCAGACTG AATAATCCCA ATGTTGCGAC ACAACATCTC AGCTTCTTCT AAGTAATGGG 1320

TCGTCAAAAT AATCGTGGTT CCCTGCGCAT TTAACTCTTG TAAAAAGGCC CACATTTCAC 1380

GGCGCAATTC AATATCGACA CCTGCAGTTG GTTCATCTAA AATCAATAAC TTCGGCTCAT 1440

GCATCAGCGC GCGAGCAATC ATCAAGCGAC GTTTCATTCC GCCAGAAAGC ATTCGCGCCC 1500

GTTCGTTCCG TTTTTCCCAT AAATTGGACT GTTTTAAATA CTTTTCACTC CGTTTCATGG 1560

CTTCTTTACG AGACACCCCA TAATAGCCCG CTTGATTGAC CACAATCTGC TGGACCGTTT 1620

CAAACGGATT GAAGTTAAAC TCCTGCGGAA CTAACCCAAT CTGTTGcTTC GCCCGCACCA 1680

TCTCCGTATC TAAATCATAC CCAAAAATTT TGACTTTCCC CGAAGTCTTA TTAACTAACG 1740

AGGTCACAAT CCCAATCGTC GTTGATTTTC CTGCGCCATT CGGTCCCAAA AGAGCATAAA 1800

AATCGCCTTC TTCGACTGTT AAATCAATCC CCyGCAACGC TTCAACACCT GAAGCGTACA 1860

CTTTGCGTAA ATCTTGTATC ACAAGAGCAT CAGTCATTCG TAACTCTCCT TAATGTATAA 1920

ATATTCATGA ATATTATTGT AACACTTCTC GATAAAAACG GGGGTTATTT TGATTAAATT 1980

TTACAGTTTT GCTTAGCTAT TTTTCAACAA CAAACTAGCC AACGATTAAC GTGTTTATTT 2040

TGATTCAAAA CTGTTTAAAC AGTTTCAAGA GCGTTAAAAC CCTTTCAAAT CAAGGATTTA 2100

GAGTTCTCTT TTTTTGGCAC AGAACTTGCT TTATATAAAA GTGAGGTGAA AAGAATGAAA 2160

AGAAAATTTT TAATTGCTAG TCACGGAAAC TTAGCGAAAG GCTTCCAAAG TTCCTTGGAT 2220

ATTCTCGCAG ACAAAGGAAA AGAGCTGGCG GTCATCAATG CCTATGTGAC TCCTGAGGAC 2280

TACACGCCAA TTATTCAAAC GTTTCTTCAG TCCCTTGGCG CAGAAGAACA AGCTATTATT 2340

TTAACAGATT TATATGGTGG TAGTGTCAAT CAAAAAATTG TTCAAGAAGT CATGACGACA 2400

AAACCAGACA ATGTTTTTAT CATTTCCAAT GCCAACTTAG CCATTGCCTT GTCGCTGATT 2460

TTCTTAAAAG AAGGCGAAAA GCTAACCAAA GAAGATATTC AAGCAGCAAT TGCTGAAGCA 2520

CAAATTCAAT TCGTTGAATT AAACCCATCA AATGAAGAAG AAAACTTTTT TTAGGAGGAA 2580

ATTAAAATGA TTACACAAGT ACGAGTAGAC GACAGATTAA TTCACGGCCA AGTAGCCGTA 2640

GTTTGGACCA AAGAATTGAA TGCCCCCTTA TTAGTTGTAG CGAATGATGA AGCAGCCAAA 2700

AACGAAGTAA TGCAAATGAC GTTAAAAATG GCGGTCCCTA ATGGTATGAA ACTATTAATT 2760

CGCTCAGTCG ATGATGCGAT TGATGTCTTC AATGATCCTC GGGGCAAAGA CAAACGAATT 2820

TTTGTAATTG TGAACAGTGT AGCGGACGCT ACTAAAATTG CGAAAAACGT GACAGATATT 2880

GAAAGTGTGA ATGTTGCCAA CGCTGGCCGC TTTGATAAAT CAGATCCAGC CACCAAAACA 2940

ATGGTTTTCC CAAGTGTCCA ATTAAACCCT GAAGAGTTAG AAGCTGCCAA AGAATTAGCC 3000

AGTTTGACTC ACGTGGAAAG CTATAACCAA GTCCTGCCAA CCAACTCAAA ATTAAGTCTA 3060

AAACAAGCCG TCAATTAAAA GGAGGAAATG AACAATGCTT ATGCATGCAA CAATGGCAGC 3120

CTTAGCTGTA TTTATCTGTT TTGCTGGGAA TTATTTAACT GGTCAAAGTA TGATGGAACG 3180

TCCCTTAGTC GTTGGATTAG TCACAGGAAT TTTAATGGGC GACATGAAAA CAGGGGTCTT 3240

AATGGGGGCC TCTTTGGAAG CAATTTTCTT AGGAAATGTT AATATCGGTG GTGTTATCGC 3300

AGCGGAACCT GTAACTGCGA CTACTTTAGC CACAACGTTT GCAATTATTT CAAATGTCGA 3360

ACAAAAAGCC GCCATGACGT TAGCCGTACC AATTGGGATG TTAGCCGCTT TCGTGGTGAT 3420

GTTTTTAAAA AATGTCTTTA TGAATATCTT TGCTCCTTCT CTTGACAAAG CAGCACGTGA 3480

AGGCAACCAA AAAATGGTTG TGACACTGCA TTACGGAACG TGGATTATTT ATTATTTAAT 3540

CATCGCTTCG ATTTCATTTA TCGGAATTyT AGCAGGAAGC GGGCCAGTAA ATCTGTTTGT 3600

TGAAAGTATT CCACAAAATT TGATGAACGG CTTAAGCGCA GCTGGCGGAC TTTTACCTGC 3660

CGTTGGGTTT GCTATGTTAA TGAAACTGTT ATGGGATAAT AAATTAGCCG TCTTTTACAT 3720

TTTAGGCTTT GTTTTAACCG CGTACCTACA ATTACCAGCC GTTGCGGTAG CGGTCATCGG 3780

GACAGTAATT TGTGTCGTGA GTGCCCAACG GGATGTAGAA TTTCGCGACA TTCTAAAAAG 3840

AAAAcTGCGG CGTCTTCTGC GGTTGAAGGC AGCGCGAAAG AAATCGAAGA GGAGGACTTT 3900

TTCGCATGAA GTTAAAAGAG AATTTATCTA AAGAAGAGAA AAAAATGATG CGCTCTGTTT 3960

TCTGGCGTTC TTGGACGATG AACGCTAGCC GGACAGGAGC TACACAATAC CACGCAGTGG 4020

GTGTCATGTA CACTTTACTC CCAGTTATCA ATCGTTTTTA TAAAACCAAG GAAGAACGGG 4080

CAGAGGCATT GGTGCGTCAC ACCACTTGGT TCAATGCAAC CATGCATATC AATAACTTCA 4140

TCATGGGGCT CGTTGCTTCG ATGGAAAAAC AAAACAGTGA AGACGAAAAT TTTGATGCTA 4200

GCTCTATTAC AGCAGTTAAA GCCTCATTAA TGGGGCCACT TTCAGGCATC GGTGATTCCT 4260

TCTTCTGGGG CATCTTACGT GTTATTGCTG CTGGGATTGG GATTTCATTA GCCAGCACTG 4320

GTTCGCCATT AGGGGCGATT GTCTTCTTAC TTTTATACAA CGTTCCTGCG TTCTTAATTC 4380

ACTATTATGC GTTGTATAGC GGCTATTCAA TCGGTGAAAG TTTTATCCAA AAAATGTACG 4440

AATCAGGCGG AATGAAGATA CTCACAAAAG TATCAAGTAT GCTAGGCTTA ATGATGATGG 4500

GCAGTATGAC TGCTTCAAAT GTAAAATTCA AAACGATTCT TGAAGTTTCC GTTAAAGGAA 4560

GCAAAGACGT AGTTAAAATC CAAGACTATT TGGATCAACT ATTCGTCGGT ATCGTACCAT 4620

TAGCCGTGAC ATTACTTGCT TTCTGGCTAT TACGCAAAAA AGTCAATGTT AACGTGGTGA 4680

TGTTTGGTAT CATGTTCTTA GGCATCTTGC TAGGATTACT AGGCATCTGT TAAAATTGAA 4740

ACAATAAACA TTTAATTATT AAACACTTAG AGCCTGAGAC AAAAATCCAA AGTGATTTTT 4800

ATCTCAGGCT TAAAAGCTGA TAAACGGCGG GAACAGAAGC AACCCCTGTC CCGACCTCTC 4860

TTTTGGAAGG ACGATGAATC ATGCGCTTGG AGACCTACCA ACACTTAGTT GATATTTTAG 4920

AAACCAGCAA CACCCCGATG AGCACACAAG AAATTGCCAC AAAAATTCAA TTAAGCCGCT 4980

CGGTAACCAG CTTATACTTA AATAAATTAT TGGAAAAAGG TGAAGTCCAA CAAACGGGAA 5040

AAAAACCCGT TTATTGGCAA TTAACCCGTG CAACAACCCC CACGACCGAT GTGTTTCGAC 5100

AATACATCGG TAGCCAAGGT AGCGCCAAAA AAGCCATCGA ACAATGCAAA GCCGCTATGT 5160

TGTATCCCCC ATTGGGTATG CCTTTATTAA TCCATGGAGC CAGTGGCGTC GGCAAAAGCT 5220

TTTTGGCAAA ACTGATTTAT GAGTATCTAA AAAACGAACA AATCATCGGC CTGGAAAAAT 5280

TCTATACCTT TAACTGTGCC GATTATGCCA ATAATCCCGA ATTGCTATCA TCTATTCTCT 5340

TCGGCCATAC CAAAGGCGCA TTCACTGGGG CTGAATCAGA AAAACAAGGA CTCTTGGCCC 5400

AAGCTAATAA TTCCGTACTC TTTTTAGATG AAGTACATCG ATTATCCAAC GAAAACCAAG 5460

AAAAACTCTT TCAGTTTATG GATACAGGGA CCTTCCGCCC TATTGGTGAA GAAGGAAAAA 5520

TGGTGCATTC CAAGGTCCGG CTACTATTTG CGACCACCGA AAACCCGAAA AAAGTCTTAT 5580

TACCAACCTT TTATCGCCGA ATTTCCGTAA TCGTCTCACT TCCGAACTTC AAAGAGCGCC 5640

CCATTCGAGA ACGGATTGCG ATTTTGAAAA ACTTATTCCA CCGAGAAGCG AAACGAATGA 5700

CCAAAGACAT CAACGTGGAC GAAGAAATCT TTACGGCATT ATTAGAAAAT GACGAACCTG 5760

GCAACGTGGG GAGTTTATCC AACAAAGTGC AGCTCCTTTG CGCGTCGCAA TTACGAAAAA 5820

CTTTGCCCAA TCAACCTGTC GTCATTGGCG ATGCAACTCA ACCAATGATT GCGATTCCTT 5880

TAGACAAAGA AGTGTTAGAA GAAGACACAC TTTCTTCCGA TATTTTCGCA ACATTTGAAG 5940

CCTTATTTAC AAAAGAAAAA ACACTGGCCC ACTTAAAAAC TGAACTGACA CAATTCATTA 6000

AATATTGTTT AGACGATGAA ATAACCTTGG AAAATGATTA TTTTTTACAA AACCTAGTCA 6060

CGGAAGTCGA TAAAGTCCGT ATAATTGTGT AAAAGTAAAA AGGCCATATA ACAGTCCTTT 6120

TACGGTACAA TGTTTTTAAC GACAAAAACA TACCCAGGAG GACTTTTACA TGaCCCAAGT 6180

ACATTTTaCA CTGrAAAGCG AAGAGATTCA AnGCATTATT GAATATTCTG tAAGGaTGAC 6240

GtTTCTaAAA tATTTaACAc gGtATTTAAT CACTATGGGG AAATCACGGA CCGAATTT 6298

(2) INFORMATION FOR SEQ ID NO: 509:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 498 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 509:

CAAATAGCCC TTGATTAAAC GCCAAGACAT CGACGCGTTC TTTTTTCATA AGTGCCTTCT 60

TTCTATCTCA ATTGTAACTG AGTTAACAAG TTTGCTAAAT GTTCTTCTGA AAAGTTGGGA 120

ACATTTGCTT TGATTTTTTC TAAAACAGCG CTACCTTCTG CTAATTGATG TGTTAGCGCA 180

TCTTTCGCGC CAgCAATCCC TAGAAGAGCT GGATACGTAC TTTTATTCAA CGCTTCATCT 240

CGGCCTACTT TTTTACCTAA ATCTGCTTCT GTACTTGTGG CATCTAATAA GTCATCACGA 300

ATTTGaAAAG CCAAGCCATA GTGAgCGCAA ACTGTGTCAG TAAACCaATG ACTTCTTCTG 360

kTTGATTGGg CTAATACGCC CCCAGCAATT AAGGsGrATT CAATTAGCGs TCCGGGTTTC 420

TTTTCATGAA CAGcCGcTAA ATCTTCTnAA GGCAAGCTGA ATTTTTCTCC TTCAAATATC 480

ACCCATTGGC CGGAAACC 498

(2) INFORMATION FOR SEQ ID NO: 510:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1923 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 510:

ATTAAAAGTG nACGACAAAA TCAAAGCCGC TTTACCTTAA CCAAAATGAA TACGTTCATC 60

AAAAACTTCC TTGGTAGCTA CATTCTGTTA AGTTTAGAAA TTTTAATTGC TGCTGATATT 120

ATTGAATCCA TTGTCAAACC AACTTTTCAA GATATTTTGA AATTGGCAAC CTTGGTGATC 180

ATCCGTACAG TTATTTCTTA TTTCTTACAT AAAGAAATTG AAGACACTAT GACCGATATG 240

GAACAACAAG AAAAAGAACC AGAGACCCCA AAAAATTAGC GGGTACGCTG GTTCTTTTTT 300

TATTTAATCG CTTTAAAATA ATGAAGTGGT TCGTTAGACA AACGCTGCAT TTTACGTTTT 360

TTATCTTCAA TAATTGCTAA AATCTCGTCT GAAAAATTTT CTAAAATAAC ACGGCCGCCT 420

TTATAGGTGT GTCCTCCCAT TTCAGCAACT TCGCTGGCCG TCAAAATGAC TTTACGATGC 480

GCAATTTTTT TGAAAAATTC TTCAATTAAG TACGTCGGTA AGTAAAAATT ATTTTGAGCA 540

GAAAATGATT CCAAATAAAT AAAATCATCG ATGCTGATGA AACAATCATC CATCGATAAG 600

TATTCAATTG ATCTGTCTAT ATATAATTCA CGCTTTTGTA AAATTTCTTT GACTTTTTTA 660

TTTAATAAAT TCAATGATTC AATTTCTTTA ACCAAATGTT CTAGTTCACG ATAACGTACT 720

TCTTTTCGCT CTTTGTAAGA tGAATGTAAG GCGGCACTTG CTGTGACTAT TAGCGCACCG 780

ACGATTACAC CGACCATTCC GGCGATAAAA ACAAACATAT TTTTCCCCTC GCTCTTTTCT 840

AATTTTCCAT TTTAGTATAA CATAATCCAA AAGTTTCGGA TAAAAAATGA CACAAAATTT 900

TTCAAGCGAC TtCGTTGACC TTTATTGACC AAAGAGGTAT ACTTTTCATA GGCATTCAAA 960

GTGCTTTGTG ATAGAATGAC TTTAAAAGTT TCTTTATATA GGAGGATTTA CTTATGAATT 1020

TAATTCCAAC AGTTATTGAA CAATCATCTC GCGGTGAACG TGCTTATGAC ATTTACTCTC 1080

GTTTATTAAA AGACCGCATC ATTATGCTAA GCGGCCCTAT CGATGATaAC GTGGCAAACT 1140

CAGTGATTGC ACAGTTATTA TTCTTAGATG CCCAAGATTC AGAAAAAGAC ATTTACTTGT 1200

ACATTAACTC TCCTGGTGGC AGCGTTTCTG CAGGCTTAGC GATTTTCGAT ACCATGAATT 1260

TCGTTAAAGC GGATGTCCAA ACGATCGTAT TAGGAATGGC AGCTTCAATG GGTAGCTTCT 1320

TATTAACAGC TGGTCAAAAA GGCAAACGTT TCGCTTTACC AAATGCGGAA ATCATGATTC 1380

ACCAACCACT TGGTGGCGCT CAAGGGCAAG CAACAGAAAT TGAAATTGCT GCTCGTCACA 1440

TTTTAGACAC TCGTCAACGT TTAAATTCAA TTTTAGCTGA ACGAACTGGC CAACCAATTG 1500

AAGTGATTGA ACGTGATACC GATCGTGATA ACTATATGAC TGCTGAACAA GCAAAAGAAT 1560

ATGGTTTAAT CGATGAAGTA ATGGAAAATA GTAGCGCCTT AAATTAAAAA TAAAAAAATG 1620

GAAACAGCGA CAGTCGCTGT TTCCATTTTT TTTaTTTTTC AATGTATCCG CGTAATTCTT 1680

CTTCAGTGAC ACTTTCGTCA AAGATATACT CACCAATGAT GATCGTTGGT ACAAATTGAA 1740

TGTGCGCCGC ATTGGGCTTC AGCAATCACT GGCGGGAnAC AAGCGTTGGC ATCTTTTTGG 1800

TTCTTTTAAG CCCAGATTTT TTTCAGCATA CGTCGGCTAC TTCTTCTAAG GGTAAGTTTC 1860

CCCATTCATC nTGCGTGAAA CCATTTTAGG CAnGCTGACA CGCTGGTTCA GGCGCTGAAT 1920

AGT 1923

(2) INFORMATION FOR SEQ ID NO: 511:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1613 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 511:

TTCTCGCTTA ATTTCTGGAA AAGGATATGT GAGTGCCGAT AGTCAAGAAA AAATTCTCAA 60

AGCAATGAAA GAATTAAATT ATACCCCGAA TTATGTGGCG CGGAATTTAC GAGCAGGTGC 120

CACAAAAACC ATTGGCTTTT TAGCCCCTAG CTATTTAGGT TCCTTAGGTA TTTTCTTGAA 180

TCGCTTTATT TCAATTGCCA AAAAATACAA TTATTCCGTA ACCTTATTTT TTACAGATGG 240

CGATCAGCAA AAAGAAATTG ATGCCTTGAA CCAATTAAAA TACAAACAAT TAGATGGCTT 300

GTTTTTATTG ACACGAACAA ATGAATGGTC GGTGATTGAG CCTTACAGCG CGTATGGTCC 360

TTTAGCAACC TGGCATCGTA TCGATTCACC TAATATATAC TCTTCCTACA TTGATCATTA 420

CTCTGGTTAT TATCGTTCCT TAGAATATTT ACAACAACAA GGATACCAAA AAATTGGTCA 480

TGTCCTCGGT AATCCCGAAA ATTTAAACAC GCAcGCACGA AAACGTGCAA TGAAAGATTT 540

TTATGAAAAT ACTCAGGCAC CTTGGCAAAC TGAGTGGCTG ATTAATGATC AGTATACTAA 600

CCATTCTGGT CGCAAAATTG CCCATTACTG GCAACAATTA ACCGACAAGC CAGAAGCTTT 660

GGCCTTTTAT AATGATCACG TAGCAGCTGA GTTTATTTCA GAACTTGAAA ATCTAAATTA 720

TGcCGTTCCT CGCGATGTCG CTGTCATTGG CTTTGATAAT AGTTATGTCA GTGAGTTGAT 780

GCATATTACG ACTGTTGATT ATGCCATTCA GTATCAAGCC GAAAACTCAT TTATTTACCT 840

ATACAACCAA TTAAATCAGA CGAATATTCC TGAAAAAAAA TTAACCGTTC GTCTAGTTGA 900

ACGACGAACG GTGCCTAAAA AAGACCCGGC AGATCAGTCA CTTTGACTGA TTTACCGGGT 960

CCTTTTATAG TTAAACAAGG ATTATTTATG CTTTTGTTCA TTTTTTAAAA TAGAAATTGA 1020

CTGTTTTAAA GATTGAATCA TGCCGCCTTC ACTGTATACA AGAACGTTTG ATTTGTATAA 1080

ACGAGCAGAG AAGATTGTTA AAATAATTCC AAAAACTGCC AAAATCACTA ATGAGACAAT 1140

TGCTTCCATT GTTGATGCCG TTTCGGTCGC CAAGCGTACT GGCATAATGT ATGACGAAAT 1200

CAGCGGAATA AAGGAAGTCA CTTTCACCAC GATATTTTGT GGATCATTCG TCCCAAACGT 1260

AATGCCGATC ATATAACCGA TTAAACCAAT ATACATGATG GGCTGAACTG CTTTTGCTGT 1320

ATCCTCTGGT TTGGAAACAA GAGAACCACA AAGTGCTGCC AAAACAGCAA AAACGAGAAT 1380

ACCAATAATG AAAAAGCCTA ATGTAAAAAT AAAGGTTGAT GAAAAAATAC TTTGTGGTGA 1440

AATACCACTT AATAGACTTT TCATAAAATC AAGATTTTTC AATTGAGAAT AGCCGAGAAC 1500

AAACGCCACT CCGTAAATGA AAATTTGTGT TAAAGCTACT AGmATAAcCC CAGTTAmTTT 1560

GCCATAAAAA TGGGTTgCGC yTTTGTACTA GAGAGAATAA CTtCCATGAT TCT 1613

(2) INFORMATION FOR SEQ ID NO: 512:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1496 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 512:

ATAAAATTTC ATATTATcTC TATCCTTTCT ATAACTATTA ATGATATGAA CTTTTAAGTG 60

TAAAATAAAA CACCTAAACA ATTAATTTCA TTTCCAAAAG TTAGATAATT ATTTAAAAAA 120

TGTTTCGTAC TAAAACATTT TTTAGTCTAA ATTTAATGTA AATTTAGAAT CAACTGACAT 180

CTGATATTTT TTAGGAGTCG TCCCTGTTAT CTTTTTAAAA ACTTTAGTGA AGTGAGACGT 240

ATCTGAAAAA CCTAATAATA GTGAAATATC ATTAATAGGA TAATTAGATT TTAACATTAC 300

TTTAGCTTCT GAAATTTTTC TTCTTAAAAT ATAATTATGT ATACTAATGT TCATTTCTTT 360

TTTAAAAGCT GTGCGTAACC TAGCTTCACT CATTGAAAAA TATTGAGCAA TCTGTCTTAC 420

TGTAATCTTT GTATACATAT TTAGACCAAT GTACTGAACA ATTGATGATA TCTGAGGAGA 480

AAGATTCCTT ACTTTAGCTT TTCCTATTTC AGAGGTATAA AATACTATAG CAGTATCTCT 540

AACTTTTAAT ACTTCTTTTA AATTTATACA TAGTTCATTT TTCCTTATAA AACTATCTCG 600

ACTCTGATAT GCATTTATTA AGTCCATTCC CATATTTATA GCTGCTTGTG CTAACTTCTC 660

AAATACAACT ATTGAATAAT TCTTTcAGAA CGTACGTTAT CCCCACTAAC CACAGGwACA 720

ACACCATTAC TTAGTTTAAA TATCATATTA CTTAATTCTT GTACATTACC ATTTACAACA 780

TATTCTAAAA TTTTATTTTC ATATAAAAAT AAATAAATTT CTGGATCATA ATTTTGAGAT 840

AATAACATAT CTATCCGTTC TATTTGGATA TCTTCTGAAA ATTTTTTCTC ATAATCATGT 900

AACGGCTTAT GAAATAAGTC TTCAATCTTT CCCGTAAAAA AATAATTTAT AAGGATAAGA 960

ATATCTCTAA TATCCCCTAA TGAAAATAAT GGTAATTCAT TTAATAATTT ATATAAGGAT 1020

TCTTTATCAG AATGCTTTAT ACTTTTATAT TGAACCATAG AAAGTCAAGT CCAGACTCCT 1080

GTGTAAAATG CTATACAATG TTTTTACCAT TTCTACTTAT CAAAATTGAT GTATTTTCTT 1140

GAAGAATAAA TCCATTCATC ATGTAGGTCC ATAAGAACGG CTCCAATTAA GCGATTGGCT 1200

GATGTTTGAT TGGGGAAGAT GCGAATAATC TTTTCTCTTC TGCGTACTTC TTGATTCAGT 1260

CGTTCAATTA GATTGGTACT CTTTAGTCGA TTGTGGGAAT TTCCTTGTAC GGTATATTGA 1320

AAGGCGTCTT CGAATCCATC ATCCAATGAT GCGCAAGCTT TTGAATATTT TGGTTGATCG 1380

ATATAATCAT GAATCAATCG ATTTTTAGCC TCACGCGCTA AGTTAATATC TGTGAACTTA 1440

AAAATTCCTT TAACAGCTCC TCTGAAAGAT TnGAATTTTT TAGGATGGGG nAAGAA 1496

(2) INFORMATION FOR SEQ ID NO: 513:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 683 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 513:

GGTACTTTAC AAGAATCACG TAAACGTGAA TTCTACGAAA AACCAAGTGT AAAGCGTAAG 60

AnAAAATCTG AAGCAGCTAG AAAACGTAAA AAATTCTAGT TTTGAATGAA TTGGAGTTGG 120

TGGTATGTCA TTGCTAACCA CATTGAATGA CGACATTAAA ACAGCGATGA AAGCTAAAGA 180

CAAAGAAACT TTAGCCGTAC TTCGTATGTT AAAGACTGCC ATTCAAAATG ATCAGATTAA 240

AGCGGGCCGT GAATTGAACG GAGAAGAAGA ACTCACTGTA TTGTCTCGTG AGATGAAACA 300

GAGAAAAGAT TCTTTATCAG AGTTTGAAAA GGCGGGACGC GACGATCTTG TTGAAAAAGT 360

AAAAGTTGAA ATAGCAATTG TTGAAAAATA TATGCCTCAA CAACTAACGG AGGAAGAAAT 420

TCGTCAACTA GTTCAAGAGG TAATGACGCA AACTGGCGCC ACTTCACCAA AAGATTTTGG 480

TAAAGTGATG GGCGCTATGA TGCCGAAAGT CAAAGGCAAA GCcGATGGcA ACCAAGTTAA 540

CGCaATTGTG aAAGAATTAT TACAAGaAAA CTAAGAGCGT TwTCGAAACT AGGACAGGAT 600

ACTGcCgawT TtTTTTTaTG sCcaaGAgAA AcCGGTAGCC TTTTTTACGG TTTCCCCTTT 660

TTTGGACCTT TCCAAGGTCC AAT 683

(2) INFORMATION FOR SEQ ID NO: 514:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 793 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 514:

CGAATAGTAT GAcGTACACG GAATAATTTC CTTCGTCAAC TGTTTTGACT TGAATAAAAA 60

AATAGTTATT TTTAGAATAC GTTTCCACGC TTTCTATCAA TAACTTGCCG ATACCTTTGC 120

GATGATACCG CTTTTTAACT GCCATACAAT GGATTTCTAC TGTATCTTTA CTTGTTTCTG 180

AAAGTGTTAT AAAACCGAGT ATTTTATTCT CTACATCCTT TGCCACCCAC ATAGGTAAAG 240

TGCTAGCTAT ATCGATATAC TTGCGGGTCT CTTTTTCAAG GCCAAACCAT TCTGGAAGAT 300

TAAACAATAC TTCTTGTGCT GCCTGAGTCT TTTTATCTGC ATTTTTTTCC CTTATAATTG 360

TAAATTCTTT CAATTTACAA CCTCCTTTTT GTTTACATAA CAAAATTATT GTATACCCAA 420

TCTTTTTTAT TAAATAATAG CATAGGTAGT TTAAAAATGT AAGTTAGTCA TTTATTCTTT 480

ATTATTTCCT TAAAACGACG CCTTTTCTCG CTAAATCTGC CCTTTCTTTA TTTTTTTCTG 540

TTAAACTAAA CGTATTAAGT CTATGAGGAG CGTGCGTATA TGGATTATCT GAACACTTAT 600

AAGCGGTTGC TAGCATTGGC AGATGCTGGT CTTTTTTACG GGAAAGATGC TTTTGATCAA 660

GAACGTTACC AAGAACTACG AGAACTTACG TkAmAACTTA TtAGCGCAAC CGGACACGAA 720

GCGAcTGTCT GGCCCGACTT AGGAAAAAAC CTTACCTAAA GAAGAnGGTA TCCAACACCG 780

GAAnGTAGAT GTT 793

(2) INFORMATION FOR SEQ ID NO: 515:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2805 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 515:

ATAAATGGCG TCGATGATGG GAAAACTTTT CTTGATCATA TCGACGATAA AATAACGTTT 60

CTGCTTCAAT AATAGAGCGA TAACGTTGAG GACTCATTGT ACGATAATCT AAAACAGACA 120

ATTCTTGTTT TAGGTGGATT GGTAAGACTT CAATCTTTTT ATTGGCTTGA TTCAATTGAT 180

AAAAGAACCA ATGATCTAAA TAGTCCAAAG GAAATTCTTC CGAAAATCCT TCCAACCAGC 240

GCAGTGTTTC TGCCGTAATG ACTGTACCAG AATTGATAGC CATCAACGGT TGGTTGGCTA 300

TCCCTGCTGT TGGCTTTGCT CCTTTAAGTC CAACATATTG ATCACTATAA ACTGGCGAAA 360

TTTGTTGTCC ATTTGCTTCT ACAATTGGAA CATAGACTGC CACAGTCGGA TCTAACGGCA 420

TGATGATCAA CGTATCAAAA TAAGAAGCTG GCACTTCTGT GTCTTGGTCA AGGAGCAACA 480

ATAATTCACA TTGATTCGTT TGACTAAAAG CAATCGCTTC ATTATAAGCG GTCGCTAGGC 540

CTGGATTATC AGGATTATGT CGATAAGTAA CATTTTGTTG TAAAAATAAT GCATCTTCTT 600

GAGGAAGTGG GCTGTTGTCA TAAATAAATA AGTGCAATTC GGGGTGATCT ACCACTTCTT 660

TTAATAACAA ATAATTCGGC GTATCAGCCA TTTTCATTTG ATATAAGACG ATGACTACCT 720

CTAATCTTTG ACTCATTAAA AGAACCTCCA GCCCCATTTG TTAAAAAATT TTCCCATTGA 780

TTGCATAAAG ATTTTAAACA ATTTTCGACT TTTATGGGCG CCCTTTTCAT ACATGTGAAC 840

AACCGTTTCA AAAGGCGTAT AGAGAATCCG ATAGCCTGCT TTGCCAAAAC GTAAACATAA 900

ATCGTTGTCT TCaAAGTACA TGaAGAAACG TTcATCGAAC CCACCAATTT CAACGAATTT 960

```
TTCACGATCA ATCAACATAA AACAGCCTGA GCCCATTTTA ATATCCGTTG TTTCTGTATC      1020

CGACAAATCG CGACATTCAT AAATACTCAA ACGTTTATCA AAAATTTTCT TTACAAATTG      1080

AAAGGGAATA AAACGTAACA TATAATCGAA AACATCTAAT TTTTGACGAA CTAAATATTG      1140

CGTCGTGCCA TCTTCATTTA GCACTTTAGG GCTAACGACT GCAATGTTCT TATCTATTTT      1200

GATACGGTTT AATAGACGAT CAAGCACGTC TTTAGTAACC AAAACATCGG GATTAAAAAT      1260

AATTGCATAC TTTGTCGAAG CATTGAATAA CACTTGATTA TGACCATGAC CAAACCCTTG      1320

ATTTTCTTCA GCGCGATGGA TAGTAATAAA TGGTTCATAT GTTGTTAATT TTTCAAGATA      1380

CGCTGTTTCA GAATGATTGT CATAGATATG AATATCATAG ATACTATCAG TACCCaGTTC      1440

GGCTTTTAAT TGGTCTAGTA CATTAAAAAT GTAACGACTA TTATAAGTGA CAATCGAGAT      1500

TGTTACCATT TCTTGCATGC TCATTtCTCC TTACTACTGA AACATCGTCT TAAAAAAAAA      1560

GCATATGTTT CACACATATA CTTTTTTAAG TTTAATGGAA TTATATCATA GGTTTACGCT      1620

CTCGCAAAGC AACCTAAAGA TTTCTTTACT TTCTCTTAGA ATGCTTGCCA AGTCGCTTTT      1680

TCATTTGCTC CTGACGATAT TCACGATTCC CTAAAATCCG TAACAAATAC ATCAACGGTT      1740

GATGTCCTTC GCCAACTAGA CCGATTAGTT CAATGAATAG TTCTAAGCCA ATTAAACAAA      1800

AGACAATTAA TAAAATTGAT GCTACTGTAC TTGAATAGCT GAACAATAAA GAGACAAAGG      1860

AAAAAACTAA CGCTAATGCA TAAATAGTCA TGACCGCCCC TTTATGGGTA AAACCTAAAG      1920

ATAACAAGCG GTGATGCAAA TGCATTTTAT CTGCTGAGGA AATGGGCTTC TTGTTCAATA      1980

GCCGTCGAAT AATTGCATAA ACCGTATCCG TAATTGGCAC ACCTAAAATC ACCATTGGCG      2040

TAATTACCGT AATAAACGTA GCATTTTTCA AGCCCTGTAA CGACATTACT GCAATCATAA      2100

ACCCGAGGAA TAACGCCCCG GTATCTCCTA GAAATATTTT AGCCGGATAA AAATTGTATG      2160

GGAAAAATCC CGCAATGCTC GCAACTAAAA TAAAAATAAC AATTGGGATA TAGACCGTTT      2220

TAGCATGTAG GAAAAAATAC CCTGTAATAC CAATCGTGGT TAATCCAATA ATGGATACGC      2280

CTGATGCTAA ACCATCCAAA CCATCAATTA AATTTACTGC ATTCGTAATC GCTAAAATCC      2340

ACAATAAAGT TAGTGGTAAA CTAAACCAGC GCAAATCAAT CATTCCAACA ACTGGCAACG      2400

TCACAAAATC TATCCGAATT CCAGCAACAA AATAAATAAC TAATGCTGCT AACAAAATAC      2460

CGATTGTTTT TTTCATTGGA GTAATTTCTT TAATATCATC AATGAGGCCT GTCAAAACAA      2520

CCATTCCACC AGCCAAAATA ATCGGCCAAA TATAATCTTG TGGGATAATC GAACGAAAAA      2580

TCAATAAACA TGAACTAGCA AAAGCnATGT AGATTGCAAG TCCTCCkGCT GTAGGCATAT      2640

TTTTGGTATT GaTTCGTCGC TCTCCTGGTG cATCATAGGC GCCAATCCGA AACGCTAGAA      2700

GCTTAACAaG TGGTGTTAAT AGCAAAGATA AGATAAATGT CATAAATAAA CGTATCAAAA      2760

CTTCATAAAT AAATAGTGAC ATAGTCGAAC CCACTTTCTT CTCGT                     2805
```

(2) INFORMATION FOR SEQ ID NO: 516:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5101 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 516:

AAAGGTGGCG TCATCGAAGG AAAAGTTTCT TCAGTGGAAC AAATTACTGC GTTGGCAAAA 60

CTACCAAACC GCGAAGGTTT ACTATCTATG CTATTATCTG TATTACAAGC GCCAGTCCGC 120

AACGTGGCTT ACGCTGTCAA AGCAGTGGCA GAGAAAAACG AAGAAGTTGC TTAATTGCAG 180

TCTAATTCAA ACCAATAAAA AACAAACAAT TTTAATTTAA ATGGAGGAAA TTCAAAATGG 240

CATTAAACAT TGAAAACATC GTTGCTGAGT TAGAAACAGC AACTATCTTA GAACTAAGCG 300

AATTAGTTAA AGCTATCGAA GAAAAATTTG ACGTATCAGC AGCAGCTCCT GtAGCAGTAG 360

CTGGACCAGC AGCAGGTGGA GCAGCAGAAG AACAAACTGA ATTCACTGTT GAATTAACAG 420

CAGCTGGCGA TCAAAAAGTT AAAGTTATCA AAGCAGTTCG TGAAGCAACT GGCTTAGGCT 480

TGAAAGAAgC TAAAGCTGTA GTTGATGGCG CTCCTGCACC AGTTAAAGAA GCTGTTTCTA 540

AAGAAGAAGC AGAAGCTTTA AAAGCTGCTT TAGAAGAAGT TGGCGCTTCA GTAACAGTAA 600

AATAATTCAT TCAACTTTTG AATGTTCGAG AACCCTTGTG TTGCAAGGGT TCTTTTTTAT 660

TAAGCTTTTA AATGGTGGTA AAGTTAGTGA ATTGATTGTC AAAATAATCT TcGTTTAATA 720

AAGtgCtTCC TAGCTAAggA AAATCCAGAG GATCTTAACC TCTGGATTTG ATAATTTTCT 780

ATAAATTCTC GTTACCAATA TTATTTGAtA AAAAGCAAAT TATTTAAATA GTGACTGAAT 840

AGTAGGTCTA TTTCTTTTTA TnTTTTATAA TGAATATACT AACTAAAGAT AGTAAGATGA 900

TTCCAATGAT AGTAAATAAA ACTGATTTTT TTTCGCCAGT ACTGGGAAAT ATCCCCTTTT 960

TGTGCTCGGT TGAACTTGAT GTAGTAGCTT TCTTCGAACT GCTCGTAGTT GTTTGTAATG 1020

TAGATTCTGT ATTAGTAGTA TCCGAAGAAA CACTTTCAGC ACTTTGATTA ATTGACGTAT 1080

AATCGTTTGC TTTTATTTTA TCTGCCATTG CTGCTAAACC ATAATCACCA CGGACATCAG 1140

ATGGGTAATT TTGGTTAGCG TAGTCTACCT TGACTTGATT AATTAAATCG ACGAGTGTTT 1200

GATCGCTTAA ATTATTtACA ACATCACTAT TTAAATcATA TACATAAATT AACGAAAGTC 1260

GAATTTGAGG TAAATAATTT TTTAAGTCAT CAATCGTTAA tGTTGaGTAA TCTGTAGAGT 1320

AGGAGTCCTC CGGGATAGGT TTCTCACCAT ATAATTTAGT AATCATTCTC GCTGTTCCAC 1380

TGATGTCTTG GTTTACAAAA TGAAACGAAA CTTTTTTTGC GTAGTCCAGT TCGTCATCAG 1440

ATAAACGATT TATCAAGGTT TCTGTTAAAC CATACATACC TACGTAATTA AGAAGTGTTT 1500

GACGAAGCGA TGGAAGTTCC TCGCTTTCAG CTGTAAGTGC TGTTGAATCA GTTGAAGCAT 1560

TAATTGTTGT TGCTGAAGTT TCTTGAGTTA AAGAAGTAAC TGTTGACTTT TCCACAGAAG 1620

CTTCATTAGC CGAATCTTCT GCCCATACAT TAGAAACTGA TAGTGTGGTA ATCGAAACAA 1680

AACAAAAGAG TCCTAAAAAC ATTTTTTTCA TTTTAATCAT TCCTTTCGTT ACCTATTGTA 1740

ACATTTTTGA AATATTTTTG TAACAATTAA ATATTAAGAA AAGCTTAAGA ATATATACTA 1800

GGTGATAAAA AATATAAATT TTAAACTATT AATCTATCAA AAATACGATC CGTAGAGACT 1860

ATTACATTAG TGTCAAAATA TTTCAAAGGT GTTGGTTAAA GAAGTACTGT ATTTGCTTAA 1920

TGGTAGTTTC GAGTAAATTA GTCATTATAT AAAAGATTGA CAAGTATGCT GCATAGCGAA 1980

AACTTCGTCA ATCTTTTAAA AGGATAAAAA TACAGATTTA AGAATAATGA CAATTTATTT 2040

TAAGTCAAAA ACTGTTGACG ATATTGTCGA GGTGTCATCT CTTTTTTCTG GAAGAAAAAT 2100

TTTGAAAAAT AAGAAGGTGA AGAAAAACCT GTTTCAAAAC TGATTTCCGT GATGGTTAAG 2160

GAACTTTTTC TTAATAGAGA ACAGGCTTTT TGCAATTTTA AGTTTTGTAC ATAAATACTA 2220

GGATTTGTTT GTTCATAACG CTGAAATAGC TTAAATAAAT AGCTACGATC AATGTTATAC 2280

TCAGAAGCTA ATTGACTAAT TGAAAGATTG TTTGAAAAAT TTTCGTGAAT AGCCATTTTT 2340

ATTTGGTTAA ATAATTGTTG GCTTTGATTT TCGTGCTGTG TAACGGAATG ATGCTCTAGA 2400

TAATTCAACA ATTGAAAGCT TTCACTTAAG CTGCCAAAAT GATTATGTTT CTTCGCTTCT 2460

TGGTAAATAT TTTTAAATAA ACTTTCTAAT TTCTTATCAA CAGTTCCTTT TAAAATACAA 2520

TTTTCGGTAG TCAGGCCTAT TTTTTCTAAA TACTGAGACG AAAGATCTCC AACAAAACCG 2580

ACCCAATGAT AGGTCCAAGG ATTTTGCCGA TTAGGATAAT AAATATTGCG ACAATTGGCT 2640

GGGATCATAA ACAAATCGCC AGCGGTAATG TGTTGTTTTT CCATAAAGGG TGTTGAAAAA 2700

TAGCCTTCAC CATCGCTACA ATAATGCAAA ACCCAATAAT TAATGGGGGC AAACTCAGTT 2760

GGTTTTGTCG GACGGCATTT TTCATAGCCA GCTTCATAAA GATGAAATTC TTGGCTAATA 2820

GGTTGATCAT TAATCAGGAT GCTGACATTA TTTGTAAAAA CATTTTGCAT GCTATCACCT 2880

CTTTATAAGA GTATGTAAAG GCTTTCTAAA AGTCAACATC TTTACAAGAA ATAGGCAACA 2940

TTTTTTATTT TTAAAAAACT GTAAGCGTTT TACAATGTGT GAGTACTAAT TAATTTGATA 3000

AAGAGAGGAT GTAGTAAAAT GGCAACACAT TCACGGGAAC GATTAGGTTC TGTCGCATTG 3060

ATGCTCATGA CTTTTTCAGC GGTTTTTGCC TTTCCTAGCA TTGTTAACAA CAGTATTCAG 3120

ATTGGCCTAG CAACTATTCC AGGCTATTTA TTCGGCTCTA TTTTTTACTT TTTACCATTT 3180

ATTCTAATGA TTGCAGAATT TGCTTCTGCT AATTCTGAAA ATGAATCAGG TGTTCACAGT 3240

TGGTTGGAAT CTGTGCTAGG ACCTAAATGG GCATTTTTAG GCGCTTGGTC GTATTTCTTT 3300

GTCAACTTGT TTTTCTTCTG TTCATTGCTA CCGAATACAT TGATTTATGG TTCATACGCT 3360

TTTTTAGGTC AAAACGTTTT TCAAGGAAAC CATAGCACTA AAATAATTGC AGTGATTTCT 3420

ATCCTGTTAT TTTGGTTAAT GACATGGGTT TGTATTAAAG GGGTTTCGTG GATTTCGAAA 3480

GTGACTAGCT TAGCGGGTGG TGCTCGTTTA TTTATGGGCG TAGCTTTTGT GGTTTTAGCC 3540

TTTGTCGTCG TGTTTGGTTT TGGCAATGAG CCAGCACAAG AGTTTACCAC GACGTCAATT 3600

ATGCCAACTT TTAATTGGAC TTTTTTCATG ACGATGGCGT GGATTTTACA AGCTGTTGGC 3660

GGTGGTGAAA GTATCGGTGT ATATATTAAG GATGTTAAAG GCGGCAATAA AACGTTTGTC 3720

CGCACGATAA TTGGTGCAAC GATTGCGGTC GGTATCATGT ATATTCTTGG TGCCGTTGCT 3780

GTAGGTTTGG TAGTGCCaAC GGACGTCTTA AAAGGAAATT TTTCAAATGG TATTTTTGAT 3840

ATTTTTAAAA TTTTAGGTAC GTATTTCCAT ATTCCCGCAG CAATGATGGT GCGCTTAGTA 3900

GGAATTATCT TATTTGTCGG AAGTCTTGGT TCTTTGGCTC TGTGGACGGC AGCACCAGTT 3960

AAAGTATTTT TCTCAGAAAT TCCTGATGGG GTCTTCGGTA AGTGGCTAGT GAAAACGAAC 4020

GAAGAGGGAA ATCCAACCAA TGCTTTGTTA GTTCAAGGAA TTATTGTGAC GATTTTAGTT 4080

GCCATTCCAG CATTAGGAAT TGGAAATATG GATAGCTTTT TAGAAACGTT AATTAATATG 4140

ACAGCTTCTA CGTCATTAGT GCCTGTCTTG TTCTTGTTAA TCGCCTATAT TGGGCTTCGT 4200

TGGAAAAAAG AAACAATGCC GAGAAGTTTC CGCTTTGGTA ATCGAACCTT TGGCTTAATT 4260

GCGGGAATTT TCTTATTAGC CATTTTTATC TTTGTATTCT TTATGTCAAC TGTGCCGGAT 4320

CCAAAATTAA TCATGGAAGA AATCAATTGC ACGTTACCGA AAGGAACCGC TAGTCCACTG 4380

GGCATGTTGG CGTATAACAT CATTGGGTTA ATTGTCTTTA TGGGCTTTGC TTGGATTTGT 4440

TGGAAACGAT ATGAAACAAA AGAAAAAAAT GAAGTTGGTA AAGGAGAAAT GGATTATGAA 4500

GACTTGGGAA AATTATAAAG TAGATAGTAT TAATCGGTTG CCTGGACGAG CACACTTTTC 4560

GAGCTTTCCT TCGAAAGAAA CAGCGTTATT AAATGAAAAT AAATATACAC AAGCATACAA 4620

AAATTTAAAC GGATGCTGGC ATTTTCTCTT TTTAGAAGCG CCAGAATACA GCCCAGAAAA 4680

CTTTTTTGCG ACGGACTTTG ATACAAGCCA AATGGATCAA ATCACTGTTC CTGGAAATTG 4740

GCAAGTCCAA GGATATGGCA AAATGCATTA CTCTGATTTA TGGTACAATT TTCCCATTAA 4800

TCCGCCGTAT GTACCAACGG AAAACCCAAC AGGTATTTAC AAACGGACAT TTGCTATAGA 4860

TGAAACATTT CACGATAAAA AAATCATTTT ACGTTTTTGT GGCGTAGATT CAGCTTATCA 4920

kGTTtGGGTG AACGGCCATG AAGTTGGCTA CAGTAAAGGt GCaCGAAATG AAGCAGAaTT 4980

TGgACATTAC GCCTTACGCT AAAAtTGGgT GaACCAAATG aTTTamCGGg CgAGTCTATC 5040

aATGGtCaGA TGGkACgTAT TTwrGAGGGT CCAAGATATT nTGGnnGGGT TAAGTGGGGA 5100

T 5101

(2) INFORMATION FOR SEQ ID NO: 517:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1084 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 517:

ACAATTAGAT GGTACCCGAA TCGGCTTTAT CCTTAAAGGC ATCGTCGTCG GAATTCTTGT 60

CGGCTTTGTT GTCAGCTTAT TTCGCTTAGG CATTGAAAAA ATTGGCGAAC AAGTCGTTCA 120

TCTGTATCAA ACGTTTCACG AAAAAACCTT TTGGATGATT CCTTGGTTCT TATTTTCTCT 180

CGTGCTTGCT TATTTATTAG GACGACTGAT CAAAAGCGAA CCAGCAATTA AAGGCAGTGG 240

CATTCCCCAA GTGGAAGGAC AACTTCAAGG TCAATTAGAA ATTCACTGGT TTCCAGTGCT 300

TTGGAAAAAA TTTATCGGTG GGATCTTAGC GATTAGTCCT GGGCTATTTT TGGGACGTGA 360

AGGTCCTTCA ATTCAACTAG GTGCAGTGGT TGGACAAGGC TATAGTCAAT GGCGGCAAAG 420

CACCAAGTCT GAAGAAAAAA TTTTAATCAG TAGTGGTGCC AGTGCTGGTT TAGCCGCCGC 480

TTTTAATGCA CCAATCGCAG GACTTTTATT TGTTCTAGAA GAGGTTCACC ATTCCTTTTC 540

ACCTTTAGTT TGGTTAACCT CTTTTTCTGC GGCCATTTCC GCTAACTTTG TTTCTTTGCA 600

TTTCTTCGGA TTGCAGCCAG TTCTTTATAT TGGTCCGGTT AAATCTTTGC CTTTAGAGTA 660

CTATTGGACA TTAGTACTGc TGGGCGTTTT GCTAGGATTA TTAGGTTGGA TATATCAAAA 720

AACATTGTTG TCCTTACCAA AAGTATATGG AAAAATCAAA GGCCTTTCTT CTAATTATTA 780

TGGTTTTGTT TCTTTTATTT TGATTTTACC GATTGGTTAT TTCTTTCCAC ATTTACTTGG 840

CGGTGGGAAT CAAATTGTTT TAGCTTTAGG CAATCAACCG ACTACGATTT GGGCTTTAGT 900

CGGCTTACTG GTTTTACGTT TTGTTTTTTC AATGGTTCTT ACGGTCGAAT CTCCCTGGTG 960

GATTTCTGCC ATTCTAACTT TAGGGCATAT CGGACATGAT GGAGCACTGT CCAAACGAGG 1020

AGGAnCATTT GGAAACTTAA TTTGCAnGCG GATTACGGTG GAAnACGTAC ACGTATTAGA 1080

CAGA 1084

(2) INFORMATION FOR SEQ ID NO: 518:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2677 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 518:

CCACCTTCCA ATTTTGGTCC AAGTTGGGGT TAGCGGGArG CACGCAACrG CACAGCTTGA 60

AAAAAGCCGC TGATTATTAT GGCAACGGCA CACGCAATCC CTATCTGCGG ATCAAGAAAA 120

CCCAAGCGAA TTGGAGCTTA ACGGCGCAAC TGTCGCAACC GAAATCAGCG ACAGACAGCT 180

TGCCAACAGC GACCCGCTTA TTATTAGGGG CGGCTCCTGT TTCGAGTTTC TCAAATTACA 240

ACCAACCTAC CGAGTTGAAA AATGCGGTCG GTACTACGAG TGCCATTAGT TTGAATGCCA 300

ACAACACGGC CACTAGGATT ATTGCCAATC AACAATTCAC TGGGAGTAAT ATCTATCAGT 360

TGGACTTCAC CTTCAACAAT GTCAAACTTG AAGTGCCAGC CAACCAAGGG GTCAAGGGGC 420

AACAATACCA AGCCGCAATT ACGTGGAACC TAGTTACAGG TCCTTAAAAT AGTTAAGTTA 480

TCATGAGGAA AGCTCAAAAT GTTAGTATCA AACACTAACA TTTTGaGCTT TTGTTTATGG 540

AAATTAAGGA ATTTGGCTGA AAATCAGAAG CAAAATGGCA GATTCTCTTA ATTTTCGAGT 600

GTTTTTCTAG TGAAAAAGGA CTCTTCTACT TTGGTATAAT AGGAATCGTC GGAAAAGTAG 660

CTTAGCTGTA GCAAAATTTG AAATTTCAGG GCCACCTTTT TTTTGATAGA ATAATGGAGA 720

AAATTAAAAG CTGTACATAT ACCAATGAAC AATAAATCAA TACATAAAAA AGGAGAATAG 780

```
CATGAATCAA AAATGGCAAA AGTTACTTCC AACAAGGAAC TCACTAAGAG GTGGTTCAGC     840
ACAAGCTATC ACGACTGACC CTTGCGCCCT GTATGAGGGA ACAAGGGGCA AGCAAATCTT     900
AGCATTCTCA GCATAGCCTT CATCGGCTTG GCGGGAATGG TACATAGAAA GAAGGGACGA     960
CATGAAGAGA ATTAACTGGC AACGCTTAGC AACCATTGGC TTGTGTAGTT CTTTAGTGAT    1020
TAACGCCTTT TTTGGCGTAA CAGCCATTGC GGAAACAGTC ACGAGTGAAA GTAGCCCGAC    1080
AGTGGCAAAT AGTATTAAGG AAGCAACACC AGCAAGTAGC GCAAGGCAAG AAAGTACGGA    1140
AAGCAGTCAA GAAACGACTG AAACGAGTCG GGAAAAAGTA ACACAGGAAA CAGCGAAACA    1200
AGCAGAGACA AAAGCGCCAC TGGAAACACC AGAGATAAAA ACAAACACAA CAGCACAACC    1260
TGCAGAAAAT AGTAGTGAAA ATGACCGACA AACGAGTATT TCGCCACGTT GGGTGCCGAA    1320
TAATCAAGTG ATTTATCCGA GAATAACCTT GGCTTTCACT AATGAATCAG GGCAAGCTTT    1380
TGCTGATGAA AACTTAACGT TATCAGGGAC GTATACTAAA TTTGGATATG AAGACGCAAC    1440
TAGTTTATGG CGACCATATG CGGTTAATTA TCCAATTACT TCTTCTAATG CAGGTGGTGG    1500
TCAATATCGA GTAGCTGTCA ATGCAACGGT TCCTTTACCA TACGATTTTT TCACAGATTT    1560
ACCAGCAAAT TATAGGATGA ATATATACGC GGTTGAACAA TTAAAGATTG ATAACACGCT    1620
AAAGTATGTC GACAGTATTG TTGCTGTGCC CGATTCTACA CTGAGGATGA CCAGATATAG    1680
TGCGCAACAA ACAACTGAGT TACCGCAGTT TGGAAATCAT ATGACTCCtT TTGtATTACG    1740
CACAAGAGAT AGCATATTTT CTAAAAGCAG TGAAAATGTT TTCTCTCTAA ACACGGATGG    1800
TAATGAATTT TTAAGAATCC TGAGTGTCGC CCCCATTGGC ACTCCAGTAG AACCAGTATT    1860
ATACGGATTT GATGGCACTG CAACGTACCG AGAAACACTT AATTACGTAG TCACTCGGAA    1920
ACAAGTCACC GAGAAGTTCG TAGATGCTAA CGGCGTAgcT ATCACGCCGC CAACAGGTTT    1980
TACCCAAAAC AAGAAAACAC CAATGACCAG CAATGACTTC ACCTTCAAAC AAGCTAGTAC    2040
TTTGCCGGAC ACCTATCAAG CAGACGGCAA AACCTACAAA TTCAAAGGCT GGTACAAAGG    2100
CAAAACCAAG CCAAATACCT TGACCACCAC CAAAGCGCCA AGCTATGCGG TGACCTATGA    2160
TGGGAATGAC GATTTGACGG TGGTGTATGA GGCGATTCAA GAAGGAAAAG GAATGTCATC    2220
TGTTCAGGTG ATGAATAAGG CGATTCAGAA TCCAGTGACT TTGCCTGATC GCATGTATCG    2280
TGTCCTATTT ATCAATGAAA CAGGGGGGCT GACATTTGCT GATCGTTATG GCTTTACTGG    2340
AGAATTAGTT GAGGTTGCGG ATGGCAAAGT GACGCCAATT GGCCCAGTCC CGACTAACAA    2400
TAATAGCGCA GTAAAAGAAA TCACTATACC AGGTCGGGTT TTCGATACAG AAAGACCTAC    2460
GCCGTTGCAA TATGGAGTTC GCAATGCAAC CTTTACGTTG CCCAAAATGT ATAAAaCCGT    2520
AACGTATCAA CCTGGCCCAA ATTATACAGG TACAGCCTAC TCTATTCCAG AAACGTATCG    2580
tCTTAGTTTA AGAAGTGGTG TCmAAmGTGA CATCGATACm ACTGGATGGC GAATTGGCTT    2640
GACCCGACAA CAGATCCACA AAAGTTTACG ATGAGCG                             2677
```

(2) INFORMATION FOR SEQ ID NO: 519:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1494 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 519:

AGATTGTCCn AATTTTAATT CCCAATCCAG TGCCCAGTAT GGCnGGGATG TTTTCCnTCC 60

TTGTGATTAA AATCCCnAAC CAGTAAATGT GTCCAAATTT GGACCACCCA GTCCAAGTGA 120

CCAAATATGA GGGGTATGTT TTCCTCCATT GTTGGGAATT ACCAAGCCAT TAGATGTAAC 180

TAAGTTTGAT ACATCCCAAA GTGACAGATA TGAGTGATAT GTTTTATGGT TGTAAGGAAT 240

TACAAACATT AGATGTGTCG AATTTTGATA CGAGTCAAGT GATATATATG AGTGACATGT 300

TTGGTGAGTG TAGCGGATTA CAAACATTAG ATGTGTCGAA TTTTGATACG AGTCAAGTGA 360

TATATATGAG TGACATGTTT GGTGAGTGTA GAGGATTACA AACATTAGAT GTGTCGAATT 420

TTGATACGAG TCAAGTGATA AATATGAGTA GTATGTTTGC TGATTGTAGA GGATTACAAA 480

CGTTAGATAT AACTAATTTT GATACGAGTC AAGTGACAAA TATGAGTAGT ATGTTTGCTG 540

ATTGTAGAGG ATTACAAACG TTAGATATAA CTAATTTTGA TACGAGTCAA GTGACAAATA 600

TGAGTAATAT GTTTGATGgT TGTAGAGGAT TACAAACATT AGATGTATCA AAGTTTGATA 660

CAAGTAAAGT AACCGGTATG ATTTATATGT TTAGAGATTG TAGCGGATTA CAAACATTAG 720

ATGTAACTAA GTTTAATACC AGTCAAGTAA CATATATGTG GAATATGTTT TCTGGTTGTA 780

GTGGATTACA AACGTTAGAC TTATCCAATT TTGATACCAG TCAAGTAACA AATACGGATG 840

AGATGTTTGA TAATTGTGAT GCCCTAAAAA AGATAACATT GGGGGCTAAA TCAATATTTG 900

GGACTAAGAC TAATACTAAC TTGCCGTCTA TAGCAGrTAC TTCTTTGTAT ACTGGAAGAT 960

GGATTGGAGT AAACACGAGT AACACTTATA GTGATAGTAA TACATTTATG TCTAATTATG 1020

aCGGATCaGT TCCaGATACG TATGTATGGG aAAAAGCTAG CGTTCTAAAC aGTACTTTaG 1080

AGCCTTCTTC aGTAAGGGtT CATTCaGAAA GTGAAGTAGA ATGGACTTGG AAAATAACTA 1140

ATTCGTCGTC TAAGTCAGCC GAGAATGTTT ACTCAGATAT AACCTTACCT GAGGGTCTTA 1200

AAATTGACAA GAATAGTGTA AAAAAGAACA ATTTGCCTGT GTCGGTAGAT GATATTAATG 1260

GTATGAACAA TTTAGGAACA CTATCTAGTA ACGAAACAGT TACATTTACG TTTAAGACAA 1320

TAGTATCAGG GAAACCAGAT AAAtGGTTGG AATTAATGGG AAAAGTGACT kGGGAAGATA 1380

ATGcmTCCGC ACAGTCAATA GTTCTAATAA AGTGrAAATC ATCGATGAGG AGCAAAAAGA 1440

TAAAGGGAAT CAAACAAATG ATTTAGAGTT ATTATCAGTT CCAGTTGGGT TTCG 1494

(2) INFORMATION FOR SEQ ID NO: 520:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 884 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 520:

TATTTATGCT GATTCGTTAC TGGAACAACG TATTAAAGCA ATTAAACGTA TGGAAAGAAC 60

TATCGCTTAA AAATTAAAAA AATTAGTCTA TTCTTTTGAT AGGCTTTTTT TTTTAACCGA 120

AAGAAAACGC TTACGCATTT TTCTATCTTT TTTTTCAAAT CATATTATTC AATTAAAAAT 180

GATTCTAATT TCAAGATTAT TTTTTTCATT TCTTTTTTTT CGCCGTCAAA TACCGTCAAA 240

CGTTTTAATA AACATCTGTT ATTTCAAGGT AGTATATAGT TTTTCATAGT TTTCAATAAA 300

CTAGGATAAA AAACTTGTTT ATCCATTACA TTTTACCATC AATCTAATCT TTTTTGAAAA 360

AATTGGTATT AACTTCTTAA GTAAGTTTAT ACATATAGTT AGTTATTTTA AAACAAAGAG 420

GAGAGAAAAG AGATGAAAAA AAATTTAGTA GTAGGTTTCG TATCAAGTGT CGCTTTAGTA 480

GGAATTGCTA TCGCAGGAGG GGGTAATGTT CAAGCAACTG TTACTGATAG TGCAAGTACT 540

GATATAGGAA TTGGATTTTC AGGACATGGA CCAGGAACGA CACCGGGACC GTTGGATATT 600

CAATGGGCTC CAATAGGATT AGATTTTAGT AATTCGAACA TGGTAAATAC TACAGCCGTA 660

CAAGCATTTC CTGAAACCAC TGGAACCAAT AAATATGTTG TTGTAAGTGA GAAACGTTCC 720

AATGAGCCAA CTAGAGAATG GAGTTTGACT GCGCAATTAT CAGACTTGAC GAATGCGGGA 780

AGGTCAGAnA CACTTACTGG GGnTGGACTA AAAGTTTGTG CTGCTGTAAA AAGGATACAC 840

TGGAACGAnT ACACCAGAAA GCCCAGGATC AATTGTGAAA nAAG 884

(2) INFORMATION FOR SEQ ID NO: 521:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2583 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 521:

GAAACGTACA GAAACAAGAC AATTAGAAGT ACTTAAATAT ATATACGAAC AAGTCGAATT 60

AAAGGGCTAT CCGCCAACCG TTCGCGAAAT TGGGAAGGCT GTTGATTTAT CCTCAACTTC 120

TACCGTTCAC GGAcATTTAG CTCGTCTTGA AAAAAAGGGT CTAATCTTAC GTGACCCAAC 180

AAAACCGCGA GCAATCGAAT TAACACCAGA AGGCTTGGAA AAGATCGGAA TTCAGCCGAC 240

AACTATTCCG ATGTTAGGTG TCGTAACTGC TGGTGAACCA ATTTTAGCGG TGGAAGAAGC 300

ATCAGATTTC TTCCCGTTAC CTCCkGATTt ACGTACAGAA GAAAATGCTT TATTwATGTT 360

GACGATTCGA gGAGAAAGTA TGATTAACGC AGgTATTCtA GATGGCGACC AAGTGATTGT 420

TCGTAAGCAA TCGAATGCCA ATAATGGCGA CATCGTAATC GCaATGACCG aTGAAGATGA 480

AGCAACCTGT AAACGATTCT TCCGAGAAGT AGATCATATC CGCTTGCAGC CTGAAAATGA 540

TGCATTAGCG CCTATTCTTT TAGATAACGT TACAATTTTA GGTAAAGTtG TTGGTTTATA 600

TCGCAATCAT ATTTAATAAA AAAAGGAAGT CACTCCGTTG AGCGGCTTCC TTTTTTGCTT 660

GTTAACAgTG AAtAGCCTTT TGGGCTTCCT CGATAACCTC TTCTTCAGTT TCTACAGAAA 720

GTAGATAATT TAAAGCTTCT AAACTTTGCT CGGGTTCyCT TTTTGcGAGC TGACCTaACG 780

ACCAAGCAGc TGTTCCGCGA ATGACAGGAC GCTGATCATT TAAACAGAGA ATAATTTGTG 840

GGATAGCTTC ACGGCCGCCC AAATTcGtAA CGCAATTAAA GCATTCCGTT GCAAtGGTTT 900

TTTACCACGC CAAGAGCCTG CTAAGTGTCC GAATTGCTGT TTAAACTCTT TATTGGAAAT 960

AGTTAwTAAC GGTGCTAATT TGGGATAAAC TTCTTCAATC TTTGGTTCCA TCTCTTCATG 1020

AAAATGAAAA TCTTTTCCTT TATTATAAGG GCAAACCAGC TGACAAATGT CGCAACCATA 1080

AATGaCGTTA CGCATTTTTT TCCGATAGTC TTCAGGCATC ATGCCTTTTG TTTGCGTTTG 1140

ATAAGACAAA CATTTTTGAG CATTCATTCG ACCGTCCCCT AACAAAGCCC CTGTAGGACA 1200

CCCTGTGATA CAACGAGTAC AATCACCGCA ACCATTTGGA ACAGGTTCAT CTGGTTCAAA 1260

TTGAATATTG GTGGTCACCT CACCTAAAtA TACAAAGGAA CCAAATTCTT CGGTAATCAG 1320

TAGACCATTT TTACCAATGA ATCCTAAACC CGCCCGCTGA GCAACAGCGG TGTCAACATA 1380

TTCGCCAGTA TCTACTTGCG GCCGAAAGCG CCAGTGTyCT TTTTCTGCTT CAGTGGCAGC 1440

TTGCTCTTCA ATAAAAGCAA TTAATTTTTG CAACCGATCC TGCAAAATGT GATGATAATC 1500

AATTCCCCAA GAGGCTCGCG CAAATTGTCC ACGCTTTTCA TCTCTAGGCA TTTTTTCCTG 1560

AGCTTTGGTG GGATAAGCTA AAGCAATGGA GATAATACTT TTAGGATGCT CAAACGTTTT 1620

TTCTGGGTAA ATTCGTTCTT CCACATTTTG ATGTTCAAAT CCCGAGTTAT ATCCTCGCTC 1680

TCGTTGTTCA TGCAGAGAAT CTTCTAATTC AATAAATGGT TCAGCATGAG TAAAGCCGAT 1740

TTTATCGATG CCTAAACGTT GACTCTCTTG AATAATTTTT TCTTTCAATG TCGGCATTGC 1800

TCATGACTCC CTTCATTTTA TTATTTATCG TTTCGATGAT ACGTTTCATA GTCATGTGCA 1860

TAGAGATTTT TCTCGTCCAC GACGCCGGGA ACAGTGGCAA CTTTTTCCCA AACAGACCAG 1920

TCGATTTCTG GAAAGAAGGT GTCGCCTTCA AACGTTTCAT GAATCATCGT CCGATATAAG 1980

ACATCGCATG CGGGAATCAA TTCTTTAAAA ACGACAGAAC CTCCACCAAT AACGGTCACG 2040

CCTTCATACT TGTCAGCATA CGCAAGAATT TCCTCTACGG AATGCATAAC CAAAACGCCT 2100

TCAGCTTGGT AATCCGGATT ACTGGTAAGA ACAATCGTTG TTCGATTTGG TAACGGACGA 2160

CATCCCATAC CTTCGAAAGT TGCACGTCCT AAAACCAATG TATTATGAAT TGTTTTTTCC 2220

TTGAAAAATT TCAAGTCGTT GGGTAAATGC CAAGGCAATT TGCCTTCTTT ACCAATCACT 2280

CCTTGTTCAT CTTGGGCCCA AATGGCTGCT AACATAGCTT CTCTCTCCTT TTTTATACGG 2340

CAATCGGCGC TTTAATCGTT GGATGTGGGT CATAGCCTTC TACTTTAATA TCTTCCATAT 2400

CAAAATCAAA AACAGAAGCC TTGTCTGGAT TCAAAACGAG CGTTGGGaAA GAACGAACTT 2460

CTCGTGATAA TTGGTTCTTG GCATTTGGTT CCACATGGAT TTTTGATATA AGTGGGGGCA 2520

TCTCCTAGCG TGnTGGTACA AACTCCCCCA CTnCTAAAnC TGTTTCATGC GGCAATCAGA 2580

TGT 2583

(2) INFORMATION FOR SEQ ID NO: 522:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 965 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 522:

GATGGTTTTT CGATCAAATT TCTTAnnCnC TGTAGCAGTA TAAATTAGTA ATAACTTACT 60

CAGAGATTTT ATAGAGAAAG ACGTTGTTGT AGCTTCTCTT AATTGCTTTA ACTGCTTTTG 120

CAATTTTAAC GGATTTTTAT CTTTTAATAA TTTAATTTTC GTTAAACGAA AAATGTGTTC 180

CTGATGCTCA AATGAACGTT CAGCATCTGA ATGACTAAAA CGAAGATATT CTGAAAGCAC 240

CTCTTGATTT ACAGCAACTT TTCCTTTTGG TAAAAACCAT TCTTTGCCAT TTTGATCTAT 300

TAGATGCCGC TCACTTTGAT TTGAAATAAC AATTTCTTTA TCGTCTTCAA AAAGATACGA 360

AAAAAGCATT GCTTTTTTTA TAATTATGTA CTCAGTTACT TCATATGAAG ACAATTTTCT 420

TAATTCTTCT ATCAAACTAT TCCATTGATT ATTTTGAACA AATTTACTTA GTTTGTGAAT 480

AGGTGGTACT CGTAAAATGA ACTTCTCATC TTTCTCTATT TTCACGATAT AAAAAGTTGT 540

GTCTAAGTCA TTTTCCCAAA TAAACAAAAT AATCATTTAC TGCCCCCTAG ACTTGTTGTT 600

CCGATACTTA ATAGTCAAAT AAATATATTT AGGAATTTGC ATTAACCGAT CGTATCTCTT 660

CATTGTCAAA CAACGATATA ACCATTCCAA ATTGTGATTT ATATAAAATT CTGGCGCTCT 720

TTTTACTGAT CCAGCTAATA CGTCAAAAGT AmCGCcTACG TCTAAAAATA CATTAGCTGT 780

CAGGTGATCA ATTGTTTGTT CCAAAAAAAT CTCTTGTTtC GGTGAACCTA ATGCTACaAA 840

GACAAACCTT GGkTCAGCTT TGkTAAwTTG kTCCACAATT TCTTGCTGTT TTTTCTTTGT 900

ATAGCCATGT AAAATACCGC TAATAACAAT CCCTGGATAT TGTCGTTGAA TATTTTTAGC 960

TGCTG 965

(2) INFORMATION FOR SEQ ID NO: 523:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1108 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 523:

CATnCCGGGG ATACAATTGG TGGATTGGCT TTAAACTATA TTCATTTTGC GACAAGTGAA 60

AGTAAGTTAT ACCGTGCGCT TTATTTAGAA GAATATGGTG GCGGCAAAAA AATGCCAAGA 120

ATTCTCTTAT CACTATTTTA CTGAAGCGGT AAAAGCTGAT CCTGAGTATG CTAATTTAAA 180

CGATGTTCAA ATTGACTCTT TACACATGGG AACATGGGTG GTCGCTACTG GTGTGGCTGC 240

ATTGATGACT TCTGGAATTA TCCATCCTTC AGAAGAACAA ATCATTCATT TAATGAAAGA 300

TAGTATTGAA GCAATTCTTG AACGAGATGA ACCAATTGAT ATTGACCTAT AAAAATTAAA 360

AAAAGTGTAA CCTCATGCTG ATAACTTCTG TGAGGTTACA GCAAAAGACT GTTTCTTAAA 420

AGAACGGTCT TTTTTTGTTA AGTGTATTAA TGGTAGATTA TTCGCAACCA TGCTACAGTA 480

AAGAAGAGAA ATAGAAaGGG GCAAAAAACA TGAACCGAGT GAAAGATAAG CGTTTATGGC 540

TAAGTATTGT AGGAATTACC GGACTAGGGC TCTTGTCGGG CTTATTTTCA GTAAATGCGA 600

GAGAATATTA TCAAGCGTTG AGATTACCAA GTTTTGCTCC GCCAGGCTGG TTATTTGGAC 660

CAGTGTGGTT GGTGTTGTAT ATTTTTATGG GCATTACTTT TTATTTCATC TTAGTTCATC 720

CAAATAAACA ACAGAAAAAA CGGATGATTA CGCTATTTAT TTTACAATTT ATTGCGAACT 780

TTTTCTGGAC ATTTTTCTTT TTTTCTTTAC AAAATAATTT ATTATCTGTG ATTGATATCT 840

CATTGTTATG GCTATTATyA ATAATTCAAC AGTGGGACTA CTTGCGTTAT AAATTAATAA 900

CGCGTGGCTG GTTGATGATT CCTTATATTC TTTGGGTTAC TTTTGCGGCA GCATTAAATT 960

ACAGTATTTT ATTTTtAAAT TAGGAATATC AAAACACCAG AGAAAAATGG GAACTAGCCA 1020

TTTTTTTCTG GTGTTTTTTT TATTTATTTC AACGTAGTTA TTCTGCGGCA GTTGCTAATT 1080

TTTTCGCAAA TTCTTCTAAA CGTTGGAT 1108

(2) INFORMATION FOR SEQ ID NO: 524:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2349 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 524:

TTAGTAATTT CCTGTACTTC TCTCCCCATA CAATTCATAA TATCCTCACA TTGTTTCGTT 60

GTACGTACAT CATTCCATAA AATAGCAGGA CGAATTATTT CATTATCCTT GTCTAATAAT 120

ACTAAACTAT GCATTTGTCC AGATATACTA ATCCCTTTCA ATTCTTTTTT CATTTCGGGA 180

AATTCACTCA ATAAATCATT AATTACATTT TCAATTGCTT TAATCCAATC GTGTGGATTT 240

TGTTCACTAT ATCCTGGTCT TGAAGAAAAC AATGGATAAT CTGATGAAGA CGTAGCAATG 300

ACATCCCCTA TTTCAGATAC CAATAATCCT TTTAATGAAC TTGTACCTAA ATCTATACCT 360

AAAACATAAC TCATAATTAT TAAATCCTTT CAAAAAAgCT GAAATACCTA CAAAAGATAT 420

TTCAGCTCAT TTTTCTTAAT ACmGaTaATC aTTTAAAATA TTTTTTACAT ATTCTATATG 480

ACTTGATTCT ATTATTGGTT CATCATTTTT TAGGGCATAA TTAGTTAAAC TTTCTAAATT 540

TTCTTTTCCA GAAATAATTT TTGCTCCAAT ACCTTTTTTA AACGATGAAT AACGCTGTTC 600

TTTTAGATCC TCAAAGAAAC GATCCTCTTT TAATTTCATC GCACTCTTTA ATCCTCTCGC 660

AAATGTATCC ATTCCAGCAA TGTGAGCAAG TAATAAATCT TCCATTTCAA AAGATGAaCG 720

TCTAACTTTA GAATCAAAGT TTATTCCCCC AGGTTCAATA CCCCCATTTT CTAGTATTTC 780

ATACATAGCT AATGTAACAT CATAAACATT AGTAGGAAAT TCATCTGTAT CCCAGCCAAG 840

CAGAACATCA CCTTGATTAG CATCAATTGA ACCCAAGGCA CCATAGTTTC GAGCTACATT 900

AAGTTCATGT TCAAATGTAT GTCCAGCTAA AGTTGCATGG TTAGCCTCTA GATTTAGCTT 960

AAAATCCTTT TCTAACCCGT ATTTTAAAAT AAATGCCATA GTAGTTGCAG CATCGAAATC 1020

GTATTGATGT TTTGAAGGCT CTTTGGGTTT TGGTTCAATC AAAAATTGTG GCTTGTGTCC 1080

AATTTTTTCA CCGTAAAATA TAGCCATTTT GAATAGACGC GCAATATTAT CTTGTTCAAA 1140

CTTCATATCA GTATTTAGTA ATGTTTCATA TCCTTCACGT CCACCCCAAA AAACATAATT 1200

TTCTCCACCT AATTTTTTTG AAACaTCTAA ACCTTTTTTT ACCTGAGCAG CTGCGATtAc 1260

GTATACGTTk GCATTATTTG TAGATGCTGC GCCGTTGACA TAACGAGGAT TTGAAAACaT 1320

ATTTGCAGTA TTCCACAATA GTTTAATACC AGTtTCaTCC ATCTTTTCTT TTATTAAATC 1380

TGTAATTTCA TCAATATTAG AAAAGAATTC TTTTAAAGAG TTACCTTCTG GAGCAATATC 1440

AATGTCATGA AAACAAAAAT ATTCAACATC TAGCTTTTCT AAAATTTCAA AAAAAGCTAT 1500

AACTCTTTTT TTTGCAGTCT CCATAGGAGA TTCTCCCAAC CaACTTCGCT TATTAACkGC 1560

TTTTCCAAAA GGATCTGAAC CGTCTTGTGT CATAGTGTGC CAGTATGCTA CAGCAAATCG 1620

TAAATGTTCT tTCATAGTCT TTCCCATAAT TATTTCTTCT GAGTTATAGT GTCGAAACGC 1680

GTACATATTT TTTGTGTTCG TTCCTTCATA TTTAATTTTT TCGATTTTTG GGAAGTAAAC 1740

CATGTTAAAA GTCTCCTTAT CATCTTTAAA ATTGTTATTT CAATTCGATT AATATTTTTA 1800

TAATTTATAT ACTATTGAAA ATTAATTAGT CAGAGGTAGT TTTAATTTTA CTCTTCCTCT 1860

TCTTCAGCAA TTTGTAACTT CGGTATTTTA ATACTTTCGG TACCTATTTT CAATAAATGT 1920

TCTACAAGAT TTTCCTCTTT TATAGCAAAG AATGATTCTA AAACCATTCC CAAATTAACT 1980

CCAGTTATTA CCTGAATTCT TGAGTCTTCA CTTGCTAGCA TCATGGCAGT ATTTGCTGGA 2040

GTCCCACCCA TCAGATCTGC TAAAATCACA ATTCTTTTTG CATCATTTAA AGTACCmATa 2100

ATTTTATCTA ATTTATTAAA GGTCCCTTCA ATACCATCAT CAGGCGTCAT AGAAACaACA 2160

GGATAaTTTA TTTGATCACC TAwGATCATT tGAGCCGAAT TCAAAATTTC TTTCGCCATA 2220

AGGCCATGGC TCATTAAAAT TACTTCGTTG GTCATAATGA TCTCCTTTAC AATATATCTT 2280

TTAAAAAAGT CTTTTTATCA CCGGGGGCCA TAATnGTGGT AATTCATATC TCGACTTAAT 2340

AGTTCCnCA 2349

(2) INFORMATION FOR SEQ ID NO: 525:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1735 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 525:

```
TACGGCGTCG AAAGAAAAAT AAATAGACCG TACTAAGAAT AAAAAAGCGA CTGAACTTTT      60
CAAAAGATCA GTCGCTTTTT TGTGCTTATT TTTAATTTTA AGCCCGTTTA AGCCGTTTTA     120
AGTATTTTTC TAGTAAAAAA GTACTTAGAA ATTAAAAACG GCTCTTAAAA CGCAAATATG     180
AGCCAAATAA ATAAATTTTA ATTTTTATCA GTACGAGCGA TTTTTTACGT ACTGCCCATT     240
TACATGCAAA TTTAAAAATT AGCAAAAAAT CGTTCCGAAT TTGCATTTTA AATGGGTAAA     300
CTTGCAGTTT AATCGCATAG TTCTCGCAGA GGCTAAAAAA TAAAAAAGCA CTGAAATATT     360
TCAGTGCTTT TTTTATTTTT CTTCTTCAAT TTCGTTGATT GTCTGATGAT TAATTTTTCT     420
TTTTTCTTCT TCGGTTAATT CGTGAAGTTC ACTCGCCAGG TACTCTTTCT GATTCCAAAT     480
ATAATAAATT TGATACAAAT ATTCAGTTGG ATCAATTTCT TTGAGATAAT CCAATTCTCC     540
GTTTTTTAGT TTCTTTCTCG CCTCTTTAAA AAGTCCTGAA TAGACTAAAA TCTGCTTACC     600
TTTTAAAGAT TTATAATACG CGTCAAATAT TTTTTGATTG ATAAGATAAT CGCTATCTTT     660
ACCCGAATAC TTCGCCATTT CGTATAATTC TTTGTTGTTA TTTTGCTTAA TTTTTTGAAC     720
ATGGACTTGC GTGATCTCCG GCATTCCTGT GACATCTCGC CACATTTCGA GCCATTCTTT     780
TTGTGAAATA TAAACCTTAG AATCTTTAAA ATAAGATTTA TTTACAGCAA TTAATACATG     840
AAAGTGCGGA TTAAAATCAT CTCGTTGcTT GTTGTATGTG aTTTCTAATT TTCTGACGTA     900
ACCTTTCATC ATCGCTTTAA CATTTTTTCT TTTCACTAAC TTTCCAAACG ATAAATTAtA     960
TTTTTTTATT TCaTCTTCTA ATTCTTCAGC TTTCACATTT GGTGTCGTTA AGGTTAAGAA    1020
AATAAACTCT TTTTTCTCTT CTTGCTTAAT GTACTGCATC ATCAACGATA GACCTAACGC    1080
ATCTTTTCtC GCTCGTCGCC ACGCACACAT CGGACAGAAA CGATTTTTAC ACATATTATT    1140
TTTATACAGT TTCTTTTTTG TCTGCTCTTT ATCCGTTACA AACGCTAAAA ATGTGTTACA    1200
GTCCTTAATC AGATCCATCT GTTTTTCTCC CACAAATTTA TTAATAAAGT TCTGAAACAC    1260
TTGATTTCTC TTCTTTTTCT CAGTATACTT TTCCATATCA GATATATAAA ACAACTGGTT    1320
CTGTTGCAAA GTTTTAAATA AAGAATAAAA TCCCTTACGG TATCTATGAT TTAAGCTGGG    1380
ATTCCCAATA ATACCTTGAT TTCAGTACAG ACCGAAAACC CGAAGAGAGT GCCTTCTTTT    1440
CGGGTTyTCT TATATAATCC TCGAATGGCT TCCATGCCTT TAATCGTGGT AGAGGCAGTG    1500
CGTAAACTTC GATAGAATTT ATTGCGTCTC TTTACTGGAC GATGGTCTTG TTCAATCAAA    1560
TTATTCAGGT ATTTAATGGT ACGATGTTCT GTCCCTTGAT AAAAGCCGTA TTCTTTTAGT    1620
TTCTTAAAGG CACTTGTAAT AGAGGGGGCT TTATCTGTGA cTACAACCTT CGGTTCATCA    1680
AACTGCTTCA CTAACCGCTT AAGAAAAGCA TAGGCTGCTT GTGTGnCCGT TTTTT         1735
```

(2) INFORMATION FOR SEQ ID NO: 526:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 744 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 526:

AAGATTTAAA AGAGTCTGTC AACAGTTTAG AAGATAAGTT TTTAACAGTG GATGAATTAC 60
GAGCTTTTTT AAACTATGGG ATAGTTTATG AAGAGTTACC CATGTCTGTT CTTTTTCATG 120
TTTTGTTTTA TACTGGGTGT CGTGTGAGTG AAGCGCTTGC TTTACAGCCT CAAGaTATAG 180
ATTTTGAACG AAATGAAATT TTGTTTTATA AACAGACAGC AGTAAAAGGA AAAAGTAAAG 240
ATTTTCGAAT AGAAACTACA AAAACTGTTA GTTCTGCAAG AAGAGTACCC GTCACACCAT 300
TAGTAATGGA GAAATTACAA GAATTAATTG ATGTATTAAA TAAAACAAAA AGAAATAGTA 360
ATTTTGTTGT AGATGAAACT TATTTATTTG TTTATCTCGA TCCTGGAAAg CGTGGAGTTC 420
CTTATCGTCG TGAATATGTA AACGATCATG TTAAACGTTG TGTAGAGCGA TGTGGAATTA 480
ACAAAGATTT TCACACCCAT TTGGCTCGTC ATACTATGGC TAGTTTAGTG GCAGAgtAtT 540
GCmGTkGGGA TGkTTwAAAG GATCGtTTAG GTCATACnGA CAGTACAACC TCAAAAATAT 600
ATCGCCATTT AACATCAACT GAAAAGTTAA AGCCTCTGAC TGCTTTTAAT nCTTTAGAAG 660
GTTAAAAAGA TGGTGAAAGT ACTGGTATTA CAGGATTCTT GAAAAACATT TTTTTAATAG 720
GTTAGTCAAA GTGTAGTCAG AGGC 744

(2) INFORMATION FOR SEQ ID NO: 527:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7107 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 527:

TGAAGAAGTT ATATTGGTAG TTTCnCTCTC nGAAAaAAAG TTArGCAAAa CAAtGGATaA 60
aGAGTATCmA GaATGCTAAT CCtTATTAGT TAAAGTTCTA TCTGGAAAAm GCaAAAGTAT 120
TTGAGACACA ACACGCAGCC TATAGAGATT CAAGTGATCT GTGTAAAGAA TTAAAATAAA 180
AAACGAACGT TCATGTTGAA ATTAGAAATT AAATGTTTTA TCATACTAAT AGTATAAATT 240
ATCTGTAATG GATAATTATT TTTTATAGAT TGAATAAAAT GTTCGTTCAT TTTATTGTTT 300
TGTTACTTAA ATAGTTTTGG AATTTTAAGG AGGAGAAAAA ATGTTTTTTA CTGGTAAAGA 360
AAGGAATTTG AAATTTGCTA TTAAGAAGAA ACACGGTGGC GGTGGTGCAG CATCAGTAAT 420
TATCGGATGT ATTATTTTCG GAAGTATTGC GTTTGGTGGG ACAACATTAn CTTACGCTGA 480
TGAAGTGCAT AATAGTATAA ATCAGGATAT ACAAGATTCT GGTAGTACAA TTATTGGAGA 540
AAATGATTCT TCTACCAAAT CAGCTGAGTA TAAAATGATT CATGAAATTG ATGGAACTAA 600
AATTAGTAAC GGTGAAAaTA GTnAaGAAaC AaCTACAAGT TCAGGAACTA TACTGGCTGA 660
AGAAGCAATA GAAAGTTCAA ATCAAAAAAA TTCAAAGACA AGTGAAGTCG AACAGGATCT 720
TCATAAAGAT GTATCAGGAT CTGAATCAGT AAAACAAGTA GAAACTTCTG ATTCTATAAA 780

AAAATCTGAA GAATCAGCTG TTAAAACATT AAATCTGGAT GATTCACAAG AGAATACTAA 840

TTCAATAACT ACCAAGGCAG AAAATGATGC GCTATCTACA GTTAATGATG AAAAAGTATT 900

AAATGAAAGT GATAGTATTA TCAAATCAAT TCCTTCGGAA ACAGAGAATG TCGATAATAA 960

TAGTAAAAGT GATAATAGGG ACTTAGTAGA AAACTCAGTG GAAACTACTA AGGAAATAAT 1020

GAAAGAGGCT GATAAAACTG AAATAGCTAA AAGGAATGAC AATATCTCTG TATTGAATGC 1080

TGGAACTGAA TCATTTAATG ATAGTGATAA AATAAGAAGA TCTAAAAGAG ATAATGAATC 1140

TCCTGTTCCA AGTTTGAGTA TTCAAATGCC AGTCACAAAA GAGATTTACG AAAATGATAG 1200

TACTATTAAA GGAAAGTCTG AACCAGGAGT AACGATAAAA ATTTATAAAA ATGGTATTGA 1260

ACTGGGGTCT ATTAAAGCAG ACGATGATGG AAATTTTGAA TATCCTCTAG AATCACAAGC 1320

TATCAAAAGT GATAACTACT CTTTCGTTGC GGAAAAAGAT GGAGCGAAAA GTATAGAGAG 1380

CTCTGTAAAT GTAGTAGAAA ATAAAGAAAT AGTAGGTAAT CAACGAGATG CGTATACTCC 1440

ATATGGGTTT GGTATATATA ATCGAGGTTT AGTTACTACT GATGATGGAA AGAAACATTA 1500

TACAGTTGAT ATGGAGGGGA ATTTrGCATC TGAAGCGTAC TTTGATCTTG CAGGCGGAAC 1560

AATGTACTAT TCAATTGATG AAAAATTTGC ACCATATGTT GAAAAAATAG TTCTAGATGG 1620

TAAAACAATT ATTTCTAAAA ATCCGTATGA ATTGCCCAAT AAAACCAATA TTTGGTCTTC 1680

TGGTATTCTT AGTACAGATA AAAGAGGAGG ATTGGTTCGT GCTGCTTTAG TGGGTCGTAC 1740

TAACGGAACA ATAGATATTT ATTTTAAAGA TGATACTCCT CAAGAAGTTC TCAATAGTGA 1800

AATACCTTTC CAAGTTTGGG CTAGATTTAA AGAGAAAAAA GAGGGAGTAA AAAATGATCT 1860

AATTTCTGAT TTTGATCTAA ATATTTTACT TCAAAACAAT AAATTAATAT TAAGTAAAGA 1920

TACCTTTTTC AAAGATGAAG ATATTGTCAA TAATAATGGT CCTAAAATTA GTTCTAATTT 1980

TAACTATACC AATAATACTA TTGATATTAA TTATGGAGTA AATACTGGTA ATTATGGCTT 2040

AACGATAGGG AAACGTAGTC CTTATAACTT GCATATTAAT ACAGGGAAAG AATTAAGTGA 2100

TTTAGTTGAT ACGATAAAGA TTGACAATAA AAATTATGAG TTCGAAAAAC TTCCTGATGG 2160

TTCGTTAGTG ATTAAAGATA TATACAAAAA GGGAATTTTA GGTGGAGCAT TACTTAACCG 2220

CCAAGAGATT ACAATCAGTT TAGGATTAAT AGCGAAAAAA AATCTGGGCG ATCTCATTGA 2280

TTTCAAACAA GAATTATTAA AACTTGATGT ATCTATTAGA GATAATACAA CCTTTAAGCT 2340

AGTAAAAGGA GAAGCTACAG AAGCTATCCG TAAAATAGAT TCCTCATTCG ATACATTTAA 2400

ACAGGAATTA ACGAAATGGA TTGAAAAAAG AGATAGTAAT CCACAGGAAC CTATAACTAT 2460

TAATTCTGGA GACAATACTT CTATAGTGGA TCCTAAAGAG ATGGTTGAAT TTGCAGTAAA 2520

AAATCCAACC ACAGAAAATA TTAATAAAGC ACGTAGTTTG GTTTCTAAAA TGTGGCTTTG 2580

GGACAAAAAC AAATATAATC CAGTTCTTAA TGCTGCTGAA GAAGTAAACA GATATCTTTC 2640

AACTACGTTA AATGGTTCAT CAAATGATGA TGGCTTGATA TTAAATAATA CTCATGCGGC 2700

AACTTTCGTT AAATTAATTG ATACGGATAA AGATGGGATA CTTGATAGAT ACGAAGCAAA 2760

CATGGGAATT GGAACCAACT ACACAAATGC TGATTCTGAT GGTGATGGAA AAAGTGACGG 2820

ATTTGAAATT TTTAATGATA CTGATCCTTT GGTGAGTCCG TACGATTGGT TTGATAAAAA 2880

TGGAGAAAAA ATAAAAATAG TAACAACCGA TACTGATACG ATAAGTGGAA GAATAGGAAA 2940

TAACAACTAT GATACTGAAA ATGTCTACCC TAGAACTGTT CAATTAATAA AGGTTACTGA 3000

TAACGGAGAA AATCTGATTT CTGAAATTAG TTCTTCTGAA GATAGAAAAG GGACTTTTGA 3060

ATTTACTGGA TTGAGTGGTA AATTAAGCAA GGGAGATAGA CTCGTTGTTA AAATTATTAC 3120

TAATGAAGTA CAAAGAGAAA TTGATCGACA AAGAAATGTT GTTCAAGTTG GTTATGATAA 3180

TCCAGAGGTA TCTGAAGAAG TAATTGTTCA GGGCGCACAA GTAACAGTCT CTTTTGACTT 3240

TAATTATAAT CAAACTGATC CTGACTTTGA CCGTCAACCA GTTACTCAAG TAGTAGAATT 3300

AGAAAAAGGT TCAAGGCTTT TAAATAGTTA TTTTACTCCA GAACGAAAAG GCTATAAATT 3360

TATTGGATGG AATACTGATA AACTTGGAGT AGGACAAACT GTCAATGATG GTTCTAGTTT 3420

TAATGAAGAT ATAACGGTAT TTGCTCAATG GGAAAAAGAA CCTAACCTAG CTGGAGAAGT 3480

ACATGCTCCA CAAGGTCCTA TTGAAGAAGG TAAAAGTATT GTTCTAAATA ATGTAGTACA 3540

GTCCAATAAA TCAGGGTCTG TAATTGTAAC TGGAGAACTT CCAAAAGGGC TCTTTGGGTT 3600

ATCTATAAAT GAAACAGGTA ATCTAATTGG TACTCCACTT ATCAATGACT GGATTGATGG 3660

AGAAAATAGT CGGGAAGTAA AGATTCCAGT TACAATTAGT AATGGTGACG AAAAGGTTAT 3720

GGTGGAAGTT CCATTAACGA TCCTGCATGA TACTCGAGCG GCAGACGCAG TAAAAGCAGC 3780

AGAAGACGCA GGCAAAGCTG GAGCAGATAA GAAAGCCGAA GTGGAAACCG ACGGTTTAGT 3840

GACTCCAGAG GAAAAAGCGG CAGTGGATGG CTTGAATGAC ACGACTACCG CGAAGAAAGA 3900

AGACGCAAGC AAGTTAGTGG ATGCCTTGCC AGAAGGACCA GTAAAAGATA GCTTGAAAGA 3960

TCGCTTAGAT AAAGTGACAA CCTCTGAAGT AACAGTGAAT GATGCAGATA GCAATGGCAA 4020

AGCGGACGAT GTAGATTTAG CTGAAAAAGC AGCGGCAGAC GCAGTAAAAG CAGCAGAAGA 4080

CGCAGGyAAA GCTGGAGCAG ATAAGAAAGC CGAAGTGGAA ACCGACGGTT TAGTGACTCC 4140

AGAGGAAAAA GCGGCAGTGG ATGGCTTGAA TGACACGACT ACCGCGAAGA AAGAAGACGC 4200

AAGCAAGTTA GTGGATGCCT TGCCAGAAGG ACCAGTAAAA GATAGCTTGA AAGATCGCTT 4260

AGATAAAGTG ACAACCTCTG AAGTAACAGT GAATGATGCA GATAGCAATG GCAAAGCGGA 4320

CGATGTAGAT TTAGCTGAAA AAGCAGCGGC AGACGCAGTA AAAGCAGCAG AAGACGCAGG 4380

CAAAGCTGGA GCAGATAAGA AAGCCGAAGT GGAAACCGAC GGTTTAGTGA CTCCAGAGGA 4440

AAAAGCGGCA GTGGATGGCT TGAATGACAC GACTACCGCG AAGAAAGAAG ACGCAAGCAA 4500

GTTAGTGGAT GCCTTGCCAG AAGGACCAGT AAAAGATAGC TTGAAAGATC GCTTAGATAA 4560

AGTGACAACC TCTGAAGTAA CAGTGAATGA TGCAGATAGC AATGGCAAAG CGGACGATGT 4620

AGATTTAGCT GAAAAAGCAG CGGCAGACGC AGTAAAAGCA GCAGAAGACG CAGGTAAAGC 4680

TGGAGCAGAT AAGAAAGCCG AAGTGGAAAC CGACGGTTTA GTGACTCCAG AGGAAAAAGC 4740

GGCAGTGGAT GGCTTGTTAG AAATAAAACA GTCTTCATTT ATGCCGTTTG AAAATTTATT 4800

TTCGACTACA AATGATTACT CACAGTTTCC TAAAACTGGT GAAAAATCTG ATTCTATTTT 4860

AACCATTTAT GGAGGTTTAT TATTCTTAAG TAGTATAGGA TTATTAGGAA TAAAAAAAAG 4920

AAAAAATAAT ACGAATTAAG TTTGTTTGTA TTCTTCTTTA AGAAAGGATA GGTGTATAAA 4980

TTTTATCAAT AAAAAGCTGA CTATTTGTCA AATAGAGTTG ATAGAATGAT AATAAAAGAT 5040

CATCTAAGGT AGGATTTCTC TTTTCTACTT TAGaTGACTT TTTGTTTAAA TATCAGATAA 5100

ATTTTGATTG GACCTGTAGT TGAATATAGA TCTGTCGTAG TTACAGAAGG TAAATCGCAT 5160

ACTTTTCAAT TCTAGAGGGG AAGTCTTGTT GAGTATTAGC TATTAAGCAT CATAGCCCAG 5220

TTTTGAATTG CCTAACCACT TCGCTTGTAA TTCTTTCATG CGTAAACAAA GTAGTTTTAG 5280

GAGTACATTC TCATTAGTTA AGTTGAGTAG CATTATCTAA TGCAAAAAGG TTTGGCTTTT 5340

AGCTCATAAT AATCTTTACT AAGTCTCCTC CTTTCTTCGA GTATTATTTC ATTTGTGTAT 5400

AAATTGCATC TTTTTCTTCT TCTAGTTTCA TGATTAACAA CTAAAAATTT TTTTAGAAGG 5460

ACACAAGATA ATACATTTAT GCTGTAATAG ATAAAGAGGA TAATTATGTT TAATCAAAAG 5520

GAAGAGAAGC AGTTCAAATA TGCAATCAAA AAGAAAAAGG GGGCGGTAGA GCATTTTTTT 5580

GTATTATCGG AGTAGTATTG TAAGAAAACT AAAGATTAAA AATAGCAAAC GGArATTATT 5640

TCCAkTTtCT TATTTTTGCA AAAAATGAAT GTTCATGtGT ATGAAAGTAA CCAATAAATT 5700

TATCTTTTCy CgAAGGAAGT TTATTTTTAA TAATTTGAGA GGGTGATTAT ATTGTGTTTT 5760

GAAAATAAAG GGTATTTTTT TGAAATCTCA GaCATTGCTG AATTTGTTAA AGTAGAAAGA 5820

AGGAAGCAAC GCGGCTCAGC TACATTAAAT ACTTATGACT ATATTTTGAT TAATATTGTA 5880

AAACAATTGT GTAAGCAGCA AAAAAAATTG ATTTACTTAA TAAACTCAAA GAATTTCAAG 5940

ACTATATTAA GCAATTTAAT TGAAGTGGAT GCCAAAATTT CTGATTATCT TTTTTTCTTG 6000

ATTAATTGTA GATTTGAAAC TGATAAAGAA GTAATAGAGT GTGTAGAAAG TGTAGAAAAA 6060

GATGCTCTTA GAGACTACTA TTGTGGGTTA AATTTCAGCG ATAACATAAA TAACTATAAA 6120

TTAATAGTTC ATACAACTAA ATATTAATTT TTTTTAGAAC CATGTTAAAG GACTATGTGA 6180

TATTTACAAT TCCATCTCGT ATGTGCTAAC CTTTTATCGT TAAAGCTATT TCGTTTATGT 6240

GTAGAAACGT GAGTTTTTGT TTCGTACGTT ACACGAACGA AAATGATTAA AGTCAATAAT 6300

GAAAATGGGA CGAGAATGGG ACAATACTCA GAAAAAAATC TGATTATTGC CCCATTTTTT 6360

TGATTTCAAA GTACACAAAA AGCACTCATT CCTGTAAGAT TAAAGTGACG AAACCCAATC 6420

AGAAAGGAAG CGAGTGCTTT ATTTTTAAAA TTATACCATC AATCAAACAA CTTTACCATT 6480

AGATATGGAA CAATATATTC CTGAAACAGA TGTCGCTTTT GCGGTCAATG ATTTAGTCGA 6540

AGCCATGCCA CAAACACTTT TTCTTAAATT AGAAGAACAA CTAGGATGAC TGGCCTACCA 6600

CCCTAAAATA ATGTTAAAAA TTATTCTATG TGCTTACACG CAGCGTGTTT TTTCAGATCG 6660

TAAAATTGGA TTTCTATTAG ATGACAGCTA TCGTATGCGT TGGCTAGCAA ACCACGAACA 6720

AGTCAGTTAC CGGACCATCA ATCGCTTCCA AAGTCAGGAA ACAACTGCTC ACCTTTTAGC 6780

AGAAGCCTTT GTCTTATTCC GCTGTCAATT AATCACCAAT CAGGTAATTG ACAATGAAGC 6840

GCTATATATT GATGGCACAA AAATCGAAGC GGATGCCAAT AAATTCAGTT TTGTCTGGCG 6900

CAAAGCAACA AATCAGTATG AAGCTTCTTT GGATAAACAA TCAAACGAAT TCTATCAAAC 6960

CCTTTATAAA GAAGAAATCT TGCCTTCTTT AAAAGAAGAA AACCAAAGTG ATAGATTGAC 7020

AAGCAACTAA TTGGAAGAAA TCGTGCATCA CTTGGAGACA GAATTAGTGG CGACAGAAGA 7080

ACAACTACAA AATGAGAAAA GGATAGA 7107

(2) INFORMATION FOR SEQ ID NO: 528:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1533 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 528:

TGAATGGATT TGGCAATATG CTAAAAAAAA TAAGAATAAG TTACTGCTAG CTGTTATTTT 60

TGTCGTTATT AACGCTATTT TAATTGTTAT TATTCCATTG TTATCTGGCA GAATTGTAGA 120

CCAAGTAATC AATCAGGGGA AAACTGAAGA GCTTGTCCCA ATTCTTGGGA TTATGATTGG 180

AGTGACCGTT ATTCGGACAG TTGTTCGTTA TAGTTATCAA GTCTTGTTTG AATCGGTTGG 240

ACAATCTGCG TTATTTGACT TAAGAAAAGA TATGTATGTC AAACTTCAAG AATTAGATTT 300

TGATTTCTTT AATCATACAC GTGTTGGCGA CATTATGGCG CGAATGACAG GGGATACAGA 360

CGCTATCCGC CATTTTGTwT CTTGGGTCAC TTATAACCTC TTAGAATGTA TTTnATGGkT 420

CATCGCAGCC GTGATTGTGA TGGGGACGAT CAATTGGCAA TTGATGCTAG CATtAATTTT 480

AATTACACCT TTGATTTTTC TGTTAACACA AAAAATGTCG AAGGAATCGC ATCCGGTATT 540

TTTTGAGATC CGGGAAAGTT TTTCTCGGCT AAATTCGATG GTCGAAGAAA ATATTGGTGG 600

AAATCGGGTG GTTAAAGCTT TTTCTCAAGA ACCTTATGAA ATTGAAAAGT TTAACAAGCA 660

CAATGAAGAC TATAAAAAAT GGAATATGGC TTCTGCAGAT GTCTCAAAAA AATATTTACC 720

GTGGTTAGAC GGTTTTGCAG GTAGTTTAAA TGTTATTGCC TTAGTTCTTG GTGGCCTCTT 780

TGTAATTCAA GGTCGGATGA CAATTGGTGA TTTAGTGGCC TTTAACGGTT TTCTCTGGAT 840

GTTGAACATG CCGATGCGAA TGAGTGGCTG GTTGATCAAT GATGTGCAAC GCTTTACGGC 900

ATCGTCCATT AAAATTCGTC AGTTATTAGC AACGAAACCG AAGATTCCCA TTGCTCGGGA 960

AAAAGAAGCA GAACCAATCC AAGGCTATGT GACGTTTGAT CACGTGTCCT TTCATTTTTC 1020

AGATGATCCA GAAACCCCCA TTTTATCGGA TGTTTCATTT TCAGCAAAAC CTGGTGAAAC 1080

AATTGGTATT TTAGGGGAAA CTGGTTCAGG CAAAACCACT TTAGTTAACT TAATCGCTCG 1140

GTTCTATGAT CCAACAAAAG GCACTGTTTA TATTGATGGA AAAGATGCTA AAAGCTATTC 1200

TGTTCGTAAA TTGCGAGAAA ATATCTCCAT GGTTATGCAA GATGTCTTTC TTTTTTCCAA 1260

TACAATAGAA GATAACATTG CTTTTGGTAA CACATTGGCT AACTTGGAAC AAGTGCAAAT 1320

GATGGCTGAA ATTGCTGATG CtCATAGCTT TATTTCCAGA ATGCCCGAAG GTTATGCAAC 1380

AATTGTTGGT GAACGCGGCG TTGGCCTATC TGGTGGCCAA AAACAGCGGA TTTCACTGGC 1440

CCGTGCCTTA ACCAAAGATC CAGCTATTTT AATTTTAGAT GACACAACTT CCGCTGTTGA 1500

TATGGAAACA GAATCAAAAA TTCAAAAAGA ACT 1533

(2) INFORMATION FOR SEQ ID NO: 529:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 895 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 529:

TTCTAGGCGG CGTACCCTTT TCCGGCTAAA AACGGTCCGG AAACTCTTGG AAACCCCTTG 60

GTTTTAACTT GGGTCGGAAA TACTTGGGTC nTAGTTTCCT CCAAACCGTC AATCCCCTGG 120

TTCAAAATTG GATTTAAATA ACACATCCAT ACGGAGTTTT TTCTTCTTTT TGCCCAATAA 180

GGAATGAACT CTTTTTGGgA ACGCAACGAC CTTCGTTAAG GTTTCTTCTA ACAAGCCGAA 240

ATCTTGTGCA GCGCGTGCTT TTAGCCAATC TGTATGTGCT TGGTTTAATA CTTTAACATA 300

TTCTTGCATA CGTTCAGCTG GAACATTTTT ATTCAAAGCA TATTCTTCTT TTACTTTTTC 360

AAAAACCAGT TTCCCTAATT CAGATAATTC TTCAGGGCGT GTCTCAAAAT AGGCCAATGC 420

TTCTTGAATA ACTGGGCCAA TACTTCGTTC AAAATATAAT CCGGTTAAAT ACCCTTCTAC 480

TTCTCCTCGA AAAGGACTAC TTTTTTCTGG CATCCCAGTT GATGAATCCC AGTCTAATAA 540

TGTCAAGGCA TTTTTTAGTA ATTGAATTTC TTTCACTTCT TGTAAAAATA CAGCTTCCTT 600

CATTTTTTTA CACCAGCTTT CTTAGTCTTC CCGTGGTTTT CTTGGTGGAC GTGTTCCCCA 660

GTATTGGAAT AAATCTGTTC GTAAGGCTCC GTTGTAAAGC TTcCGTTTTT TAGTAGCTTk 720

GGCACCATAA TATTCTTCAA AAGCAAGATC ACTCGTCAGT ATATACTTGc TCCAAGTAGT 780

TAACGGACGG AAAACATGGn CCATTCTTTA TACAACGGGG AACACTTCCT CTTCTCCTAA 840

ACGCTCACCA TAAGCGGGTA GCTACCATAC ACCATATCCT CTCGGTGGAA AACTT 895

(2) INFORMATION FOR SEQ ID NO: 530:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 532 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 530:

TATGAATTTT GnATAAAAAT GGTGCCATGT ACAAGAATGC TTGGACAAAC AGCGATAAAT 60

ATTCATATTA TAGTACATCA GATGGCTCAT TAGCGGTTGG ACATCAACAA ATTGGTGAAA 120

ATAAATATAT TTTCGATAAA AATGGTGCAC TATATAAAAA TATGTGGGTA GATTTAGACG 180

GAAATCGTTA CTACAGTAAA GAAGATGGTG CTCTCGCTAT TGGAGAACAA ATCATTGATG 240

GAATCACTTA CATGTTTGAT GAAAGTGGGA AACTAGTTAA TAATTAATTT TTGAGAATTA 300

GAAAATTTAT TTTAGAATAG AGATGTAAAT CTGTTGCAAA GTTTAAGAAG ATATTCGGAT 360

TGAGATAGAT AACACAGCrA TrAAAACATA CCAAGATTaT GAATTAATCT CGGTATGTTT 420

TTATTTTTTA AAACCAAAAC AAACGTTTTT GTGTGGTATA AACCATATCA ATACTTCGTT 480

TGCCTAAAAT CTCGCCAtCT GCGTGAATGG TTTGTGGAAT CGTCGAAACA AT 532

(2) INFORMATION FOR SEQ ID NO: 531:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1129 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 531:

AGTATCCAGT TTAATGCTTT GTCAATCGAA ATACAAGAAA GTGGnACAGC CGTTTGACGA 60

ATGTCCCACT TTTCTGTTTT ACTATTAACA GGTTTGTCTC AATGTTTGTG TAAAATGCTT 120

GGATGGTGTT ATAAAAGATT CCTGCGAACA TAAAGATTAA CCAACTAAAC CGCCAtCCCC 180

TAAAACAAAA CTmATAAATA GATAGAGTgC CACAACTAAG GGCCAATAAA TGGATGCAAT 240

AATTGCACCG ACAGGATTTC CCACAGAACC TTCCTCTGTT TTTGTTTGCT TAGGGGCATA 300

CTCAGGATGA CCAACATGTT CTGAATAAAA ATAAGTTTCA TTCAACAATT TATTAAAACT 360

TCCTTTAATT ACTCCACCAA ATATAAACAA GAAGACACCA ACCGATACTA AAAAGAAGAG 420

CCAACTAATC CCCATCGTTC CATTATAATA TTGTCCGCCC ATCATTTGAT GCATGGAACG 480

GCGCATCATC AGATGAAGGC CACTACTAAA GACAACAACT GGGATTAAAC CCAAAATACA 540

AAAAACTACG CCCGCTGCCA TACAAAAAAT AAATGATCGT TGAAAACTTT GTTCTCTTTG 600

TGCAACTTCT TTTTTTAACT TAATCGGAAT CAATCGTTGC CCCAGCTGTT TGTCCATTGC 660

AACAAACCGA TTACCAGCAA CAATAAATAA TGGAACGCCG ATTGCCACAC CTATGAAAAG 720

ACAAACTAAA CCAAGACCGT TAGAAAGAAC ACCACGCATA CCAACCAGCA AGCCAACACC 780

CACCATGATT GCAACAACAC CAAGACCAAT TTGAATCGCA CCTTTACGTT GGACAGCTAA 840

ATAGTCAAGT CCTTCCTCCA CTGrAATTTC TGGCAAACCG TTCATTGAAT CGTTTGGTTC 900

ATCCACTTGT TTATTTCTA AATTCAATTC TTCCATTAAT TCATCAATAG TACCAAATTC 960

AGArATAATT GTACCArTTG ckTCATTTtC CGATTTTCCT TGGCTGAGCA AGTCTTCATA 1020

ACGATCCTCA GCATTTGCTA ATrAGTCCGm CTTTAATTGT TTTGTTTCAG ATGTTTCAAC 1080

AACGCCTGAA AATAAGCTAT CAATAAATTC TTCTATAATT TTCATTTTC 1129

(2) INFORMATION FOR SEQ ID NO: 532:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1283 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 532:

```
AACCCCCGAT AATTACTCTT TACTCGCAAT TAAGCGAAAA ATTGATCGTT CATTTTTACC      60

TGATCATATT TCTTCCGTCT TATTAGATAC AGAGACGTTT GGTAGCAAGT TAGCCTTATT     120

ATTGCTTAAC AAACGCAAAG AAAAAACAGC CTTGATTAAT GAGGCTGAAA AATTGGTGCT     180

AGATCACAAA AATACATTAG GCATGTCTAG TGTTAATCCA CATAGCAAAA TGaTTGTAGT     240

GGGTAGTTTA AACGTTGATA ATTAtCTGTA TTCAACAAAT TTACCACATa ACGGCaAAAC     300

AAATTTTTTG tCTAGCTATG CTAAATTCCC CGGCGGAAAA GGATTAAATC AAGCTGTTGG     360

CTTAACCAAA CTGGGCCATC AGGCCACCTT AATTGGTTGT TTAGGCTCAG ACACAGATGC     420

GAATTACTTA tATAAAGAAT TAGAAAAATA CCACGTTACG ACAGATGGTA TCACACGTAT     480

TCAAGATACT GAGACTGGTC AAGCTTATAT TTATGTTGAG ACAAGCGGCG ACTCCATGAT     540

TTCAATCTTA CCTGGTGCAA ACACAGCGCT TACGCCTAAA AAAATCGCAC AGCAAAAACA     600

CCTATTTATG GATGCCAGTT TTTGTCTCAT TCAAACAGAA ATTCCTTTGT CTGCTGTAGA     660

GAAAGCTTGT GAAATCGCCC AACATTCAGG GGTACCGATT ATTTTAAAAC CAGCTGCTAT     720

TCATCATATT CCAGTGAATA TACTAGAAAA GGTTGATTTT TTCATTCCCA ATGAAGATGA     780

ATTATTGGAA CTTCAACCAG ATACTGGTAC ATTAGAAGAG AAAGCGGCGT ACTTCTTAGA     840

AATGGGCGTA AAAAATGTCA TTGTTACTTT AGGAAAAAAA GGCGTGTTAT TAAAAACACC     900

CCAAGTGTGC CACTATTTCC CTGCAACAGA AAATATAGCT GTGGATAGCA CTGGTGCCAG     960

TGACAGCTTT ATCAGCGCGC TCGCCTCCTA TCTTTCGAAA GgTTATCCGA CTGAAGCAGC    1020

CATTCAAATA GCAATTCAGG CTGCAGGATT TTCAGTTTCT AAAGAAGGGG TGATTGATTC    1080

TTTAGTTGAT CACGTCACTT TGGAAAATTA TTTGATAAAA AAAGAACCCG CTCTTTTTGC    1140

TCATCGTAAC ACCTGTATCG ACTGACACAG GnGTTTTCCT GCTGAATATC TTGTnATTTT    1200

CTCATnAACC AATGAGAATT GTGCTTTTTA TTACAATTTT ATTAGGCGAA CTATTATAAT    1260

AAAAGAGAAG TTATAACTAA GGG                                            1283
```

(2) INFORMATION FOR SEQ ID NO: 533:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6470 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 533:

```
TTTTCGGGAT TCCTAGAAAA GCTTGGACTG TAACTTTAAC TTGGTATTAA CGCGGGACTT      60
```

CTTTTTATTC ACTGAAATCC ACGTCGTTTC AATTCTTGGG TTACCCTTCT ATAGCCATAA 120

TGTTGATGTT TTTTACGAAT CGCTTGAATT TCTTCTTCAA TGATTTGTGT TGAAGCTGTT 180

CTGTCGAGGC GTTTTTGCCA GTACATGTAA GTAGATTTTG GAAACCTTAG GGTTTCCAAA 240

ATATCTTTTA ATCGAAAGTG TCTTCGGAGA CTGTGAATGA TTCGTGCGAT TCGTTCATTT 300

TTTGTTCGTC CTCTAAACGC AGTCTCCTCA ATTCTTTTAA ATAGGCATTC TCAATTTTTA 360

ACATTCTATT TTCTTTTTCT ArTTTyTCTA CTTGAGATmG TTCACTTrtA rTGkTTkGGG 420

TTCATTTCGT TTCTTyTTAG ACATGGTAGG TGGACACCCT TTCTATTTAG AGAGTCCCTT 480

ACGCTAGACT TCCGATACGT TCTAAGCCAA TTCGCAATCA ATGATGGATT GTTCGTTTTT 540

AAAAGATTAG CAACTTCTTA ATAGGACAAC TTTATCGCAT CTAGCTTAAA TTGAACAGAA 600

TAATTCTTAC TTTTTTGACT ACGAAGTAGT CCTTTTTCAC CAAGTCTTTA TAGACATTAA 660

TCCATTTGTT TATTTGAGAA CAAATTTAAG TTTAAATTCA AAATTTTATT TTGCCATACA 720

AAAAAACTCA AAAGTTAGAT TTCTAGGTCT AACTTTGGGG GAGCACTACA GAACTGTAAT 780

TCTCTTAAGC GCATTTAATA TAGATTCAAT CTTTTGTAAA AATCCCTATA TTTTGTTTCT 840

TGAAAAATTT AATCATTTTA ATGAACTAGG GATGATCGAT TTCGAGAACA TAAAAAACAT 900

ATTAATACAG GATTGCTTTA TCCTTGATAC GTTTTAAGAG TACTCCGGTT TGTTGTATCT 960

TTTATTACTC CAAACGATGA TCCCGCAAAC GATAGCCCAC TCCAATCTCA GTTATCAAAT 1020

ATTCTGGTTT CAATGTATTA CTTTCGATCT TTTTACGAAT ATTAGACATA TTCACCCGCA 1080

ATGTCTGCGA GTCATTGCTG TAGGGGCCCC ATACTTTTGT CATCAAGGTT TGATACGCAA 1140

CTACCTTTCC CAACTGACGA AACATCAGCG CTAAAATTCG ATATTCGTTT TTTGTAAGAT 1200

GGATTTCTTG TCCCTTTTTA TAAACTAACT GTTTTTCGAA ATCAATTTTC AAGTCTTTGT 1260

TTTCCACGAC AGTTAGTGAT TCCGTTTGAA CGTTTTGATG GCGCAAAGCT GTACGAATCC 1320

TCGCAAGTAA TTCATTCGTT CCAAATGGTT TAGTTACATA ATCATCCGCA CCTAAATCCA 1380

ACGCTTTAAC TTTCTCCTCT TCATTATTTC TAGCAGAAAT AATGATGATG GGCAAATGCA 1440

TTCGCTTGCG AAGGATTTTT AACAAATCAA TCCCATCAAT GTCTGGTAAT CCTAAGTCTA 1500

ACAAAACTAA ATCAAACGAT GTTGTTTGAA AAGCAGTCAA TGCCTCCATT CCAGTACTAG 1560

CAATCGTTAA CTGATAGCGT TCCTGTCCCA AAACAACTTC CATAAAATCC GTAATATTTG 1620

GATCATCTTC AATCAATAAA ATGGATTGCT TCATAATTCT TCCCCTCACT TCAAATAAAT 1680

TCGAACCAAT GTTTTGCCCC TACCTATCGT CATTTCCATT TTACCATTGT GCGCATGGAC 1740

GATTGTTTTT ACAATACTCA ATCCGATTCC TAAACCATTT TTTGAGTCAA CCGGTACTTC 1800

ATTTGTGCCG GACAAGTTTG ATTGGATCTT TtGAAATTGT TTTAATGGAA TTTCGCCGTG 1860

ATTTTCAATT TCAAAAACTG TTTGTTCTTT TTCCTGATAG ACGTTTAGTT TGACTGGTAG 1920

ATCATTTTCT CCATGACGAA ACGCATTTTC TATCAGATTA AACAACGCTT GTTCAATCAA 1980

AATGGGGTCC GCTTGAATAA AAATGACCTC TTCTGGCAGG TGAATATCCA CTTGGCCGTC 2040

TGGATAGACC TTCCTCAAAT GCTTATAAAC AGCCTCGATG ACCTCTTCTA CTGGTTCAGC 2100

AGTTTTATTA ACCTTCATAG TATCCATGTT GATTCGCGTG ATGGATAAAA GATTTTCAAC 2160

CATCCGAATC AGCCATTGTG ACTCTTCTTG AATATCCTTT AATAATTTTC GCTGTGTCTC 2220

TTCTTTTAAA TCATTTCCAA TGCCTAATGT TTCAGCAATC CCTGAGATAA CAGTTAATGG 2280

TGTTCTCAGA TCATGAGATA CGGCTCGTAA TAAATTGCTG CGGACTTTTT CCCGTTCATT 2340

TTCTAACTCA ACTTGCTCTT TCTCATCTTT CAATTCTGTT TGTTCCAAGA TGACCGCAAT 2400

TTGAGTCAGC ACCAATTTTA AGTAATTGAG CTGATCATTT TCCAAATCAA GATCTGCATT 2460

TCGTTCAATG CCTAATACCG CCAACGTCTT CCTGCCTGCA GCAATCGGCA AATAGAATCC 2520

TTTAGCTCCA TTTAGTGTGT CTGTACCATT CCCAGCTTCT TTTTGGTTCT TTGCTGCCCA 2580

AAAGGCTACA GCTGCTTCGT CTTCATTATT TAATATTGAT TTTTTTTCGT TGATACAATG 2640

AACACTCTCT GTCTTCACTT GACGATCAAA AATAATCACT TCTCGTTCCA ATAATCTTGA 2700

TAAATAAGTG GCAGAAATAT CAAGAATTTG TTTTCGACTT TCCACTAAAA CATATCGTTT 2760

GTTCAATTCG TACAAAATTT CCATTTGGTG CTCTTTTTTC ATAGAAGTAT CTGCTTGTTT 2820

CTTTAATCGA ATCATTAAAT TACTGCTCAT CAGGGCAACA ACTAGCATGA GCAATAAAGT 2880

AAAAGGATAT CCCTGTTTAT ATACTGTCAG AGAATATAAT GGTTCGACAA AAAACCAGTT 2940

AAAAGACAAA ACACTTAAAA TAGAGGATAA TGAGCTCCAA AAATAACCGG ATGTCGTTCG 3000

AGCGACTAAT AATACGAATG AAATATAAAT AAGCATCAGA TTTTGATCGC CAACGTGGAT 3060

ATATTGCATC AATTCCGTCA CGACCGTCGC TAGAAAAACA CCACCAATTG CAATCACTAA 3120

ATCTTTTCCG CCACCCTCAA TGACTTTACG AGTCTTAAAA AATAAAGACC GCTTTTCTTC 3180

ATTGAAAGGA ATTAAATGAA GTTCCGTATC TTTCAATCGT TTTAACAAGC GATCATCAAA 3240

CGCTTCAATA AAGATTTTTT CATACCACGG CTGCCTTAGA TTTTTCCCCA TAATAATATC 3300

AGTAACACCC GTCATTTTTG CGAACTCGAC AATCGTCTCA AAACTATCAT CCTCTTCAAT 3360

ACTAATAACC TCCGCACCTA ATTTATCGGC TAAAGGAATG TTCGTTGGTG TATTTTCTTG 3420

CGTACGAACT TGAATCACTG TCCAATCGAC AACTAGACCT TGCGCTAAAC GTGCTGTCCA 3480

GCGAATACAT TTTTCGGTCA TTTTTGGAAA GGCATCGTTC ACCACCGTCA ATAATTTACT 3540

TTGTATCCCA ATTGTCTTTC CACTAATTCG ATTAATATGA TCCGACGCTC GCTGAATTGC 3600

TAATCCACGT AATTGGTCCA ATTTTTGTGG AATAAAAAAA TTTTTCAAGG CACGTTTCGC 3660

GTTTTCATTC GCATATATTT TTCCTTGTTC AAGTCGTTCG ATCAGCTCAT CGGGCTCCAC 3720

ATCAATCACT CTTATTGTCG CTTGCCGTAG AAAAGTGTCC GGCACCGTTT CCTTGACCTC 3780

AATTCCGGTA ACTTCTTCAA CGATATCATT CAAACTTTCG ATATGTTGAA CATTGACCGT 3840

AGTAAAAACA TCGATTCCTG CATTCAACAG TTCATCAATA TCCTGGTATC GTTTACGATT 3900

TCGAGAACCT TCCGCATTGG AATGAGCCAG TTCATCAATC AACACGATTT CGGGCTTCTG 3960

TTGAATGATT TGATCAATAT CCGGCTCTGT TAAAACCATT TGTTTATAAA AAATATTCTT 4020

AGGAGGAATT TGCGGTAATC CTTCTAATAA TCGATTGGTA TCTGGACGAT CATGAGGTTC 4080

GATATAGCCG ACAACAACAT GTTTTCCCAT CATCAATAAT TCATGAGCCT CTGTCAACAT 4140

ACCGTAGnTT TTCCCAACAC CAGCCGCAAA TCCAAAAAAT ATCCGCAATC GACCTCGGCG 4200

TCCATCGCCA GATTGCATTT CTTCTGTTGT CTGCATCAAC AGTTCTCCTC TCTTTGTATG 4260

CACAACAATT CCCTAACCAT TTTTGGTTAG AGAATTGTTC TATATTATAT TCCTAGTTTA 4320

TCTAATGCTA GGTTTAACTG CAAAACATTG ACATAAAATC GGTCTGAAAA CCAATCTTTT 4380

TGGCTATGTT TATCGATGAG TTTACGAATT GATTGTTGAG ACAAGCCACG TTCTTTAGAA 4440

ATTCGGGCAA CTTGAAAATG AGCTGCTGCC AAACTAATAT CCGGATCCAC ACCACTAGCA 4500

GATGCTGTCA CGAGATCATT CGGAACTTTT TTCTCTGCTG GATTTTTTTC CAATTCTGCA 4560

ACGGTGCGCT TCTCAACTAA TTGCTTTTGC TCCGCCGAAA CAGGTGATAG CTGACTAACT 4620

TCCTCACTAC GTCCTGAAAA ATATTTTGCC TGTTCAAAAG TTTGCCCAAT CAACTTTGAA 4680

CCAACAACTT GATTATTCAC TGAAACCTTA CTGCCGTTAG CCTGATCTGA AAAGAACAAT 4740

TGACCAATGC CAGTAACTGC AACGGTATAT AGCCCACCAA AAACGATCAG GCTAAAAATC 4800

AAAAAACGCA AACTGCCTAG AATACTTTTT TTCATTGTTT CTACTCCTTT ATAAAATGAT 4860

CAGCGAGAGA ATCAGATCTA TAATCTTGAT AAAAATAAAT GGTGCAATAA TTCCCCCAAG 4920

TCCATAAATC AGCAGATTAT GACTCAAAAT TTGACTGGCC GGTTTTTCTT GATACCTCAC 4980

ACCTTTTAAT GCTAGTGGAA TCAAGGCCAC GATAACGACT GCATTATAAA TGACAGCGGA 5040

TAGAATCGCA GTCAGTGGAC TACCTAATCC CATAATATTC AATCGATCCA GTTGCGGATA 5100

AATACTGTAA AATAGTACTG GAATCACCGC AAAATACTTA GCGATATCAT TTGCGATACT 5160

AAACGTAGTT AACGCGCCGC GAGTCATCAA CAATTGTTTC CCAATCTGTA CGACTTGCAG 5220

CAATTTTGTT GGACTGGAAT CCAAATCAAT CATGTTCCCC GCTTCCTTCG CGGCTTGTGT 5280

TCCTGTATTC ATCGCCATCG CAACGTCGGC TTGCGCTAAT GCTGGTGCAT CGTTTGTCCC 5340

ATCACCAGTC ATTGCAACTA AGTGACCCTT TTCTTGATAC TCGCGAATCA GATTCATTTT 5400

ATTTTCTGGG GTAGCTTCTG CCAAGAAATC ATCTACTCCA GCTTCCGCTG CTATCGCTGC 5460

CGCTGTTAAG GGGTTATCCC CAGTGATCAT GATGGTCTTG ATACCCATCT TACGCATATC 5520

AGCAAATTTT TCCTTTACAC CATTCTTAAC GATGTCTTTC AAATAGACGA CACCCATCAC 5580

ACGATTATTT TTTATAACTA CTAGAGGTGT TCCACCTGCT CGAGCGATTT TATCAACAAT 5640

CTTGTCACAT TCAGATGGAT AGTCCTCTCC TTTAGATTGT ACATATTTTT TCATGGTGTC 5700

TGCGGCACCT TTGCGAATCA CGTCACCGCG ATAATCAATT CCGCTCATGC GTGTTTTGGC 5760

ACTGAAATCG ATAAATTTAA CCTCTGATTG CTGGAATTCT CTTTCCCGCA AGTTGAAACG 5820

TTCTTTTGCT AAAATCACGA TACTTCGACC TTCTGCTGTC TCATCAGCTA AAGAAGATAA 5880

TTGCGCCGCA TCCGCTAACT GTTCTTCACT AACACCATGG ACTGGCAAAA AATCGCTTGC 5940

TCGTCGATTT CCCAATGTTA TGGTTCCTGT TTTGTCGAGT AAAAGAACAT CCACGTCACC 6000

AGCTGCTTCG ATCGCACGAC CACTCATGGC AATGACGTTT TCCTTCGTCA AGCGGCTCAT 6060

ACCTGCAATC CCGATCGAAG AGATTAAGGC CCCAATCGTC GTTGGCGCCA GACAGATTAG 6120

TAAAGCAATC ACAATAACCA GTGATAAAGC TTCTCCTTTA CCAGATAACT GACTGCTGAA 6180

ATCAGTAAAT GGGACTAAAG TAATTGATAC AGTTAAGAAA ATAATGGTCA AAGTAATCAA 6240

GAAAATCTGC AAACCAATCT CATTCGGAGT CTTTTTTCGT TGAGTTCCCT CTACCATGGC 6300

AATCATTTTA TCTAAGAAGG ATTGACCGTT TTCCGAGGTA ACACGAATCA CTAAATAATC 6360

GGAAACGACG GTTGTTCCGC CAGTAACCGC ACTGCGATCC CCACCAGATT CACGAATTAC 6420

TGGCGCTGAT TCGCCAGTAA TGGCACTTTC ATCTACTGAT GCAGCCCCCT 6470

(2) INFORMATION FOR SEQ ID NO: 534:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1685 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 534:

CAATTAATTT GGAGATTTCA GCCATAGTAA TGCCTTTTTG AAATAATTGG ATAATGGTCG 60

TTTCTAATGA ATCATTGGTT CGCTTATATG CTGGTAGTGT TTGTTGAGAA AGTTCGCCAT 120

TCCGATCTCT AGGAATAGTT AAATTCAACT CGCCATATTC TGTTTTAAAT GTACGTGAAT 180

AATTCCCGTT ACGGGAATTT CCTGAATTGA ATCCATtTCA ATCATATTTT TCGTAATCTA 240

AAAAAGCAGT TaGCTCTGCT TGTAGAAGTG AATTAATCGC TCGCTCTAAA TGATCCTGAA 300

ATAATTCATC TAAATCCCCT TTATTGAGTA GTGTTTGCAT AATTTCTATA GTAAATTTAG 360

TCATGAGAAA GTCCTCCTAT AAAATTTCTG TGTGGTAACT TTAATTTTAC AGAATGGACT 420

TTTTTTTCTA CCCTAAATTT CTATTTACAC AAAATATTTT ACGCTATCTT TTATTTTAAA 480

ATATGGGATA TTATGATAAA CTAACTATGG ATGATATTTG AAAAAACCAT ACTGTATGTT 540

TGATTTAATT TTTATTTTTA AGCTCATATC GTAAGAATGG GACTTGTAGT CTCACTTCTT 600

TATGGAGGTG ATAATTATGA ACGTTGGTTT TATTGGAGCC GGCAAGGTAG GTTGTTCATT 660

CGGGAAGTAT TTCCAGGAAC ATAAAATACA AGTAACAGGA TTTTATAGTA AAAGCGAAGA 720

CTCGTCATTA GCAGCTTCGA ATTTTACATC TAGTAAACAA TATTTAAACT TAAGAGAATT 780

AGTAGATGAA AATGATACTA TTTTCATTAC GACTCCTGAT GGTCAAATTC AGGAGGTATG 840

GCAAGAAATT AAAAATTATC AAATCAAAAA CAAATTAATC TGCCATTGCA GTGGTGCAAT 900

ATCTTCAGAT ATCTTTTCAA ATATACAGGA CTATGGTGCT TATGGCTATT CTGTCCATCC 960

TATGTTTTCA ATATCAGATA AATATAATTC ATACAAAAAA TTAAAAAAAT CTTTTATAAC 1020

AATTGAAGGC GATGAAAAAT ATGCAAAGTT TTTATGTGAT TTCTTCAAAA AGTTAGGTAA 1080

TAGTACTGTA ATTGTATCTA AGGAGAATAA ATCTCTTTAT CATGCCGCAT CAGTAGTATC 1140

TAGTAACCTT GTGCTAGGAC TTATTAATAC AAGCGTTACA TATCTAATTC AATGTGGATT 1200

TACTGAAAAA ATGGCAATTG AAGCTTTATA TCCATTAATT GAATTTAATA TTAGAAATAT 1260

AAAAGAAAAA GGAATTATAG AAAGTTTGAC AGGACCAGTT GAACGTTGTG ATATTTCAAC 1320

AATTAAAGGT CACTGTGAAG TTTTATCAGA TGAAGACAGA GCTTTATATA TGCTATTATC 1380

AAAAAATGTA TTAGAGATAG CTATGTkrAA AAATATTAAT AGAGATTATA GTGAATTAGA 1440

AAAGTATTTA GGAGAGAGTT AGGAGAGAGA TAGATGAAAA ATACAGCAGT TACATTTAAA 1500

GAATCTAAAT TAAGAAATGA AAAATTAACA ATGCTTACAG CTTATGACTA TTCAACAGCA 1560

AAAATAATTG ATGAAGCTGG GaTAAATGGA ATATTAGTAG GaGATTCTCT TGGTATGGTA 1620

TGTCCAGGAC ATGAAGGATA nCACTTTCAG TAACCATGGG AAGCCTGATT CCACCATACT 1680

AGAGC 1685

(2) INFORMATION FOR SEQ ID NO: 535:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1566 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 535:

GGGAAAAtAA akGyCCCCTT AtcCCTTTkG GTAtTCCsCc AgtGATAAAC yCCtGTTAAG 60

aTAGTGGCtA TATAAaAAAG GcmGrTaCTT AAGaAGaGTA TATTCaTGaT TyCCCyCCTA 120

GTTTAAAtGa AAGTCaTGAA TGaTTGTACC aAGATTTTTT AGAGAAGTTA AGATTGAATA 180

GAAAAGTAAA TACGATAAAA TTTTTATAGA CTGTTACGTT GAAATGTTCT TCAGAGACTA 240

TTTCTTCCAA TACCGCCTAC TAGATAGGTG GTTTTAATCA TTATTTTCGG AGGTTGTGTT 300

TTGGCCAAAA AGTAAAGATT TAATGGATGT AGGTTCATTT GTGAGCCCAT CTAGATAGAA 360

CGTTTCAGGA CTTTAGTGAG GCATGAAACG ATTCGATGCA AGCATTATCT GAAGGtGTCC 420

CTTTTCGGGA CCTACTCATG GTAATGCCTT TTTTTATTTG GTTCTGGTAG TCAAAAGAAG 480

TATAAACAGA GCCCTGATCA CTATGCAATA GACAATCCGT TGTTTGAGGC AATTGATCAA 540

ATGTATCCAG AACAAAATCC GTATCTTGCT TATCACCGAT GGTATAGGCG ATAAATTCAC 600

CATTGTATAA ATCCATAATT GAAGACAAAT ACAATTGCTT TTTACCAAAA GGCAGATAAG 660

TAATATCTGT TACCAATTTC TTTAAAGGAG CGTCTGAAGA AAATTACTTA TTTAGCTTGT 720

TCTCGACAAC CGTGTAAGGT TGTCCCACTT TACTGCGTTT CTTCATTTTC ACGCGACAAT 780

TCTAATGATA TTTCCGCATA ATCTTTTGTA CGGTGTTTTT ATTTACTTTG TATTCCTGAT 840

TAATCAACGC AtAATCTTTC GATAGCCATA TCGAAACTTA TGCTGGAAGC AAAGTCGACG 900

AACAGCTTCt TCAAGAGGCG TTAACTTTCC CAAGTCTTTT CGATGAGTCC AACGATAATA 960

AGTAGCTCGC GAAATGCCGA AAAGTTGACA CAATGCAGTA ATGGGTAATT TTCCTTTCAT 1020

TCTTTTAACT AATTCTACAA ATACTTCTGG CTCCACGTTC TTTCCAATTC TTTGAACTTT 1080

TTAAGCACCT CAATGTGCAT TTTTAAATAG TTATTTTCAT TTTTAAGCTG CTGTTCATTT 1140

GATAGATTTT CTGGACCTTT TCCATAAGCA TATTTTTTTC AACGGGTTGA ATCAGACGAT 1200

GATGATTCTC GCCATTACGA TACCCCAACG CCACCATGTG TCAATTTGTG TTTTGTTTTT 1260

AATTCCTAAT TCTTCCATAA TCGTTTTTAC AGGAATATTC TCTTTTTTCA TCTGTATAGC 1320

CATTTCTTTG ATGTAAATAG GATAAGATAC TCAATTTCTC ATTGAAAAAG CACCTCCTAG 1380

TTAAATTCAT TTTAACGTGA ATTCAACAAG AGGTGCTTTT yTTGTCTCAG TTTATTGGGT 1440

CAGTTCCCTA TTTGTTTGGT CACTTTTATT TTAGCAAAGG TGTTCTTATT TTTATACAAA 1500

AAATAATCCA GATGTCAATy AATCTTATTA ATGGTATCGG AATTTCAAGC nCTGTTGnAC 1560

AAAnTT 1566

(2) INFORMATION FOR SEQ ID NO: 536:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1378 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 536:

CAATAAAGGT TTCGCTTAAC GTCCCAAATA AAACTGATAA AAtmAGAATC ATCcAGAAGA 60

TTAAATAAGC CAGAAAAATA AACGGTTAGA TAATACTTTG TCCATTGTGT TTTTCCTGAT 120

TGATCTTCCT GATCCATATC CATAATATTT TTTAGAATCC AATCGATTAG CATATCGCCC 180

TCCTcTTCAA GTATTGCACT AAATTATAAC GAATATAGCG ACTAATAGAT AGATAATTCC 240

TTAATTAGTC ATTATCCCTC CCTAAGTAAG AAATAGTGCC TTGACCTGAG TATAATATCT 300

TCTTGAGACT GGTACTTCCG TTCCGTCTCT CATAGTTAAA AGATGCGTTT TATAATTTAA 360

TTTTGACGCA TATTTTTTAT TGATGATGAT TGATTGATGG ACTCTAATCA AATCTGGATG 420

CAGTTGTTCA ATTTTTGATA ACGTTCCATA AAATTCTCTT TTAAAATTTT TACCAACCAT 480

TGACACACGA TGATCGTATT CCGTTTGAAA GAAATAAATA TCATTAATAT CCATTTTTAT 540

AGAAGACGCT TTTAAGTTAA TGATAATATA TTCACTTTTT GTCTTTATAT TTTCTTTTTG 600

TTGAATATAT GCATCGACAC ATGTTTCGAC CTCTTTTTGG AAATCTAGAA AATCCACGTT 660

CTTTTGAATA AATTGAAGAG CTGAAACCAT ATATTTGTAG GAAGTTAAAA CTAATTCTGT 720

GTGAGTAGAA ATAAATACAA TCAGACTTTC ACTGTCTGTT TTTCTGATTT CTTTAGCAAC 780

CTCAAATCCT GCCTGGATAT GATTGTTCAT TTGAATATCT AGGAAAAAGA TATTTAGTTG 840

GGCGTCCTGG TGTGCCACTT GTAATAATTC ATCTGAACGT GAAAATAAAT GGATTGTTTC 900

ATTATAAATA TTTCTTGAAT GTAGGATATT TGTTAACAAT ACTTCTAACG CTTTTGCTTG 960

AATAATTTGG TCTTCTAATA TATAAATAGC CATTTGTTCA CTCATCCCTT TCTCAAGAAG 1020

ATATAAAATT ACACAACAAA AGACAAATGA TTATAATGTT CAATCAATTT ATACACCAAT 1080

AAAAGAGATA AAAAATTGAT TGTCTGTTAG TATATTCACA TAAATTTAGT ATAGCGATTT 1140

CAGTAGTTTC CTGCAAACAA AAACATAAAA TAACAAAGGA TCAAGAAAGA GTACTTTGGA 1200

CTATTCATCT TTTTCTAACG AATCATGAGG ACTATTGTCT CTTCGTTGAT GAAAAACAAA 1260

GAATCAACCG CCAAACTGTT CTTGTCATTT TATACATGAC AAAATCaATT CATTGGTCGA 1320

TTCTGGTTAA TTTAATTTGT TAGTcACACA TGGGATCAtt GTACTAAtTt TtTACTTA 1378

(2) INFORMATION FOR SEQ ID NO: 537:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7696 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 537:

TAAAATATCT AAACCACGAC GACCTCTCAC CCGTTTTAAA CTGTCTCCAG AATCAGCAAT 60

GTATTGTAAT TCATATAAAA CAAAGTTAGG GATGAGCAAA ACACCTTCTA AAAAGCCTGT 120

TTTAGCGATA TCGTAGATTC TGCCATCAAT GATGACACTG GTGTCTAAAA TTTTGTATTT 180

ATGGAAGTTG TCTTCCACTT TACGTTCCAA CATTTGTCCT TCGGTGTCTT CCGTATTTGC 240

TTTCTTTTGG CGAGGTGTAA AAACTTTTTT CCACTCCTCA ATCCGTGTTG TTCCCATTCG 300

GAAACCTAAG TAACCCAGAA TAATCATAAC TAGAATTGGC AGTAAACTAT TCACGACAGG 360

AATTTGTAAG TTGTACATTG GAATGGAAAT AATGACACCA ATTGTTAAAC CAATGATGGC 420

ACCAATACTT CCAAATAATA AATAGGTTAA ACTCATTTCA TTTAAAGCTG CTTCGATTTT 480

TTTGACAGCG GAAACAATAT ACTTTGCTAA CCCTAAAGAT AAAATAAAGA AAATAATTGC 540

ACCGATTAAT CCGTTCGTAA AGTTATTGTT TAACCAACTG TTATTGCCTT GCTTAGCCAT 600

TTCCCAAGCC ATTGGGAAAA GAGAAATTCC TAAACTTGCA CCAGCAATGA TCATCAGCAA 660

CGTGATGATT CGTTTTTGCA TATAGCCATC CTCCTCGTTT ATTTAATAAA GACGTCTAGT 720

GTTGTGGGGC GTACAAAAAT CATACGCCAC CAAAAACCTA AACGGATTAT TTCTACTTTC 780

ATTATTTAAA AACTTTTCTT AACGTTTCAC CAATTGTTGA AACGCCAATA ATTTCAATCC 840

CTTCAGGCGC TTCCCAGCCA CCAAGGTTAT TTTTAGGGAC ATAGACTTTT GTAAAGCCTA 900

GTTTTTGCGC TTCACGCACT CGTTGTTCAA TACTATTCAC GCGGCGGATT TCACCAGTTA 960

AACCAATCTC ACCAATGAAA CATTCCGAAG ATGAGGTTCC TTTTTCTTTA TAGCTAGAAG 1020

CAATACTGAT TGcTAACGCT AAATCAATAG CTGGTTCATT AATTTTGACT CCACCAGCCG 1080

CTTTTAGATA CGCATCCTGG TTTTGTAACA AAAGGCCTGC CCGTTTTTCT AAAACCGCCA 1140

TGATTAAAGA AACGCGATTA AAGTCCAACC CGGTGGTCGT TCGCTTGGCA TTGCCAAACA 1200

TCGTTGGGGT AACTAACGCT TGGATTTCCA CCAAAATAGG GCGTGAACCT TCCATTGCTA 1260

CCACAATTGA AGATCCGGTG GCTCCTTCTA AACGTTCTTC TAAAAAGACC TGAGAAGGAT 1320

TCATAACTTC AACCAATCCA TGTGTCTGCA TTTCAAAAAT ACCAATCTCA TTCGTTGAGC 1380

CAAACCGATT TTTGACAGCC CGTAAGATTC TAAAGGTGTG ATGCTTATCT CCTTCAAAAT 1440

AAAGAACCGT GTCTACCATA TGTTCCAACA TCCGTGGCCC TGCGATAGAC CCTTCCTTCG 1500

TTACGTGTCC GACAATAAAA ATGGCAATCC CATTCGTTTT TGCCAGTTTC AATAATTCGG 1560

CAGTTGTTTC *TCGGACTTGA CTGACACTAC CAGCAACGCT GgTAACATCT GGCTGCGTCA* 1620

TTGTTTGAAT CGAGTCAATA ATGACATAGT CTGGTTCTAA TTTTTCAATG GCTCGTGATA 1680

TTTCGTGCAT ATCTGTTTCT GCGTATAAGT AAAATGTCTC GTTGACGGTG CCTAanGCnt 1740

CCGCTCGTAA TTTAATTTGC TCAGCACTTT CTTCACCAGA AACATATAAA ACCGTACCAC 1800

CTGTGGCTGC TAATTGTTGC GATACTTGTA AGAGCAATGT TGACTTACCT ATTCCGGGAT 1860

CGCCACCGAT TAAAACTAAC GAACCTGGAA CAACGCCACC ACCTAAAACC CGATTCAGCT 1920

CCACTAATTC AGTTTTTACA CGAGGCTCTT TTTTAGGTAT GACTTCAGAT AGCCGTTGTG 1980

GTTGTGTCTT TTTACCAGTC AAACTTACTC TAGCACGACG ATCAGAAGTA TCTTGAATAA 2040

CCTCTTCCAC CATCGAATTC CATTGTCCGC AATTAGGACA GCGGCCGAGA TATTTTGGAG 2100

ACACATACCC ACAGCTTTGA CATTCAAATT GAACTTTTGC TTTTTTTGCC ATGTAGAATC 2160

TCCTTATCTG TTGTTTATTT TAACATTTAT TAAATGGTTT GAATAACGTT GCCTACTGAA 2220

AAAATGGGCT TTATTTTTGT CCAGACGAAC CAAAACCGCC CGTTCTTTCT GCCTCGACTT 2280

GGTCATCATC CGCTAATAAA AACGGCTTAA AAATCCCTTG GCCAATTCGT TCACCTTTTT 2340

CAATGACAAC ATCTTCATAT CCAAAGTTTA AAAATTGGAA CATGATATGT CCTTCGTTAT 2400

TCTCGTTATT ATAATAATCA CTGTCAATGA CTCCGACACC ATTCGCTAAA AGTAGGAATC 2460

GTTTTAGCGG GTTCGAAGAA CGACTTACCA ATTCTAAATA TTCAGAATCA GGCATATACG 2520

CCTTAATTCC TGTTTTAACT AAAATTGGTT TGGGATTTTC TCCAGCTTGT TGTAACTTCC 2580

AAAGACTTGG TACAACGACC GTTTCGGCTG CTTGAAAATC ATACCCCGCG GCACCTTTTG 2640

TTGCTCGTTG TGGGACACTC AAGCCCTCTT CTTGATACTT AGAAATAATC TCAAAACCAC 2700

GTTGTTTCAT ATTCATTCTC CCCTACTGTC TAAACTACTC CTATTTTACC ATAATCTCTG 2760

ATACTCTGGT TGTTTTTAGT AAGTTTCTCA TTTCTAAAAG CCCAGTTAAT CTAGACTACG 2820

CAAAAAAGAG GAGAATTTTT GTTTATTTTC CTTCTTTTAG TCTATCATCT TCTCTTAAAA 2880

TAAACAGACC AAATCCCAAA AAGTCCCAAG ATTCCCTCAA AAAGTTCACA CTTTCATTAA 2940

ACTTTCTTTA CTATTCATTA TTTAGCACTT TCTAGTAGCC GCTGATTTTG TTATAATGAG 3000

TGTCAAATTT TACTGTAAAA AGAAACGAGG CACTTATGAG AAACAAACTC CAACAGGCCA 3060

AACGAATTGT GATCAAAGTC GGAACCAGTT CTCTGATTTA TCCCAATGGC AATATTAACT 3120

TAAAAGCAAT TGATCAATTA GCCTTTACTT TGAGTGATTT AAGCAACCAA GGCAAAGAAA 3180

TTATTTTAGT TTCTTCTGGC GCAATTGGTG TCGGACTGAA TAAATTAAAT CTATCTGTAC 3240

GCCCAACGAC AATTCCCGAA CAACAGGCAG TCGCTGCAGT CGGCCAAGCA GAATTAATGA 3300

ACATTTATAA TCAGCGGTTT TCAACCTATA GCCAACAAAT GGCACAAGTT TTATTAACTC 3360

GTGATGTTAT TGAATATCCT GAAAGTCGCA ATAATGTTAC CAATACTTTT GAACAACTGT 3420

TAAAAATGAA TATTATTCCT ATCGTCAATG AAAATGATAC TGTCGCAATT GAAGAATTGG 3480

ATCACTTAAC CAAGTTCGGG GATAACGATC AATTGTCTGC CATTGTCTGC CAAATTGTTC 3540

AAGCTGACTT GTTAGTTATG TTATCAGATA TTGATGGTTT CTTTTCAGAC AACCCTACTG 3600

TCAACAAAGA AGCCACGTTA TTTTCTGAAA TCAATGAAAT TAATGAAGAC TTGTTCCAAT 3660

TAGCCGGTGG TAAGGGAAGC CGCTTTGGCA CTGGTGGGAT GTCTAGTAAA TTAAAAGCTG 3720

CTGAACGTGT GTTAGCTAAT CAACAGGCGA TGATTTTAGC AAACGGAAAG CAGCCAAAAA 3780

TCATTTTTGA AATTTTAGAA GGAAAAGATA TCGGTACGTT ATTTATTAAA GGAGGGCATG 3840

AAAGTGACTG ATTTAAAGCA ACTAGGGCAA CAAGCTAAAG AAGCCTCTTA TACGCTAGGC 3900

TTGATGGATA CTCGTCAAAA AAATACTTTG TTAAACAAAA TGGCCGCAGC CATTGAAGCC 3960

AATGCACCCC GCATTTTACA AGCAAATGCA TTAGATTTAG AGCAAGCCGC TACCCACGGA 4020

ATTTCTGAAA CAATGCAAGA CCGCTTACGT TTAACAGAAG AACGGATCAC AGCGATGGCT 4080

GAGGGCATAC GCCAAGTCGC TACTTTACCC GATCCAATTG GGGAAGTAGA TAAAATGTGG 4140

CGCAATGAAG CGGGACTGTT AATTGGGCAA CAACGCGTGC CTTTAGGTGT TATTGGAATC 4200

ATTTATGAAT CGCGGCCAAA CGTGACAACA GATGCCGCTA GCCTTTGTTT CAAATCGGGA 4260

AATGCTGTTA TTCTTCGAGG CGGAAAAGAA GCGTTTCACT CCAATCAGAT ATTGGTGACA 4320

ATTTTACAAG AAGCGCTAAT CCAAGAAGCC GTTTCTCCCC ATCTCATTCA ATTTGTTGAT 4380

GATACTTCTA GAGAAACAGC CCAACAACTT ATGCGCTTAA ATGACTACTT GGATGTCCTC 4440

ATTCCGCGTG GCGGCGCCAA CTTAATCAAA ACTGTTTTGA CAACAGCCAC TGTTCCTGTT 4500

ATTGAAACGG GCACGGGCAA TTGTCATATT TACGTGGACA AAGACGCCCA GTTAACAATG 4560

GCAACTGAAA TTATTGTCAA TGCTAAATGT CAACGTCCTT CTGTCTGTAA CGCCGCCGAA 4620

ACGTTATTAA TTCATCAAGA GGTGGCTGAA GCCTTTTTAC CAACAATTGA AAAAGCCTTG 4680

AAGGAGTTTC ACGTGGAACT TCGGGCAGAT GAACGTGCGT TAGCAATTTT TGAAGAAGCC 4740

ATTCCAGCGA CTGAACAAGA TTGGGAAACG GAATTTCTCG ACTTTATCTT AGCAGTGAAA 4800

GTCGTTGATT CTTTGGACGA AGCCATCCAG CATATTAATC GCTACAATAC AAAACATTCT 4860

GAAAGTATTA TTAGTGATAA TTACTTTGCC ACACAACAAT TTTTACAACA AGTGGATGCA 4920

GCCGCTGTTT ATGCGAATGC TTCAACTCGT TTCACGGACG GTTTTGAATT TGGATTTGGC 4980

GCCGAAATCG GCATTAGTAC GCAAAAATTA CATGCCCGTG GACCTATGGG TCTAGCAGAG 5040

CTTACTTCGA CAAAATATGT TATTTATGGA AATGGGCAAG CACGTTCTTA GAAAGAAAAA 5100

CCCTTGATTT AAAAATCAAG GGTTTTTCTT TCtCCTCTCT AATGTGCATT TTTACACATG 5160

AGATTGTTCA CAGTTCCACA ATCTATTCTT GATATTTTTT TTGAATCCTC ATTATACTTT 5220

CTACATATGC TTGTTTCACT TGGACTGGAT CTAAAATTGT CACTTCTTCA CCAAAGCCCA 5280

TAAAATAGTT GGCTAAAAAT TGTGTTTCTG TTGGATGAAT CTGTCCAACT AAAAAACACT 5340

```
GCGTGGGCGT TTCTTCCAAC TGAATATTGT CAAAGCGATG TTTATAAAAA AATTCTTTGC      5400
CTTTTTCTGT AATGATTGCT TTAAATGGAA TCGTACGTGC CGTCGTTTGT TGATGGCGAT      5460
ACGATTCTTT CAATTCTTCA GTCGTAAAAT CCAGTGTTTG GTTTAATTGC TGAATAGCCT      5520
CCATAAAATC ACAGCGGTAA GTACGCCAAG CCTTTTTAGT CACATCATAG CCTGAACAAT      5580
ACCAATACCC TTCTAAAATA GTTAACCGGG TTGGTAAGAT TTTTTTCGTT GTTGCTTCAT      5640
ATCTGGTATA GTTAAACTGA ATTACTTGTT GCGCCAAAAT ACAATGAAAC AGCTGCTCTA      5700
AATTGCCTGG TGCTTCTACT TGTTCAATTC CTTCGTAATA GACGACCTTA CTCATTTGAC      5760
TGATTTTAAA TTGGGTTGGT TCATCTAAAG AATGCAATAA TTTTTGCTTA ATTTGCTGAT      5820
AGGAATGACC AAACGGTGAC TCTGCGACTA ATTTCAGTAA TTGTAACGAA AAAAAGATTG      5880
CTAACACTTC ATTTTCGTTA AAATAAATCG GCGGCAACAA CGGTTTGTTT AAAACTCGAT      5940
AGCCACCATA ACGACCATTC TCCACATATA GCGGCACGCC TAAATTTTCC AAAGCTTCAA      6000
TATCCCGCAA AGCTGTGCTT TTAGAAATCC CAAAAGTTTT CATTAACTGA TTTAAATTAA      6060
ACTGAGGATG GGTTCTCAAG AAAAAAAGTT CCTGATTTAA ACGTTCTGAT TTTTTCATTT      6120
TTCCCTCCAT CGGTGTCACT TTTTGACACC TTTTATTGTT ATGCTTATCT TATCAATCTT      6180
AGGAGGTTCA AACAATGAAA AACACACAAT TTGCCCTTTT TATGGTCTTA AACGGaCGAG      6240
CAACGGaAGC CTTGACGTTT TACCaAGAAG CCTTTGCTGG TGATATTCTT TTTAAAATTA      6300
CTAATCAAGA ATTTAAAGAG CGCTTAAATC CTTCTTTGAC CATTCCCGCT GGCGAAGAAC      6360
ACTGGCTTTC TCACTCGATT TTACAAATGG ACACTTTTCA ATTACAACTA GCTGATAATC      6420
CTTTGTTTGC AGGAATGAAT ACCACGCCCA CAGAACAATT GACCCTCAGC TTAACCGTCA      6480
ATAGCCGTCA CTCTGCCAAA ACAATTTTTG AAAAACTCAC GTGCCATCCC GAATCCTGTA      6540
TTATCCAAGA ACCAATTGAA AATGAATTTG CAACATTTTA TGCGATTGTG AAAGATCCTT      6600
TTGGTTTAGT CACTCAAATT ACTCATGAAA AACAAGCAGA TCCTACCAAG AAAGGAAGTT      6660
AGCCGAAGAA TGCGAACCTA TGAATCAAAA GAAGCCTTGA TTGAGGCCAT TCAAATAGCT      6720
TCACAAAAAT ATTTAGCTGA ATTTGCAGAA ATTCCTGAAA CACTTAAAGA TCACCGAATT      6780
GAAACAGTAG CTAAAACACC TTCAGAGAAC TTAGCCTATC AATTAGGTTG GCTCAACTTG      6840
CTGCTTTCTT GGGAAGAACA AGAACAACGT GGTCTGACCG TTCAAACGCC AGCTGAAGGC      6900
TATAAATGGA ATCAACTGGG CGCGCTCTAT CAATCATTTT ATCAAACCTA TGGACAAATG      6960
AGTTTAGAAA GTCAGCTGAT TGCGTTGCAA GACACCTTAG AAAAATTACT TCATTGGATT      7020
GACTCGCTTT CCGAAGACGA ATTATTTTTA CCTCAACAAC GGGCTTGGGC GACCACCAAA      7080
GCACAATGGC CTCTTTGGAA ATGGATTCAC ATTAATAGCG TTGCCCCTTT TACTAGTTTC      7140
CGAACGCAAA TTCGCAAATG GAAAAAAGCT TGTCTTTAAA AAGTTATCCA CATTTTTTCA      7200
AAAAAAGCTT GACAAGTATT TTTAAATTCT CCACAGAAAA CTTTGTGGAT GAAATTTCTT      7260
ATTTTTTCCA AAAAACTTTT CCACAATGCT TATTTGTTCG ACATTAAATC TTTGTAAATT      7320
```

CTAGAAAAAA GAAGATGAAA ACGTTGTTTT AATAACGTTT TCATCTTCTT TCAACTTTAG 7380

GGATTCCTTT ACTTTTAGCA AACAGGACTT CTGAACTTAC GTCTGTTTTT TTTAGCAACT 7440

CGAAAACTTA TCCTTGTGGA TAAACCGAGA TAAAAAGCTT TTTTGTGGTT TtCAACAGAG 7500

TTATGCACAA AGTTATCCAC AGGGCGTTTT CAGCAATTTT TCGGAGAATA TTCGTTGTTT 7560

AATCTGGTGA TTTAGCATTT GAATGTCAAT CCCTTACTAC TAATTCTTCT tCACTCCACC 7620

CTTAGGATCC TAATAAAATT TTAAAGTkTT cTcTAtTGGT ACATCArATA AAGAGTATGr 7680

kAAACTATAG GTACTT 7696

(2) INFORMATION FOR SEQ ID NO: 538:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 987 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 538:

CTCCTATACA AACATAGGCT ATTGCTCAGA AATGTCGAGA GACACCAATG AGTAGAACAA 60

GACTGGTCGA GGAGAAGGCT TTTCTCAAAA TAGCTAAGAG TGATCTTTAC GTTGTATAGT 120

GCTAAAGCTC AACGTTCATG TTGATGTATC TTTTTAGTCT CTAGTCAATC AGTTAAGTAA 180

GTACTTTTAT TTAAAATTAA GAGAGGAGTG AAAGTATGAC TACTATGTTA GCAATCATTG 240

GAACTGTTGC AATCATCACA GTTCAAGTTA TTGAATATAA TATGACAGAA ATTCATATTC 300

CTACTGCAGG AACATTTAAT GACGATGATG CAGAACCCCC ACATTCTTCT GTAACCCTTG 360

TATTTATTCG CACTAtACyT AtCACTCTCT CACTTAACTT TCCATCTCAt TTTTTTGAGA 420

wGGkTTTTTT ATGkATyCGr GAAAATwATr AACmAAAAAG ACyGtTAAAy CAATGATTTA 480

ACAGTCTTTA GTATACCGGC GGCCGGGGTC GAACCGGCAC TCCCGTGAAG GAACTGGATT 540

TTGAGTCCAG CGCGTCTGCC AATTCCGCCA CGCCGGCAAG ATATTTTGGA AGGCGGTAAC 600

CGGATTTGAA CCGGTGATGA AGGTTTTGCA GACCTCTGCC TTACCACTTG GCTATACCGC 660

CATATTCCAT TTTTGTTACm AAAAAACTGG GGTAGCTGGA TtCGAACCAA CGCATGACGG 720

AGTCAAAGTC CGTTGCCTTA CCGCTTGGCT ATACCCCAAT AAAATTACAA AGGGCGACCG 780

ATGGGAATCG AACCCACGAA TGCCAGAGCC ACAATCTGGT GTGTTAACCA CTTCACCACG 840

ATCGCCGTAA TCTTAAAAAG TGTATCTATT AAATAATTAA ATGGCAGGGG CAGCAGGAAT 900

TGAACCCACA CCAACGGTTT TGGAGACCGT TGTTCTACCT TTAAACTATG CCCCTAAAAT 960

GGAnGAAAGT GGATTCGAAC CACTGAA 987

(2) INFORMATION FOR SEQ ID NO: 539:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 759 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 539:

```
CGCACCTCAC GATCGTAAGC AACAAATTTT TCCGTAAATA GCAAGGGATT TTCCAAGCCT     60
AACGCTTTAG CAAAGGTAAA CAAGCTTGTT TCTTTTAAAG AAGCTGCTTG AATTAATTTT    120
TCATCTGTTA AAACAGCAAC TGGAGTCGCA TCTGCTAGTA AACGTCCATC ACAAACTACT    180
AACGCACGTG TTGTATATTC CAACATTAAG TGCATATCGT GAGTGATCAT TAAAATAGTC    240
ACACCTAAGC GGTTCAATTC TTCTAAAAAT GTCATCATTT CTGTATAATG TTTAAAATCT    300
TGCCCAGCTG TTGGTTCATC TAAAATTAAT AATTCAGGTT CTAAAACTAA GATTGATGCA    360
ATAGTTACCC GTTTTTTCTG CCCAAAACTC AACGCAGAAA TCGGCCAGTT CCGGAAGGGA    420
TACAGCCCAC AAATATGTAA AATATTCGTC ACACGTTCCT CAATTTCTGC TTGTGGCACA    480
TCTCGTAAAA CTAATCCTAA AGCTACTTCT TCAAAAATCA TTTTTTTAGA AATCATTTGA    540
TTGGGATTTT GCATCACATA ACCAATTTTA TCTGCACGTT CTTtGATTGA ATAATTGGAA    600
AAATCCTGTC CTTCCCATAA AATCTTGCCA GATTGAGGCG TGATGAACCC GCAAATTATT    660
TTAGAAAGTG TAGACTTCCC GGcACcATTT TTTCCTACAA TGCtGATCAT TTCtCCAwGA    720
TGaATAGTTA CwGAGAAGTC ATCTAACACT TTTTCACCG                           759
```

(2) INFORMATION FOR SEQ ID NO: 540:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1171 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 540:

```
ACCGACGCCT AAATCCTTTT GCTTCAAAGT AGACTCCATT GCCATTCGCG CATGAAGTTT     60
CGTATTATTT TCTTTGCTAA GATGACCTAA ATAAATGCGT TTTGTTTTAT CACCTAACAC    120
ATCCGCCATC ACAAGAGCAC CATCATCATT GGATAAATGG CCTTTATCTC CTAAAATCCG    180
TTGTTTAAGA CTCCATGGAT AAGGTCCTGC TCGCAAAATT TCAATTTCAT GATTGCTTTC    240
GACTAAATAA GCATCTGCAT TTTCAATCGT ACCACGAATA TGATCGCTAC AATAACCTGT    300
ATCCGTTAGA ACAACAAAGG AACGATTATT GCGATGAAAA CGATAAAATT GAGGCGCAAT    360
CGCATCATGT GAGACACCAA AACTTTCAAT GTCCATATCT CCCAAGGTCA AAACTTTGCC    420
CATTTCAAAA ATATGTTTTT GAGCGACATC GACTTTGCCA ATCATGGGAT CCATCGCTTG    480
CCACGTTTTT TCATTGGCAT AGACATCTAA CTTATATTTT CGCGCTAACA CGCCAACGCC    540
GTGAATGTGA TCCCGATGTT CATGTGTAAC TAAAATCGCA TCTAAATCTT CTGGTTTGCG    600
ATTGACTTCT GCTAACAGTG AAGTGATTTT TTTGCCACTT AAACCCGCAT CTACTAATAC    660
TTTCTTTTGG TTGGTTTCAA TAAAAAGGGA ATTGCCTGTA CTGCCACTGG CAAGAACACT    720
GATATTAAAA GCAAACTCTG AATCCATTTA AAACTCTTCC CTTCTACATT ACAACATTAA    780
```

CGTGTTTCAT TATACAACGG ACCTTAATGT TTTTCCACCT TTGGCACAGT ATTATTTGTG 840

ATGATGGTAT TACTTACGGC ATTCACCTGT TCAATTTCAA CTGAATCTTC ACCAGTAGAA 900

ATTCCTACAA ACCAGACTGG AACGTATACA TTTTTCTCAC GAATTTTATA AATTCTTGaA 960

TAGGCCAATT TAAtGAATGT TATTTTGGAA TTACTAgGAA TCCGGTTATT CATATAGAGC 1020

GTGTCTATCG CATCACGATT GGAATATAAA TCCGTTTTGT CTCGTAACTC TTCAATATTT 1080

TCAATGTGCG TTTGCGAATA TTTACTAATT TGCCATAATT CACCTGACGA ATCTGCTAAA 1140

ATACTTAnTT TTGCCGGGTC GnnCCCGAAA A 1171

(2) INFORMATION FOR SEQ ID NO: 541:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4134 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 541:

GCAATCATTG TATTGGGATA GAGGTTAAGn CCTAGGGCAT CCCCATCTGG AAATGCTGTA 60

TGGTCATGGn CATTCCTAAC CAGGATTAGG AnGGAGGGAT CGCAATTyCC ATCTCGTTCC 120

ATTTAATTTT GCTTAAGGAT TGTGCCATCA AAACACCAAC AATAATCAAC GCTGGTGCTG 180

TAACTTGCGA AGTGACAACT GCTAACAAAG GAGAAAATAG CAAACTGAAA ATAAATAAAA 240

TCCCGGTTGT AATAGCTGTG ATTCCTGAAC GACCACCCAC AGCAATCCCC GCAGATGATT 300

CGACATATGC GCCAACTGGC GATGTCCCTA ATAGAGAACC AGCTAACATC GCTGTTGAAT 360

CAGCGGCTAA AGCTTTCCCC ACACGCGGCA TTTTATTGTC CTTCATAAAA CCTGCTTGAT 420

TGGCTAAGCC CACCAACGTA CCAGCCGTAT CAAAGAATGT TACTAATAGG AACGTTAAAA 480

CCACGACCCA CATTTGAaTC GTATTgATAT CTTTCACATG TCCTAACGCC ACTAAAAATG 540

TTGGTTCTAA aCTTGgCcGC TTGAAATAAt CTTATCAGGC ATTGGAATAA CGCCTGTCGC 600

TAAACCAACA ATCGTTGAAa TCGCCATCCC AATAAAAaTA CCACCCGGTA CTCGACGAAC 660

TAACAAAATT GCAGTTACGA CTAAGCCGAC AATGGTAACC CACGTTGAAC CAACATTCAA 720

AGGTCCTAAG CCAACGAGTG TCGCATCATT AGCCACAATT AAGCCACCTT GGCTTAAACC 780

AAGAAAAGCA ATAAACAACC CGATCCCACC TGaAATGGCA AATTTCAAGT CTGCAGGAAT 840

TGCATCAATA ATTAGCTCGC GTAATTTAAA AATTGTAATT AAAATAAATA TCAATGAAGC 900

CACAAAAACG CCCGCTAATG TCGTTTGCCA AGGAATCCCC ATACCAACAG AGACTGAATA 960

AGCAAAGAAT GCATTAATTC CTAAAGCTGG TGCTGTCGCA ATCGGATATT TCGCCAACAC 1020

ACCCATTAAA ATACAACCTA AGGCACTCGC TAAAGCGGTG GCAGTAAAAA CAGCGCCTTC 1080

GTCCATGCCA GAAGCACCAA GTACGGTTGG ATTAACGAAC AAAATATACG CCATTGAGAT 1140

AAACGTGGTG AAACCAGCTA ACATTTCTCG TTTCATATTC GTATTCAATT GTTCAAGTTC 1200

```
AAAATATGAA CTTATTTTTT CTTTCATGTT CCGTCCTCCT AAAATCTAAA TATTTTACAA    1260
TGTTCGTAAA ATATTCAATT TATTAATGAA ACAACGTGCT TCCTTGAGTG GATATTATTA    1320
AGCACGAATA GCCCCATTTT ACTTGATACA ATAGGAATTT GCAACAAAAC ATAATTAGTT    1380
AAATACACAT AGTTTTGATC AAAAGCACAG ACCAAAGCAC AGACCTTTTT TAATCCATCA    1440
CGCCAATAGT TACATATTAC GTTTACTGTT TTTTAAACCC AACGGGAAAA GAATAGTTAT    1500
ACCTGATATA CCTATACTGA TAAAAAAATA AATGAGGTAA CTTGTCCAGT CAGCGGTACC    1560
CCCATCTATA TAACCTGAAA TTGAAAAATA TGAAAaCAAA ACTATCACAC TGACTAACCA    1620
CTGATATCTT AACCTTAATT TTAATTTTAT TGTCATTACG TAATACATCA AATACACGCC    1680
TATCAGTACC AACCATACGA TGTAACTACT GATAGAATAA TGACCATCGA ATAAGcCCAC    1740
ACTAGATTAA TCAAAATAAT CAATATTATT GGTAGTAATA CTTGACATAA AGTGTTAGAT    1800
CGATAGATGC ATACTATTTT ATCGATAATT TTTTTCATTG TTTGCTCTCC AAAAGTTTAT    1860
TTTTAATATA AGCTAATAAA GTTCCTAATT CAATTACTAC CCATGCTGTT AACAGAATAG    1920
TCCAGAATAA ACTTCGTACA TAGAATAAAG CTATAAATGT AAATATTAAT ATTGCCAACT    1980
GACTGATTAT CACTAATAAA TCTTTTTTTT GGGAAAATTT ATAAGCTAAA AGACTGCACA    2040
CTATGCCAAT TACTGGGATA AAACGAATAT AACTAGGAAC CCCTTCCATT GGTACAGTAA    2100
ACAATAGAAT TATAGAATAC CAAAGTTCAT CAGGAACCTT CGACGATAAT GCCCTTATTC    2160
TTCCCCAATT CATAAGTTAA TCCCCCTTTT TCATTCTCTT ATACTTGTCT TCTATACGCT    2220
TATCCTGTTT TTTCCCGCCT GCATGAAAAG CTAGAATAAT CATTGCCTCT GTAATAACAA    2280
ACTCCCATAC TGACCAATGA AAAATAACTG TCATCCCTAG AATTACTAAC ACAGCTGATG    2340
CTAGAATAGC ATCTTTTGTG CCAAAATCTA TCCCTATAGG TAACATAAAA AACCTCCAAT    2400
CTCACTTTTT TTATCCTAAA GACATTGCCA AAATAAATAA TATCCAAGGA ACTAAAAAAC    2460
TAAAAGCACT TAAAAGATTA GCGACAAGCA ATATAAGTTT TAACCCTAAG TGCTTTATGT    2520
ACATACTAGG AATAGCTCCT ATCAAGCCAA TTGGTAACGT CCACAACAAA TCAAAGGCAC    2580
CATGATAGGG ATTGTAGAAA AAGAGGTACT TAATCCATAC AAGTATATTC AAAAGAACGA    2640
ATAGGACACT CGCAACAACT GCTGATCGAT TATTTAGTTC TATTTTCATA GTTACTGACC    2700
TCTTTGGTTT TTATTTTAAT TATATCAGAA AAGACTTCAC TGATAAAAAT CTTTTCTTCC    2760
CCTCAAACAA ACGAGCTATA ATAAAGTAAC TAGAAAACCA GCAAAAATAA TTTGTAATAA    2820
TTTTAAATTA TTGTGCTTTT TTATATTTTA GGTATAAAAA AGGTAGCCGA CTATCCGACT    2880
ACCTTTTAAA CACTATTCAA TTACAGTAAC ATTTACCGCT TTTGTTCGCC CTTCGTATTC    2940
TGCGGTTTCG TACTCCACTG CTTGGCCTTC AAGCAATTTT CTAAAGCCTG GCATACAAAT    3000
ACCTGTTTGA TGTACAAAAA TCTCTGATCC ATCTTCGGCA ATAATAAAGC CATATCCTCT    3060
TTTATGATTA AAACTTTTAA CTGTTCCAGT TTTCACTTAT ACTTCCCCCG TATTATGATT    3120
TAAGTTGCCC ACGATCAAAT TTCATAGCTC TTTCTACTAG ATTAGGAGCT GAAACATAGC    3180
```

TTGAAATGCC TATTACACTT CTGTTATGGC GAGTAACTCC GATGATTTTA TTTCCTTCGT 3240

CTACTACCTC ACGATTATAG ATTGCATTAG CTTTTACTTT AACTGATTTT ATAAACATAG 3300

CCTTTTCATT TAATTGTTTA AGCAAAAGTG CTATTTCTTG TTCCACGGGC TCTAGTTGTT 3360

TTTCAATTGC TTGTCTGTAT TGAGAGTCAA TATCTCCTAC TTTATTAGGA AAATGAGTAA 3420

AACAAGATTG CTGTAGCTCC TCGCGTTCTT TCTTCATTTG TGTTAGTAAA TGATTATTTT 3480

CwGTTAAAAG CCTGTTTATT TCAGGTTTTC TCTTAAGACT ATCTACCGTA AAAGTTTTCG 3540

TATTTTCTAG CGTTTCTAAC TCTTTCTCTA AGACAGATTT TTCCTGtTCT AAATCATTAA 3600

TCACTTTATC TACTCCGCTA AAAATTTCTT TTAGCTCGGA AATAAGCTCA ACTTTTCCCG 3660

GAATATCGTA GTTAATTTCT ATCATTTTAA TAGCACCCCT TATTTATCAT ATTTAGACAT 3720

GATTTTTTTA AATGGATCAG GTTTATCTTC TTTCTTATCC TCTTTATCTG ATTCCGTAGA 3780

ATCACCTAAG CTTTCTTTAA TCAACTCTTC TAATGCTTCC GAAGCTTTAT CAATGTCAAA 3840

TTGCCCTTCT TCATCTACAA ATTCCTCTAA AACTTTTTTA ATATCTTCCG CTGTCATATT 3900

TCTTACCTCT TTCATAATTT TTTATATATT TTGGTGAAAT TTCACCGAAA TAAACATATT 3960

CTGCTTGATA CATGATATTG CCTAATTGAA AAGCGTCATT TACTTTGTCC TTGTTTGTAC 4020

ATAAGTGTGT ATGGAATTGC nGnATCCAAT AAACTAGCAG CCGTGTGATC TGGTTCGAAT 4080

ATTAGCAATG GCAGCATTCC GATAACACAA TATTACTTAT TAACGGCACn CCAT 4134

(2) INFORMATION FOR SEQ ID NO: 542:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1004 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 542:

AATTTTTTTA GTTCGTCTGT GGCGTTCTAA GTGCAGCAAA TAATCGTCTT GCAGTTTTAG 60

AAGGTCTGAA TATACCTGAT CCATAGTCTT GCTGGCCGAC AAAGAGAGAG AATAAATAAA 120

TGAAATAGGA AGTGTTTCAA TTTTyTTGTT GCGAGTAATC CAAGAAGAGA CAGTTCCTTG 180

AGAGAAAGCA TGTAATTCAC AAAATTCTTC TACAGTAATC CCTAGTTGCT TGATAATGTA 240

TGCGTTGaTA GGGTGTGGAT ATACAAAGGT TTTTCTmGGC ATGGTGGGTA AGCTTCCTTT 300

CTTAGGTTAT TAATGCGTTT GCAGTAATAT TGTACTAAAA AACATGCAAT AAGGGAAGGG 360

CGTACCTGCT TCTTTTTTGT TGACCTTCTT TAATAATGTG TTATACTAAT AATGCGGAAG 420

ACATAAGTTC GATATGAmCA GCACAAkAGm CCAAACcTGT GACGAGGTTT GGGCTATTTT 480

TTTCGTCTTG CGGATAGACG TGTGGGCTAG GACAACTTGC ATGGTTATCT GCTAATCATC 540

TAACCAATGG TCAACTAGTA AGATTACTAG ACCCACAAAA AGTGGGGCAA TAACCAAAGA 600

CATAAGTTCG TACATAAACA GCACCCCCAA TCGGAGGCAA GTTGCCGAAA TAATTATAAC 660

ATAAATTCGA ATATACCTAT ACAGTAGTAA ACCTTCAGTG TTTTACTTCA ACTTTTAGCT 720

CTCTTTTTAG TACGTATACT CGTTTACATA ACGAACATTT GTACTTACCT GAATCACTTG 780

ACACTCTGCT CGCAAGACAC CCaCACGCAC ACTGTACGTA TGCAACGCCA CTTATTTTAT 840

TGTnGTCyTT AATCACTGcA CCCtGGTAGT TAAaTTyCTT CATAAAATTC TCCATCATCG 900

TAATAATCTT TCATAAAAGC ATCGATTGAT yCTtGTTCTT CATCTGATAA TTCTGGAAGT 960

ACGACGTCTG CGAAAGGATT TTCCAATGTA TCTACAGCAG ACCA 1004

(2) INFORMATION FOR SEQ ID NO: 543:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2170 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 543:

CCCCTTyCAA ATAAgCtCGT TCgTGCTGGT CGGGCCakGA ATGCCgGTGT TTAcTTTGTC 60

aCGCaAAATA GTGaCGATTT aCTGGATGaA CGCTTGAAGA ATAATATCGG ATTGAAATTT 120

GCTTTTCGTA GTACCGATAT TCaTGAAATC aAGAAGACTT TGGAATTTTT TGGGTTGGAT 180

CAAGAAGATG AAAGTAATCA AAAACGCCTA CkAGATTTGG AAAACGGCCA ATGCTTGATG 240

CAGGATTTGT ATGGTCGGGT AGGTGTGGTT CAGGTTCACC CGGTCTTTGA AGAACTCTTC 300

CATGCCTTTG ATACTCGCCC ACCGGTTCAA GAGAGAATGG AGTGAAGGAG ATGAAAAGGA 360

AAATTTTACG GATTGTAGGC ATTGTTTTGA CTAGTCTATT TTTCCTCATC ATTCTTTTGA 420

GTTTAGTTGG AAAGGTAGCG GAAGCGACTG GATTAGTCGA TGATACAGTG AAAGCTGGTA 480

ATCTCTATTC TCAATATTCT CTTGGTAACT ACCAACTAGA TTTTTTTGTA GACAATTCTT 540

GGGACTGGCT ACCTTGGAAC TGGGGGGATG GTTTAGGAAA AAGCGTGATG TATGGTCTCT 600

ATGCCATCAC AAATTTCATT TGGACCGTAA GCTTGTATTT GTCGAATGCC ACGGGTTATG 660

TGGTTCAAGA AGCTTATAAA TTGGATTTTA TTTCAGATAC GGCAGAAAGT ATCGGGAAGA 720

ATATTCAAAC TCTAGCCGGG ATTAACGAAA ATGGACTTCA GCCATCTGGT TTTTATTTTG 780

GCTTTTTGAT GTTAATGATT TTGGTTTTAG GCATTTACGT AGTCTATACA GGATTACTTA 840

AACGGGAGAC TACCAAAGCA GTGCGAGCTG TTATTAATTT TGTAGCGATT TTTCTACTTT 900

CGGGTTCCTT TATCGCCTAT GCCCCTAACT ACATTACGAA AATCAATGAC TTTAGTTCGG 960

ATATTAGTGA ATCTGCATTG AGTCTTGGCA CTAAGATTGT CGTACCAAAT TCTGAAAGTC 1020

ACGGTAAAGA TAGTGTTGAC TTAATTCGTG ACAGCTTATT CTCAATTCAA GTCCAACAGC 1080

CGTGGCTCTT GTTGCAATTT GATGATTCCA ATGTAGAAGA AATTGGCGAA GAGCGAGTTA 1140

ATAAAATTCT ATCCGTGAGC CCGGATGAAA ATAAAGGGAA GGATCGCGAG GAGGCGGTTA 1200

AAGCCGAAAT TGAAGACAAT GAAAATGCAA ATTTGAGTAT AACAAAGACC ATGAGCCGAT 1260

TAGGGACCGT GGTTTTTTTA GTATTGTTTA ATATTGGAAT TTCCTTCTTT GTCTTTCTAT 1320

TGACAGGAAT AATGCTATTT TCCCAAATTC TTTTCATTAT CTTTGCGATG TTTCTCCCAA 1380

TTAGTTTCCT GTTGTCCATG TTACCGACCT ATGAAAGTTT GGGAAAGAAA GCGATTGTTC 1440

GCTTATTTAA CACGATTATG ATGCGAGCGG GAGTGACATT GGTCATCACG ACAGCATTTA 1500

GTATCTCTAC GATGTTTTTC AATATTTCAG CAACCTATCC GTTCTTTATG GTAGCATTTC 1560

TACAAATTGT AACCTTTGCC GGGATTTATT TCAAATTAGG CGACATTATG AGTATGTTCA 1620

ATCTCCAAAG TAACGATAGT CAGTCTATGA GTAGGCGTGT GATGCGAAAA CCTCAAATGT 1680

TAATGAACCG TAAATTTCGT CAACTCAATC GAAATGTTGG CCGAGCGTTG GCTTTTGATG 1740

GCGGTGTAGC AGTTGCAGGT AAATTAGCAA CTGATTCTAC AAAAGAAAAC AGAGGTTCAA 1800

AAAGCAAAAG ATTATCTAAT CGAAGAAACC AATCGATACA AAGTAATCAT GAGCAAAATA 1860

CTGATCCGAA GAAACAAGAA AAGACTACTA ATAGTAGAAC CACTCGATTG AATCTTATGG 1920

GTAGGAAAGC TGGAAAGGCA ATGGATTCAA AATCTATCGT GAAGGACAAA GTAAAACAAG 1980

CAAAAGACCA TGTAATTGAT ACGCCAACAA ATTTGAAATA TAACCTGCAT AGAGGAATTG 2040

AAAAGGCGAA AAATGCCCCA GGGGAATTTA AACGCGGGCT TGTGCAGGAG AAAAAAGAAC 2100

GAGAAGAACA AAGAAAAGCT AATCACCAAC AACGGCAGAA AAGAATGGAT GAAAAACGCC 2160

AAGTGTTAGG 2170

(2) INFORMATION FOR SEQ ID NO: 544:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1141 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 544:

CAAAGAATAT TGGAGGACAA AATGATCAAT CAACCAAATT CAGTTGATCA ACCAATCCTG 60

ACGAACCAAG AAAAAATGGT CAAAGGCTCA GCTTGGATGA CAGCTAGTAA TATTATCTCC 120

CGCATGTTAG GGGCGATTTA CATCATCCCC TGGTACGCTT GGATGGGCGA GCACGGAAAC 180

GAAGCGAACA GTTTGTTTTC AATGGGGTAT ACGATTTATG CCTTATTCCT AATGATCTCT 240

ACCGCCGGAA TTCCAGGTGC GATTGCCAAG CAGACCTCGC ATTATAATTC ATTGAACGAA 300

TATAAAwTTA GTCGACAGCT TTTTwATCGk GCCTTACaAC TAATGGGtGG CTTAGGTGTG 360

GkTTTTGCTA ttGtGATGTA TTtGGCATCT CCGGCATTAG CAGCTTTGTC TGGCGGAGGA 420

CCTGAATTAG TTCCAACAAT GCGTyCATTA AGTTTAGCCG TTTtAGTTTT CCCArGCATG 480

AGTGTCATTC GCGGTTATTT CCAAGGAAAT CAAGAAATGA TGCCCTTTGC TTTATCACAA 540

ATTGTTGAAC AAGTAGCACG AGTTTTTTAC ATGTTATTAA CTGCTTTTAT TATCATGAAA 600

GTCTTTGAAG GTAATTACGT TACAGCGGTG ACCCAAGCCA CTTTTGCGGC ATTTATTGGG 660

ATGCTCGCAA GTTTTGCGGT TTTAGGTTAT TACATGTACA AACAAAAACC ATTATTTGAT 720

TATTTAGAAG AACACAGCGC CAATGAACAT GAAATTGCTC CAAAAGAGTT ATTGATTGAA 780

ACTTTTAAAG AGGCGATTCC TTTTATCGTA GTCGGCTCAG GAGTGACTGT CTTTAAACTT 840

GTCGATCAGT TTACTTTCTC AAACTTTATG CGTTTGTTTA CGGaTTATTC AGATACCCAA 900

TTACGAGAGC TTTTCGGTAT TTTTAATGCC AATCCGGACA AATTAACCAT GATTGTCGTG 960

GCTTTAGCGA CTTCGATTTC GGCTACTGGa TTACCGTTAA TTACAGAAGC GGTCACCTTA 1020

AAAaATCACC GTGAATTAGC GCGTTTGAtT AGTAATAATT TACmACyTTT TGtGkTTGTA 1080

ATGtGCCAGC AACGgTCsGt ATGATGTTTT AGCGnCACCA CTATACACTG GTTTTTACAT 1140

G 1141

(2) INFORMATION FOR SEQ ID NO: 545:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 965 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 545:

TTAGATATGC CCTTATCAAG TCAGGCTCTG TATTTTCATT TAGCGATGCG TGCCGATGAT 60

GATGGATTTA TCAATAATCC CAAAAAATTG CAGCGAATGG TCGGTTGTGG GGAAGACGAT 120

CTAAAATTGT TGATGGTTAn AAAATTTATT CTAGTATTTG AAAGTGGTGT GATCGTTATC 180

AAACATTGGA AAATTCATAA TTATATTCGC AGTGATCGTT ACAAACCAAC CTTGTATCAA 240

GAAGAGArAA ATCAGATTGT TGAAAAAAAT AGCAAAGCTT ATACGTTTAA AGCAGAATCG 300

TCTGTCGGTG GTCAACCAGC TGACTACCAA CGGTTACCAC AGGAAAGCAT AGTCCAGTCT 360

AAGTTAGGTC AGAGTCAAGG GCAGTAGTTC AGAAAACGAT TGTTTAAAGA TGATTTATCA 420

TTTTaTGAGG AAAACGGCTT gGTACACTGG CCTCAAAAAC AAGCCArGAT TTTaAGTATT 480

GGTTGCAAGA TTTTATACAA AAAGGGGCTA GCCAAGAGGA AGCATGCCAA TTAATCTTGC 540

ATGcTTTAGG AATTGCCGTC GATCGAAATA AACGGAATTA CGGCTATGTA AATGCTATTT 600

TGAAAAGTTG GGAGCAACAA AATTATTTAT CCGTACATGA AGTTCTGGTA AATGATAAAA 660

AACAAGTGTT GGAGCATGCG CCGCAAATGA CAGAAGAATA TCAAGAGTTA GGTTTTTAAA 720

GAAAGGAGGA AATCAGTATG CATGCGACAG ATCnAACTTT TCCAATACTA TTGAGTCAAT 780

TGTTAGAAAA AGTTGAAGAC CGTTGTCCTG AATGTGGCAG TGAACAATAT GTTTGGCCAC 840

AAAAAAATAA AGATGGCACm GAACGTTGTG CCCCAACTTG TTGGTCGTGT GGGTAtAAAA 900

TGCTAAAAAA ACATGAACAA nGAGCCAATC AACAACGTTC TCAAGAGAGT TTTATGGCAC 960

GTAAC 965

(2) INFORMATION FOR SEQ ID NO: 546:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 716 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 546:

GACAAATGCT GCTGCCGCGA AAGGATCGAG AGACCCTTAG TTGGTTGAAC AGGAATAATC 60

GTGAAGGTTA TTCATAATGT AGCTGATACA TGTTATCTAC GCCCTATAGT GCTAAAGmTC 120

AACGATAGAG AATCTTATTT AAAGATGGTA TTTTGGGACG TACTCTATTG TTGAGTAGGT 180

CCTTTTTTGT GTGTTTGGCA AAAAAGCAAT TATTCACTGG AAGTATAAAT AAGAAAAAGA 240

AAGGATGACT GTCATGGAAG ACCCTGGAAG TGAACAGAGA AACAAAATAC AGTCACCAAT 300

GAAAGGAGAA GATTTTAGCG CTCTTTTCGG CAGGTGACCA CAACCTTGGA GGAAAAGAAA 360

ATGATGAACA AAAAAACAAT GGaTATGAGA AAAGCAACCA AAGaTCAAGa ATTACGTAAA 420

TTGGCTTTGC TTTCTGAACG GGAATTAATG ATGGAATTAC GGACATCTGA AAAAGGACTC 480

TCGAATGAAG ATGCCGAAAA AAGATTAGAG GAATTTGGAC CAAACGAaGT TTCTGCACAA 540

AAGCCAACCC CAGCAATTAT TTTATTTTTA AGTGCATTCA AAGATCCGTT TGTCTATGTC 600

TTAGCCTTAT TGATGGTCGT TTCAACTTTA ACGAAAGATT TCGAAGCCGC TATCGTGATG 660

GGCGTGATGA TTTTAGCCAG TGTCTTGAnT GCTTnATACA AGAGTATCGT nCTCAA 716

(2) INFORMATION FOR SEQ ID NO: 547:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 963 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 547:

TAGACAATTG GTCCCCACTA CAACTATCCA GTTATAACnA AGAGCGCCCA CTCGAACCTT 60

TTTCTGTAAT ACGCCTAATG TCATTACCAT CATATTGCCT GTTGCTAACT CACCACCAAC 120

AAAAGTAAGC GCTATCAACC CAACTGGAAA TAAAGCTCCA CCCAGAAAGT TGTTAAAcTT 180

CCCCmTTCTG TGGGGGCAGT TCCaGTGATA CGAATAAACG CCAAGTAACC CAgTGCAaTA 240

CAAAAACCAG CCaTGATGCC TAAAATACAG AGTCGAACAA ACGATAATTT CCCTTTCATT 300

TCTGCCTTAT CACCTAAGCT TGCAACTATT TCTTCACTAC TGAAAAAATC CATTACGATT 360

CCTCCTTTTT ATCGATACCA CTGAGACAAT TCTTCTTTTC GATTGACATT TTTAAGTTGA 420

TTCTCTTTTG ATAGTTGCAC CGTTTTTACA GATAGACGCT CAAATTCTTG CCGAATCTGA 480

TAGTTATTTT GCGCAAGCTG TTTTTGAAAT GTTGGTAAAC AAGAACATCG ATAAAAGCCA 540

AATAATGGTT CTTGACGACC TTCAAAGTCA AAGATTGTTA CTTGATTTTG TCCACGCTTA 600

GAGGCTAACT GATACACAAG TTTAGGAGAA AAAGTTGGTA TATCACAGGC AATTACAAAC 660

ACTTCAGCTT TCTGAGTATA GTTCATTGCT GTCACAATCC CACCTAATGG CCCCTTTCCA 720

AGATACATAT CTTGAATGAT TGTGCATTTT CTAAACGCCT TAGGAAATTT TTGACTATCA 780

TTCGTTACTA AAATAACCTC TTCGAACAGT TCCATTAACT TTTCAGCTGT TTGCAAGATC 840

ACGTATTGCC CTTGAATTTT CAATAATGCT TTATCAAATC CCATCCGTGA ACTTTTCCCG 900

CCGCATAACA GCACTGckGT ACAATCAGCA ATAATTTTCG CCACCCCTTT TACCATAATn 960

Tnn 963

(2) INFORMATION FOR SEQ ID NO: 548:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2744 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 548:

CACATCCTCT CTAAGTTTTC TTCTACATCG TCGGAAAGCT CGTTTTCTTC TAATTCTATA 60

TTTTCCATCA AACAAATAGC TGTTTCTATC TCTTCAATCT CATCTAATAC GCTCTCTATA 120

AGCCTGTGAG GGCAATTAAC CTCTTCTTTA TACTTGAATC GCCTAATGAG CAATTGTTGC 180

TCTAAAGGCT TAAGAGTAGC CCAAAACCGC TCAAAATGTT TTTGTCGAAA TAATGTGCGT 240

TCGATTCGTT TATCTATTAG CTCTAGCTTT AACAAAAGTT TTTCCACTTC TTTATCTGGT 300

CGGAAGCCTT TTGTAGTCAT TCCTAACTCA TCAAACACAA TATGCGTTGC CATATTTTGG 360

CTGAAAAATT CAATACGCCT TAATTTTTTT ATTGAAAGCA ACTTTTGCAA CACTAACGGA 420

AATGAGAGAT AGTTTTTTAT AACAATTTCA TCAAATGGAT CCATTGACAA CGTTCCTCTC 480

TTTCCTGAAT TGTCTTTTTT AGGCTGCCGA AAGTGCCGAC TTTTTTAGGT GTCTGATTTG 540

TCAGATTTTA AACAAAAAAG AGGACACAAG ACCTCACTCA ATTAAGAGTA AGCTGTCTCA 600

TATCCTCTGG TTCTTCCTAG TAAGATACAT AGCGCTAGAA TAGCAAGATA TACTTATTTA 660

ACTGTCTTTA TTATAGCACT TTTTTCGCCT AATAGATAGT TTAGACTGCT AATTCTTAGG 720

TAAAACATGC GATGTAATTG TATCGTATTT TACTACCTTC CCATCTTTTA CAATCAACTT 780

AACTTCCCCA AATTTTGGTA GTTCTACCAC TTCCGTTTTT TTACTGTTAT CAATGTATAC 840

TTTATCCATG TTAGCCTCCT GTTAATAAAA ATCTTTTAAA TTCTTTAGTT TTCTTAATAG 900

TAAAAATAGC TGTGCACGTT GTTAAGTAGC TCTTTAGCAC TCCTTAAGAT GCATAATTGT 960

GTTTGTTTTG AAATATTTTA ATTATACAGC GTTAGTAATT GTTAGTATTC ATACTTTTTC 1020

ACACCTAATT AAGGAAAAAG TCAAGTTTAG TCAAGTTTTT ATATAGAATA AAACGAGATC 1080

TTGTAAGGkT CTTTTTTAGG ACGGAAGAAG CGGAAGATTT CATGCCGTTA TTTCTTCTTC 1140

AATCTTTTTA GTTTCTTCGA TAATCGCCGT TTTTCAATCT CTTTAGCGTT GTTCCCTTTA 1200

GAAGCTCTAC GGTCTTTAGA CTGCACATTT AATGCATGTT TTGTGCCTGC TGGTTTTTTC 1260

TCTCTAATCA TGACTATTAT TCCTCCTCAT ACCAAATAGA AATAGTTGTA ACTGACGTAA 1320

CCTCCTCAGa ATCTCCGTAG CTTTCCCCaG TCTGTTTAAT GTGTTTGaTT TCAATATACG 1380

GGTTCTCCTC TAAAAAATCG TTTAATTCAC GCTCTAACTT GTTGATATTT CCtCTTTGAA 1440

AAATTTTAAC TTTCATAATC GCTACTCTCC TTGTCTTATA ATGTACAATT TGTAGTATTT 1500

TGTAGTATTT AAAATTTGGG TTGACTATCA AGATACATTA TAAACTCATC CGAACATAAC 1560

AGAAGGAAGG TAAATTTATC AAAAAGATCT ACGTACAAGC TCTCCTGATT CTCAGTAAGA 1620

TTATCTTTTG CTAAGGCAAA GTAAAGCCGA ATCAATCGCG CTAATTCTCC TTTAAGCTGT 1680

TCGACACTCG AAAAATTTTC ATCTTCTAAT AAAGACTGTA CAATTACTCC TTGCATGCCT 1740

CCTGCAGCTT CTTCTAACCA ATTAGGGTCT TTTTCCATTT CCGCTAAAAC GATAGGGGAA 1800

AAATGAGAAA TATCTCCAGT AAAATGACCT CGTGTAAAAG CCAGCATATT GGATAAAAAA 1860

CTTCTTGTAA TTTCTTCATA TTCAAAATTC GGTTCTTTTT CTGTTGATAA TATAAGCATA 1920

AGATTTCCTC TTTTCTTTAT CTTTTTTTGT TTTATTCGTT TGGTAACTAA TTCTAAAAAA 1980

AACTGGTTAC CGCTTAAAGC ATTGGGGGAG TAAGGATAGA GCCATTTGGT AACTAGTTAC 2040

CAATTCTTTT TACTATTCCT CAAACTTTTA TTAAAACTTT ATTTACAAAT TTCAATAGAT 2100

AAAATTAGTA ACTGGTTACC AAAACGCTCC TCCCTCACTC TCCCAAGGGG TGGGAGTGGT 2160

AACCGATTTA TTTTGAATTG GTTACCATTT TTTTATAAAT TTCTCTTCCC TCTTACTCTC 2220

CCAAGGGATT TGGTCGGTTA CCAATTTATT TTTAAATCGG TAATTTAACT TATCAGCTTT 2280

AAATCGTTTC TTACTAAAGA TTGGTAAGGC ATTCCATTTT CTGGTGGTTC TTCATAGCTG 2340

AAAAAATCTC CCAATGCTTT TATTTTTTCT ATCTCATTCA CACTATATTT TATTTTTTTA 2400

TTTATCCAAC TATCCCCTAA TAAATATGTG AACTGTTTAG AAAATTTAAT TTTACTCAAT 2460

GGTTTAAAAT TATTTTCCTT GCAAAACTCT TTATATTGAG AGTAAACCAT ATAAAAAGGG 2520

ATTTTGTTTA TTCCTAAAGG CTTGAATACT GTTTCTCTAA AATCAACTAG AGGGTCATTT 2580

TCTTGTTTAT ATTCTTCTAA AAGTTTTTTG GAAACATCAG GTACTATAAA ACGTTCAAAA 2640

TCCATGTTTA CCGCTTTATG CAAAACGTAT TCGAGCACCT CTTTTCGGTT GAtATACTCT 2700

TCTTTTATTC TCCAATCATC TTTAGCACCT TCAAATGAAG CATT 2744

(2) INFORMATION FOR SEQ ID NO: 549:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2165 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 549:

ATTnATCTTT TTCnGGCTCG GAACCAATAG ACCCATTTTG CCATTCTAAA CTGATCTAAA 60

ACGTAATAAA ACAGTTTGAT CCATCTGGAT TGATTTCAAC CCATTTAGGT AGTTTGTAAG 120

AACGTTTAGC CAAATTTCAA TACTTCTTTA ATTGGCAAGC CCTACGTTCC TATAGATAAA 180

CTTTTTGCTT TTAATTGTAC ATTACCATTC GCCATCACGT AAGGATCAAA ATATAAGTAA 240

AAAGGAATGT CATGCCCTAA CACCTGAAAT GTTCCATTTA ATAAAGCATC ATTTTCTAAA 300

```
TAAAATTTAT ATTTGATCTC CGAACCTTTT TGAAAGTCGG CTAAATAAAA GTCAATCAGT    360
TTGTTCACTT GTTGTTTTTT TGACTGAATG GTGACGACTG GCTCGCCTTC TTTTTCAACA    420
ATTGCTGGTA TTTTCTTTAA ATCTGGTTCG CGAACTTGTG TTGCTCGAAA TGTCACGAAA    480
GCTACACTAC CAATGATTAA ACCCACTAGA ACAAGAAAAG CGATTTTCCA GGGATTTCGT    540
TTAAGATTGG TTGGTTGTGT CTTTTTAATT GATTTACTTG TTTTTGGCTT TTCTTCATTC    600
ATTTTATTCA CTACCTTCTG ATTTGGTTAT CCATTCTTTC TCTGTCTTGA CCATTTCATC    660
CCGGACAGCT CCGGCCATAA TTTGGTAGCC AAGATTATTC GGATGGAAGC GATCTTCTTC    720
ATATAACAAA TTGTTTAAAT CCTCTTTGCT AGCACTACTT CCTGTAGTTT CTGAGTCACC    780
ACCAGTTACG CCAACTTCAT CGCCACGACC TTTATAGAGT AAATCATTGA TAGGAATGAA    840
ATAGGCTCGT TTTTGCTCTT GAACCATTTC TTCGGTTGCy TGATTCCAAT TATCAACGAT    900
TTCTTGCATT TCTGTAATTT CGGAAAAGTT TAAGTAGAAG GGATTGTAAA TTCCTAAGAC    960
ATAAATAGGG GCCTTCTCGT TGTACTCTCG GATTTCTTCA AGTAGTCGTC TTACTCGGCG   1020
TTGATAGGCT TTTtGTGGAC GATTGAACGA GCTGACTTTC AAGTCGAAAA TATTACTACT   1080
AATTACTTTC ATTAAGTCAT TTCCACCAAC CGTTAGCGTA ATCACGTCCG CAGAGGCAAG   1140
GCCCTTTTGA ATTTCAGGTT TCTCTTTGAT TCGTTTTAAA ATTTGATCAC TACGATCCCC   1200
ATTTTTCCCA AAATTGTCTG TTTGAACACC GTTCAAGTTG TAGTGTTCTT TTAAATCATC   1260
TGCCACAATA GGAACAAAAC CTCCACTATT CGTCAAATCG CCAATTCCTT CAGTTAAAGA   1320
ATCACCGATA GCTGTATAAT GAATGACTTC TTTTTGATTT TTTTGAGCAG TTGTGGCCAC   1380
TTTCTCTTGC TTTAATAGTG GCTTTGCTTT GGGGATGGCC ACACTTAACA ATGTAAAAAC   1440
ACCTAGCGCA ATGAGGATAG GTGTTAAGAC GGTCAGCAAA ATATGCTGTG TTTGTTTTTT   1500
CATCTTCGTC ACTTCCTTTT AAATAAACCA ACAAGAAATT TTCCTTTTCT TGTTAGAAAG   1560
AGGTTGCGTC TACTCGTTTA AGAGTGAGAC CCAACCTCGT TATTTCTTAG TCTGTATAGT   1620
ACATAATAGC AAAGGCATTT TTcCTGTATG TGTTGCAATG ACTGGaTTCG TATGTAATAC   1680
GGGaATATCC ATGTCTTTaA AGATTGCTTG TAATCCTTCT TtGAATCCAT TtGCTAGTTC   1740
TAGCCCATCC gCATGAGAAA TGCCAATTTG TCGAACATTT GGAATCTTAC TTAATTCTGA   1800
TTTCAATTCG TCAAACCATT TATTAAACGT TTTAACGCCT CGGCCTTTCG CTACAGGAAT   1860
CAATTCCGTA TTTTCAAAGT CCATGACAAC TTTCATATTA AAAATGTTTG ATAATAATCC   1920
TGTTGTACGG CTGATTCGTC CACCTTTAAC CAAATTATCC AATGTTGAAA TGCCAATGTA   1980
TAATTTAGTA TTTTGTTTGA CGCGTTCAAT TTCAGCTAAA ATTTCTGGAA CACnCGCTCC   2040
CGCTTGTGCC AATTTTGCTG CTTGAATGAC TTGGAAAGAC AAACCTTGGT CTGTAAAATC   2100
GCTATCAATC ACCGTTACTT TACTTGATGA TAAGTTGCTA GCTTGACGGG CCGCTTCAAn   2160
CCGTT                                                               2165
```

(2) INFORMATION FOR SEQ ID NO: 550:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9797 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 550:

GGTGAAGCAA CCCTTCTGTT GCTCCACCAA CTCCTAACTA GACCGTTTCT CGATGCCTTA 60

GATTATTTTG ATTCTTTTTC GCCTGCTTCT AAGTAGCTGA CATCATTGAC TTTATCCAAA 120

GTATATTCGC CATTTTTGTA ATGAATTGTT GTGACACTAG CATTCTTCAA ACCGCCTTCT 180

GGGACTTTAA AATCATCAAA TAAAGTTGCT AACAACGCTG AAATACTCAA GCCATGAGAG 240

ACCACTAAAA CATTGCCATT CCCAGAATTG GCTGATTCTG TGGCAACAAT TTTATCTAAG 300

CCTTTTTTCA AACGTTTAGT AATTGTAGCA TAGTCTTCTG CAGGCCAGTT ATTCTTGCTT 360

TCCTCGCGCT GTTGATCCAG TTTAGCTACA CTATTGGCAA AGGATTCAGG AGTCATGTTT 420

TTCATAAATT CTTCTAAGGA AACACCTTGA TCATCAGCAA TATCCTGCCA CATTGTCTTA 480

TTTAAATCCC CTTCATAGCT ACCAAAATTA AATTCACGTA AATCTGGGTC ACGCACGACT 540

TCAAGGTCTT TGCCTGCTTT ATTTTGATCT AAAATAAGTT GAGCAGTTTG CAAGGCGCGG 600

CCACTATCAC TACTATATGC ATTTTGAAAG GCAACATCTT TCAGTCCAAT CCCAGTTGCT 660

GTCACAACTT TTTCACCTTC TGGTGTTAGG ACCGCATCTG ACCATCCTTG TACGCGGTCC 720

GTCGTATTTA ACATGGTTTT TCCGTGGCGC ACAATGTAAA GAGTTAGTTC TTCAGGTTTT 780

GTCTCtTTTT CTTTTGTCGT TGCTTGCTGA TTTTTACCAC AACCAACCAT TACTAAAATG 840

GCCAGAAAAA GCATCCCCAC AATCGTCAAT CGTTTTTTCA TTTTTTATTT CCCCTCATCT 900

TCTACTCTGT TGTTAAGTTT TCACCATTAC TAGCGATAAC TTGTTTGTAC CAATTAAATG 960

ATTTTTTCTT AGTTCGTTTC AAGGTTCCCT GACCTTGGTC ATCCATGTCT ACATAGATAA 1020

AGCCGTAACG TTTTTTCATT TCGCCAGTCC CTGCTGAAAC TAAGTCAATA CAACCCCATG 1080

GCGTATAACC AATTAAGTTA ACCCCGTCTT CTTCTATTGC GGTTCTCATG GCTTCAATAT 1140

GTTGTTTAAA ATAATCAATG CGGTAATCGT CGACAATCGA GCCATCCTCT TGCAATGTAT 1200

CCACCGCGCC AAACCCGTTT TCAACCACAA ACAATGGcAT ATTGTAACGA TCTGACAACT 1260

GACACAATGA AATTCTTAAC CCTAAAGGAT CGACAGCCCA GCCCCATTCA GATTCTTGCA 1320

AATAAGGATT TTTAACCGCA GGCATTTGGT TGCCTCCGAC ATATTTTACT TCTTCGGGAT 1380

AAGCAGTCGA AACAGTCGAC ATATAATAAC TAAACCCGAT ATAATCAACT GTTCCTTTTT 1440

GAATAATCGC TAAATCATCT GACTCCATTT GAATTGTCAT TCCTTCACGC TCAAATTCTT 1500

TGAGCTTATG CGCTGGATAA TAGCCCTTAG CCTGGACATC CATATAAAAT TCTTGCTCTC 1560

GGTTAAATTG AACAGCTGCT ATCACGTCCT CAGGTTTGCA GGACAAAGGA TAACTTGGAA 1620

TATAAGCCAC CATCATACCA ATTTGGAAAT CTGGATTGAT GGAATGTCCT AATTCAACTG 1680

CTTTCGCACT TGCTACAAAT AGATGATGCG CTGCTTGATA TTTACTTTGT TTATCATTAT 1740

```
TGTGAATCCC AATTTGCGTC CAACTTCTTA AAAAATTAAT TTCATTAAAA GTCATCCAAT     1800
ACTTCACTTT ATCCTGATAA CGTGTAAAAA TGGTTTCGCA AAAGAAAACA AAGAAATCAA     1860
TCACTTCACG ATTTTTCCAA CCATCATATT CATCGGCCAA ATACATCGGA ATATCAAAAT     1920
GGTTAATCGT CACAACGGGC TCAATTCCAT GGCGCAGCAA AGCATCAAAA ACTTTATCAT     1980
AAAAGGCTAA GCCTTCTTCA TTAATCTCTT TTGTTCCATT TGGGCAAATC CGGCTCCAGG     2040
CGATAGACAG CCGAAAACAT TTGAAACCCA TTTCTGCAAA TAAGGCGATA TCTTCTTCGT     2100
AATGATGGTA AAAGTCCGTC GCTACATGAC TAGGGTAGTA TAATTCTTCA TCGATATAGC     2160
CAATCGCTCC TTCGGGTAAT GCTTGATCGC GCGTACACGA AGCTTTTTGG CCTTCTTTAG     2220
TTAAATACGT AATCCGTCGC GGTTCTGTTT GACTACCACC AGTAATTGCA TCTAATGTAC     2280
TCAGACCTTT GCCTCCTGAA AGgTAGCCAC CTTCATATTG GTTAGCAGCT GTTGCGCCGC     2340
CCCATAAAAA ATTTTCTGGA AATCCCATAA AATCCCTCTT TCCCTCTGTT TTTAGGCATC     2400
ACAAAGAAAC CATCAAGGGA ATTGACGCTG GCAGAAGAGT CCTTGTTTTT TCTACCAACG     2460
TCAAACGCCC TTTAATAGTT GCTAATTTTA TTAATTACTT AAATCTTCAC CATTTGTTGC     2520
GATTACTTTT TTGTACCAAT CAAACGATTT CTTCTTGCTT CGTTTTAACG TACCATTTCC     2580
TTCGTTATCG CGATCCACAT AGATAAAGCC ATAGCGTTTT TTCATTTCAC CTGTTCCAGC     2640
TGATACTAAA TCAATACAGC CCCAAGTCGT ATAGCCTAAC AAGTCGACAC CATCTTGTTC     2700
AACAGCATCT TTCATCGCTT GAATATGCGC CGCCAAATAA TCGATACGAT AATCATCGAC     2760
TACATCACCA TTTTCGTCTG GTGTGTCTAC CGCTCCTAAG CCATTTTCAA CAATGAATAG     2820
TGGTTTTTGA TAACGATCAT ATAAGTCATT CATCGTAATT CTCAAACCAA GCGGATCAAT     2880
TTGCCAACCC CATTCACTGG CTTTTAAATA AGGATTTTTC ACAGAGGCAA AGATGTTACC     2940
AGCTGTTTGT TCATTAATTT CTGGATCTGT TGATGTGACA CGAGAAGAAT AATAAGAGAA     3000
CGAAATGAAG TCCACCGTAT GTTCTTTTAA AAGTTCTTCG TCGCCTTCTT CCATTTCAAT     3060
TTGAATACCT TTACGGGCCA TTTCTTTTAA AGCATACGCA GGGTATTCCC CACGTGATTG     3120
GACATCAATA AAGAAATAAT TTTCACGATC TGCTTGTCGA GCAGCGAAAA CATCTTCTGG     3180
TTTACAAGTA TAAGCGTAGT TGGAACCAGC CGCTAACATA CAACCTACTT GGTTTTCAGG     3240
ATCTACCTCA TGAGCAATTT TTGTCGCAAT TGCACTTGCT AGTAGTTCAT GATGGGCCGC     3300
TTGGTATTTC ACTTGTTCTT TGTTTTCGCC TTCTTCAAAG TATAGACCTG CTCCCATAAA     3360
TGGTGCATGT AAAATCATAT TAATTTCGTT AAATGTTAAC CAATATTTGA CTAATCCTTT     3420
ATATCGGtTG AAAATTACAC GGCAAAGATT TTCATAGAAA CCTACTAACT TACGACTACG     3480
CCAAGCGCCA TATTCTTCCA CTAAGTGCAT TGGGcAGTCA AAGTGTGTAA TCGTTACTAG     3540
TGGCTCAATC CCATATTTAT GACATTCTTT GAATAAGTCT TCATAGAACT TCAAGCCTTC     3600
TTCATTCGGC TCTGTTTCAT CTCCCATTGG AAAAATTCGA CTCCAAGCAA TGGATAAACG     3660
GTATGTTTTA AAACCCATTT CACCAAACAA GGCGATATCT TCTTTATAGC GATGATACAT     3720
```

ATCAATCGCC TCTTGTGCAG GATAGAAATG TTCCTCATCA AAGTTAAACA TTTTCTTTTC 3780

ACCAGTGATT ACTGGAAAAC GATCTGGACC AGTTGGCGCT AAATCGACGT TTGCCAAACC 3840

ACGTCCGCCT TCGTTATAGC CACCTTCACA TTGATTGGCT GCTGTTGCGC CACCCCATAA 3900

AAAGTCTTTT CTAAATGCCA TAAATTCTCT TCCTCCACAT CTTTTTATAT TAAAGCGGTC 3960

AGCAAATCAT CACTTGTTGA AACTTCATTA CTGCCGACTT CCAAAATATC TAAATAATCC 4020

GCTGTATTCG TCACAACGAT TGGAATTTGT GTACTAAACC CAGCTTTTTT AATTCCTTCT 4080

AAATCAAAGG TTACTAATGT TTGACCTTTT TTCACGCGGT CACCTTGTTT TACAAAAGCT 4140

TCAAAGCCTT CGCCTTCTAA TTGAACTGTA TCAATCCCAA TATGAATCAA TAATTCTGTG 4200

CCGTTATCTG AAATCAAACC TAATGCATGT TTTGTTGGGA ATAGTGTGAC AATCGTTCCA 4260

TCAAAGGGTG CAACCACTTT ACCTTCCGTT GGTTCAATCA CGACCCCATT TCCTAATGCT 4320

CCATTTGCAA AAGCAGGATC CTCTGCATTT TTTAAAGATA ACACTCGTCC TTTGACTGGA 4380

CTTGTAATAT TTTCTTTTTT AATCGTTGTT TTATCAATAA TCACTTCTTC TTCCTCAACT 4440

GTATTGTCTT TCCAGAAGAA GAATGTTAGA CCAAAACCAA CGACCGCTGC AATGGCAATT 4500

GCAATAAAGG ACTGGATCAT CCCACTTGCA TCATCACCAT TGATAAAGTT TAGCACACCA 4560

AAAATTCCTA AACCACCCAT TGTATAAGCT GTCACATTGT TGATCATCAA ATATAAACCA 4620

CCGACAGCGC CACCAATCAT TGAATAAATA AATGGCCATT TTTTTGGTAA AGTGATTCCA 4680

TAAATTGCTG GCTCTGTTAC ACCAAAAATA CCAGAGATAA TTGCTGGCGG ACACAAAGCT 4740

TTTAACTTTT TATCTTTCAA TTTGAAGAAC ATCGCTAAAA CAACCGCTGT TTGCGCAAAA 4800

CTAGCTGCAA ATGATCCCGT TAATACTTGG CTTGAACCTT CTTGTGTCAC TTGCATAATA 4860

GCTAATGGCA CAACGCTCCA ATGTAAACCA AAAATAACTA AGACTTGCCA GAAGAAACCA 4920

AGAATCAAAC CATACAAAGC TGGTGAGAAA CTCATTAATG CTGTAAAGCC AGCGCTTAAT 4980

AAATCTGTTA GCATGCTAAC GATTGGCCCA ATTACTAAGA AACCAATTGG TAAGGCAATC 5040

AACAAGACGA AAAACGGTAC TAAGAACGTT TGAACAACTT CTGGAATAAT ACGTTTAAAC 5100

ACTTTTTGAA CTTGTGCCGC AAAAGCAATA ATGAAGATGA TTGGTACGAC ACTACTTGTA 5160

TAGTTCGCAC CGACCCAAGG AATCCCCATA AATGTATTAT AGGCAGGTAA ACCGAATAAA 5220

TTATACGGTG CTGCGGCTCC GGCACCCGCT GTTGTTTCAA AGGCCGTTTG TAATGCACTT 5280

CCTTGAATCG TTGGGTAACA TAAAGctGCG CCAATCACGA TCCCTACCAT TGGATGCAAC 5340

CGGAATTTTT TAGCTGCTGT ATAACCTAAA ATAACTGGCA TAAAGTAGAA AATTGCATCG 5400

CCAATCCCAT TTAACATGGT ATAAGTCCCT GACGTAGCCG TATAGAGCTT TAAGAAAACC 5460

AATAAGGCAT TTAACCCTTT GACCATCCCA GCCGCTGctA ACGCCCCTAA AAAGGGTTGG 5520

AAACAACCAC TTAAAATATC TATTAAGCGA TCAAATAAAT TCCCACTGGA AGCTTCTTCT 5580

TCACGTTCTC CTGACAAGCC TGCAATACTA ACTACTTCTT CGTAAACGGC TGGAACATGG 5640

TTCCCAATAA CGACTTGGTA TTGTCCACCG CTTTTCATAA CAGTTACCAC ACCGTCCATA 5700

TTTTTTAAGA CATCATCATT CGCTTGACTC TCATCTTTTA ATTTGAAGCG AAGACGTGTG 5760

ATACAATGTG TCAAACTATT AATATTCTCT TGCCCACCGA CATTTTTTAC AATTTTTTCT 5820

GCTAGTTCAT GATATTTTCC CATCTTTTTT TCCTCCAAAA ATTTATTTTC ATTTGAAGGT 5880

TTAGCCAACT GAATGTCACT ATCCAAAGTG GTGCGTTTAT TGTTCCGTTG TTTGTACAAT 5940

TCGTGCAATG TGAATCGTTA GGTACAAGCG TTCTTCATTT GAAACTTGAT AGTCATAGTT 6000

CTCATTTAAA TACAAGCTGA TCTTTTCCAC GCAACGATAG GCGTTGAGAT ATTTCTTTTT 6060

GATTACCGCA TACAGTTCTT CATCATCATT TTCGACAACT TCACGTTGGT GTAATAAGCG 6120

ATAACAGAAA AATTTTAATG TGTGCTAAAT CGATAAAAGT AGACAGAGTC TTCATCAAAA 6180

CTTACCTTAA AATAGTATTT GACGATATTT GTAATCTCTT GCATCAACTT TGTCAATTCG 6240

TAAATATTAC CGACTTCTTC CTCCATCTCC GCATTTACCA AGTGCAATGC GATAAAACCT 6300

GCCTCATCCG TTGACAATTC CATTCCTACG CGCTTTTGTA CCATTTTTAA AGCATTTACT 6360

CCTACTTCGT ACTCTTCAGG AAAAAAACGT TTAATATCCC ATAATAATAA ATTCTTAATT 6420

GCAATTCCTT GTTTTTTTCT TTCTAAAGCT GTGTACAAAT GATCCGTCAG AGAAATATAA 6480

ATGGTATCAT TCAAACTCTT TTTCAACGTT GATTTTGCAT AAACAATAAT GTCATCCGAG 6540

AGTTTCATCA TCTCTAAAGG AATATTTCCT AACAATTCTT GAAACTTAAG TGATGTTTCT 6600

TGATTGGCTA AATAAAAAAC TTGATCGATA CTTTCATCAG GAACCCAATC CCCAATGCGT 6660

TTCTTAAAAG CAATTCCTTT GCCCATAACA ATTTGCTCTC GTTTATCATC ATCTAATGTA 6720

ATAACCACAT TATTGTTTAA AATTTTGTGA ATAATCAATG CCCCACCTCC TTTGTGTTTC 6780

GTTGAAAAAA GAAAAACCCA ATTTCTTCAA AAAGAACGAC ACACCTGTCG CTTCATGAAA 6840

AAATTAGGTT TAGCTTGTCT TTAAACAATC ACTATCCATC TATTTACTTA tCCCtACaTC 6900

CaAAGAATAA CATATTTACA ACAAATGTCa ACGCTTTCAA CGACATTTTT TTGTAATAAA 6960

AATAAAAAAA GCATACAAAC ATTGTTTGTA TGCTTCACTA TGaCCCGTAC GGGACTCGAA 7020

CCCGTGTTAC CGCCGTGaAA GGGCGGTGTC TTAACCGCTT GACCAACGGG CCATTAAAAA 7080

GAAAAAACGG AGAAGGAGGG ATTTGAACCC TCGCGCCGGT TTCCCGACCT ACACCCTTAG 7140

CAGGGGCGCC TCTTCAGCCA CTTGAGTACT TCCCCAAGAA ATAAATGGGC CTAAATGGAC 7200

TCGAACCATC GACCTCACGC TTATCAGGCG TGCGCTCTAA CCAGCTGAGC TATAGGCCCC 7260

AAATAAGAAA AAAGCGGGTG ACGAGAATCG AACTCGCGAC AACAGCTTGG AAGGCTGTGG 7320

TTTTACCACT AAACTACACC CGCAAATTTT TCTTGGATGG CGCGGGACAG AATCGAACTG 7380

CCGACACATG GAGCTTCAAT CCATTGCTCT ACCAACTGAG CTACCGAGCC AAAAACGGTC 7440

TGGACGGGAC TCGAACCCGC GACCTCCTGC GTGACAGGCA GGCATTCTAA CCAGCTGAAC 7500

TACCAAACCA ATTCGTTTTT GCTTTATGCA AATGTATGCT TTAAAAAATA AAAGCAATTG 7560

CGGGGGCAGG ATTTGAACCT ACGACcTTCG GGTTATGAGC CCGACGAGCT ACCTGACTGC 7620

TCCACCCCGC GATAATTAGA CCAAGGAAAC TTAATTCCTA AGAAAGGAGG ATAAGGGATT 7680

CGAACCCTTG CACGGTTTTA CCCGCCTGAC GGTTTTCAAG ACCGTTCCCT TCAGCCGGAC 7740

TTGGGTAATC CTCCGTAACA AAACTATGAC CCGTACGGGA CTCGAACCCG TGTTACCGCC 7800

GTGAAAGGGC GGTGTCTTAA CCGCTTGACC AACGGGCCAT TTTATAAAAT TATGGGCCTA 7860

AATGGACTCG AACCATCGAC CTCACGCTTA TCAGGCGTGC GCTCTAACCA GCTGAGCTAT 7920

AGGCCCAAAA AGAAAAGCGG GTGACGAGAA TCGAACTCGC GACAACAGCT TGGAAGGCTG 7980

TGGTTTTACC ACTAAACTAC ACCCGCATGG CGGTCTGGAC GGGACTCGAA CCCGCGACCT 8040

CCTGCGTGAC AGGCAGGCAT TCTAACCAGC TGAACTACCA AACCATACTT TTCTTTTTAT 8100

GTGCTTTTGA AATCAAAAGC AATTGCGGGG GCAGGATTTG AACCTACGAC CTTCGGGTTA 8160

TGAGCCCGAC GAGCTACCTG ACTGCTCCAC CCCGCGATAA TTAGACCAAG GAAATTTAAT 8220

TCCTAAGAAA GGAGGATAAG GGATTCGAAC CCTTGCACGG TTTTACCCGC CTGACGGTTT 8280

TCAAGACCGT TCCCTTCAGC CGGACTTGGG TAATCCTCCA TGATAAAAAT AAAACTATAA 8340

ACAACGTCAC AATGGACCTT GTAGGACTCG AACCTACGAC CGGACGGTTA TGAGCCGTCT 8400

GCTCTAACCA ACTGAGCTAA AGGTCCAGGC TCTTCCAAAT GTATAATCGC GGCGGAGGGG 8460

ATCGAACCCC CGACCTCCCG GGTATGAACC GGACGCTCTA GCCAGCTGAG CTACACCGCG 8520

AAACTCTTAT ATTAAAATAA GAATGGAGCC TAGCGGGATC GAACCGCTGA CCTCCTGCGT 8580

GCAAAGCAGG CGCTCTCCCA GCTGAGCTAA GGCCCCATAA AAATTATTAT CGGGAAGACA 8640

GGATTCGAAC CTGCGACCCC TTGGTCCCAA ACCAAGTGCT CTACCAAGCT GAGCTACTTC 8700

CCGTATATAA CTAATGCACC CAAAGGGAGT CGAACCCCTA ACCTTTTGAT TCGTAGTCAA 8760

ACACTCTATC CAGTTGAGCT ATGGGTGCTA AATATAAAAC AATGCCGAGG ACCGGAATCG 8820

AACCGGTACG GTGATCACTC ACCGCAGGAT TTTAAGTCCT GTGCGTCTGC CAGTTCCGCC 8880

ACCCCGGCGT AATGAATTGG CTAAGCGGAA AACGGGGTTC GAACCCGCGA CCCCCACCTT 8940

GGCAAGGTGG TGCTCTACCA CTGAGCTATT TCCGCAAAAT GGTGCCGGCT AAAGGACTTG 9000

AACCCTCGAC CCTCTGATTA CAAATCAGAT GCTCTACCAA CTGAGCTAAG CCGGCTAAAT 9060

ATGCGGGTGA AGGGACTTGA ACCCCcCAnC CTTGCGGCGC TAGATCCTAA ATCTAGTGCG 9120

TCTGCCAATT CCGCCACACC CGCAAATGAA ATATGAATAT GAGCCGTACA GGGCTCGAAC 9180

CTGTGACCCT CTGATTAAAA GTCAGATGCT CTACCAACTG AGCTAACGGC TCAAAAAATG 9240

GAGGTTAACG GGATCGAACC GCTGACCCCC TGCTTGTAAG GCAGGTGCTC TCCCAGCTGA 9300

GCTAAACCTC CAAAAACAAC AATTCGCGTG GCGACGTCCT ACTCTCACAA AGGGAAACCC 9360

TTCACTACAA TCGGCGCTAA GAAGCTTAAC TTCTGTGTTC GGCATGGGAA CAGGTGTATC 9420

CTTCTCGctA TCGCCACCAC ACTGGGTGTT GTTTCTTATT GAGTTGAATC TTCATTCACT 9480

CAAAACTGGA TTGAAGTTTG AATCAAAATA ACCAAGTTGC TTTTACTTAT CCATTCTTTG 9540

GTTAAGTCCT CGACCGATTA GTATTGGTCC GCTCCAACTA TCACTAGCCT TCCACTTCCA 9600

ACCTATCTAC CTAATCATCT CTCAGGGGTC TTACTTTCTT AAAGAAATGG GAAATCTCAT 9660

CTTGAGGTGG GCTTCACACT TAGATGCTTT CAGCGTTTAT CCCTTCCCTA CATAGCTACC 9720

CAGCAATGCC CTTGGCAGAA CAACTGGTAC ACCAGCGGTA AGTCCATCCC GGTCCTCTCG 9780

TACTAAGGAC AGCTCCT 9797

(2) INFORMATION FOR SEQ ID NO: 551:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1807 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 551:

CCTCTTTCGT ATATGTnnCT GTTGTAACGT AATCTCTTTT CAkGATGTGT TCyTCcTAAA 60

AGTTGTTTTT GTTTACACAT TGACTTTACC AAGTTTAGCG AAGAGTTTAC TATCCACCAA 120

AAGTTAGAAA ACGCTTTCTT TTTCGGCGTA ATTTTTCATT AAAAAAAGTT TATTTATGGT 180

TATAAACCGT TGGTACACCA GGGGTTTATA AAAGTTTAGG CGGATTTCTC TCTAATTTCT 240

GACCAATCAT AGATTTGAAA ATGGTATAAT TATTTTGAAA AATCATTCAC TTAAAATAAA 300

AGATGAAAGG GGTTTCCAAT CGTGACATTT GTTTTTATTT ATAACATCTT GTTGATTATT 360

TTATATTCGT TTACTTCAAC CTTTACCTTA AATCTTTACT TGAAAAATAA ACAACCGATT 420

TTTTTATTGC TTCTCTTTTT AATGGTTATT TTTATTTGTG ACAATGTGAT TGTTTATATG 480

ACTGAGTTCA TTAATTCGTT TGcAACCGaa TATaATCaGa CCTTTATGAC CGCACCTTTT 540

TTAAAgACCA TCATTTTTAy TTGkTGtAAT TTTGCCTAnC TGGCTATTAT CAATACTATT 600

AGTGGTCGTC CTTTTAAAAA TTATCAATTT GTCTGGTTAT TTTTAATTGG TTTATGGATG 660

CTAGCTATTC CATTTTCACA AAATTCCGCC TTAAAAGTTT GGTTGTACTA TTTACCCAAT 720

CAGCTATTTT TAATTTACCT TGGATGCTAT GCGCTCTATC AACTACGAAT AGATCCTTTA 780

TCAGCCCTTG CGAAAAAATA TCTTCGTTTT ATTGGTTGGC TAAGCATTGG TTTCGGTGTG 840

GCGATTCTTT TGGAAGATAC CtTCGTTATT TTTAACATTG ACCAATATTC TGATATTGTC 900

TTTAAAATTA ACAACCGAAA CGTTTCTGAA GATATTTATA CAATCATTCT CTCCATCGCC 960

ATTATCTACT TCTGTAATCG TGATTTTCCT CTGTCGGTTC TTGAAAAAGA CGCTGCTAAA 1020

TTGGAAGAAA ATCAATCAGA TGAACCCGTA CTTTTGGCGC CTTTTTGTGA TGCCTATCAA 1080

TTAACGCAGC GAGAACGAGA AGTGCTTTCG TTACTTTTAG AGTGCAAAAC AAATCAAGAT 1140

ATAGCCAACG AACTATTCCT ATCAATCGGG ACAGTAAAAA CACATATCCA CAATATTTTC 1200

GTTAAATTAG AAGTAAACAA AAGAGCAGAA GTTTTTGTCA GTTATCAACT TTTTTCTCAA 1260

CAACAAACTG AACATTTAGC GAGATAATTA AAAAGGAAGC CTCTCCGTGA TCGGTGGCTT 1320

CCTTTTTGAT TAACTATTTT CTTGTACTTG TCGAACTCTC TTTTCAATTT CTTGAAAGTT 1380

AATAATTTTA GCTTGGCGAT GAATTTCTTT TCCTTCATAG TATAAGATAA TGGTCGGTGC 1440

AGAAAAAGCA TTTAATTGTC CAATTATTTC TGGACTTTTA TCAGCTTCTG CAATGTATGT 1500

AGGAAATTCA TATTTTGTTG CAATTTTTTC GACAATCGGT TTATCAACTG TACAAACACT 1560

ACAATTTTTT GCTGTCAAAA AGACGGCTTG TAGACGATtG TCTGTAATTT TTTCTTCTAA 1620

TcyTCtAATG TTGTtACATA TTCCaCTGaA TAGCCCTCCT TGTCyTTTGt CyAAAATAAC 1680

AGATTTACyA GCAAAATAAC AACTTAtTGt TTGGcyAGGa AkGCyTTAAC aAAAaAGTTG 1740

tCGcCCaTTT TAtaAAkGGG GsGaCAACTT TTTACTGGAT AACCGCTTAT TAAnTAAAAA 1800

ATTTAnG 1807

(2) INFORMATION FOR SEQ ID NO: 552:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 862 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 552:

TGAAGTCTGG TTCACCATTG TTTGAACAAA TTGATAAATC CATCGACAAG AAAGTGAATT 60

TATTCTCAAA CAGCTTTTCC CGCTACGCTG TTCGTGCAAT GCTTGCTTGT CTATTTTTAA 120

CGTTGGGAAC AGCCGTTGCT TTTGCAATCG CAATTAAAGG AGAAGGAATT TCTCATAGTT 180

TAGGGAAAAT GCTTTACGCC TTCATGTTCA GTTGGtCACT TGTGaTGATT TTATATATGA 240

ATGCCGAACT TGGtACCTCG AATATGTTGT ACATGACTGT TGGGGTCTAT CGTAAAAAAG 300

TCAACTTTTC CTTAgCTGCA AAAATTTTAT TTACCTGTAT CTTATTCAAT TTAATTGGTG 360

GCGTCTTGTT TGGCTTCTTG GTATCACTTA CGGTGCCGTT CCAAGATTTG CCAAAAGATA 420

GTTTCTTCTT TACATCGATT GCTGGAAAAT TAGAAAAAAC CACCCTACAG ATTTTAGTTG 480

AAGCTATGTT CGCCAATATT GTAGTTAATA CAGCCGTATT AGTGAGTATG CGAATGAAAG 540

ATGACGCCGG CAAAGTCGCG GCCATTATTT TCATTATCTT TATCTTTGCG TTCCTTGGTT 600

TTGAACACGT GATTGCCAAC TTCCCAGCCT TAAGTTTAGC TTACTTTGCT TCGAACGGCG 660

CAATTGAAGC TTTCACAGCT GGTAATGTTG TTCATAATCT GTTCTGGGCC TTTATCGGTA 720

ACTTTATCGG TGGTGGCTTA ATTATGGGTC TCGGCTATGC cTGGTTAGAT AAAGACACAA 780

AAaCTTAACG TATTTkGATT AATTGACCAA CCAACTATAC TAAAACATGG ACACTGTTTT 840

AGTATAGTTT TTTTATnATA AG 862

(2) INFORMATION FOR SEQ ID NO: 553:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 472 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 553:

GAGGTTGTCG TTGTCCCATG ACACTCACTC CTCTATAAAT GAATnTATTA CaCTTATTAT 60

GTTACTACTA TTTTACAAGG TGTCAAATAT TTTCATCGTG CTTTCTGAAA ACCATTTCAA 120

TGATAACTCC ATTAAAAACA ACGTATCACT TGATTGTCTC TTTTTTGGTG CAAAATTGTG 180

CCTTTTTTGT ATAAATAAGT CCCTCGTTTT CCCACAAAAA AACCGCTACT TTTTCATCTA 240

GTTTATAAAA AAGTAACAGT TATAAATTAT TTTTTCTTTT TCCCAATCGA GAGAATCCCT 300

CTTTTTATGA CATCAAAGAA ACTCAAAGAT TGGACCATAT GATCTGGGTC AACATATTCA 360

CGAATAgctC GTAACACTCG CkTCGGCkCT TTTGAAGCAA ACGTATACGT CCCATTTTTC 420

TTGGTTTGAA TCGCATAACG AGGaATCCAT TTTCCTTTGa ACaTTACCGG AA 472

(2) INFORMATION FOR SEQ ID NO: 554:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1062 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 554:

TATAnTTTTC AGCAATCGTT TCAAATAACT TAAATAATGT CATGAAAATT TTAACATCCT 60

TAACAATTGT CCTGACCATT CCGACCATCA TCGGAGGAAT TTACGGCATG AATGTGAAAC 120

TCCCTTTTGC AGAGCATGAA TATTCTTTTT GGATTATTTT TGCCATCACT ACCTTAATCT 180

GCGTAATTAG TATTCGAATT TTAAAGAAAA AGAATCTCTT ATAACTTTAG GTTGCTGTTT 240

TCAGGAAAAG GTTGTACACT ATTAGCAGAA ATTAAAAAGG ACGTGTCATT TATGAACGAC 300

AAAGAACAAA GTCGCTTTTC AAAAATCACG AAAGTCGTGG TTTGGCTAAT GCTGATCGCA 360

ATTGTAGGTT CAACTGTTTT AACAGCGATT CTTAGTCTTT AAACTGGAGA ACTTTTTCAC 420

ACATTAAAAG TAAAAAAACA GCCCCTTCTA TCAATTGAAA GGGCTGTTTT TTTACTTGCT 480

CAAATAGAGC GGCTATTTTT TCGCCATCGC CAACAGCTCT TGTGCATGTT CAATGGTTAA 540

TTTCGTAATT TCACTTCCTG CGAGCATACG CGCAATTTCG TTTACTCGCT CTTTTTCAGA 600

TAAAATTCGG ACGCTTGTTT CTGTCCGCCC CGCCACAATT TCTTTTTCAA TAAAATAATG 660

TTCATCGGCA ACGGCCGCCA CTTGTGGCAA GTGCGTGATA CACAACACCT GCGAATTTTC 720

TGAAATTTGA TAAATTTTAT CGGCAATCGC CTGTGCTACT CGGCCACTAA CTCCTGTATC 780

CACTTCATCA AAAACAATAC TAGTGATCCC TTGCGTTTGA GAAAAGATTG TTTTCATAGC 840

CAACATCACT CGCGAAAGTT CTCCGCCAGA AGCCACCCGA ACTAACGGTT TTAATGGTTC 900

CCCTGGGTTA GTAGTAATAT AAAATTCTAC TCCGTCTAAG CCATTTTCTT GTAAATGCTC 960

AAGTTCTGTA AAGCGGACTT CAAATTCAGT TCGCTCCAAA TATAGTTCTT TTAATTCGGT 1020

TAAAATTTGT TGTTCGAGCT CTTTGGCTAG GCGCTTTCGT TC 1062

(2) INFORMATION FOR SEQ ID NO: 555:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1237 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 555:

TTCCTGATTG AAGCAAAGAG AGGTGTAGCG CTATGGAAAT GGAACATATT AATGAAAATA 60

CGATTCGTGT GTTAATTGGC AACGAAGACC TGGCGGATAG AGGGATTACT TTTCTTGACT 120

TACTAGGAAA CCACAAAGAT GTTGAAAATT TCTTTTATAG TATTTTAGAA GAAGTTGATG 180

TCGAAGATGA GTTCCAAGGC AGTGAAGCAG TCACTTTCCA AGTCTTGCCT AAAAATGATG 240

GTTTAGAACT GTTTATTAGT AAAAATGTCG CTATGGACGA CTTGTCTTCC TTAGAAGGTC 300

TGAGCGAAGT CaATGCAGAT GTTAGCGAAT TGATTCGCAA ACAAATTGAA GCAGACAAAG 360

CAGCCGCCGA TGAACTTGAC GAGATGGAAG CCACAGATGA AACAAACAGA AATGTAATCT 420

TTGAATTAGA TAATTTTGAA GCGATGATTC AATTATCCAA AGAAGTCTTC ATGCAATCTG 480

TTTTGACTAA CTTATACACG TACAACGATC GTTATTACTT ACAAGTTTTG TTCTTAACCG 540

ATGAATTAGA AAAAACAAAT GTTGACAATG AAATTGCACA GATTTTAGAA TTTGCTCATA 600

AAACAACTGT CACTCAGGAT ACTTTAGTCG AATATGGCAC TTGTATTATG GAACGCAGTG 660

CATTGGAATT AACACGTTAT TACTTTAATG ATTAAGCAAA AAAGCAAAAA CTTTCGAGTT 720

TcTTkTTkTT tGGaTGAAAA TTAAGAAACT TTGCGTAGTT TATTTATGAG AGGTGAAAAT 780

CATGTTAACA GCTTTAACGG CAGACAACCA AGAATTATTT TCTTTAATCG ATTGCTCGAT 840

TGATGAATTA ACACAGATAC GAAGCGAACA AGCTTTTGTG TGTCCAATGT GTCATCAATC 900

GGTCATTTTG aAGGCAGGCC CCATTAAAAT ACCCCATTTT GCTCATCGTA AAAAGAATAG 960

TTGTTGGtAC GAAGCAGAAG CTGAAACAGA AGAACACCTC CGCTTGAAGC AGCTTTTTGC 1020

GGAAAAGTGT TTACGCGAAA AGCTGTCCTT CCAAGTTGAG GCATACTTAC CCACCTTGAA 1080

GCAACGCCCA GATTTACTCA TTGGGGAAGA TAGCCATTGA AATTCCATGG TAGTCCATTG 1140

GCCATTAAAn CGTTTGGTAG AACCGGACGG AGACTTAATC AAACTCCATG GCTAATnGAG 1200

TGGTTTGGGA TTnTGGTGGA CCGATTAAGG CCTAnAG 1237

(2) INFORMATION FOR SEQ ID NO: 556:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1306 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 556:

GCTATTCAAG ATGAnTTAAG CGATACTAAA GACGAAATAC GTGAAGTGAC GAAACAAGCG 60

AGAGAAGAAA CAnAAGAGTA TAAAGAAATC TTAGAAAGTA AAATAAAAGA CCATAACAAA 120

GTGGTTGATA AATCGAATAC AGCGCTAAAA GTGATGACAA AAGGCGTTAC AAATATATTT 180

TTTGTGCTAA TTATTTTTGT ACTTGTAATG CTTGTGACAG GTCCGATAGG TCATTTCTTT 240

GGTATTGAAC ATTTATATAG CTTTATTAAT GGTTTTATTG ATGACCACGA AAGTGCATGG 300

CGATATTTGA TGTTGATACT TTATGCAGTA CCCTATGTAT TCTTCGCTAT CATTCTGTGG 360

CTCATAGCAT TGTTTTTTAG AGTATTTAAT GACCTTTAGG TAGATTTATA TATAAAAAGG 420

GCTAAATACG CAAAATACCC CCTTAAAATG CAAAATAAAT GCAAATAAGG GGGTAACTCC 480

TTATGAAAAA ACAACTAATA GTATTATAAC ACGTTTTTAT GCCGAGAAAA CTTATTGGTT 540

GGAATGGGCT ATGTGTTAGC TAACTTGTTA GCGAGTTGGT TGGACTTGAA TTGGGATTAA 600

TCCCAAGAAA GTACCAACTC AACAACACAT AAAGCCCTGT AGGTTCCGAC CAATAAGGAA 660

ATTGGAATAA AGCAACGAAA GGAGTTGAGT AAAATGAGTG AGAAAACGAA GAAAATCATT 720

CGAGTATCAT TAGCTTTTTC ACAAGTAGTA GTGCAATCAT TGATAGTTTA TTATACTTAC 780

CAGCAATATA AATTATCTAA AGAGAATTAA GTTTTGATTA TGTAATAAGT TTTGGTACTA 840

AAAAAGACTT GATCTAATTA GACCAAGTCT TTTGTTAGTG TTATATTAAT AACAAAATAA 900

AAATAAGAAG TCGCTCACTC CCCGACCAAA GTTTGTGAGC AAAATAAACA GTCGATGTTT 960

GTTTCTTAAT TGTATATCTT GATATTAAAC GATTTAAGAT TATTCTTCAA GATATATATT 1020

CGCAAATGTA ATTGGAAAAG CGACTTCTTA AACAAAATTA AGGAGTCGCT TTTTATGGCA 1080

AAAATAACGr wAAATGAGCA AGAAAATTTG GTGGAAAATT ACACAAAAAA ACCGGCGTAC 1140

TCCAAACTAG TCAAGTCCAG ACTCCTGTGT AAAATGCTAT ACAATGTTTT TACCATTTnT 1200

ACTTATCAAA ATTGATGTAT TTTCTTGGAG AATAAATCCA TTCATCATGT AGGTCCATAA 1260

GAACGGCTCC AATTAAGCGA TTGGCTGATG TTTGATTGGG GAAGAT 1306

(2) INFORMATION FOR SEQ ID NO: 557:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1268 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 557:

CTATTTAATC TTGCTCACGT TGTTTTCTCT TGGCTGCTCG TTCTGCGCGG nGGGCAGCCC 60

TTCTTGCTTC CGCTTCCGGA TCAACTACGC GTTCTTTTTT CACTCGTGGC TTCTCGTTCT 120

CTGGCGTATT TATTTCTGGT TCACTTTGTT TTTCAAGTTT TGCTCGAATT TTTTCTTCTT 180

GACCAGCTAA ACGAGCATAA TTATAGAACC GTCTCTTAGC ATCATTAATT GTTTTATTAA 240

ATAACAATTC CGCTGTATCT GGTTGACTGG ATTCTAAGGA CGCAAAGCGA ACTTGTTGGC 300

GCATAAACTC TTTCATTAAT GAAAAATCTG GTTTTTTGAA ATCCAATATC ATTGGTTCTT 360

TGCCTTTTTC TCGCAATAAT GGATTATAGC GATATAAAGA CCAGTAACCT GAATGAACCG 420

CGTCTTTCGC TTCTTTCAAC GTTTGACTCA TCCCACCTGC TAAACCATGT GTAATACATG 480

GTGTaTAAGC AATAATAATC GATGGTCCAG gGAATTTTTC TGCTTCTTCA AAAGCTTTAA 540

TCGTCTGCAT TTGGTTTGCA CCTGAGGCAA TTTGAGCGAC ATAGACATTT TCATAAGTCA 600

TTGCCATCAT CCCTAAATCT TTTTTAGAGG CATACTTCCC ACTGGCAGCA AATTTCGCAA 660

TCGCTGATGC CGGAGTTGCT TTGGAAGTTT GGCCACCTGT ATTGGAATAG ACTTCATTAT 720

CTAGAACTAA CATATTTACA TCCGCGCCAC TAGCTAACAC GTGGTCAATA CCGCCGTAGC 780

CAATATCGTA AGCCCAGCCG TCTCCACCAA TCATCCACTG ACTTGGTTTA ACAAAAAGAT 840

CTTGATCGTC ATAAATTGCT TCCAATAAAG GTTGATTCGT TTTTTCTTCA AGCAAAGCTG 900

CACGTAATTT AGCTGCACGT TGTTGCGTAC CTTCACTTTC GGACAGATGT GCTATCCAAT 960

CTTCCATCAA TAAACGTAAA GAATCACTGG CTACAGAAAA AGCTTTGGTC ATTTTACTAG 1020

CTAAACGTTC GCGACGCGCT TGCGTTGCTa ACAACATGCC GTAACCAAAC TCGGCATTAT 1080

CCTCTAAAAG AGAATTTGAC CAAGCTGGAC CTTGTCCTTG TTCATTAGTA GTGTAAGGTG 1140

TCACACCCGC TGCAGCACCC CAAATCGATG AGCAACCTGT TGCATTAGCA tCAGCATTCG 1200

GTCACCAAAC ATTTGCGTCA ATAATTTAAC ATAAGGCGTT TCGCcGCAAC CAGAACAAGC 1260

ACCAGAAA 1268

(2) INFORMATION FOR SEQ ID NO: 558:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1060 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 558:

GTTTTACCTT AAAACAAAGG GCAATCAGTG TTAAAACAAT AGTCATTAAA GGAGCAGCAC 60

CTAAATCAAT TAAAGGTTTA AAAATTGTAT TTGCTAGGTC AATAATCGTA TTCATTTATT 120

TTCTCCTTCA TAAGCTAAGA TAACAAAATT TTTATCTTAG TTTTGTTATT GTTGACAGTG 180

AAGCGTAAGT AATTAGTCTG AAAAGGAAAT CATTTGTTTA AAGAGTGAAT GATACTTTCC 240

ATTTGTTGAT AAACAGGTTC AGCCATTGCT GGAATTCGAT AAAGAATTGC GCCTGCATCA 300

ACTACAGGAA TCGTGATTGG GAAACCCAAA TCTGTTTTGG CGATTTGGGC GAAAATATCA 360

TAATGGGATA GCATCTCTTG ATTGATATCT TTAATCATTA CGGCATCGCA AGAAACTTGG 420

TAGCCTCTTT TAGATAATTC TGTTTCTAAT GCAGATTTGA TTTGATGGCT AGAGTTTACC 480

CCAGCCCCAC ATGCAGCTAA AATTTTGATC ATGTTAAATC TCTCCTTTAA AATTTTTTTG 540

TAAAAAACGA TAAACTTCTT CTTTATCTTC ACATTGAAAA AAAGCCAATA GTGATGAACG 600

GTCTGTCGTG TTAATAAAAT CCaTGATAGT CGCTAAAAGA CCCGTTTGTT CAATACCGTT 660

TTCATTTAAA ATCATGAAAA GAAAGGrAAC aGAGAGCGTT yCATCTGGGr TAATCATGtT 720

AtGAAATgAT AAGGATTGTT TcAATTTAAT nCGGGATAAT TCTTGTCGTC CGTACAAAGG 780

TACTTTCGGT ATGCGGAATG GCGATATTCG GTAGTGCAGG ATCAATCGGT GTTAAACTTA 840

AACCAGTCGG ATAGTTGTGT TCTCGTTCCA ATAAATTAGT TAGAAAATnC AGGTGTAACT 900

AATTTCTTTT TTAACAAGTC TAAATAAACT TCTTCAAAAA CTTCTTCTTG TGTTTTTTTG 960

TTTGAAATAT AGGTAGTATC TAATGAAAAA ATCGATGTTT TAATAACCAT AGAACCCCTC 1020

GTTTCTTTAA TATATTATTA nTTATTTACA AGTATAATAT 1060

(2) INFORMATION FOR SEQ ID NO: 559:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1565 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 559:

AnATnCCAAA AGAATTAAAA GACAAAAATT TACTTTATGA AAAACGTATT CCTCCGAAAA 60

AACTCTCTAC TGGTGAATTC AAAAACTATC CAAACAGATT GTATTTAGGT AAGCTGGAAG 120

TTTCAGCAAC AGATGTTTAC GCCATCAGTA ATGCTAGCTA TACAAATGAG TTTCCCGAAA 180

GTGGGAAAAA CCAACCTTCG GATAAACATG ATAAAATCGG CGTTTCTTTC GAAAGTGGGA 240

AAAACCAACG TCCAGAAGAA TCCAATAATA CCAGTACTTT GTTCGAAAGT GGGAAAAACC 300

AACCCAATCT ATATCTAACT AATTCTTTAG ATACAATAGA TACAATAGAT ACTGAAAAAG 360

AACGATTGCA ACAACAATTA TTGCTCGATC AATTTTCAGA AGTTCAAGAG CATACATTTC 420

TTAGCAAAGA TAGCTTAAAA TTTATTGCTG CCTTTTCTGA TACTATTCAA GAAGCCCATG 480

AAATGGTwGG AACAATTATT CGTGCTAAAA CTAAAGTAGA AAAAGAATAC AATmTCGTTC 540

TAATTGGTGA AGATTATCAA GAAGAaATAG ACAAATGTTT GCGACGAGTG ATGCATAAaA 600

TCAAGACGGA TTCTACTGTT AAAAGTCCTA AAGGCTTATT TTACAAGTCT TTTTACAATT 660

TATTTGTTGA ATGTGCACTA GAGAAAAAAA GTCAATTGAA CAAGAAAAGT AATTCTAACA 720

CACCAGGAGT AATCACTCAT AACTGGGTTG AAAATCaATa AAGTaGCTGT TTaGTaTGaA 780

TTTAAATTTA TGGAGGAAAA TTATATGTCT TTTGAATCTG AAAATAAAAA CGTAGAAAAA 840

ATAATTGTGT CTAATAAAAC GTATTATCTT GTTGGGACAA GTCACATTTC AGAGAATAGT 900

GTTAAACTTG TTAAAGAGGT TATTGAAAGA GTTCAACCAG ATACTGTTAG TATCGAGCTA 960

GATAAAAAAA GATATGAAAA ATACACTAAC TCTAACCAAT GGGGAAATAC AGATATAATa 1020

AAAATTATTA AAGAAAAAnA ATtAGTTGTT CTAATATCCA ATATTGTTTA CAGTGCATAT 1080

CAAAAAAAAT TAGCAAATAC TAAAGGAACT ACTCAAGCTG GTGAATTAAT TCAGGCAATT 1140

AAAAGTGCTA AGGAAATAGG AGCAAACATT CAATTAATCG ATCGTGATAT CCAAGTTACT 1200

TTTAAAAGAA TGTGGCGTCA TTTAAGCTTT TTGGAAAAGC CAAAACTTTT TATGACATTT 1260

TTTACTGAAT TTGATGATAT TGAACAAGAT AAATTAGAAG AATATTTAGA GTCTGATAGT 1320

TTTGACAAAG TATTTATTTC GTTATCAAAA AAATATCCTT CGCTGTATCA GGATATGATC 1380

ACGGATAGAG ATAAGTATAT GTCTACTAAA TTGAmAAATA ATTCAAGTCA AGTGAATGTT 1440

GTGGTGGTTG GGAAAGCTCA TATGAaAGGa ATTAAAGAGA AGTTAGAAAA AAGAACAGAG 1500

TTTTCCTTAG ATAATTTAAA TGAAATTCCA CCAAAAAAAT TArGTmCTAA ATtATtAGaA 1560

TTnCC 1565

(2) INFORMATION FOR SEQ ID NO: 560:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 780 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 560:

CCGTGGAACC TTCTAAGTCA CCACCAATGA TTCCAGGTAG CCCCATTAAT AAAATATCGT 60

TGGATAATAG GCTCTTTTAA CnCGATCAGC ATCCCACGTC CCATAATTAA TAATGCAAGC 120

CTAAAACAAT ACTCGGAATT TGAATAATAA TATGTGAACG GAAGGGGTTA ATATCTCTTA 180

ACCAATGCAC TTGAGTGAAG GCTTCTGGAA TAGTGGCTAG GAGCACCATC ATAATTCCTG 240

AAAAATAAAG TAAAACCACT TCTAGCTTGT GGGCAATTTC CAAAGTAATT TGTTTTAGCA 300

GTTGACGATA ATGAGTATCC AACTTTTGGC TTAAGTGTTT TGTAAAGAAC AAACAGCCAA 360

TTAAGAACGG AATAATATAA TAGAATAAAC GATAGAGCAA TAACCACAAA ACAATAATTT 420

CACGATCGAC ACCTAAGTTG CTCAGTCCTA AAATCATCAT GACATCAAAA CTACCTAAAG 480

CCCCTGGAAT CATCGAAGCA ATCCCAATAA TCGAAGCTGC TACATATAAA GGAACAATAT 540

CAATTAAAGG AATCGGTACA TCTAATAAGT AACCAATACT GATAAAGGTA ATTAAAACAC 600

CTGTCCATTC TAAAAACGAA GGAAACCACT AGAGATAAGC GTTTTTTGCA TCGAGGGTAC 660

CAAnGGACTA TTTTTTTGGG AATGTAAATT AGTTAAGTAC nGGAATTAAG CCGCACCTAA 720

TTACCnATCC CATATTGCTG TAATAAGCCT CTTTGnTCAA CTGACCAGGG AAAGCAAGAT 780

(2) INFORMATION FOR SEQ ID NO: 561:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1610 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 561:

GAATGTGGAT TTAATTGATC GTGATGTGTT TGGTTATGCT GTTGATAAAG GCTGGATGTG 60

TGTCCAAGTA TTCTTTGTCA GACAAGGAAA ATTAATCGAA CGTGATGTAT CGATTTTTCC 120

ATTTTATGAT GATGCCAGTG AAGCATTTTT AACTTTTATT GGACAATTTT ATCAAGAGAA 180

TGAGCATTTT GTGCCAAAAG AAGTCTTGAT TCCTGATGAC ATTGATAAAG AAAGCGTCGA 240

AGCTTTATTA GCTACGAAAG TCTTGCAGCC GCAACGGGGA GAAAAGAAAA AACTGGTTAA 300

ATTAGCCAGT AAAAATGCTG CAGTTGCTTT AAATGAAAAA TTCGATTTGA TTGTACGGAA 360

ACAAGAACGA ACAATTGGCG CAGTTGAAAA ATTAGGCAAT GCGATGAACA TACCTGCGCC 420

CATTCGAATT GAAGCATTTG ATAATTCAAA TATTATGGGA ACAAATCCTG TTTCTGCGAT 480

GGTTGTGTTn ATTGACGGAC GGCCAGCTAA AAATGAGTAT CGAAAATATA AAATAAAAAC 540

CGTCCAAGGT CCTGATGATT ACGCTTCTAT GAGAGAAGTC ATTTATCGCA GATATTCACG 600

AGTTTTAAAA GAAGGCTTGC CTTTTCCTGa TTTAATTTTa ATCGATGGTG GTAAAGGTCA 660

AGTCGATGTT GCCAAAGATG TGTTAGCCAA TCAACTAGGA GTCGATATTC CTGTAGCTGG 720

ATTAGCCAAG AACGACAAGC ACAAAACGAG TGAACTTTTA TTTGGTCCAA ACTTAGAAGT 780

TGTACCATTG GAAAGAAACT CACAAGAATT TTTCTTGTTG CAACGAATCC AAGATGAAGT 840

CCATCGTTTC GCCATCACAT TCCATCGTCA ATTACGTAGT AAAAATAGTT TTGCTTCCAA 900

ATTAGATAAC ATTGAAGGCT TGGGACCGAA ACGCAAAAAA AATCTTTTAA AAGAATTTAA 960

ATCGCTTAAA AATATTACGG CAGCCAGTGT TGAAGAATTA AGAAAAGCTG GGTTACCTGa 1020

AACTGTCGCA AAGAATGTGT ATCGTCATTT GcATCAAGAA ACAACCTCTG AGATAGAAAA 1080

ATAAATGAAA ATGaaCCAGa CCmATTaGGT GCTTTTCCTA ATTTGtCTGG TTTTTTTATT 1140

CAATTTTTTC ATGAAAAGTT GAACTCTCAA GAAAATATGA TAACGATTTC TTTCGTGATA 1200

TACTAATACT AACTTAAAAG AGAGGACGTC TGATCATGAA AATACTAACA CCCGTTTATC 1260

AAATAAAATT TAAAAAAGTA CCAAATCAAT TACTGATTGA TGAGCGTTTA ATGAATGAAT 1320

TGGGTAGATG GCTGAATCGT ACAGGTCGAA TTTGGGCCTG TCAATCATCA AAATCGGCAA 1380

AGGAATTTAA AAATACATTT TATGAAAGTA CAGGCTTGTC CGCAAATGAA GTGTATATTT 1440

CTGCCGAGAC GGATGGCCTT TTTCGTTaGC AGAAATAAAG CGCTGTTAAA GTGCACATCT 1500

TTCCATAACA ATGTCATGnC AGCCACACGG AAGTACGTTG GTACTAACAC GGTTGAAAAn 1560

GAAGGGGCGT TAGTGTAnGG AAAGGTAGCA ACCTTGGGAT TAACnGTGAC 1610

(2) INFORMATION FOR SEQ ID NO: 562:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1227 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 562:

TGAATAATAA ATAAAGGAGA GAAACAAAAT GAAAGATACC TTTCGCTTAG AAAACCAGAC 60

GATTTATTTT GGTACAGAAC GGGCAATTTC AGCCAGCCCT CAAACGATTT GGCGCTACTT 120

AACTGAGACA GACAAGCTAA AACAATGGTT TCCAGAATTA GAAATTGGTG AATTAGGTGT 180

TAACGGGTTT TGGCGGTTTA TTCTGCCCGA TTTTGAAGAA ACAATGCCTT TCACTGATTA 240

TGCAGAGGAA AAATATCTGG GCGTGACTTG GGATACTGGT ATCATCTATT TTGACTTAAA 300

```
AGAACAAGCA CCTCATCAGA CCCTGTTAGT CTTTTCAGAA TCTCTCCCAG AAAATTTTAC     360
AACACCTCGC CATAAAGATA TCGCTGGTTG GTCAATTGTT CTCAACCGCT TGAAGCAAGT     420
AGTTGAAACC CCTGATGCAG CACCAGAAAA AATAGATTTT CCGCAAATAG AAAATCATTA     480
CTTGGaAAAA TTAACGAATT TAGAAAATTA AAAAGTGTGG CACAACGGGT AAATCACCCT     540
CTGTGcCACA CTTTTTAAAA ACGTAAAATA CGCTGATAAA CAAATAAACC TAAATCAATT     600
AAGCAAACAA CAAAAATTGC CAATAATACA AAAGAGACCC GCAACTCTGC GGCGTACATC     660
CCTCGTAAAC TAATTGTTGC CAAACATAAC CCAAAAATCG TTGCCATTGT CACTTGTAAC     720
ACGGTTCCTG GTTTTAGGGC ATAAACAAAA ATGCCAATCA AACTAGAAAC AATGAGGTGT     780
GTCAGTGGCA aTTGAAGCCA ATTGACAACA AACAGACTAA GGAAACATCC TAAAAGAGTC     840
CCCCAAAAAC GTGTCTTAAT ACGCGTGTTC ATTTCTGCTT TCACTGGATG CAAGAGTAGA     900
AAAACATTCA TCGGAATCCA AATTATTTTA ATGCCTCCTA CACTCCACAT TAAAATAAAA     960
CTCAGTGTAA TGGTGAAACT CATACGAATA GCTAAACGAA ATCGTgCACC CTTTCGTCAG    1020
GACTTTCTAA AAATTGTTTG AAACTTGTAA TCCCAAAAAA AATCTGTTGT AAGTCGTGCC    1080
AGAAATTAAT AACTGCCAAT TTTTGTTGCT TCATATTTAA TTGAATGCTT TTAAAAAAAC    1140
GAACATGGCC CCTACTCTTT AAGCTAGTTA GCTGACGTAA TCCTGTTTGC CACCAAGCGT    1200
TTAATGTTAA AGCCAATAAA AAAAGAC                                        1227
```

(2) INFORMATION FOR SEQ ID NO: 563:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1619 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 563:

```
AGGATCCCGC ATATTTAAAG CGCTCGTGAT nCAGTCAATT AAATCTTGAG GTGTGACTAC      60
CATTGTTGCT CCAATATCAC GAGCTTTTTG TTCTTGTTTC CTCTTTTCAT TTGTATCTTG     120
ATCTATTATA TCGCCAACCG ACGTGAGGTG ATGTCCAATT TCCTCTGCAA CCGTACTAGT     180
TAATTCTCTA GGATGTTGTT GAGGATTTAA ATACACAACA TTGTTAATAT ACAATCCTTT     240
TTGCTTTTCG GGCATGTTTG GTTCAAACTT ATACGTTAAA TCTGAAAAGC GAGCCATCAG     300
TTCTTCTGAA GTTACCATTT AAACACCTAC TACTTTAATT ATTTGTGATC GCGTTTCTTA     360
ATATAATCAA TGAAAGAGAG AATCTCATTC ATTTCTGTAT CTGAAACGTT ATCGTCTATA     420
TGAGCAGCGA CGGTTAATTG TTTTTTTGAT AAGGAAGTAT CTTCAAAACT TTCTCTCCCA     480
TGTAAATAAT CTAAACTTAC ATTGAAATAG TCAGCAATCT TATTTTGGAT ATCTATATCT     540
GGAGTTCTTC TTCCTTGTTC ATAGGAAGAG TAAGTCGTTC TAGCTACACC AAGTGCATTT     600
GCAACATCAG TTTGTGTTAA TTTTTTTTGT TTCCTTAATT CTGTTAAACG TGTACCGaAC     660
ATAATGTTAC CTCCTACGAT TCGTTATGTA CATATAATAC TACGCAATCT GAGTAGTTTG     720
```

AAGTTAAATA TATTAAATGT GTCAAAAAGA GTAATTAATT GTTGACATGT GTCAGAATGC 780

GTAgwATTAT AAAAGTACGC AATACGACAC ATAAAGGAGG ACACAAAATG AGAGATTGGC 840

TATTGGAATT GCGTATAAAA AATAAAATGA CTcAAGAAGA GGTTgCCgTT aAAgCGGaAA 900

TTTcAAGAAC CaCGTATGCA TCTATTGaAC AGGGAAGAAG AAGGCCATCT GTGGaAAGTG 960

CTATGCGAAT CGCATCAATC TTaGATTTTg aTTgGACAAT TTTTTTTtGa GTAAcTATGA 1020

CTCGTTATGA CTCTTTATAT TTGGATAGGA GGTGAAAATA TGCCAGAAAA ACAAAAAATG 1080

ATTGATTTCA TTTTGGAAAT CATGCCAAAC GTAAAGCTTA TTGTTGAACT AGCmACAGAT 1140

GATCAGTTAG TTCGTCTGTA TGrACAAGCA CAGGATAAGT TGAACTATCm ATTGGATCTA 1200

ATCTAATTGT ACaAAAAAAT ACATCATATA AAAAGATGGT TAATTAAGAA AGGAAGAAGT 1260

AAAAGTAAAT GTCaATGTCa GTAATATTAA GAAAGTCGTT AATCGAAGTT CTTAATAAGA 1320

AAAATGAAAA GAAGAAAGAT GTAGCaCGTG AAATCAGTAC ATCTCAACAA TCATTGAGTG 1380

ATTGGACTTC TCAGAATAAT GTTAAACCAG TAACAATTGA AAATGCTCTA AGGCTTAGTG 1440

ATCACTTTAG AGATTCGACT TTCACAATGG AAGTTATTCA TCAGTTTTTT GGAATGTTTA 1500

AAACTTGCGA TGGAGATGTT TATCGAAAAG ACCCTTCGTC ATTGGACCGA TTACAGAAAA 1560

TTGAATCTAA TGAGAGAnnA GCATTGAAGC ATGACATAGA AAGAATTGTT TTGAAAAAT 1619

(2) INFORMATION FOR SEQ ID NO: 564:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2082 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 564:

ATCCCGCAAG TCGTCGGTnC TAAAGGAAAA CTAGGACGAA ACAATGTAAT TTTTACTTTA 60

TCTTTGGCTT ACTATGCGTC AAAAATTGAA CAAGACCGCT GTTTTGATGA TATGGATGTA 120

TTCAACAGTT CTTTGAATTA TCCTCTATCA GTTTCAGAAA TAAAGAGAAT TGTTAAGAGT 180

GCTTACTCAG GTAAATATAA AGGCGCAAcA AAACCTATGT GCAAGAACTA GTGTCAAATT 240

GGGGATTAGA ATCGTTATCT GAGGAAAAAA TGTTTACACA GCAAAGAAAT ACTTGGTACA 300

AATTTAAGAA GGAACGTCAT AAAAGAAAAA AATCTCACCT TTCCGAATGG AAAAAAGATA 360

TTTTAGATTA CCTTGAAACC CAGTGTTATA GATATCGACC TGAAGTTAGT ATGAAGAAAA 420

CAGATTTGCA AGCGGCTGTT ACTTTCAATG GGCAATCGAT TCCTAAAAGA TCACTTGACC 480

GTGCACTACA AGAATTAGTG GCCGAAGGTA AATTATTTGT ACAAACAAAA GCCGGTAGGG 540

GTGGAGGACT AGTAGTTGCT ACTCGGAAAG CATTAATTCG TACAGTTATT CAAGTAAAAC 600

AGCAAGTCAA ATATGCTTAT AAACAGGGTA TTAAGACATT CTTTAAAGAA GCTGATATGT 660

TAGTAAGACT GTTTGAACAA ACTGATAAAn ACAGTCAAAG AGCCAGAGAA TACAAGCAAT 720

TAAAACTATG GAACACTGGA TAAATTTAAA CATGTCAACT GCCAATACAT TATATTTCTT 780

TTATACTTAA CAGCTATATA ATTTTTGTAT TATTTTTTTT AGTTATTTGT TAAAATTATT 840

TTATACAAGC CTTATTATTT GGTTAGTAAC TTTTAACAAC GGAGGGTGAT AAAAATGAAG 900

AAGTACAAAA AGTTTTGTTT TTTAGGTGTT TGGTTATTAC CTTTGGTTTT AGCTAGTTGT 960

GGTACAAATA CTGCTACAAA TGATTCACAA GATATAACCG AAAAAAAAGT AGAACAGGTA 1020

GCGACTTTGG CTGCAGGGAC ACCTGTTCAA AGTTTAGATC CAGCAACTGC TGTTGATCAA 1080

ACGAGTATGA CTTTATTATC CAACGTGATG GAAGGTTTAT ATCGATTAGA TGAAAAAAAT 1140

CAACCGCAAC CAGCAATTGC AGCTGGGCAA CCAAAAGTGA CGAATAATGG CAAAACATAT 1200

ACCATTGTGA TTAGAGATGG CGCTAAGTGG GCTGATGGTA CACAAATAAC TGCCAGTGAT 1260

TTTGTGGCTG CGTGTCAAAG AGTTGTAGAT CCTAAAACAG CTTCTCCAAA TGTGGAACTG 1320

TTTTCTGCTA TAAAAAATGC CAAAGAAATT GCTTCAGGAA AACAAGCAAA AGATACTTTA 1380

GCAGTGAAAA GTATTGGTGA GAAAACATTA GAAATTGAAT TAGTTGAACC AACACCTTAT 1440

TTTACTGATC TGTTATCCTT AACGGCTTAC TATCCAGTAC AGCAGAAAGC AATTAAAGAG 1500

TATGGGAAAG ACTATGGTAC TTCTCAAAAA TCAATTGTAA CAAATGGAGC ATTTAACTTA 1560

ACAAATTTAG AGGGAGTAGG CACTTCTGAT AAGTGGACGA TTTCTAAAAA TAAAGAGTAC 1620

TGGGATCAAA AAGATGTTTC TATGGATAAA ATTAACTTCC AAGTAGTCAA AGAAATTAAT 1680

ACAGGAATAA ATTTGTATAA TGATGGACAA CTAGATGATG CACCTTTGGC TGGGGAATAT 1740

GCAAAACAAT ATAAAAAAGA TAAAGAATAT TCAACAACAT TAATGGCCAA TACAATGTTT 1800

TTAGAAATGA ACCAAACTGG GGAAAATAAA CTTTTACAAA ACAAAAATGC CCGAAAAGCG 1860

ATTAGCTATG CTATCGACCG GGAAAGTCTA GTTGAAAAAT TACTAGATAA TGGGTCCGTT 1920

GCTTCTGTTG GCGTAGTACC AAAAGAAATG GCTTTTAATC CAGTAAATAA nAAAGATTTT 1980

GCTAATGAAA AATTAGTGGA ATTTnACAAA AAAAAAGCAG AAGAATATTG GGGATAAGGC 2040

AAAAAAAGAA ATTGATTTAT CCAAAAATAC TCTTTAGGTT Tn 2082

(2) INFORMATION FOR SEQ ID NO: 565:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1091 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 565:

AAATGACTTG ACTGTTGCTT CAGTTGnCAG CATTCACCTT GCTTTTAGCA GGCTGTAACA 60

CAACCAATAA TAAAGGCACG ACCGCAACGA GTGATTCTGT CGCTAAATCC AGTGAAAGCA 120

AAACAAAGGA TTCGGTCGAC TTAAATTCAC TAGATTTACC ACAACTAGAT GACAAAGTTG 180

CCGAAGATCA AGACTTAGTC CAAATGGTTA CTTCAATGGG AAATATTGAA ATCAAACTAT 240

TTCCTAAACA GGCTCCAAAA ACTGTCGAGA ACTTTATGAA ACATGCCAAA GATGGCTATT 300

ATGATGGCTT GACTTTCCAC CGCGTCATCA ATAACTTTAT GATCCAAGGT GGCGATCCTA 360

GCGGTGATGG aACTGGCGGA AACAGCATCT GGaATAAACC ATTTGAAGAT GAATTTTCAA 420

ACCAGTTATA TAACATTCGT GGCGCGTTAT CAATGGCGAA TGCaGGrCCT AATACTAACG 480

GCAGTCAGTT CTTTATTGTC CAAAATACAG ATGATCAATC TGATGGTCTT TTATATGATG 540

ATTATCCAAA AGCAATCATT GaTGCTTACa AAAAAGGCGG CTATCCTTCT CtTGATAAAA 600

AACaTACTGT TTTCGGTCAA GTAACCAAAG GAATGGATAT TGTCGATAAA ATTGCCAAAG 660

TCGAAGTTGG TGCCAACGAT AAACCAAAAA CTGATGTTAA AATTGAAAAA ATCAATATTC 720

TTCAAGAAGC AAAAAAATAA CAAAAAGTCA GAACGTACGC TGAAGATACG TTCTGACTTT 780

TTGTTATTTT TGAATCGCTT GTAATGCCTT TcTGGtACTT GGkCyTGkTT GACCTCyTCa 840

TACGTATTGA TGTGGGcCCC TTwATTkGrG mCTTTtAGaT AATGrTCTGk TTTTAAwTCa 900

kGGTCCGCTG tGaAAGkTAw TTCyCgCGkT TTGCCTTCCT CATCTGCGGC TACTTGCGTA 960

TAACTAGCAA TTCCATAGCC ATTAACGCTA ATCGGCTTTC CTTGTTTTAA CATACACATT 1020

TCCTGCTTTA AnCCAATGGG ATTTAAATGA TCCTAnCACA CCAGCTAAAT TCGCCAGAGG 1080

TCCnCTCCTG C 1091

(2) INFORMATION FOR SEQ ID NO: 566:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1575 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 566:

ATTTCGTTCA CTGGTAAATT GAGATATTTC AAAACACCTT CTCGATTTTG AACTTCATCA 60

AAACAAATCA CTGTACCAAA ACCAGTACCA TCTTCACTAT CATCCAAATC ATTTGTTGAT 120

TGTGTAATCA ATACTAAGAA ATTATTAAAC GAACGTCCAA TACGTTTCAT TGATTTTAAA 180

ATCGCTTTAC CTTCTGGCGA AGATTGAAAA ATCCAAGATT CATCAAAGAA CGTCACTGTT 240

TCTTCTTTTG GATTTCTTGA ACCAAACTTC GAACAAAATG TTCCTAAAGC AAACATCAAT 300

GTAACAGAAA GACGTTTATT TTCATCTAAA TTTTCATTTT TATCAGTTGG TAAGTCTAAG 360

TCCTCAATTT CTAAAATTGT TACTTTCTTA TCGAATGATA GACCCTCAAC ATCACCATGA 420

GAAAAAGCTA ATTCTAAAAC TGAACCTTTA ACAGTCGACA AAATAAATCT ACCCATATCT 480

TTCACGTCAG AAATTTCACT TTCCACTAAT CGCTCAATTA CATGCCAAAA TCCAACCTTT 540

TTGCCTTCTT TCCGTTCTAA AACAACTTTC GCAATCGCTT CATTAATTTC TGTTTTTTGC 600

TTTAAATTCC ATGTTCCTTC GTATATATTG GTTAACATTG CTTTAGTTGT TGCAATCGCT 660

TCTGTTTCAT CAAATAAAAC AATCGGGTCA AGCACGCCAT GGTTTTTCGT ATCACGAACA 720

TCTAAAGTGA CGAAATTAAA GGACTTCACa AAATCAACAT CTAAAGGATA TTTCTTTTGA 780

TATTCAGGGT CATTAATTGT GTACATAAAT TGCTTTCGCA TTTCTTTTTT TGGATCAATG 840

TATAACACTC TGTCTTTTAA AAGTGCCATT TGCATAAATA AGGTTTTAGC CGTCACTGAT 900

TTTCCATTCC CAGTATCACC TGTAAGCGCC CAATGTGGGT TGTTTGTTTT CTTTCCATCA 960

ACCCCTTCTT TATTTGCTAA TAGCATGTTC ATAAAAATTA GATTTTTTGA ACTAGAAATA 1020

ATAGATTTTC TATCTTCTTT TTCTTCCAAA GTTGTATCCA CTCGTCCAAG GTAAAAGCCT 1080

AACTTCGTAC CTGCATGAAG TGAAGTAAAG AAATTCAACT CGGCAAACGA TTCAATCGTT 1140

GACGTGTGTT GCCAACGAGA AGTCTGCATT CGCAAAAAAT TACCAAATAA GGTTGATTGA 1200

AATAAATAAG GATTATCAAA CGTTGCTTTA ATAAGTGGAA TATCTAAACT ATCAAAACGA 1260

TTGATTAAAT ATTTTTTTCT TGCACGCATT TGTTTTTTCG TGCTACCACT AACAACAATC 1320

ACTGCTTTCC AGTCAATAAT ATCCACACCA TCATCAATTT TTTGGTCTAA ATCGTCaAGT 1380

GATAATTGAC CTTCAACAAT CTTTCGTTTT TGTTTAGAAC CAGCTAAGTG TGCTTCCTGC 1440

ATGATATTCT TAGTGCGGGT TCTTGCACGT GACGAACGCC CAGACAGTGC TAATTGCCCT 1500

TTACTTTCTG CAAAAGTCGC TTGAATCTTC ACTTCGACTG GGAAAGGCAT TTTTTGAACG 1560

ACCTCTAAAA AATGT 1575

(2) INFORMATION FOR SEQ ID NO: 567:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1158 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 567:

ATGGTCATTT AAATTTGATT GAACGCATGC GnAAGTTATT TGACGAATcA TTTAtTGGCG 60

TTTTTATAAA TACAAGTAAG CAAACATTAT TTACGCCAGA AGAGAAAAAA TATTTAATTG 120

AAGAAGCAAC TAAAGAAATG CCAAATGTGC GTGTTATTAT GCAAGAAACC CAGTTAACTG 180

TGGAAAGTGC AAAATCATTA GGGGCTAATT TTTTGATTCG TGGCATTAGA AACGTCAAAG 240

ACTACGAGTA CGAGAAAGAC ATTGCTAAAA TGAATCAACA TCTGGCGCCA GAAATTGAAA 300

CGGTTTTTCT ATTAGCAGAA GAACCATATG CACATGTAAG TTCCAGTTTA CTAAAAGAAG 360

TTTTACGTTT TGGGGGCGAT GTGTCTGACT ATCTTCCTCC CAATATTTAT CATGCGTTAA 420

AACAAAAGAA GAATGACTGG aGCTAAAATT ATGAATAagA AAACAGAAAA ATTTTCGTTT 480

AAATCaCTCA TTCCAaTGGT ATTGGCATtG CtGCTAATTG GTTTGTTTAT CGTCCCAATT 540

CCTTATTATA TTGAAGGACc AGGAACGACA GAAAATTTAA AAGAGTTTGT AACTGTCGAT 600

GGAAAAAAAG ATACGCAGTC TgGTGCTTTT TATTTAACAA CGGTGGGCAT TCGTTCGGCC 660

ACTATTTTTT CTGCCATAAA AGCAAACTTT TCGGATTTTC AAGAAGTCAT GTCCAAAAAA 720

GAATTAATGG GTGATAGTAG TAATAGCGAA TATAATCGAA TTCAACAATA CTATATGGAT 780

TCCTCAAAAA ACGCCGCAAT TGAGCAAGCC TTAAAATTAG CGAAGGTTCC ATATGAGATG 840

AAATTTAAAG GCGTCTATGT GTTAGCGATG GAAGACAACT CAAGTTTTAA AGGAAAAATT 900

GAAGTTGGCG ATACCGTGAC AGGCGTGGAT GGTAAATCTT TTAAAAGCAG TGAAGAATTA 960

ATGAATTACA TTAAAGCACA AAAAGTGAAT CAAAAAGTCA CTGTTCAATT TATTCAAGAT 1020

GGCAAAGCAA AAGAAGCGAC CGGCAAATTA ATTGAATTGC CAACAGATAA AAAAGCGGGG 1080

ATTGGTATTG GTTTGACCGA TCaTACAGAA ATTGATTCTT CTnTCCCAGT ATCTATTGAA 1140

GCGGGTGATA TTGGTGGG 1158

(2) INFORMATION FOR SEQ ID NO: 568:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2108 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 568:

AATCAATTCG TGACTGTTCG GAATGCCTTG AATTGGTCTT TCAATATCCC TGTCGTTCAT 60

GTTAATAATG AGCTAAGAAA AAAAATGGGT GATGaTAATT tCyCTTACAA TCACTACTTA 120

AGTAAAaTGA ACTATCCAGC TTCAGATGCT TGGGCTTACG AATCTGCGCC ATTGGGTTCA 180

GTAGAAACAA ATGTTGTAAC ACAAACAAAC GGCTTCCAAG CGCTTGCTAA CAAAGGTAAA 240

TATCAAAAAG CCTACATGAT TGAAAAAATC ACAGACAATA GCGGACATGT GGTCTACGAA 300

CACAAAGATG AAGGAACGCA AGTTTATTCA CCGGCGACCG CTTCTATTAT GAATGACTTG 360

TTGCGCTCAG TGGTGGATTC TGCGAATACG ACAAAATTCA AACCAACGTT AGCTGGCTTG 420

AATCCTCATC TAGCTAGTGC TGATTGGGTC GGAAAAACTG GGACAACTGA TGAGTTTAAA 480

GATTCATGGC TGATTGTCTC TACCCCAACT GTGACACTAA GtTCTTGGGC AGGACATGAT 540

CTACCAGCAC CAaTGACGAT GACCAGCGGC GACAATAATG GGAATTATAT GGCGAACTTA 600

GCCAACGCGC TGTACTATGC GAATCCAGAA TTGTTTGGCA TTGGTCAAAA ATTTGAATTA 660

GATCCAAGCG TTATTAAATC AAAAGTTTCT GaATtTACTG GTGAAAAACC TGGCTCAATT 720

ACCTATAATG GTGCCAAGTT CAACACaCCT GGCAAAACAA CGACTTCATA CTATGCCAAA 780

GATGGTGCAC CACAAAGTAC CTATAAATTT GGGATCGGTG GAACAGACAG CAATTACGCT 840

TCTTACTGGG GAAATCTTGC TCCACGTGCC ACTACAAATA ATAACAACAA TAACAATAAA 900

AACAACGATA ATAAAAAGAA TAATAATTAA CTCGCTAAAA TGAGCCAGAA aaCGTATTTT 960

cTGGCTCATT TTTAGGTATC AAAAAcGGCT AAACTAAGAG TTAaCGCTCT TAGTTTAGCC 1020

sTTtAATTCA ATTTATTTTG TTGAACCTTT TTGATCAATT CCATAAGGTA AAACAcTGTT 1080

TTTTCTTCAA TTtCTTCTTT GTTCATTAGT TTTGTTAATA AACGCATTGA CACGGCACCG 1140

ATATCATATA AAGGTTGTGT AATACTTGAT AGACGTGGAC GAGAGACTTC TGTAAGCAAT 1200

GAGTTATTGC TTGTAATGAT TTCAAATTCT TCTGGAACTT TCACACCAGC ATCTAGCATC 1260

CCATCTAATA AGCCAATTGC TAGCTCATCA TCTGTAACAA AGGCAGCAGT TGCCCCACTA 1320

TTACGGACAC GCTCAGCCAA ATTAATGCCT GCTTTAAATT TATATTCAGA TTCAAAAACT 1380

AAGCCTTCAT TATAAGATAA GCCATTTTCG GCTAAAGCTT CTTTGTAGCC TTTCATTCGG 1440

TTTTGACCAT TAATTGGATC GATTAATGCA CCACTAACAA AAGCAATTTT TTTGTTTCCA 1500

TTTTTCGCTA AAGTCGCTGT TGCATCTTTC GTTGCTTCTG TGTAATCAAT ATTAACGCTA 1560

CCAACTTGTT CGTCTGGATC AATTGAACCT GCTAAGACAA CAGGTGTTTT TGAACGTGAA 1620

AATTCACCAC GAATATCATC TGTAATGCGA TGACCCATAA AGATAATCCC ATCTACTTGC 1680

TTGGCAAGTA AATTGTTTAA TACAGTAACT TCTTTTTGAT CGTCTCCATC TGAATTAGCT 1740

AAAATGATAT TGTATTTATA CATTGTAGCT ACGTCGTCAA TGCCACGTGC TAATGAAGCA 1800

AAAAATGCAT TACTAACATC TGGAATAATG ACACCCACTG TTGTCGTTTT TTTACTTGCT 1860

AAACCACGTG CGACTGCGTT AGGACGGTAA TCTAAGCGAT CAATCACTTC TAAGACTTTT 1920

TTACGCGTTG CTGGTTTTAC ATTGGGATTA CCATTGACAA CACGAGAAAC AGTAGCCATA 1980

GATACATTTG CCTCTCTAGC AACATCATAA ATTGGTAATT GTTTGTnTTT CCATGTTTCT 2040

TCTCCCnATA AATATGAATT TGGnTCTCAT TTTCTGAAAA CAnCTGGTTA CCATTTCTGT 2100

CCTAGCCC 2108

(2) INFORMATION FOR SEQ ID NO: 569:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1151 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 569:

TTTTGACTTT TAATCACTAT TTTGCAAAAA AnTCAGACTT TTTTCATTAT-TTAAGAGACn 60

TCCCGCAAAA AGAACTTTTC TTTTTTTGCT ATTTCTAGTA AAATAATTCA TTGAGAGCAA 120

AACTATACGG AGAGGAGAGA GTTTCAATGG CAACAGGTTG GGTAGTATTA ATCGCAGTAA 180

TCGCTTTATT AGTAGGCGCT GCAGGGGGCT TTTTCCTTGC ACGCAAATAT ATGCAAGATT 240

ATTTCAAAAA AAATCCTCCC GTTAATGAGG ATATGTTACG CATGATGATG GCATCAATGG 300

GTCAAAAACC GTCAGAGAAA AAGGTTAGAC AAATGATGCA ACAAATGAAA AACCAAGGAA 360

AATAGTAAAG GTTATTCAGG TAATCATGAA TAAGTGTAAA AACGGGCTGA TTTTCAGTCC 420

GTTTTTTTGT TGTTTTGCGG GAGAATAGGA TAATCTAGGA GGAAAAGAGC GAAACGTTTA 480

ATCAAAAACA ACAATAACGA AGAAACCTTC TAGCTCGAAA GTAGTCTTCG TAGAACAACA 540

AAGTAAAAGG AGGGACAACT GAATGTCCAT ATTCAAAAAA TTAGGCTGGT TTTTTAAGCA 600

AGAGAAAAAA AGTTATATTA TTGGGGtTTT CTCATTAATG ATGGTCGCTC TTGTTCAATT 660

AGTCCCGCCC AAAGTTATTG GCGTCGTTGT AGATGAAATC GTTaACAAAG AAATTCGCTT 720

AACGAAAATT ATCGTGTGGG TTGCACTCTT GATTGGTGCT GGGCTTGCCC AATATCTTTT 780

TCGCTATATT TGGCGGATGC ATATTTGGGG GAGTGCGGCT CGTTTGGAAA AAGAGCTACG 840

GACTCAATTA TTTCATCATT TCACAAAAAT GGATAGCATC TTTTATCAGA AATATCGGAC 900

AGGTGACTTG ATGGCGCATG CAACCAATGA TTTAAATGCC ATCCAAAATG TTGCTGGAGC 960

TGGGATTTTA ACGTTTGCCG ACTCTGTGAT TACGGGAGGm ACAACGATTA TCGCAATGGT 1020

TCTATTTGTC GATTGGCGCT TAACATTAAT TGCTTTATTA CCGCTGCCTT TATTAGCCGT 1080

CACTTCACGA GTTCTAGGTT CTAAGTTGCA TGATGCCTTT CGAGATTCAC AAGCTGCTTT 1140

TTCAGCGATT A 1151

(2) INFORMATION FOR SEQ ID NO: 570:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 736 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 570:

AAAATAATGA AGGAGAATAG ATATGGGAAA ATTTCAAGTA ATTGATCATC CGCTAATTCA 60

ACACAAACTA ACAATGATTC GCGAAAAGAA CTGTGGGACA AAGGTTTTCC GCGAaGTTGT 120

CAACGAAATT GCTATGTTAA TGGCTTATGA AGTATCACGG GATATGCCTT TAGAGGATGT 180

TGTCATCGAA ACACCGATGG GTAAATCAAC ACAAAAAACA CTTTCAGGGA AAAAAGTAGC 240

AATCATTCCT ATTTTACGTG CCGGCATTGG CATGGTTGAT GGGATTTTAG AATTAATTCC 300

AGCAGCGAAA GTTGGACACG TTGGACTTTA TCGTGATGAA GAAACATTGC AACCGCATGA 360

ATACTTCGTG AAATTACCTG AAGACATTGC AAGTCGTCAA TTATTTGTTG TTGACCCAaT 420

GCTAGCAACA GGTGGTTCCG CAATTATGGC CATCGATTCA TTAAAAGAAC GTGGTGCTAG 480

CAACATTAAA TTTGTATGTC TAGTGGCCGT TCCAGAAGGT GTGAAAGCTT TACAAGAAGC 540

GCACCCAGAT GTAGATATTT ATACAGCAGC TTTAGATGAG CGCTTAAACG AAGACGGCTA 600

TATCGTTCCT GGTTTAGGGG ATGCTGGTGA CCGTTTATTC GGAACAAAAT AATAACAAAA 660

AACCGTTAGg TAACCGATTA TGGkTACTAA CGGTTTcTTT TGCyTAAAGA mCTTycTAGG 720

TGTwTAtAAA AAtaaA 736

(2) INFORMATION FOR SEQ ID NO: 571:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5621 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 571:

CGAAAGCTAG ATACGGCGGT AAGTAAGCAA TCGATTATCG ATTGACAACT ACGAGCAAAA 60

GCTGAGACAC ATGTCTCCAG CTGACTATCA TAGGATTTAG taTGTGGGAC AGCAGTCAAG 120

CATGACTTTT GTCCCACATA TTTTCTTTTA TATAATATGG TGGCTCAACT GGAAAAGAAT 180

ACTTCTTTTC CGTTCATTAT GTTATGATAG ACAAGAAAAA AACACCAAAG GAGAGAGAAG 240

AAAATGACGA TCAAAGAATT CATTAAGAAA AGTCGCTTGA AAACATGGGA CGTTATTATT 300

ATTGCTATTT TAGTTCTTGG TTCTTTTCTG CCCCTTGTTG TTTTTGCCAT GCAAAACCGT 360

GGTCAAGAAG CGGCTACGTA TCAAGCAGTC TTGAAAGTAG ACAACAAAGT AATCAAAGTC 420

TTTGACTTAA AAAAGGACGG CCCGCATTAT ACATATAAAT ACGAAGCAAA AGACGGTGAT 480

TACAACCTGA TTGAAGTAGA TGGTGATCGT ATCCGCGTGA AGGAAGCCAA CTGTGCTGAC 540

TTAGTGGATG TTCGACGAGG CTGGATTTCC AAACCTGGTG AAACCCCGAT TGCCTGTCTA 600

CCCCACAATT TGTTTATCAC CGTGGAAGCT TCTGATGGGA GTGAAGACGG TAGCTTAATT 660

TATTAAGCAC GACAACTATG AGCAAATTAC ACAAGAATAT TTATATTGCT ATGCTCGTTG 720

CCCAAGGGGT CATTATTGGC TTAATCGAAA ATATGATTCC TTACCCTTTC GCATTTGCTC 780

CTGGGGCAAA GCTCGGTCTA GCAAATTTAA TTACCATTAT TGCGATTTTC ACCATGCGTA 840

AgcGTGATAG TTTTTTCCTG GTCTGTTTAC GTTTGATTTT AACGACATTA CTGGGCGGCA 900

CACTTTCGAC CTTTTTCTAT AGTGCCAGTG GTGCCTTATT AAGTTATTTC GGCATGTTGC 960

TTGTCAAGCA GCTGGGGCCA AAACGGGTCA GCATTATCGG GATTAGTGCC GTTGGTGGTT 1020

TCTTACACAA CGTTGGTCAG TTGCTAACGA CCTCTTTCTT TGCTCATTCG TGGGCGCCGA 1080

TGCTGTACTT ACCAnTTTTA TCTCTGTTTg GGGTtACTCT CGGGGaTGCC ATTGGGgATT 1140

GCCGCTaATn ATtACyACAA CATGTcCAGA CGTTACGAAA ATTCCAACTG AATTATGAAC 1200

AAACAACGAA ACATAACTGG AGTCAATTTT TCTAATTGAA AACCTGTCCG CCACAATTAA 1260

AGGAGTCATC TTATGAATAT ACATCCTATG TGGAATACCT ATCCTACGTT ATCAAAAGAA 1320

TTAACAACTA CTCTTCGCTT GATGGAACAA GCGGTTCAAA TCGATAATCA AGAGATTCAA 1380

GCTGCTGTCC ATGACATGAT TCATTCTGGC GGCAAATTGC TACGTCCAGC CTACCAACTG 1440

TTGTTCTCTT ATTTTGGTGA ACAGCGGGAC CCTAAAAAGG CGACAGCGTT AGCGrCTTCC 1500

ATTGAATTAC TGCATACCGC CACGTTAGTT CACGACGATA TTGTCGATGA AGCTGATACG 1560

CGTCGTGGCT TGCCAACATT GCGTTCGCGC TTTGGAAATA GCACAGCCGT GTATACAGGC 1620

GATTATCTTT TTGTTTGCTG TTTCAAGCTC TTATCTGATT ATTCTTCTTC ATTAAAAAGC 1680

ATTCAATTAA ATTCGCGCAG TATGGAAAAA GTCCTGACTG GTGAGCTTGG TCAAATGGAC 1740

AATCGCTATA ATTTCGAGGT AACGATTGAC CAATATTTAA AAAATATTTC CGGGAAAACA 1800

GCGGAATTGT TTGCGCTAAG TTGTTTTGTC GGTGCCTATG AAAGTGGCAC CTCCCAACGT 1860

TTTGCCAAAC GTTGTGGTGA AATCGGCGAA AATATCGGAC TGGCTTTTCA AATTATTGAT 1920

GATATTTTGG ACTATACACA AACTGTGACG CTATCGGCAA ACCCGTCTTA GAAGaCGTAC 1980

GTCAAGGCGT GTATTCTTTG CCCTTGATTT ATGCACTTGA AGCCAATCGT GAAGTCTTGT 2040

TGCCACTTTT AATGAAAAAA GAAGCCTTGA CCGATGAAGA AACACAGGAA ATTTATCGTC 2100

TTGTCCATGA ACTAGGTGGC GTAGAAAAGG CCCAAGAATT AGCCACTCAT TACACAGAAA 2160

AAGCGCTGAA AGAAATCAGC AAACTTCCTG AAACAAAGGC CCAAGCGAAA GAACAATTGT 2220

ATGAAATTAC ACAAACGATT TTGACAAGAG AAAATTAAAA AAGCCGCTGG ATCCTATTGA 2280

GGATTTAGCG GCTAGCTTTT TAGGTGATAT CTTGGTATAA TAAACACGAA GAGAGCGTAC 2340

TAGACATACG CTCTCGGGTA GACCGTTTAA GACGGTAGCT TGATTTTGAT ACGGAAATAA 2400

ACCCGTCAAT CCTTGCTAAA GATTATGACG GGTTTATTTT TTGTCGTTTT TATTACTGAT 2460

TTTGACAATT AATGTTGTTA CTTATACAAC TTAACTACTT TTGATGCTTC ATTCCATCTC 2520

TAATTTGTAT TACCTAGTGA ATACATGTTA TACTAATTAA AAAATGGAGG TGTTAGCTAT 2580

GACATCAACA AAAACAAGAA AACAAGGAAA CTCATTAGTT ATAACAATTC CTGCTACTTT 2640

AGGAGTTAAA GAAGGGGAAG AATTTGTTAT TCTTCGCAAA AATAATGGCT CCATCGCATT 2700

GATTCCAAAA GTAGAAGATT TTTTTGAAAA TACTGCTGAA GGAGAGTTTT ATCTTCCTGA 2760

ATTAGCGATT GATTATTCAC CTAGCGGAGG AGAAGTTGAT GGGCTATAAA CCAACTCAAA 2820

GAGATATTGT TATTATCGAC TTTGCGCCTT CTAAAGGTTA TGAAATCAGA AAACGACGAC 2880

CTGCTCTCGT AATGAGCAAA GACAGTTACA ATATTTCAAC AAATTTAGTT ATTGTCTGTC 2940

CAATAACCTC GTTGGATAAG GAACGTCCTT TTCTTGTCCC TATTTATAGC GAAAAACTTC 3000

ATACCTCCGA TAATGCAGTA AGTAAAGTGA ATACGCTCCa AGTCTATTCA CTCGACTATA 3060

CAGAACAGGC ACAACGCAGA ATAAAATATG TAGATACGTT AGACGAAGAA ACATTTTACG 3120

AAATTGCTCA AAAGTTTTTA CAAAACTTTT CTTTTGCTGT TTAATTCGGC TTCAACTGAA 3180

CACGCATATA TGATTTACTA TTCGACTTTT CCGTTTCAAC TTGCTATAAT AAACACGAAG 3240

AGAGCGTACT AGACATACGC TCTCGGGTAG ACCGTTTAAG ACGGTAGCTT TAATTTATTG 3300

GTtAATAAAC CCGTCAATCT TCGCTAAAAG TTATGACGGG TTTATTTTTT GTCGTTTTTA 3360

TTACTGATTT TGACAATTAA TGTTGCGAAA gcAATCGCAA GCATCAATGC TTCAAAGACA 3420

GTCACTACTA TTCCTTTCTA GGGAAAGAGC AAACAAGCAA AAACATTGGC AGTCCCCCTT 3480

TTTTTGAGAA ATTGCCACCG TCATAAACTT CCTACTTGTG TAGTATAACA CATTTGTACC 3540

TTCCTAGAAT ACGCTTACTT TCCTGATTAT TCAACAGCCG CTTTCTTTTG TAATAAAAAT 3600

CGCGCACCGT TTACTGGGAT AAAACGATTC AGCCAACTAA AATTTTTTAG ACTAAACAGA 3660

CCGACCATCT TTTTTGATGG TCGGTCTGTT TAGTCTATTG CTCCGCCTTT ACTCCTTCTC 3720

CGACAATGCT CGCCAGACAT CAAATGTTTG TTTTAAGCCT TGGACATTAT GAACACGTAA 3780

AATTTCTACG CCTTGTTGTG CTGCATACAA GGAAGCAGCG GCTGAGCCAA AATCACGGTT 3840

GGCTGGTTCT GGTTCGTTCG TTAAAGCGCC AATGGTTCGT TTGCGAGAAA CGCCGTATAG 3900

TAACGGATAA TCTTGGTAGC GAAAGGCGTT AGGATCTTTT AACAGTTGAA GGTTTTCTGC 3960

GACAGTTTTA TGAAAGCCAA TCCCTGGATC GAAACAAATT TTTTCTAAGG AAAGCCCATA 4020

TTTTTGGCAT TGTTCAATTT TTTCTGTATA AAATTGGTGC AACTCTTCTT TTAAAGAAAG 4080

CGATTGGCGT TTTCTAGAAT GCATAATCAC AACTTGGACT TCAGGATATT GTGCGAGGAC 4140

TTCTGCCATC CCCGGAGTGT CTAATCCTTT GATGTCATtA ATCATTGTGG CGCCAgcTTG 4200

GATCGCGGCT TCCGCTACTT CGGGATAGTA GGTGTCCACA GAAATCGGCG CTAGAGAAGT 4260

CTCACGTATT TTTTTTATTA ATGGTAATAC TCGGTCTGCC TCAACTTGTG GAGAAACTTC 4320

TTCGTAGCCA GGACGTGTGG ATTGGCCGCC AATGTCTAAA ATATCGGCCC CCGCTTCTAG 4380

TAGATGAAGC GCATGTTGAT AGGCCACGTC GACCGTTGTG TAGCTACCGC CATCAGAAAA 4440

AGAGTCGGGG GTCGTGTTGA CAATCCCCAT AATTTGGTAA TTCTTTTTTA AATCAATCAT 4500

TGTCACATTG CTCCTTTATC TGCTGAATCA TCTTTCGTAA AGCCCGTATT CTAGGACTTA 4560

TCTTCATTCG TTCACTCGTG GACAACTCCG CTAAGGTTTT GCCTCGGTCT GGTAAATAAA 4620

TAATCGGATC AAAGCCATAA CCGCCTGTAC CGCGCGGTTC AACTAATTCA CCAGTCAAAG 4680

TGGCTTCCGT TTGCAACAGT TTGTCGTCAT TGAGTACATA GACCAACGTG GCTGATAAGG 4740

TTAACTCTCG GGTCGACTGT TGCCCTTCAA ATAAGTGCAA CAGTTCGCGG TTCTGTTCTT 4800

CCGGATTGGC TGAATGGAAA AAGCGGCTTG TATGAATACC TAATAAATCA GGAAAAGCAG 4860

TTAACGTTAA GCCGCCATCA TCACCTAACA CTGGGCGGCC AATTAATTGT TGGAAAAAGC 4920

GAGCTTTTAG ATAAGCATTT TCAGCATACG TAGTGCCAGT TTCTGCAGGT TGTTCTTGAC 4980

TGGTTGTATA CTTTCGATAA GAGACAAGTC GAATAGCGGG GTCCTTAAGT CCTGATTGCA 5040

TTTCTTTTAA CTTACCTTGG TTATTCGTGC CAACAATGAT TTCCATTTTC TAACGTATCT 5100

CCATTAAAAA TTGATTTTTC CACTCTTGTT CTTTAAATAG TCCGGTAAAT TGAAATGTTT 5160

TGGTGCTACT TTGTGGTGTT TTCACGCCGC GCATCGTCAT ACACATATGT TCCGCTTCAA 5220

TGGCTACTGC AATGCCTTTG ACGGGAATAT TTTCTTGGAG TTTGCGGGCA ATTGTCACTG 5280

TTAAATCTTC TTGAACTGTG GGACGTTTGG CACAATGTTC TACTAATCTT GGTAACTTGC 5340

TTAGTCCCAA TACTTTGCCA CCTTCTGGCA GATAAGCGAC ATGTACTTTT CCATAAAATG 5400

GCAATAAATG GTGTTCACAC ATTGAATAAA AGGCAATATC TTTTACTAGA ACCATTTCGC 5460

CTTCATTTAA GCTATCAAAT AATTTATAAT CGTCAAATTC AGGTTCGGTT AATCCACTAA 5520

AAACTTCTGC ATACATTTTA GTACGCGTTT GGGTGTATCA AGGAGTCCTG CCCGTTGAGT 5580

ATCTTCCCCC ACTGCTTCTA AGATTGTGGT AACTGCCTGC n 5621

(2) INFORMATION FOR SEQ ID NO: 572:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3807 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 572:

CAnATATCCA ACAATGCTTG GAATCTCCTC CGCTTTCCAG TGTAAAATGT ACTTGGGTCA 60

TGGTAAAAGT CCCCCTGGGG TATGGTTTTT GGTCGTTAAA AACCATGTAC CGTAAAAGGA 120

CTGTTATATG GCCTTTTTAC TTTTACACAA TTATACGGAC TTTATCTAAT GATTCCCTAA 180

CAGTATTTTA CAATTACTTT TnTATTTGCC GCCTTTCCCC AATAAGGCGG TATCTTTTTG 240

CATAAAAATA TTGAATTAAA ATGGTATTGT CTACGTACGC ACTTTTGGAC ATTTTCACAT 300

AGATCTCCTT TTTACCGAGA AGGTGGCCAT TTTTTTGTTG AAAATTGAAA ATGAGTAAAG 360

TrATATTATT ATATAtTAAG CTATAACTAT AACCAGTAAG AACTATTTTA CCTTAATAAG 420

CTAGCTCTTG CTTGGACCAT TAGCTCAGCT GGTTAGAGCA AACGGCTCAT AACCgTTCGG 480

CCACAGGTTC GAGTCCTGTA TGGTCCATAA AATATTTAGT GAGTAACAAG GAGTATTTCT 540

CATTTGAGAG ATAAGCCTTG TTTTAGTTGT AGTAATCCAC CTTTATATAT TCGAAGTTGG 600

ATATCATTTG AAAACCAAAT TAGTGATACT CAGACTTTTT TAGTTAATAT CAAAATCATA 660

AAATCTCTAT TTGACAGTGT TTGTTCACGA AAAAATAAAA AGAGTGTTCA TTTATAACTT 720

CTTGAAAGAG GAAGGTTATT TGTCTTAAAT GAATTTATAG GATAACGGTT CTTAGTATAG 780

AATGTAAAAA AAGCCAGCCA GTATTTGGCT GGCTTCTGTG GTAAAGCGCA GCCGTTTTCC 840

GCCAGTTGCA CTTTAGGTGC GTGTTAGCAC TTGCTTCAAT AAAGAATTAA GAATTACTAT 900

TATTATGCCA TAAAAGAAAA ACTTTGAATA GATATGGATA GCAAAAAACC AGTGAATAAA 960

ATTCGACTGG CAAGATAGAT TTTGGATAGA ATTCCTTGAG GAAAAGTTAG AATTATTATG 1020

CCACCATGGA GGAATAGTTG CTATGATGTT GCTATCAGAA ATTGATTACA AAAAGACAAA 1080

AGTAATGTTC ATCGATTATA AAAAATGTTG AATTTTACAA CGTATGTGAT GTGTATAAGT 1140

ACATTTGCAA TCTCCTATAT TATTTGATAT TCCACTTTGT GAATGTTAAT GGAGCATAGA 1200

TTAATTGGTA GAGCAGCGGT CTCCAAAACC ATCAGTATAG GTTCAATCCC CATTGTTCGT 1260

AGCATTGCTA CTTAAATAAA AGAATCGCCA ATAGATGTTT CTGCTGAATT CAGGATGAGT 1320

CACTAGATTG CAAGTGTTCC TTTTTTATTT GGCAGAAGAA TGAAGAATTA TGTATAATAA 1380

CTTTTGCAGA AAAGTGAAAA GATAATGGCT AATCTCTTGA ATAAAGGGGT GATGCCTATG 1440

GTTCATAGCT TTATCTCTAG ATGTAGTAGA CATGTCTGGA TTCGAAATAT TATTGCTAAT 1500

GGTTGCATTC GTAACGCTAG TTTTGCTGGG GATAGACCGT ATGAACACAA AAAATAACCA 1560

TCTCACTTGA CGGGGCGTAg CTGGTTATTA AAACTAATTT ATGACAGTAA CCGTCTTTTT 1620

AACGGTTCTA CATAAGGATA TGTTAATAGC ACATTTTTTT CTATATATAT TATACCATGT 1680

AACAAAATAA ATTCATTCAA GATTACTCAA CATAAGTGAG TGGTCTTTTT TGTGAATAAA 1740

AAAGCCACTA GGACAGGGAA CCTAGTGGTC AGGGTAGCAA ATATTATTAG TCGTAAGTGT 1800

ACAATAAAAA TAATTTTTAT AGGAATATTT AATCAACATC ATAATACTAT TATTATAAGG 1860

AGTTGCTACT CAGAATTATT ATATCAAATA CTAAAATAAT TTTGCAATAA AAAAGTTATA 1920

CACTAAAAAG CCACTAGGCT GTGGGCCCTA GTGACTAGGT AGCAAGCATT TTCAATTGGA 1980

GTTTTACACA AAATTTGAGG ATTATTATAA TACTTATCTG AATCGATAAa TTATCATATA 2040

TATAGGGAGT TGCTACCCAA TGTTATTTTA GCAGAAAGCA ATTTACTAAA TCAAGTACAT 2100

AGAATTTAGA GAAAGTAAAG AAATACATAT GTTAACAAAG GATTAGCTAG AAATGTGGTG 2160

AGGCCATGGC AAAATATACA AAGTGGCTAA CTGATGAAGG ATTAATTAAA ATTGAAAGGG 2220

TGAGCACGAG AGGGTCTCGC AGATGaACGA ATAGTACAAA ATATTGGGGA TTAAAAGACC 2280

AACCTATGAT TGaAAAAATA AATATTCTGa CATTACTGAC ACCCTAAAAA GaGAAAAAGA 2340

AGTTGTAGAT TGCCAAGTGG AAAACGCACT ATTTAAAAGA GGTGCAGGCT ATGAATACAC 2400

CGAAGTTACA AAAGAACTGA CAGAAAATGG GATAAAAATC ACAAAAAAAG GTTACAAAAC 2460

AATTGAATTA TATTGAGATT TAAGGTTTTT AATGGAAAGC ATGTCACTTT TAAATTTTTG 2520

ATCTGAAGGC ATAGCATAGA AAATTATTTT AGAAAATTTT GTGTAGCCAA CAGGAAAGCA 2580

TCAAAAACGC GATCTTCAAA AGTATGCTGC ATGTGTTGTA GTTAACTTAT TACCATTTGG 2640

TGGCGTTGTT TGGGATATAC TTAATTGTGA AAATCTAATG CAGGTTTTAA TTAATCCGTT 2700

GCAATGAAAA AAAAATTATG GTTATGATGT TAAAATACTT TTAGATGTAG GGAAAAGAAG 2760

TTTAACTCCT GAACAATTTG CAAAATTCAA TGCTTTAACT ATTGCAGCTG GTATAGCAAT 2820

TACTGCTGTG GCTTATTAGG GAATTAATTA GTTGAGCGGT TATGTGAGTA TATACGATTG 2880

TTAGAAGAAA AAATAATATC ATTTTTAAAG TTTTTAAAAG TTAACCCTCA GAATAATTGT 2940

TTTACATGTT ATAGTTATTC ACACCCTATT TTAAATAGTC TTCCTTTTAG CCTTTTGGTG 3000

CAGTGTTTTG AGATAAAGGA CTATTCTGCT TTAGGGTTGG ATTCTGGATA TTTATTTACA 3060

ATAGAAGCTG AGraAAAATG CTATTATCAT ATTGCTAACG TAACAGGAAA TGATTTTTCT 3120

ACTAGAAATT TTTTCGAACG GTATTGAAGG AACmATAAAT TGAATAAAAT ACTGGAAGAA 3180

ACAAGTGACA TACGAAATAT TATTATATTA ATAGACCTAA TGGGAGGAAC ACCTGCAAAT 3240

GTAACTATGA TCAAAGCAAG TATGGATAGA mGAATTCAAG TAGTAACTGG AATGAATCTA 3300

AGAATGGTAT TAGAATCTTA CTTTAAAATT AATGAAGAAA AACTTGCTGA GCATTTAATA 3360

GAAATAAGTA TAAGTAGTAT TAAAATACCT GAGATACAGA TTAGTGGCGG AGAAGAGTAA 3420

TTATGATGTA ATAATCTTAG CAGCGGTGGG GTGTGATCTA TATACAAATA TTAATATTTG 3480

TGTTAAACAT TTTTTTGATA TAAGAATGTT GTCCAATCAA ATTCTAGAAA TATTTTAGAC 3540

TATAAAATTT ATTATACGGG GAAAATACAA CTTTTTAGAT AAAGTCCGTA TAATTGTGTA 3600

AAAGTAAAAA GGCCATATAA CAGTCCTTTT ACGGTACAAT GTTTTTAACG ACAAAAACAT 3660

ACCCCAGGAG GGACTTTTAC ATGGACCCCA AGTACATTTT ACACTGGAAA GCGAAGGAnG 3720

ATTCCAAAGC ATTAnTGGAA TAATTCCGGT AAAGGGATGA CGGTTCCTAA AAAAATATTT 3780

TAACCAACGG GAATTTAATC CACCAAT 3807

(2) INFORMATION FOR SEQ ID NO: 573:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1406 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 573:

CGTTGCACGT GTCACACATA CTTGnACTTT GCnGGCACTC TTTAGAATCC TTGCnATGAC 60

AAGATGnTGA TTGCTGTGGT CCGTGCACTT CAGATAACGA GGAAGCTCGA ATGACnGCGC 120

TTCCCATCGT GACAACATCA AATAATTTTT TCGTCCGTAA ACGGTTTCTA ATCTTTTGGT 180

CTTGTTCCAC GTTGGCGTTT GCGTTCTTCA CGTTCAATGC GTAGACGTTC GCGTTTTTCT 240

TCAGGATTTG GATCACTGAT CACTGTTCCT GCTTCAATTT CTTCTAACGC ACGACCTACA 300

CTTTTTACAG AATCAAAGGA TTCAACGGTT GGTTGAACGC CTTCATCTAA TTCATGCGCA 360

CGTTTGCTTG CTAAAaTAmC CAAAGAaTAT TTkGAAGGGA CTtCTTTTAA CAATGAGTCA 420

ATTGACGGTT TTmACATCAT AAGGCTACAT CTCCTCTAAC ATTTTAATAT ATTTTCCAAT 480

TACACGGTCA ACACGGAAGT GCTnCACTCG CGATAATGTC TTTAATTCGT TGCACGGCTr 540

AAGGAACTTC ATcATTCACT ACCGCATAAT CATACAGCGC CATCATTTCG ATTtCTTCTC 600

TGGCCACAGC CATTCGkTCT TCAwTCaCAG ACATTTCGTC TGTtCCGCGG CCAaTAATTC 660

TkGATTTTAA TTCCGCTAAA TCGGGGGGTG TCAAAAAGAT AAATACGCCA TCAGGCACTT 720

TGTCTTTGAC TTGTTTAGCC CCTTGGACTT CAATTTCTAA AAAGACGTCT TTGCCTTCAT 780

CTAACGTCTG GTTAACATAC GTTAATGGTG TCCCATAATA GTTACCAACA TATTCCGCAT 840

ATTCCAACAT TTCTCCAGCT TCAATCATCG CTTCAAATTC TTCTCTTGAA CGGAAATAGT 900

AATCTACTCC ATCCACTTCG CCTTCTCGCT GTTTCCGAGT GGTCATTGAG ATTGAATATT 960

GAAAATCATT TTCTTCACTG TCGAAAATAG CCTTTCTTAC AGTCCCTTTC CCCACTCCTG 1020

aAGGACCTGA CAATACGATT AATAACCCGC GCTCTGaCAT AACGACTTCC TTTCAAAACa 1080

TAATTTCACA GTATTAGTTC yAwTTTGcAC TTTTTTTCTG CAATTTTCAA GACTTyCTtG 1140

GAGATtCACT GAAAAAATAm CATTTCTTCT TTAATTTGGC TTTCATAACT TGTTCTTTTT 1200

TTAAACAAAT TACCTAGAAT TCCTATACAG GCAGCCATTT TTTAAGTAAA ATGAAGTGAT 1260

GACTCCTTTA GGATGAAAAA ACGTGAATAA GAAAGGCTGA GAAACATCGC ACGAATTAAA 1320

CCATTTACAT TGTTTCTGCT AGGCTTAGTT GTCAGTTGTG GGGCTnTTTT GCCCAACCTA 1380

CTGTGACTCA CGCAGAAGAG ATATAC 1406

(2) INFORMATION FOR SEQ ID NO: 574:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 564 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 574:

CATTAATCGG CCTACACACC CATAAnGCCn TTTCTCTATA AGCAGTAAAC GCCCTAACTG 60

ATTATTTTAC GkTGCGtATc TGtGaCCAAT tAGGTtATAT AaTAACTGTt GGTACGTtCC 120

tTTTCCGTTg ATCCATCACC CATAAGCCAG ATTGATTTTT yCCTTGAATA GkTCCCGTTA 180

ATCGCATCAC TTATTACCTG CCATAACAGA GTAGTTATTT GTGCATCTCC GCAAGCAATT 240

TCGTTCAACC AGCTCTCTAC ATCCCAATCA TTAAATTTAG GTAACGACGG ATTTTCTACA 300

TATGCGGTTG CAATTTTTGA TGTGAATACA TAATTGGGCG TAAACGGTTC TAGTTGTTTC 360

GTTTTAAGGT TAAATACCCC GTTTTTCACT GGTATAAGAT ATCTTGAGAC CGTTTTAGGC 420

TTTATTTTAG CTTCTTTCCA AATGTGAAAA ATAACATCAT TTACsGCTCn AGCATTGGGG 480

TTCGGGTCAG CCACCCAATC ACTCGGTTAA TTGGGTTCAT TGGGGAAAAA GCCTCTTTGG 540

CGAACAAGCA ACGnGATTCT CCAA 564

(2) INFORMATION FOR SEQ ID NO: 575:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 601 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 575:

AACTATGCTG AGTCTAGGGT TATTATTAAC TGGGTGCGGA CTCAAGGAGG CACAACTGAA 60

AAATCCAGCG AACGAATTAG TAACAAACCA GTTAAAATTG TCGCTGTTGG TTCCACCGCT 120

TTACAACCGT TAGTAGATGC TGCGAAAGAT CAATTTACAC AAGAACATCC TAATTATACA 180

GTTTCTGTCC AAGGTGGCGG CAGCGGAACA GGACTTTCGC AAGTTGCAGA TGGCGCAGTG 240

ACCATTGGAA ATTCCGATGT GTTTGCCGAA GAGAAAGATG GCGTGGATGC CTCTAAATTA 300

GTCGATCATC GGGTGGCCGT GGTCGGTATG GGACCAGTGG TCAACAAAGA AGTCGGCGTG 360

AAGAATTTAA CGAAGCAACA ATTGATTGAT GTCTTTACTG GCAAAGTCAA AAACTGGAAA 420

GAAGTGGGCG GCAAAGATCA AGAAATCGTC GTAATAAACC GCGCAAACGG AAGTGGCACC 480

CGAGCAACAT TTkAAAAATG GGGCTTAGAT GGAGCTAAAC CAGTTCmATC ACmAGrACAA 540

GtTCTTCgGG AcAGTTCGTA AAATTGTTGA ACAAACACCA GGaGCAtCAG CTATTTAGCT 600

T 601

(2) INFORMATION FOR SEQ ID NO: 576:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 708 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 576:

TCATGGAAAC TGCAAGAAAG CnTAAnAAAT TCCTGAAGTG GAAAGAATTC TTTAGACTTT 60

TCTCGTTTTA ATTTAACTTT TCAATAAACC ACGATCATTT TGACAATTTC CGCATAGAAC 120

ATATTGCCAC CATTTTGATA AAGGTAACCC CCATTGTAAA CTAACGCTTG CCAACGATAA 180

TAACGATATT TCGTATTTTG GTCATTTCCT AGCAACGGTG AAAGGACTTC TCTTGAATAT 240

TCTTCTGCTA ACTCTGTCGT GTGTTTCCCG CCGCGTTCTT TAACAAACTG AATATAATCC 300

AGTCCATAAT TATATGCTTG TACGCCTGTC CACTCATCAC AGCCAGCTTC TTTTGCTTTC 360

GCTAACGCTT CCGCGAGAAA TTTGACGCCA TGTGCGATAC TTTCCTCTGG TGTTTCAATC 420

ACTTCTTGTT GTCCATAAAC ACTTTCACTG CTCTGCATAA TATCGCCAGA TTTCCCTTTA 480

GATTCTGTAT AGATAATCGC TAAAGCTAAT GTTTTGTATT CAGGTATTTT ATTTTTTTTG 540

ACTTCTGCTT CTACTTGTTT TTCAAAGGTC ATTACCGTTT CAACATTTTG ATGTATTTTA 600

TAGCCAAGAT AGCCAAGACC AGCAACTAAA CCAACAAACA AAAGTAGCAT GATTTnATAA 660

ATTATTTnTC TTACTCTACG CATCGAATCA TCCAAACTTT TTCTTTTA 708

(2) INFORMATION FOR SEQ ID NO: 577:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3656 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 577:

TTGAAAAAAT CACAATTTAA TGATTTATCC ATTACGGAGA TGCACACAAT CGAAGCAATT 60

GGTATGTACA AGAAAAAAAC GTCTTCAGAA GTAGCCAAGG AACTGTCAAT TACCGTCGGA 120

ACACTTACTG TAGCGATTAA CAACTTAGTT AAAAAAGGCT ATGTTGAACG TTTGCGTAGT 180

GAAGATGATC GACGTGTAGT AAAGCTTGGC TTAACCAAAA AGGGAAAATT ACTTTTTAGA 240

GTCCATCAGC ATTTTCATCG TGAAATGGTT AAAAACATTT TGAAAGGCAT GGAGCAAGAA 300

GAAGAACAAG CGCTGTTAAG AGCACTAAAA AATCTTCATG ATTTCTTGCA AGAATACAAA 360

TAAAAAGTGA GGATCCCCAT GaAGAATTAT GCACGAATTA GTTGTACAAG CCGTTATGTT 420

CCAGAGAATT GTGTCACGAA TCACCAACTG AGTGAAATGA TGGaTACCTC TGATGAATGG 480

ATACATTCGC GAACAGGCAT TAGTGAACGT CGCATCGTTA CGCAAGAAAA TACGTCTGAT 540

TTATGTCATC AAGTCGCAAA ACAGTTATTG GAAAAATCTG GAAAGCAAGC AAGTGAGATT 600

GACTTTATTT TAGTTGCAAC AGTGACACCA GATTTCAATA TGCCTTCTGT TGCTTGCCAA 660

GTGCAAGGCG CAATTGGCGC CACTGAGGCC TTTGCTTTTG ACATCAGTGC AGCGTGTTCT 720

GGTTTTGTCT ATGCCTTAAG CATGGCAGAA AAACTAGTCT TAAGTGGTCG TTATCAAACA 780

GGTCTAGTAA TTGGTGGCGA AACATTTTCC AAAATGTTGG ATTGGACCGA TCGCTCCACA 840

GCGGTGCTTT TCGGTGATGG GGCTGCCGGC GTCTTGATTG AAGCGGCTGA GACACCGCAT 900

TTTCTAAATG AAAAATTACA AGCAGATGGA CAACGGTGGG CGGCTTTAAC GTCAGGCTAC 960

ACGATCAACG AAAGTCCTTT TTATCAGGGG CATAAGCAGG CGAGTAAGAC GCTACAGATG 1020

GAAGGACGTA GTATTTTTGA TTTTGCGATT AAAGATGTTT CACAAAATAT TTTATCACTC 1080

GTGACGGACG AAACCGTGGA TTATCTTCTA CTACATCAAG CAAATGTGCG CATTATCGAT 1140

AAAATTGCAC GTAAGACGAA GATTTCCCGC GAGAAGTTTT TAACGAATAT GGATAAATAT 1200

GGCAATACCT CAGCTGCTAG TATTCCTATT TTGTTGGATG AAGCAGTTGA AAATGGGACC 1260

TTAATTTTAG GTTCCCAACA ACGCGTGGTG CTAACAGGAT TTGGCGGCGG CTTAACATGG 1320

GGCTCACTGC TCCTAACGCT GTAAAAAACT GTTCACATAT AAATGAAGAA AATGACTATT 1380

TTGGAGGAAT ATACACATGG TATTTGAAAA AATTCAAGCA ATTATCGTAG AAGAATTAGG 1440

AAAAGATGCT GAAGAGGTAC AATTAACTAC AAACATTCAA GAAGAATTAG ACGCGGACAG 1500

CTTAGATTTA TTCCAAATCA TTAACGAAAT CGAAGATGAA TTTGATGTGA AAATCGAAAC 1560

AGAAGATGGC ATCCAAACCG TTCAAGATTT AGTAACATAT GTAGAAAAAC AACAAGCAAA 1620

CTAATACATG AAAATAGGGG GCTGGGACAT TTAATTGCTA AATGACCCAG TCCCTAAAAA 1680

AACAAGTAAA TAAGCCGTAA AAATCCTTTT TTACTTGAAT GAAGTGTACT TATCTTAGAA 1740

CTAAAGGACG TATAAAATCA ATGAATCAAG AGTTATGTGA GTTGCTTGGA ATCAATTATC 1800

CCATTTTTCA AGGCGGTATG GCTTGGGTAG CCGATGCTTC ATTAGCAAGT GCCGTGTCAA 1860

ACGCTGGTGG ATTAGGGATT ATTGCTGGCG GCAATGCCCC AAAAGAAGTC GTAAAAAAAG 1920

AAATTAAAAA AGTTAAAGAA TTAACGGAGC AACCCTTTGG TGTCAATATT ATGTTACTTT 1980

CACCTTTTGC CGATGAAATT GTCGATTTGG TTTGTGAAGA ACAGGTTCCT GTCGTAACGA 2040

CAGGTGCAGG CAATCCACCC AAATACATGG CTCGTTTTAA AGAACATAAC ATTAAAGTAA 2100

TCCCAGTAGT TCCTTCAGTT GCTTTAGCAA AAAGAATGGA AAAAATTGGT GCCGATGCTG 2160

TCATTTTTGA AGGAATGGAA GCTGGTGGAC ATATTGGCAA GTTAACCACT ATGAGTGGCT 2220

TACCGCAAAT CGTTGACGCT GTGTCAATTC CTGTGATTGC AGCAGGTGGG ATTGGTGATG 2280

GTCGTGGTAT GGCTGCGGCC TTTATGTTAG GTGCTGAAGC AGTCCAGTTA GGCACACGTT 2340

TTTTAATTGC CAAAGAATGC AACGTTCATC CAGATTATAA ACAGAAAGTT TTAAAGGCAC 2400

GTGATTTAGA TGCAGTAATT ACCTGTCAAC ATTTTGGCCA TCCAGTGCGT ACTTTAAAAA 2460

ATAAATTAAC CGCTCAATAT AATCAATTAG AAAAGCAAGA ACTCCAAAAA GAAGTGCCTG 2520

ATTTAGAAAT GTTTGAAAAA ATTGGTCAGG GCGCTTTGCG CAAAGCTGTC GTTGACGGGG 2580

ATATGGATTA CGGTTCCGTC ATGGCGGGAC AAATTGCCGG GTTAATAAAA AAAGAAGAAA 2640

CAGCCCAAGA AATCATTGAT TCACTCATGT CTGAATGCAA AGCGATTGTA CATAAGATGA 2700

ATCAGCGTTG GGGCTAAGTG AATAATTTTA ACAAAAAATG AGGTGTCGTA TGAAAACAGC 2760

GATTTTATTT AGTGGACAAG GAGCCCAGTA TCAAGGGATG GGTGAAGAAT TATATCACCA 2820

AGAAGCGATT GTTCGGGAAA CTTTCGATGA AGCAAGTCAT ATCTTAGGTT ATGAGATGGC 2880

AGAACTTTGT TTTACTGAAA ATGAACGTTT AAATGAAACA GAATATACGC AACCTGCTAT 2940

TTTAACAGTC AGTGTCGCAT TTTACCGTCT TTTGCAACAA AAAGGACTAA CGCCTGATGT 3000

CGTAGCGGGT TTAAGCTTAG GGGAATACAG TGCTTTGGTT GCCAGCGGGG CTTTGCGCTT 3060

TTCAGAAGCA GTGGCCTTGG TCCaAAAGCG CGGTCAGTAC ATGACAGAAG CAGCACCACA 3120

AGGAACTGGC AAAATGGTTG CTGTTATGAA TGCTGAGCGT GAAGTAATTG AGAAAGCCTG 3180

CCAAGAAGCC AGTGCTTTCG GAATTGTGGC TCCAGCAAAT TATAATACAC CACAACAAAT 3240

CGTGATTGGT GGTGAGGTTG CTGCTGTTGA TCAAGCGATG ACACTTCTCA AAGAAGCTGG 3300

TGTGAAGCGA ATGATTCCGT TAAATGTGAG TGGCCCTTTC CATACGGCGC TGTTACAACC 3360

AGCATCAAAA AAATTGGCTC AGGATTTAGC AAAATTGAAC TTTCAAACGA TGCAAATTCC 3420

TGTCATTAGT AATACGACTG CCGAAATTAT GCCCCAAGAG GCAATTCAAG CGTTATTGGA 3480

AAAGCAAGTC ATGTCTGCGG TACGTTTTGA AGACAGTATC GAAACGATGA AGGCTATGAA 3540

CGTAGGAACG ATGATTGAAG TTnGgtCCAG GGrAAACATT aACyGGtTTT GtTaAAAAAA 3600

TTGACCAAAC AATTGAAATG CACCGTGTGG AAGATGTTGC CACATTAACA GAAACG 3656

(2) INFORMATION FOR SEQ ID NO: 578:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 412 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 578:

CCACACTTCa sCsTCCTGTA TAAGAGGAAT CAGATTAAAT TCTATAAACA TCTTTTAAAT 60

GTCTTTGCTC CAAAAAATTG ATGTCTAATC TTTAGAAATC ATTGTCTATC AGAATTATTC 120

TTAAAAAGTT TGATTTTTTT AATCAAGAAA GAGCTAATTT CTTCCAGAAC TCATATCATC 180

ATATTTTGCA ATGGCTAAAC AACTCAAAAT CGTTATACCA AGCAAACATC CTATAAACAT 240

GCCTAACAAA AATAAAAACA TGGTTAACGC CACCTCTACA TGaTTCAATA GACTTAAATA 300

AAAAAAATGr AGTAATCTTT CAGTATTTTT TATATTTTTT TACAAAAGTA GTTAAGTGTA 360

tTTGAACATA TCTTTTGaAC AGTaATaATA AAGAGTGGtA gACGGAGCGG TT 412

(2) INFORMATION FOR SEQ ID NO: 579:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 910 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 579:

ATCCATGACT GTTTATCAAA CGTAGCTTTA CCTGTAGCCT GTTCTTTTCG TTGCTCGTTT 60

ACTTCTCTTG ATTCTAATAG TTGGAAGCTT TCTGTGACAA CCTCAGTTAC ATAAATACGC 120

TGGCCTTGTT GATTCTCATA GTTTCTTGTT TGAATGCGAC CAGTTAAACC GATCAGAGTC 180

CCTTTTGTTG CATAATTTGC TAATGACTCT GCAGCTTTCC GCCAAATAAC ACAGTTGATA 240

AAATCTGCTT CCCTTTCATT GTTAGCATTG GTAAAGTTGC GATTAATGGC CAACGTAAAT 300

TGACCTACGG CTGTTCCACT TTGTGTATAC CTTAAATCTG GGTCTTTTGT TAATCGTCCA 360

ACTAATGTAA CGTTATTAAT CAAGATTATT TCCTCCTACT CCTATAGTCA TTAATATTTT 420

CTTGTATGGA TTtAATAAAT CATGCTATCT AATAACTCCT AATCATTATA ACATGTGCTA 480

ATCTACTTTn CTTTTATTCT ACTAGTCTGA TTTCCCCTCG CAATCTCTCT TTAAACATTT 540

TATTTTCTAA CTTTGTATTG GACTTTTTCT TTATGGAATT GGACAAGTAG CCCATGTTCA 600

TATCTTTTTT TCTTTTGATA CAAGAAGTTT CCTCGACCAA TTAGaTAATC AAGAAAGCTC 660

TTATTATCCC ATTTTAAAGA AAAGAGCGTC CACACAAGGA CGCCCACTCC TACTATAAGA 720

ATCAACCAAA AATTAGCAAT AATTGTGGTA AAAAAATTGA TTTTAGCGAC AAATGAAATA 780

ATCGCGAATA TCAATAAGAG AATAAGGATA CCAATACCTA CAAAAATTTG TTCCAATCGA 840

ATGCCGTTTG GCGCAAAAGG CAATGCTACC CCTTTGATAC TGTAAATCTT GTGAGGTTGT 900

TTAAATTCTC 910

(2) INFORMATION FOR SEQ ID NO: 580:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 787 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 580:

CCCnGAAAAA AATATTCAGT TTGATCAGAT TGATTACACA aTGTACAAAm CArGTsCCGG 60

TATCCAATTA ATCAAAmCAT CTGAATATAA TAAACTCGCA AAATTTATAG ATGAACCAAC 120

TTTAACAGGT GAGAAAGTAG TTAATTTATC ACAAAAAGAT GCTCCTTCTG CTGCATCAGT 180

AGATGTCAAA TCTGTTCAAT TGGCAAATGA TCAAGTTTTA AACATTGATA AAACAGAAGA 240

AACCAATGTA TTACCGACTT TTGGCGGTGT GGTTGTTATC CCAGATAGTA CCTCAACAGA 300

TGGAATCAAA GCGGAAACAG AAACAATTTG GCAGCCCAAA GCAGGAACCA ATCGTGAAGA 360

ATTGATAAAA GCAGAAAATA TTATTGGTGA CAAGTATCCC ATGTTAGTAG GAACAGCTGC 420

AATGTCGGCA GCCATTACGC AATATTATAC GCCAGTTCTT TTTGTCGGAC TATTTATCGG 480

TATTGTGTTC TTTGTTTCGG CGGGTAGTTT CTTATACTTC CGTTTATACA GTGATATGAA 540

TACGGACGTT GAAAAATTCA AGATGATTTA CAAACTTGGC TTGACGAAAA AAGAATTGAA 600

GAAAATGATT TACCAACAAG TAGGTATTTT ATTCTTTACC CCAATTGTTG TTTCATTTAT 660

TCATGGAGCT GTTGCTTTGA AAGCAATGTA TGCTGTCTTA GATCAGCCAA TGCAATTAGC 720

TGGCTGGGAA GTTTTAGGCG TTTTCTTATT GATTnCAGGT CGTTTACTAT TTAATTGCCA 780

GAACGTT 787

(2) INFORMATION FOR SEQ ID NO: 581:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 912 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 581:

TTCCCGGGGC CAGCACCAGA TAAATTATGG ATCGATTAAT GTCCTACCAC TCCTTGGATn 60

TTCAACTrGG CAAtCATGAT ATGATTTGGC TGGCTGCTTA TAGTGGTTCA CTGGCTTGTT 120

TAGCCAATGT AGTACGTATT TGTGCTCGCT ATGGAAATTT AGATTTGTTA GAAGAACGTT 180

ATGCTATCGA TTTAACCGCC TTAAAAAAAT TTAGTTTAGA AACGTATAAA GAAAATCCTG 240

CTTTTGCACC CAAGAAAAAT CCCTATCGTG CGTTAACTGA AGCTGAAAAA CAAGTAGCGA 300

TGCGTGTGCA CAAGCCATTG CCATTATTCA AGAAAAACTA GAAGGCCAAA TTATTGGTCG 360

ACGCCCTGaT TTTAACTTGG CACATCGGtA CGCTTAGATA AGATTCAAGG GGaAACTATT 420

ACTTTTGACG AATGTCGTTA TACATTAATT AACAGCTGTT TCCAAACCGT TTCTGAAGAG 480

CAACCTTACC AATTAACACG TGAGGaAAAA CAGATAATTG ATGATTTATT AACGCAATTT 540

CAGTCTTCAC CGCGGTTAAC AAAACACATG CGCTTTCTAA TGGAAAAAGC TTCTCTTTAC 600

CTTGTGTATA ATCAAAATTT ATTGATTCAT GGCTGTCTGC CTCTGAATGC GGACGGTACT 660

TTTCAAGCTT ATACATTTAA AGGCCACTCC TATTCAGGAA AAGCTTTAGT TGACTTCTTT 720

CAGGAAATGT TAGAAGAAGC TTACGCTCAG CCTGCTAGTA CAGATGACTA TGCAACAGAT 780

TGTTtATGGT ATCTTTGGTG TGGCGAAGGT TCCTCATTAT TTGGTAAACG AGCTATGAAA 840

ACCTTTGAGC GTTATTTTCT CGCTGAAAAA GAAACACATT ACGAAGAAAA GAATCCTtAT 900

TACTCTTTGC GT 912

(2) INFORMATION FOR SEQ ID NO: 582:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1641 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 582:

GAAATATCTT GTGCTTTTTG TTCAAGACTT GCGCATCCTG ATAATAGCAA AGTGGCTGCT 60

GTTAGCAGTA GCCACTTTTT TCCAATAAAT AGTTTCATTT CCTAACTCCC TAACTTTTTT 120

TCCGTTTTAA AGAATACTCA ATTTTCTTTT TTTCTAACCT ATCAAACGTT AATCCTTCAA 180

CAGCTCGTTC CCATTTTGTT TGTTGTTCTG TACGTTTCTT TGTACTGTCT GCATATATTT 240

TTTGTAAATC TCCATTAACC ATGAGTCACT CCTCCTTAAA ACGGTAAGTC ATGATCTGAA 300

ATCTCAATAG AACTACCTTC TAAATCTGGA AAGGGTTGTT GCGTATTGCT TTCCTGCTGC 360

CCAACTGGTG TAGGATTAGA AGCGTTAACA CCACGTTGTT GACTAACTTC TTTTGACTCT 420

AACAATTGGA AGCTTTCTAC AACCACTTCT GTCACATAAA TACGTTGTCC TTGTTGATTC 480

TCATAGTTAC GTGTTTGAAG TCTACCCACA AGACCAATCA ATGTTCCTTT TCGTGTATAG 540

TTTGCTAAAG TTTCTGCAGG TTTACGCCAA ATCACACACT GgAATATAGT CTGCTTCACG 600

TTCACCATTT TGATTCGTAA AATTCCGGTT AACkGCTAGT GtAAAAgCsG CAACAGCTAc 660

ACCaCTkGAC GTATAACGTA AATCACAATC TTTTGTAAGT CTTCCTACTA AAGTCACATT 720

ATTAATCATT TAAAAAACCT CTTTCTATTA TTTTTTTAGA AGTGATAGCA ACTGGTACTA 780

TCACTTCTAA ATATTTACTG TTTTACTAAA CCACCAAACA TATTTTTAAA CTTGATTTCA 840

CCAAACTCTG AAAGGAACGG ATCCGTTTCG TTTAAGCCGT AAGTGTATTG GTAATTCTCT 900

ACAATACTTT GATAATCCAT TTCCCAGTCC TCTTCATACG GAGATATATA AAACAAGTAC 960

AACATTGTTA TATCCCCCAT ATACTTTTGT TCAATCGCAT AGTATACAAG TATATTGTTT 1020

TCTTTCTCGA TAATTTCGAT TGCTTTTTTT CGTTCTTCGT TGATCCAATA GTTGGCGCCT 1080

AGATATGATT TTCCCGAAAA ACTAATTTTT CCTTCCTCAA TAAATTGTGC AATTGTTGGT 1140

TCATGAATAC CTAAAATCTT CATTCTTTTT CTTGCTTCCA TTTTCATTTT TtCTAAATTC 1200

AATTGAATCC CTCGTTTCTG TTTATCTTTA TAATCATTAT ACCATACATT GTATACAATG 1260

CATAGTATAA TAATGAGGAA TATTGACAAA TTTTGGkTTT TTATATACTA TATTTTGTAT 1320

AGTATATATA GGAGGGTAAT TCAATGAAAG ATCTTTCTCG TGAAGAAGTT TTAACTTATT 1380

TAGAAAACAA TGTGGTAGAT AAACAAGGTG CTGCAAAAAT TACAGGCCAA TCATTAAATG 1440

CCTTCACACA ATCCGTAAAA TTGAATGCAA TTAAGCCTTA CTTTGAAATT AAACATGTAA 1500

ACGGTGAACG ACCTACTGTA CGCTTGTATC ACGTAGACGA TTTAAAGGAA TACGCAAAAA 1560

ATAAACGAAG ATAAAAGAGA ACTCCCTTCT TTTATTCTCG TTTATTTTTT TATATAAAAT 1620

CCTTAGTTTA TTTTAnGACT A 1641

(2) INFORMATION FOR SEQ ID NO: 583:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 932 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 583:

TGTGCCCGAC CATTACGAGA GAGCGCACAC GATGCTGCCA TGGCCGATTG AATGAAAAGA 60

CGnGCAGnTT CGTTTTGATA CCAGGAACTG GAGAGAACGG nACAGCAAGT GAAGAAGTGG 120

CTAGAGGAAC AAATTGGAAA GAAGTGAATA TTTTTTTAAT GGCGTAAAAT GTTTTGAAAA 180

TTTCAGAAAA AACAGTGAAA GTCTTkGTTT ttATTAAAAa aTGGCAAaGa TTCAaCGAAG 240

AmCCAGTTCT AAaGrCCTaG TAATaGCGCA ATTTCATGtT ATAATATAGG CAACAAAAAT 300

TATACAGAAT AGAAGAGGTG AACGTTATGC GTTGTCCaAG ATGTCATCAT AATAATTCCC 360

GTGTTATTGA TAGCCGTCAA GCAGATGATG GACGTGCAAT TCGCCGTCGT CGTGAATGTG 420

AAAATTGCAG TTATCGGTTC ACTACATTTG AAAGAATTGA AGCGGCGCCA TTACTTGTAA 480

TCAAGAAAAA TGGGGATCGT GAAGAATTTA ATCGTGATAA AATTTTGCGT GGGCTCATTC 540

GCTCTGCGGA AAAAAGACCC GTTGCTATGG AACAAATGGT TCAAATTGTC GATAATGTTG 600

AAAATCGAGT GCGTAGTTTA GGGGAAAATG AAGTCTCTAC AACGCTTATC GGTGAATATG 660

TGATGGAAGA TTTAGTCAAT CTAGATGAAA TTGCCTACAT CCGCTTTGCG AGTGTTTATC 720

GACAATTCAA AGATATGAGC GTGTTTTTGA AAGAGTTACA GGACATTGTC GACAAAGCAA 780

AATCTTCTTC TCCTGACTCA GAAAATTAAA AAGGAGGTCT TTCTTTGAAA GGAAATCGGC 840

AAAGAATCCA ACCAAAAAAT ATTTTTCAAG CGGTTATTGA TTCACCGCTT TCTGATCAAG 900

AGAAAGAGGT TTTAACCTTT TATATCAACC AA 932

(2) INFORMATION FOR SEQ ID NO: 584:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 771 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 584:

TAAACCGACA GCACGGCCTG AACCGTTCAT CAAAGCAAAA GCGACAAAGA TAAACCAGAT 60

ATTTACCCAA TAGAACATGC CAGATAAACA GAAGGCAATT GTTAAACCAA TTGTAGCAGC 120

ACTTGATAAC CAAGTGAATA AACCAGCAAT CAATGCTAAA gCAATACATA CTTCAACGAT 180

TGTCATGAAT TTTTGCATGA ACAAAGCAAC TTCTTGGTTT GGCATCATAA ATTTCATTAC 240

GCTTTCGAAC CATTTAGGCA TGTGATCAAA CACTTGCATT GGTGTTTCGC CGTAAGCGTA 300

gcTTAAAcCA AAGTGCGCTG CTTGTGTTGC CGCATCGGCA CCACCGCTTG CCGCCGCATC 360

CGCCGCACCA CTTGCAGCaG TTGTTGcTGT TTCAGTTGCT TGAGAAGCAC CTGTTGTTAC 420

TTGTTCTTGT AACCAAGGGA ATGGGAAAAC GACTTTATCT GTGAACCAAG AACCGTCACC 480

AAACCAAGTG CTTGGTTTCA AGTAATCGCC GTTGCCGACA ATTTTCTTCA TTGATTCGAC 540

TAACCAAACC ATACCATAGA AGACACGTAG CGGAACACTC CACAAGACAT TGCCGTAACG 600

AGACGTATGT CCGCGAGCAA CAGAACGATC ATCTTTAATA TGGAAGATTT CGTGCATAAT 660

GTATTGGAAC ATGTAGTAAC CAGAACGAAT ATCAAAGAAG TATTTCAAGT TAACGATATG 720

TTTCATAATG ATGGCTAAGA AACCACTCAA ATGAATTTTA TCGAATAAGT T 771

(2) INFORMATION FOR SEQ ID NO: 585:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 623 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 585:

TTTCTCTTAA ATAATGAAAA AAGTCTGATT TTTTTGCAAA ATAGTGATTA AAAGTCAAAA 60

ATATACGCTC TTAAATGTCA ATAGGACCTT TAAGAGCGAT TGTTGCTTTA GTTTTGTCCT 120

TGAATTTTTT CTAATAATTC TGGATCAAAT GTTTGATTTT TTAACATCGC AATTTCAAAG 180

CCGTAAGGGg CCTTTTTATT TTTTTTGTCT TCTCCGACAT AAGGTGTTTC TAAAATTTTT 240

GGTAAAGCAG TTAATTTTTC ATGGTGAACC ACACCGTTTA AAGCTTCAAA ACCAATAGTG 300

CCAAAGCCAA TATTGGCATG ACGATCTTTA TGTGAGCCCA TAGGATTTTT AGAATCATTC 360

ACATGAATAA CTTTTAAACG ATCCAATCCG ATAATTTTAT CAAATTCAAC TAATACGCCA 420

TCAAAATCAT CTTTTACATT ATAACCTGCA TCaTTAATGT GGCAgTATCC AGAGTGACAG 480

AAAGCyTATC ATTTAGCGTA ACTCCGTCAA TAATCGCAGC TAGTTCGTCA AACGTACGGC 540

CGATTTCGGT TCCTTTACCA GCCATGGTTT CTAAAGCGAT TTGAGGGATT TGCTCTTTCC 600

nCAAGACTTC GTTTAGACCT TTA 623

(2) INFORMATION FOR SEQ ID NO: 586:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1253 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 586:

CCGCCTTTCT CTAAGnGTAA AACAAGCCAA TGTTTCATGT GAACATTGGC TTGTTTTACT 60

TACTATAATC ATCGCACTTT ATTtGyCGTt CTCTGCGGTA GTCGTATTCy TGGATACGGG 120

GACTTAACCA ACCTTTTTTG GCTGTTTTTC GTAATTGATG ATCAAATCTC CCTGGTGGTG 180

AACTATCTAG TAATTCATGG GTACGATTAT TCCGATTTGC ATAATTAAAC GATTCACCAT 240

TCAAGACTTT TTCCTGGTAT TCATTGTTTT TGGCCGCTTG aTAATAAGCT GGaTTACTAA 300

TTAATCTGCC TCGCCAGTCT TCAACTAACA CATCCCACTG GGgTCACAAA GGtCCCTTCT 360

TCcATTCaAA ATAAACTCAC TGTGaAAATT AGGTGTCAAT CGTTTCTTAt TAATTAAAGA 420

ATAAGATGCC CCCTCAGGGT ATTTGTTGTG CAAACGAGCA GAGCTTTTTT TCAGCTTCCG 480

CCCACCCATC CCCTTGCGGC TTGGCCGCAA AAACATAGCG TTTTAACGCC TGTTCATCGG 540

TGTCCCCTTC TTGTTTAAAA AAATCGCGGA TATACGTGAT ATTTTGACGA TCAATATACA 600

TTCTTAATTG ATGTATCTTT TTCGCTTTCT CTTGGTTAAA GCTCCCAAGA AGACCTAACG 660

GATAGGCTTT TTGAACTGTT TTTGCCAACA AATCCCAAAA CGTCGCATGC GGTGCAACCG 720

TcTTCGAGaA ACGATAAGGT CCTACTAATT GCAACATGCC ATGCTGATCC AACACTGCAC 780

CTAATTGCCA TAACAGTAAC GAGAGCTTTT CTTGCGGACG ATAAGCCGAA AAGCGAAATA 840

ATCGCTCATA CGCTGAAGAG CCCACTTCTT CCACTTCAGA AAACAGTTCC GCGACTGCTG 900

TTTTCGCTTG CTCAAGTGGC AAGAAACGCG TGTCTTTTAT GGCGCTTTCC ACTAATCCAT 960

AAACAGCTTG GCGGTCCCAG CCAGCGTCAA TCGCCGCCTC TACACCAGCA CTCATTGCTT 1020

GTTGTAATTC TTGCCAACGT CCCATCTTCT CGCCGCCTTT TCTTATATAC TTCTTCTATA 1080

ATGATAATCC TTTCCCCACA AAATGCCAAC TGCTTTAACG CAAAAAAACA ACCCGCCCTT 1140

TGGGGAAGGG CGGGCGGAGG AGTTAAAAAA ATGAAAAAGT GTTTTGTTAA GGGTTTTGTG 1200

TATCGGGTAT GGTTAAATAA TAGCCGTTAA AATGTGAAGA AACTATGGTA AAA 1253

(2) INFORMATION FOR SEQ ID NO: 587:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 553 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 587:

GTnTCAATGG TGATTGAAGG TCAACATTTA GTAGAAAATG GGTTGTTATA TGTGGCTTAT 60

CCGAnAAAAG GCAATAaAAT GTATCCAACC TTTGTCCACC GAGATGAAAT TTTTCCAACA 120

TTACAAGTAA ACGAGGCAGA TGGGTATATT AAAGGCTCAA CCTTAAAATT TAATCGCATG 180

GTTAGTTTAG ATGAGACTTT TACGGTTGTC GGCATGAAGA ATGTTCTTAA AAAGCTAGCG 240

AAGAATCAGA CCAGTTCTGC TGTTGGGGAT TATGTTCAAT ATTTACCGGA AATTGAAGCG 300

TTGGTACAAA CAGAACCAGC TGCTGCAGCG ACATtCGCTG CATTAACGCC TGGTTATCAA 360

AAAgAGTGGG CGCGGTATGT ATTTAGTGCA AAGCGAGCGG AGACACGGAA AAAGCGGCAA 420

ACTGAAATGT TAGTCATTTT AAAAGAAGGC TATAAAACCm AAGCGCTTaT CAGCmAAGGA 480

rGAAtAATGG CGCAGTCAAT TGAAACAGTT ACAGCATACA TCGAAACAGT GGAACCAACC 540

nGnCGAnTGn CTA 553

(2) INFORMATION FOR SEQ ID NO: 588:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 576 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 588:

AAAAAATCGT CGCATATTTT TCTTCGATAA TTTCAGCTAA GCGTTTTCGA AAGTGTTCGT 60

TGAAATTATT GTCCATGATA GTTAAGAAGA AAGATTTGTT CTCTTGAATA TTAATATAGA 120

GTTGCGTCAG AAGTTCTTCA ATGTGCTTCA CGCGGACACG GTTGCCATTA ACCAGTTGTT 180

GCGAATCAAG AATAGCGGTA AAAGCACTAA GCGTAAAGGT GAAGATTGCA TCATATAAAT 240

GCTGTTTGTC TTTAAAATGA GCATAAAAAG TCGCACGATT AATCATAGCT TCATCAGCAA 300

TATCCTGAAT CGTCACCGCT TCGTAGCCTT TCGTTTCTAC AAGTTTAAAA AAAgCTTCAA 360

TAATCATTTT tCTGGTTCGT TGTACTCTAA GGTCTGTTTT CTTTGCCATC ATAAATTCCT 420

TTCGAGATAG TCAACATTCG CGGgAAACTT tGTkGTTawC tCACaAAATA GGGTAAATkG 480

CyTGTtGCAA ATCATTTAGA ACCTTTCTAA AATAGGTATA TCnATnATAA CCAACAGTTT 540

GTnTAATAAT CCACCAACTG TGGAAAAATG AnGGGG 576

(2) INFORMATION FOR SEQ ID NO: 589:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 658 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 589:

GAGCCTGTGA CTGTTGAGGT AmACGTTCAA CTTCTCTTGA ATGATAATCA TTTGTCTCTT 60

TGATTGCTTC CGTTAACTCA CCTAGAGGAA ATGCAGTTGG CCATTTCAAC TGAAATGGGC 120

GATCTATATA ATCATTATAA GGTGTAAATT CTTTTTTAGT TTTTTTCATT GGTTTATCCT 180

CCAAGACCAG TGTTGCCACC AGCGTGACCA CCAACTAATC CGGCTCGATT CACAGCTGTT 240

GCTCCATCTA AAAGAGAAGA AGCGTGAATC AAGGCTTTGA AGCCAAATTT TCGACGTATT 300

TTATCGATCA AAAAATCTAA TTGGGTATTA ATAATTTGTT CTTCAGGAGG GGAGAAGAGA 360

TCCAATTGAA GTGTGGTATC CCAAACCAGC TTTCCmAAGG AGACTCCTAA aTTTCGAATA 420

TCCATACCAG GAACATAGTT TTCTCGAAAA AGCTTCAATA CATGTTCGGT TAACACTTTT 480

GTGTTGTTGC TACGAGCGAT TTTAAGCTGT TTTCGCCAAC CAGACCTGCC CAGTCCATCA 540

ACTTGTCCTT TAGAATAACC AACAAAAATT GAAACACACT CAGTTTGACA GTTTGAATTT 600

CTCAACCTGG TTGCTACTTG GTCTGAAAGT TCCTTCAGCA CTAATTCAAT TTGATCTC 658

(2) INFORMATION FOR SEQ ID NO: 590:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 940 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 590:

AAAGAATCTA TCAAATAGTA GAGAAATATG GATATATAGA ATAAAGATGT AGCATATAGT 60

TATGCAATTC TAAAGAAAAA GTTTTGTCAG GAGTAAATCA AAAATTTGAA CTTGGGAAGT 120

TTTATGCGAT AGTAGGAAAG TCAGGAACAG GAAAATCTAC ACTTCTTTCC TTACTTGCTG 180

GaCTTGATAA rCCTCAAACA GGAAAAATAT TGTTTAAGAA TGAAGATATA CAAAATAAAG 240

GATATAGTAA TCATAGAAAA AATAATATAT CTTTAGTATT TCAAAACTAT AATTTAATAG 300

ATTATTTATC ACCAATTGAA AATATTAGAT TAGTAAATAA ATCAGCAGAT GAAAGTATCT 360

TGTTTGAACT AGGTTTAGAT AAAAAACAAA TAAAAAGAAA TGTTATGAAA TTATCTGGTG 420

GACAGCAACA AAGAGTAGCT ATTGCTAGAG CATTGGTATC AGATGCTCCA ATAATACTAG 480

CTGATGAGCC TACTGGTAAT CTAGACAGTG TTACTGCTGG AGAAATAATT AATATATTAA 540

AGACATTAGC TAAAGATAGA AATAAATGTG TAATAGTTGT AACACATAGT AAGGAAGTAG 600

CAGATTCTGC GGATATCATT TTAGAACTAA GTGGTAAAAA GTTGAAAAAA GTAAATAAAA 660

TGAATTTGGA GGTTGAATAA TGATAAAAAA TGCATTTGCT TATGTAACTA GAAAAAGTTT 720

AAAATCATTA ATTATCATAT TAGTTATTTT ATCTATGTCA ACTTTAAGTC TCATTAGTTT 780

ATCTATTAAA GATGCTACGG ACAGAGCTTC AAAAGAAACA TTTGCTAATA TAACaAATAG 840

TTTTTCTATG GaGwkAAaTA GACAgTAAAT CCaGGAaCcC TAGAGGtGGG GGrAATGTaA 900

AGGTGaGaTt TAaAAgAaTC TCAACnGATA GTAAGACTCC 940

(2) INFORMATION FOR SEQ ID NO: 591:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1268 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 591:

TGAAAACGAC TAGGTGATTC AAACTTTCTG AAATTCTAAA TTGAATCATA CCAGAAGAAA 60

CTGGCCGAAC CAATAAAGTA AGGACAATAT ATAAGCCCAT AATGATACCA TTGGCTAAGA 120

CTGTTCTTAG CTTTCGATTG GCTTGAAACG AATTTTGCAT ACTTTTTCCT TCTTTCTCCG 180

AGTTTTTTTA CAAGGGATGG TTGATGAACC TTGGTTACAA TTGCAACAGT TCACATTCTA 240

ATATGGAAAG TCAAAATAGT CAATGAAAGC CCTGTTTTAC TTTGCACTTA TTTGCAGGTT 300

TCACCTCTCA TGCATTGTTT CGTGCCTTTT TCTAGTCATT TTTTACTTTT TTAAGCAATC 360

TTTATTCGCA GAAGTTCCTC CCTCTGGTAT AATGGACAAT AGCGATTCAA AGGAGGAGCT 420

CAATGaACAT TCACTACTAT CTTATTATCA ACGAGCAGGC AAGTAGCGGC AATGGCCGAA 480

AAGTCGCTAG AAAAGTGATT CAACAACTCA AGCAACAAGA ACTTAAATAT ACGGCTCTTT 540

ACACTGATTA CGCTGGACAT GAAAAAGAAT TGACGAAGGA ACTTGCAGAA ACAACTTTAC 600

TTCCTTGGTC TGaAGACTTA GACGTTTcAA CTTTTCCAAT CCTAGTCGTG CTAGGCGGCG 660

ATGGCACACT ACATAATGTC ATTAATTCAT TACTTCCATA TGATTCAACT ATTCCTTTAA 720

GCTATATTCC ATGCGGCTCT GGAAATGATT TTGCACGAGG GGTGGGATTA TCAAGAAATA 780

TTGATAAAGC ATTGCATCAA ATCCTGCGCA CCAGACGACC AAAAGAAATT CAAACAATTC 840

ATTATGTAGA AGCCAACCAA GAAGAAATCG GCTTAGCCAC CAATAATGTT GGCTTAGGGT 900

TGGATGCgGC AATCGTGGAA AAAACCAACG AATCGTCATC AAAAAAAGCC TTAAATAAAT 960

TTAAGCTTGG CTCGCTTTCC TATATTTCTT CAATCATcAT GTCyTTTTTA GACAAAAAGG 1020

CTTcCAATTT TAGTTGAAaT GAATGGkAAA CAATACmCaT TwAATCGTGC TTTCCTATGk 1080

ACGGTCACAA ATCaTCCTTA wTTTGGcGGG GGCGTGnnnA TTATCCGACG GCCAATCCTC 1140

GCAAAGCAGT CGTGGACTTA GTTGTCGTGG AACGAATCAA TATTTTTAAA ATTTTGTGGC 1200

TTATTTTCCT AATTGTCGCC AAAAACAAGG GAAATCTAAG CATTTTCATC ATTTTCAATC 1260

AAGCAAAA 1268

(2) INFORMATION FOR SEQ ID NO: 592:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1011 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 592:

GGATGCTTCA GCTCTAGGAC CGATGGAATT ATTCATGTTT TTCGGTAAAC CTACGACAAA 60

TTTCGTGACT TCGTATTCTT TCACCAATTC ACCTAATCGC TCAAAACCGA AGTTCTCTTC 120

CTCTTCATTG ATGCGAATAA TTTCGATGCC TTGTGCTGTC CAACCTAATA AATCACTAAC 180

AGCCACACCG ACTGTCCGTG AACCTACATC TAGCCCCATG ATCCTCATTT GATATCGATT 240

CCGTGGTTCG CTAAGTAATA TTTGGTTAAT TCTTCCATGA TTTCATCACG TTCATGACGA 300

CGAATTAAGT TTCGAGCATC TTGATAACGA GGAATGTAAG CCGGGTCCCC AGAAAGTAAA 360

TACCCTACTA TTTGGTTAAT AGGATTGTAT CCCTTTTCTT CTAATGCCCG ATATACGGTT 420

TCTAACGTTT CACTGATTGC TTTTTTTCGA TTATCATCAA AATCAAAGCG TACTGTTTCG 480

TCTGTAAAAC CCATTTTACT ACACCTCACT TCTACTTTTA TACAATTTCA TTTtACTAGA 540

AAAGAACCGA AACTACAAAC AAAATTcTTC TCCTTCTCTC tTTTTTGkTA TTCyTAGkCy 600

TTatTAATAA ArGCTtCGkA AAAAAATGAC AGCCATCGAT AATTTTAAAA TTATAAAAAC 660

AAATACACAA CTTTATAAAA CAAGAATAAA ACTCTGCTAA ATCGCCAGAA TATCACCTTT 720

CCCTCCAATT CATTTTGCGA AAAAATCGAC GTTTTTCTTT TAAAAACTTT TAAATATCTA 780

TGATTAAATA ATATCAGCCT TTTAATGTTA CAATTCGCTA ACAATCTGTT TACGATGTAA 840

ATTTTCGATT GCAGCTGTTA TTTATTTCCT CTTTTTGGTT CGTTTGTAGT ACACTAGAAG 900

AAAATACATA AAGTGTAGGC CTGCCGCAnG GACTCCGnCT CCATTTGGTT TTTACTTTTT 960

AGTAAACCnG TTTGTTAGCC GGGTAAAAAT TGnGGATCGT AAAGCGATGG C 1011

(2) INFORMATION FOR SEQ ID NO: 593:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 839 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 593:

TCTnTTTTAT TCTAAGATTG TCTACTAAGC ACTTTTGTCC TAGTAGCGTT TTGGTTTCTA 60

ACTTTATTAT TTACAGGGGA AGAAAAAATA AAGCACACCT TTACTGTGAT TCCAAAAATC 120

ACGTGTTTAT CATAATATAT TTTTGTTTCA ATCCCAAGTA TTCTGCTCAC AAAGTGTCGC 180

TTATTTAATA AACATTGCAT CGCCAAAACT AAAGAAACGA TATTTTTCCT CAATAGCATG 240

TTGGTAAGCA GCCAATGTTA AGTCTTTTCC TGCAAAAGCA CTCACAAGCA TCACTAACGT 300

AGATTTTGGT AAATGGAAGT TCGTTGAAAA AGCCTCTACG ACTTTAAATT GATACCCTGG 360

TGTGATAAAA ATATCGGTCC AGCCACTGTC AGCCTGaATT TCACCATTAA ATTTTGTGCC 420

AATTGTTTCC AACGTACGAA TAGAAGTTGT TCCAACCGCA ACAATCCGaC CGcCTGCTTG 480

GCGTACTTCA TTTAATTGCT TAGCCGCTTC TTCAGTTAAA CGATAGAATT CGCTATGCAT 540

ATGATGTTCT TCTATATTAT CaACGCTCAC GGGACGGAAT GTTCCTAAAC CGACATGTAA 600

AGTTAAATAA ACGAGATGAA CACcTTTTGC TTtGATCTCT tCTAATAGTt CTTTtGTAAA 660

ATGCaAACCC GCTGTTGGTG CTGAAGCTGG AACCCATTTT CTTCAGCCAT AAACCGTTTG 720

ATAACGATCT GGGGCCATCT AGACGGTCTT TGGATATATG GGCGGGCAAT GGGCATTTTC 780

TCCCTAAACG GATTCCTAAA AAAnTTCTAA GGAAAATTAC CAnnCAnAAT TTGAAATTC 839

(2) INFORMATION FOR SEQ ID NO: 594:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1180 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 594:

CnATTAGATG GTTCCTACTA CATTCATTAT GAGACACTAG CAGCTGGTGT CTCTTTTTTA 60

TTTGTAAGTA TTAGTTACAC TTGACCGAAC GTTCGTTCTC GTTTATAATA TAAATTGTAA 120

TCCTTATGAT TATATCTACT AGAAAGAAGT CAAGAAAAAT GCTTGTTTTT CTGTCTTCTT 180

TTTGATTTAA TTATTTTAGT AGGTATAATT AAtAcGGAGT GTTTATTAAA TGGTAAACAA 240

GAAGAGACAA GTAAGATTTG AATTTTTCCa GGTCaATGGA AAAGCTCAaG AAGGAGAAAA 300

AATTGTAAAA GGACTTTTCG ATCTATATCC aTTAGCTGAT AGGATAAATA GTATAAGTAA 360

CTATACTGAT AGAGATGTTA TTCTTTTTGG GGAAAAAGTT AGAATGGACA GATTTTTTGA 420

GGTTTCTAGT AGTCCAGAAT TATATGCAAT GCATTTTACT AGATTAAGAA ATGATAAACC 480

TGCGTATGTA GAATTAAATA ATGAAGTTTT GAAAGAAATT CCGTTAAATC CTGGAGAATA 540

TATTGCTGAG GATATTAGTT GCTTGTATGA TAGAGAGCTA TCTGTTTTAA TGGtACAAAG 600

AAATATTCAT AGTTTGTCAC CTTCAGGTAT TGrAGATTAT TTTACTGAAA TGAGTGATGA 660

TTTAGTAGAG ATTGAATTAT TGCCAGTkGT AAATAAaGAG ATTATTAGTA AnGCCTTAGC 720

TAATGAAAAA TTTCGTAAAC TAGAGTTAAG AGCAGGTTCT ATGAATACAA CTAGTGATAG 780

AAGTGGATTG AGAAAAGTAT TAGGTCCTTT TATGGAACTA TTTGAGAAAT TTGAAGGTAC 840

AAATTTTGTT ATAGAAATTA GTTCAGGGAG ATCAAAAAAA GATTTATCAG AAGATCAAAT 900

GAAAGAAGTA ATAACAGCAA TTGAACAAGA TAAGAGTTTA TTTAGTTCAG CAATTGTTTC 960

AGCAAAAAAA TCAAAAGAGG TACCTGTTGA AAAGTATGAT TTAATAAATG GAAAGCTCTA 1020

TGTGTATCGT TCTTTTGATT TGCCAGATGG TGCATTTTTA AAATCAGATA GTGTTATAGA 1080

CAATATTAAA AATTATTATT TTCATCCTAA TGAAGGTGGG TATAGAAAAC AAATTATCGA 1140

TGCTGTAAAA TAGTTTAAAG AAGAGGGTGA GATATCCTGA 1180

(2) INFORMATION FOR SEQ ID NO: 595:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1434 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 595:

GTAGACCTAT TCATCTAACT CAATATCATT CACATAGTnT TTTAATCTnA GGTTATTGTC 60

TTGAAGATTA GGATGAAGCC TATCAATCCC CGAGTCAATT AGAGCTATGG TTATATCTTT 120

CCCTAGTTGT TCTTTCACTG GATCAATATG AGATAAAATT TTTTTATATG GCCAATTAAA 180

AGACTTAAAA TCTTTGTTTA AAATATTAAT GTTTAAAATA GAACTATTAA CTTCTTCAGG 240

TTTAAAATCT GGTAATTTTT TCCCTTCCGA AACATTAAAA TATTTACCTA TAAACTTCCT 300

ATCACTATCA TCCAGGTTTT TAAAACTAAC TAATCCTATT TCCTGTATAT AATCTACTTT 360

TTCAGAAGAT AATTCAGACT CAATTTCTTC TATAGCAGTT GTTTGCGAAT TATCAATAAA 420

AAATGAAATA TTATTTGAGA GATCTGATGC ATAACCTGTT GTTCCTAATG TTAAAAAAAT 480

ATATGAGATT AAAATATAAG TCAATCCTCT TTTTTTCATT TTGTACCTCC AAATTGTTGT 540

TTTTTGTAAT ACAGTGAACG ATATAACTCG CTATTTTCAA TAAGATCTTC ATGCTTTCCA 600

ATACATACTA TTTCACCATT ATCCATTAAT ATAATTTTAT CAAAGTTTCT CACCGTACTA 660

ATTCTGTGAG CAACAGTAAT AACTGTTCTT TTCTCATCTA ATAAATTAGA AAAAACTTCA 720

AATTCTGAAA TATTATCCAT TGCaGATGTA GGTTCATCTA AAAGTAATGT ATTCACATTA 780

GAATAGAACG CTCTAGCTAA AGCAATCTTT TGCCTTTGAC CACCAGAGAA ATTACTACCA 840

TTTTCTGATA CTATAGTTTT CTCATACTGA GGAATTCCTA AAAGTACCTC ATCCATTTTT 900

GATTTACTCA TACTTTCTTT TAGTCTTTTT TTCTCATTTA TAGAAGAATT CGGTTTAAAC 960

TCTAAAGAAA TATTTTTTTC GATAGTTTCA TTAAAAATAT GCGCATTTTG ATTAACATAA 1020

AAAATGTTTC TTCTATTATT GGAGTTGTTT GATAAAGGAT AGCCTTCATA CAATAtTTCt 1080

CCATTAGAAg GTTGTAACaA TCCaGCCAAT AGTTTTAGTA AAGTAGATTT CCCTGACCCA 1140

CTTCTACCAA CAATAGCAAC TTTATCCCCT TTTCTTATGT CAAAAGAaGG GAGCGTCAAT 1200

AATTTTGTGT AAATAAATTG TCCTCCTGcA AAATAATTAG TTACTCAGTA AACATTGaAA 1260

CTAATGTATC GGTTACCTGT TGaAAACCTT TATGGcTtCT GTTTAGAAAT TTTTGaTtGT 1320

ATGTATCAAA AAkGCTGACT AGrAAGCGkT CTAGTGawTC TTCATTTTGA AACTGCTCTT 1380

TTCTACGGCT GTATCTTTTA ATTTGCTTAT TGAAAGACTC GATTAGATTG GTTG 1434

(2) INFORMATION FOR SEQ ID NO: 596:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1006 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 596:

GTAGTCCCTA CAGTTCTTCT ATTTATAAGG yTTCTCGTTT AGACAAGCCA CGTTCTAAGA 60

TTTTTTGGCT TCTAAGATAC TCTGAATTTT CTGTACTTTC AGAAAAAGAC AACTCTGTTT 120

CGATCTCCTC TGACTCAATA CAAATGAGTT CTTTGCCATA CCATTGTGCT TTATACAACA 180

ATTTAGCTAA AAATAAAGAC CAAGCCAGTT CAGAACGATC ATGTTTTGGC GGCCGCTCAT 240

TACTGTGaTG GGcTTTTTCA ATACAAATAA CATCGTAACT TGaAACCAAT TCCGTTGTAA 300

TTTTGTTTAA AATATCCGTT CGTTGGTTGG cTATTTTTTC GCGAATTTTT GCGACTTTTA 360

TTTTTTGTTT TTGATAATTT TTTGCTTGCG AAAGATCGAC ACCTTTTTCC AAAGCAATCA 420

TTTTACGCTT GTTTAATTTT CGTTCTTCTT GACGTAGTTT TTGCTTTAAA TGCTTTGAAC 480

AATAGCTTTT TTCGAAGCTT TCACCATTCG AAAGCAGCGC AAATTTTTCC CTACTGCTAC 540

AAATTCCCAC CATTTTTTGT TGCTTTGTTT TCGTTTCAAT CGGCGCTTCA CACAAAATGG 600

AAACATAATA ATTGTGACTC GCACTCATAG AAATAGTTAC AGATTTTAwC TCCCCCAGCG 660

GTTGGcGGTG TACCTTCATT TtAATTAAAG ATTtCAATTT AGGTACTTTC AAATAACGAT 720

TATCCAAAAT TTTGACCGTG CCATGTTGAt TGTTCGTAGT GTAGGAATGC ACCGTTTCtT 780

TCTTTTTCTT TTTTGGAAAG CCCATCCCTT TTtCCCGATA AAAATTTTTA AAGGCTCGAT 840

CCAAATAAAC TTGtGCATTt GCCAAGGCTA AACTATCAAC TTCTCTGAGG ACAGGGTATT 900

GTnTCTTATA TTTAGCTGGC GTTGGCATTT TAAACGTTTG TTGAGGATTT TTTCGCATCT 960

CTTTGTAGAk ATCAATCCGG GCTTGTAACA TTAAATTATA CACCAA 1006

(2) INFORMATION FOR SEQ ID NO: 597:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 832 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 597:

TGGTAGGTGA AGAGGTGATA AGGTGGATAA AAGTATGAAA ATAAAATGGA AGCAGAATTT 60

GTTAGTCGCA TGGATTGGTT GCTtTTTTAC AGGGGCGAGT ATTAGCTTAG TTATGCCGTT 120

TATTCCAGTA TATGTTGAAC AATTAGGGAC GCCTAAGAGT CAAGTAGAGT TGTTTTCTGG 180

ATTGGCAATT TCTGTCACGG CGTTTGCTGC TGCGATTGTT GCACCGATTT GGGGAAATTT 240

AGCTGATCGA AAAGGTCGGA AAATTATGAT GATTCGGGCT GCAGCAGGAA TGACGATCAC 300

AATGGGGGCG TTAcATTTGT CCCTAATGTA TACTGGCTGT TGATTATGCr TTTTATGAAC 360

GGGATTTTAT CCGGCTATAT TCCTAACGCA ACAGCGATGA TTGCGTCGCA GGCGCCAAAA 420

GAGAAAAATG GCTGGGCTCT AGGGACATTA TCAACAGGTG CAATTGcTGG GaCGTTAATT 480

GCGCCATCGA TTGGCGGTGC GTTGGCACAG TGGTTTGGAA TGGAAAATGT CTTTATTATT 540

ACCGGCGTTA TTTTATcATT ACAACGTTGT TGACTATCTT TTTAGTTAAA GAAGATTTcA 600

ACCAGTTGAG AAAAAGGATT TATTAACGAC GAAAGAAATT TTTAGTAAGA TGGATCATGT 660

TTCAGTATTG ATgGTTTGTT TGTGACTACG TTAATTTaCA ATTAGGAATT ACAAGTATCA 720

GCCCAATCTT AACGCTATAT ATTCGtCTTT AAGgGCGATA CAGAGAAgTA TTATtTGTTA 780

GTGGTTtGAT gTTcGATGCC GGCGTTCAGC AATAATTCTC TCCnACCTTn Gn 832

(2) INFORMATION FOR SEQ ID NO: 598:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1642 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 598:

CTTTTGGGAC aATnTGrTaC CTAtrGCgCC tTaCaATTrG CaCnGTACGT TTCCmTGtCG 60

CaAATAATGG caTCCgCGTG CaACCCCcAT GTGGTAGAmG GAATCTATGG CaATGATGaA 120

AATGGAGCTT TaGGAAAAAT TTTAAAAGAA ATTGAACCAA AAGTTTTGAA TAAGGTGAAT 180

ATTTCTGAAG ACCAGATAGG AATTCTTCAA CAAGGTTTCT ATAATGTGGT AAATGGAACT 240

AGTCCTTTTA CAACTGCACG TGGTTTAAAA TCAGACAAGT TTTCGATTGC TGCTAAAACA 300

GGGACCGCCG AAACGCAAGC AACAGATGCT AATGGAGTAA ACCATACAAC GGTTAATAGT 360

AACTTGGTGG CTTATGCGCC GTACGAAAAT CCTGAAATTG CAATTAGCGT TGTTTTACCA 420

CACTTAAATG ATGAAGCTTC CAAACCAAAT CAAACGATTG CCAAAGAAGT TTTAGAAGCG 480

TATATGGAAA TGTACAAAAA ATAAAATTTG TAAAGTAGAA CTACTTTACA AATTTGCCCA 540

AAAGGGTAGT CAAATGACAA AAAAAATGGT ATGATATTCA AGTCTAAGAA ATCTCTTGGA 600

AAGATCAATT ATTAGTAGGA GGGGAAAACA TGCGCGTAAA CATCACTTTA GAGTGTACTT 660

CTTGTAAAGA ACGTAACTAC TTAACAAATA AAAATAAACG TAACAATCCT GATCGTTTAG 720

AAAAACAAAA ATATTGCCCA CGTGAAAGAA AAGTTACTTT ACACCGTGAA ACTAAATAAT 780

TTATTTGAAT ATAGCCTAGA ATCCTTGCTT TATAAGGGTT CTAGGCGCTT TTTTTGTTGT 840

TAGGAATTTA CTGATTTAGT TATTGATGTA GATTTGCTAT TTTAAAAATT TACATAGCTA 900

CTGAAAGCAT CTGTTGCTTC TTTAGTAACT TCATCAGAAA CATGAGTGTA AGTATCCATA 960

GTTATTTGTA AAGAAGAATG TCCTAAACGT TCTTGGATTA TTTTAGACCT AACGTTATCT 1020

GATTCGAATA ATAATGTTGC GTGGGTATGC CGAAAACCAT GACAACCAAT AGAATGTAAG 1080

TTAGCTTTTT CTGCCAATCT TTTAGAACGT TGGTAAATGT CTTGACTTCG GGACATGGTA 1140

CCATCAATTT TTGkAAAAAT GaGTtGTGTT TTAAATCCAC CTTTTTTCAT TAAAGCTTCA 1200

TGCTGTCTAA GTTTCCATTT TTTTAAGATA TAAGCAGTCT TGTTATCAAA AGAAATTTTA 1260

CGAATAGAAT TGGGAGTTTT AGGATCGTTT ATAGTCAATC CATTTGTACT GATAGCAGTA 1320

GTTTTATTTA TATTAACTAC CTGCTTTTTT AAATCAATAT CATTCCAATT CAATGCTAAA 1380

GCTTCACCAA CACGTATACC AGTAAAAGAA AGTAAGCGAA AAATAGCACA GTCTAAGTCG 1440

GCATAGTATT TTAGAACTAA ACTTTCTTCT TTGGCTTGAT TGGCAATGCT ATCAGCTGTA 1500

TTTAAGAAAT GTTCCAGTTC GTCTTTTGTA TAGAACTTTC TTTTTGTATk CTTTtCTAAT 1560

TTCTkTAGCG AACTAGGTTT AGTTATTTTC TTAAATGGGT TTGAGTCTAT TATTTCTAAA 1620

CCAACAGCAT AGTCACAAAC AC 1642

(2) INFORMATION FOR SEQ ID NO: 599:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 694 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 599:

AGTGATCTTT ACCCTTTCCG CAGATACCGC AGCCAAAAAA GAACTACAAC GGATTCTTTT 60

GCAGTATAAT AGAAAATAGA AACGAAAGAA CAGGCACCAA CACAAAGGAG GTATTTTTAT 120

GAAAGTCACT CAnCTTTCTA GTGAAaCCCT TGACCGAGCG CACGAACGTT TTGAAGAAAC 180

ACTTGCGCAA ATGACCGTTG CCGAAGCAAA CACGATGCCC GCACCATTAA TCAAGTCTGT 240

TACTTGGCTG ATGTGGCACc AGCCAGAGAA TTAGACCTAC AAATTTCCGC ACTTAATCAT 300

AGTGATCCGT TATGGCTTAG TCAACACTGG ACAGAAAAAT TTGCCTTGGA TTTGCCAGAT 360

GAAACAGAAG ACTGGCATCA CACACCAGAG GAAGCGGCGA AAGTCGTGGT GGCAGAAAAG 420

CAATTGCTTA GTGACTATTT GGCTGCCAGT GTGGCACTTA CCAAAAGCTA TTTAGACCAA 480

ATCAAAGAAG AACAATTGTC AGACGTGATT GATAAAAATT GGACCCCGCC GGTGACTAGA 540

CAAGTCCGTT TAGTTTcAGC GATTGATGAT GCAtGATGCA TTCTGGGCAA GCCGTTTATA 600

CACGAAGATT AGTCATTGGA AAATGATAAA TAAGCAAAAA CCAAAAATGC TTCTGCTTAA 660

GCGGAGCATT TTTTATTGAG AGAAAAATAC ATAn 694

(2) INFORMATION FOR SEQ ID NO: 600:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 759 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 600:

AATTAATTTG ACATAAAACC AAGTAGAAAG ATAACATTCT TTATCTAATT CAGTTATATA 60

TCACTTTTTT TGCAGGTGCT TCAACAGTTA AGATTTCAAT AGAAACTGGA CTATTTTCTA 120

ATATACCyTC TTCTTTTGCA TGTTTCCtAC ATAAATTATT TTTTGTAGAT AGTACTCCTA 180

TGTCCTATTG AAAATATTTG TATTATAATT TCTGAATCAC AAATATTAGC CAATATTCGA 240

TAATTTCCAA TTCTGTAACG CCATTCACCA CTTCTATTTC CAGTCAAGCC TTTTCCATGT 300

TGACGAGGAT CAGTAGTTCC TTCTAAGTTT TTTCTAATCC AAGCTACTAT TATTCTTGCC 360

TGATTTCTAT CCATTTTTTT TAATTGTTTC TGCGCTTGTG GCATATATAC AACTTTATAT 420

TTCATTTAGT AAATTCTCCA TTTCCTCTTG AGAAATACCA ATCTCTTTTT TCTCTTTGGA 480

AATTTCTATT GCGTTCCGTA AGgTTTTTAA ATCAATTTCA TCTTCAATTT TATCTAATAC 540

AGCTTTTCGA TAAAGTTCAG TTAATGAAGA ATTATGAGCT TTTGCATATT CTAACACAAG 600

TTCTTTCTCA TCCTCAGGaA TTCTAAATGT AATTGkTGaT TGaGACATCt TATTACCTCC 660

ATTCGCACTT kGtAAtAACa TTGTATTAAA AACAAGGGGA TTTGGCAATA CAATGTTATT 720

ATAAATGTGA TGATACTCCA ATATATAGAA TAAACATGA 759

(2) INFORMATION FOR SEQ ID NO: 601:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 691 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 601:

ATTCATAAAA TTGGCATTAA TATTAATCAA ATGGCCCGTC TGGCCAATCA GTTTCATGAA 60

ATTTCGTCTG AAGATATTAA AGATTTGACC GATAAAGTAC AGAGTTTAAA TGCGCTGGTT 120

CAGAGCGAAT TAAACAAATT AATAAAGAGA AAGGAACAAT CGTAAATGGT GTATACAAAA 180

CATTTTGTTA TTCATACATT TGATAAATTA AACAATGCTT GCTCATATAT TGAGAATGCA 240

GAAAAAACTG AAGTCACGAA TGATAATCCG TCTGAACACT TGGAACATTT ATTTCAATAT 300

ATTGTGAATG ACGATAAGAC GTACATGAAA AAATTAGTTT CTGGGCATGG CATTGTGGAT 360

CCAACAAATC CTTATGAAGA ATTTAAATTA ACAAAATTAC AAGCAGCAAT TCAACGAAAA 420

ATCGGGTACA CATTCGATCC AAAATCAGAA CGATTGCTTC CGCCAACGTT AACAGAATTA 480

GAAAAAGGCA ACGCgctTTT AGCACACCAT TTAATCCAAT CCTTTTCTCC AGAAGATGAT 540

TTAACGCCAG caAAAAATAC ATGAAATAGG GTACAACACG GTGATGGAAT TGACGGGTGG 600

AGAGTATGAA TTTGTGATCG CCACACATGT CGATAAAGAA CATTTACmCa ATCmTAtTAt 660

TTTTAGGTTC CAACCAACTT TAAAAACCAG G 691

(2) INFORMATION FOR SEQ ID NO: 602:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3274 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 602:

TTTTACnTCT TCTATGTTAT TTATnTCTAT ATTATCTATA TATTCTTCAT AGAAATAGGT 60

ATTTCTAGTA ACTTTTTTAA CTATATAAAT ATCTGCCTCT AATTCTTTAT TTAAAAATTC 120

AAAAAAATCT CTTAATTTAT TCTCGCTATT AATCTTTGGC TTGTATACTA TCTTCTTTCC 180

ATCAGAAAAA GTTAGTGTAG AAACAGTTTT TCCGCGACTA TGAGAGTCTC CCTGTGATTC 240

TGATATAGAG TTAAGCTCTG AGGATTGTAT GTTAAAACAA TTTTGTATGC TAGGTAAATC 300

TTCAGTCACT CTTATAAGCA TTTGTTTAGT ATTATCTAAA AAATATCTCA TTCTAACAAC 360

AGTAATACGC ATCAACTCAG GATAGCATGT ATAAAAAGCT ATTATATCTT TTTTAGAGTT 420

AAATCTTTTT TTTAGATAAT AGATAAATCG CTTGCTACTA TCATTTCCCT TTAGAGGTTC 480

ATTTTTTTTA AAAGTATGCA AATCTAGCAC TAATGTTTTA CTTGTTAAGT GAATAAGCTC 540

TTGTGTTAAA GTTTCTAATA AATTAATAAT GAATTCTTTT GTACAAATAT TTAACTCTGA 600

GTTTAAATCA AGTAAAAATA AACGaGCATA CTGTAAGAAA TATCTAAAAG GATAACTAGC 660

ATCTACTTCT AATAATTTCT CTTCACTATC ATTATAAGAA TCTAAAATGG aTTCAAAAAA 720

AAtATACCAT GaTTGaTcTk gTAATGTAtT GAAtTGCAmC TTCTTTATCA GGAAAGTTTT 780

CAATAGGTGT TATTCCTAAT CCAAAATTAT CCAGTGATTC ATATTTATAT TTAATCAAGT 840

AGTCTAAATC ATCTTGTTTA AGAACACTTT TTCTTTCTTT CCAAGCTTGT AAATTTTTAA 900

TGTCGTACTT CTCTTTACCA CTTTGCTTTA GTAAAAAACA TCTTTCATTT ATGCTTAAAA 960

CATTTATTAA ATTATCTTCC ATGTAAGCAC TCCTTTTTTA TGTATAAGAG GGCTAGTTTC 1020

ACTAGCCCTC TTAATATATA AATkAGCAAA ATTTAGCTGA AAATAATGCA CCTACTCCTA 1080

AGCCTATGGT AAAACATGCT GGAGTTGTCT CAGCCTGAAC ATCTCCACTA CCTTGAAtCG 1140

CTTCCATTTC TTCTAAACTT AACTCTTCAA AAGAAGGACC AACAAGTTCT AATTTATTAG 1200

AGTAATAGTT TTCTTGATTT TCCTTATTTA GCACTGTGCT TCACCTCACT AAGTTTTATA 1260

GTATATTTTA ACAATGTTTT AAAGACACAA CTACAGTTAC TCCAGTAAAA ATACCGCCAC 1320

CAACCCAGCC ACAAGCAGCA CTACTTGCTG CAGCTGTCGC CGCaACAGCA CACACCGGTG 1380

TTGTCTCAGC CTGAACATCT CCACTACCTT GAATCGCTTC CATTTCCTCA ACACTTAGTT 1440

CTTCAAAACT AGGrACTACA CTTAAATTTT CCATAATTCT TACCCTCCAT CATAAAATTG 1500

TTGTATATTG CATAACATAA TATGTTACAA TTTAATATTA TAGAGCtTTT AATCTTTTGT 1560

CAAGTGACAC TTGACAAAAA GGAGGAATCA AAATGGCTAT TTTTATTTTT TTATTGGATA 1620

TCATTTCTGT AGTCTACTTn GTTTTTAATC TGAnTGATTC AAAAAAaGaT GmGCGAATTT 1680

TATATwCaAT TTCtTTAACA TCTACTTACA TGTACTTATT TATTAACTTT TCCATACTTA 1740

TAGTAATTGC TATTTtTCTT TTACTACCTA ATGTTATTAC TAAAAACTAT CTATTTAATT 1800

TTATCTACTA TTACTTCTCT CTATCAATGA TATTGCACTC TCTGTTACTA CTTTTTTTCA 1860

ACAAGAGGAC TTCATAAATG ATAATCAATA ACTTAAAATT AATTAGAGAG AAAAAAAAAA 1920

TTAGCCAAAG TGAATTAGCT GCTTTATTAG AAGTTAGCAG ACAAACTATT AATGGTATAG 1980

AAAAAAATAA ATATAACCCT TCTTTACAGT TAGCATTAAA AATTGCTTAC TACCTGGATA 2040

CTCCACTAGA AGATATTTTT CAATGGCAAC CTGAATAAAA AAGATTTTTA AATTTCACCA 2100

CAGAATATTT AAATTAAAAG AAAGGTGCTT CGTATGAAGA AAAAGTTTAT ATCCCTTCTG 2160

ATTTTATTTT TAGGAATAAT TGTTGGCATA GTAACAGTTT ATATTTTAAA GCATAAATAA 2220

TTATGACTGT TGTTTTATAC TTTAATCATA ACTGATTTTT TATCAGTAAA ACCCTCACTT 2280

ATGATTACAT TCATTTAAAT AAAAACTAGG CTCAAATGTA TAGTAAAGAC TATCTCTTTT 2340

TTTGGGGGAG ACAGTCTCTA AGGTTATAAC TTATTTTAGC CTTTCAATTG CCTTGATATT 2400

ACTTTTATTT CAAGATTATT TAACATGGAC AAATACTTCT GAATCTTTTA CATAAATGTC 2460

GTGGTAATAA TTAGATTTCA TTAAGTCACT TTCTTCTTTA TTCAACGCGT TAATATCAAA 2520

GATCTCCATG TTCACATTTA AGCCAGTGTA CGTTCTAAAT AAAAACTGAT TTGGCCAATA 2580

AAGGGCTTCA TTTGGTGGAA CCAGTTCTTT TAATATAGGA TAATTCTTCG AACTATTGAC 2640

AAATACTGGC GAATCCTTAA ATAAAGAGGT ACTATGcACG GTCACcGTGt CACGGTTCAC 2700

TAGATCTTTT AAATCACCGG CTAATATCAT ACTTTGTCGT TCAAAAGATT CCTTTTGGTG 2760

ACTTAACGCA CTAGCATAAA CAAATGTAAA AGATAAAAGG TAAACACAGA AAAGGCAAAC 2820

CGCTGTTTTT GACACATAAC CTACCGCGGG TTCTGTTGCA AAGTTTTAAA TAAAGAATAA 2880

AATCCCTTAC GGTATCTATG ATTTAAGCTG GGATTCCCAA TAATACCTTG ATTTCAGTAC 2940

AGACCGAAAA CCCGAAGAGA GTACCTTCTT TTCGGGTTTT CTTATATAAT CCTCGAATGG 3000

CTTCCATGCC TTTAATCGTG GTAGAGGCAG TGCGTAAACT TCGATAGAAT TTATTGCGTC 3060

TCTTTACTGG ACGATGGTCT TGTTCAATCA AATTATTCAG GTATTTAATG GTACGATGTT 3120

CTGTCCCTTG ATAAAAGCCG TATTCTTTTA GTTTCTTAAA GGCACTTGTA ATAGAGGGGG 3180

CTTTATCTGT GACTACAACc TTCGGTTCAT CAAACTGCTT CACTAAcCGC TTAAGAAAAG 3240

CATAGGCtGC TtGTGCCCGn TTTTACGnAC CAAT 3274

(2) INFORMATION FOR SEQ ID NO: 603:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2469 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 603:

ATTCCACGAG CCATAGCAGC ATTTGCAATA TCTGTTGGAT CACCAATTAA GATTnGTTCT 60

ACTAATTCAT CAGATTTTAA GCGGGCAGCA GCACCTAAGA TTCTAGGATC TGTTGCTTCA 120

GGGAAAACAA TTTTGATATT ACGGCGTACG ACTTTAAATT TTAAGCTATC GAATAATTCC 180

ACGATAAACC CTCCAGTTTT ATATTTACAA AATCCATCAT ACACTATTCT CAAAGAAAAT 240

AACAGTTTTT TCTAGCTGTA TTAGTCAAAA AAACCAATAA TCCGATAAAC TTTTCACAAA 300

CTAACTCTTC TACACTATAA CAGCATACAC TGTTACACTA CACTGTTTCA GGAAGTTGCC 360

AATCAATCGG CGTTTCCCCT AAACGTTCTA GAGCGGCATT CGCTTGAGAA AACGGACGTG 420

AGCCAAAAAA ACCACGATGT GCTGACAATG GACTTGGGTG TGGCGATTTG ATAATCACAT 480

GCCGAGTCGT ATCAATCATT TTAATTTTTT CTTGAGCAGG ACGTCCCCAT AAAATGAAAA 540

CAACAGGTTT TTCACGCTCA TTCAACTTTT CGATAATGAC ATCTGTCAGT TGCTCCCACC 600

CTTTACCTCG ATGGGAATAG GCTTGACCTG CCCGAACGGT CAACACAGTA TTTAATAATA 660

ACACGCCTTG TTTGGCCCAG CTTTCTAAAA AGCCATGGTT GACTGGTTGG TACCCTAAAT 720

CAGCTTGTAA TTCTTTGTAA ATATTCGCAA GAGATGGCGG CACTTTCACA CCTGGCTGGA 780

CAGAGAAACT TAGTCCGTGT GCTTGATTGG GTCCATGATA TGGATCTTGT CCCAAAATTA 840

CTACTTTTAC TTCTTCAAAA GGGGTTAGTT CTArCGCTGA AAATAAATGA TACATATcTG 900

GATAAATCGT TTGTGTCTGA TATTCTTTTT TCAAAAATTC CCGCAATTCT TGATAATACG 960

GTTTTTCAAA TTCAGCAGAT AAAATGTTTT GCCAACTATT GTGAATAATT TCTTTCATAA 1020

TGCTCACTCC TTCGCTCTCT AAGCATAACG AATTTCTCCA AACCTGACAA GAAATTTCTG 1080

GGTACTTTCT CTTGGAAAAA CGTGGTAAAC TATAGGTAAG CTTTAAATAG ATAATGAGGT 1140

GTAGAGAATC GATGATTAAA TTAATTGCTT CAGATATGGA TGGAACACTT TTAGATGCAA 1200

AAATGAGCAT CACAAATGAT AACGCTTCAG CAATTCGCGA AGGAwACGTT TaGGAaTCgA 1260

ATTTAtGGTT GCTACTGGCC GCGCATAtAC cGaAGCGAAA CCCGCGTTAG AAGAAgCTGG 1320

aATTGACTGC GCCATGATTA CTTTaAACGG CGCCCAAGTG TTCGACAAAG ATGGACACTC 1380

ACTTTTCACG GCTGGGATCG AAAAAGAmAC AGTAACAGAA GTACTAACTA TTTTGAGTCA 1440

ACACAATGTT TACTATGAAA TCTCTACAAA TAAAGGTATT TTTTCTGAAC ATCAAGAAAA 1500

AAGAATTGAA AACTTTGCCG CTCATATTGC AGAATCCATG CCCCATTTAA CTTATAAAGT 1560

GGCGATTGCC ATGGCTTCAG CACACCTATC TTTATTGCAT ATTACCTATG TTGATCGTTT 1620

GGACGACATT TTAAAAGATG ATTCGATTGA AGTCCTTAAA ATTATTGGCT TCAGTATGGA 1680

TGGGCCGAAA GTATTAGGTC CTGCTGGGAT GGAAGTAGAA GAACTAGATG ATTTGGTTGT 1740

TACCTCTTCT GCGCTTAATA ATATTGAAAT TAACCACCGC CTAGCACAAA AAGGGATTGC 1800

AGTTGCGCGC GTCkCTAAAG AACGTGGGAT TCCCGCTGAA CAAGTCATGA CTATTGGTGA 1860

TAATTTAAAT GACGTCAGCA TGATTCAATG GGGCTGGGGT AAGTTTTGCA ATGGGGCAAT 1920

GCCGAACTGG AATTAAAAGA ATACGCTAAA TATGAAACAG CCACTAATTa GAAAAgGTGT 1980

TGGCGArCaT CCgCGTGCAA TAAGAGAAGA TTtGTAATAA GCACTaCCAG TTACAGCACT 2040

CATTTATATA TAGAAAGGAG ATCGTGTCGT GTCTGAATTT TTTATTCAAG AACAACAGTT 2100

AGGAAAAGTT ACGCGAACCA TCGTGAAAGA CCAAGCAGGC CGTTCACTTT TTCTATTGGT 2160

GGGGCGCTGG GGCACACGAG GAGATGCGCT TTCTTTATAT GCAATGAACG GAGAAATTTT 2220

AGCCAGTATT AAACAAGTTT CTTGGACTTT TGGCACACGA TTTGAGTTAT ATCAGCGTTT 2280

TGAAAAAGTG GGTACTCTAC GGAAACTGTT CAATTTGAAT GCTGATTTCT ATTATGTTCA 2340

AGGCCTTCAC TGGGCAGTTG TTGGGGATAT CAAAGCGCAC CAATATTCAA TTTATCAAGT 2400

CCATAAAAAA ATCATGTCAA TGGATAAAAC CATGCTTTGT ACAGGCGACT ATTTTGTTTT 2460

AATGTTGTC 2469

(2) INFORMATION FOR SEQ ID NO: 604:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 797 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 604:

GAATCACCTC AGAGTTAATT TTGTCAAAGC AGGTTTTTCC TAAAAATATA GAAATTACTT 60

CTTTTTTAAA TGTAGTATTT AATGTGGAGT TTAAAAATTA TGTAATGAAA TCTAGAACTT 120

TGATTTTATC AAGAACTGTA CGTGTTATAG AAGGATGCTC AGAGAATGAA TATCAGAATT 180

ATAGAAGAAA ATTATTAAAC TTTGTAGAAG AGTATTATGA GTCGGAAGAA GTGTCAAAAA 240

ATATTTCAAA ATCTTCTATA TCGAAATGGG TAACAGGAGA ATAATAGCAT GTATAATCAC 300

GTTTATTTAA AAGAAGATAT AGATGCTTTA ATAAATGAAT TTGGAAAATT GAATGCTGAG 360

GATGaATCTT TATTTAGaTT TCTAAGTAAA AAGATTATAT TCTTAAAGGA AATAAAAATA 420

AGTATAGTAA ATACTACTGT GATATTTTAT ATAGACCAAA TGATATCGGA TTTAATGTAC 480

TTAATGGCTA GTTATCATAA GGGTGAAGTG AGATATTTTT ATTTAAATAT TAGATCAGTA 540

ATTGAGGCAT TTTCCCGTTT GTTTAGTGAG GTAGAAACAA GTACTAATAG AATAACTATG 600

ACAACTTTAT TAGATAATAT AGCTAACTAC ATTACTTTAA ATGATTTGCG TGATAGTnAA 660

GAGGrTtCTC TGGACTATCC TAGACTAAAA GGTCTTTATA GAGAATGCTG TCTTTACGTA 720

CACGGAAATA TTAATGCnAA ATATTCATTA ATTGAGTTTT ACAACGAGTT GCTTAATGAA 780

GAAATAACTA GTGCTCA 797

(2) INFORMATION FOR SEQ ID NO: 605:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 936 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 605:

TCTGCGTTAT AGAGCTTAAG AGGTAATnAT nCAAnATCAT TGCAATCAAG GTGGTAnCGC 60

GAGCAATCGT CCTTGACTGC AATGGTGTTT TTTGTGTACA CAATCTGCCA CTTAAGAGGG 120

TACAAAATAT AAATAGTTAA TTATTAATAG AAAGTAGGAA AATCAATGAA AGAATTAACA 180

AGTAGTCAAG TTCGCCAAAT GTATTTAGAC TTCTTTAAGT CTAAAGGACA TTCAGTAGAA 240

CCAAGCGCTT CTTTAGTACC AGTAAACGAT CCAACATTAT TATGGATTAA CTCTGGTGTT 300

GCAACCTTGA AAAAATATTT TGATGGCTCT GTTGTGCCAG AAAATCCAAG AATTACGAAT 360

GCTCAAAAAT CTATTCGTAC AAATGATATT GAGAATGTTG GGaAAACGGC TCGTCATCAT 420

ACAATGTTTG AAATGTTAGG GAACTTCTCA ATTGGAGATT ATTTTAAAAA TGAAGCGATT 480

CACTGGGCTT GGGAATTTTT AACTGGTGCT GAATGGCTTG CTTTTGACCC AGAAAAATTA 540

TATGTGACTG TTTATCCGAA AGATACAGAA GCAAAACGCA TTTGGCGTGA TGAAGTTGGT 600

TTATCTGAAG ATCATATTAT TGATGTAGAA GATAACTTCT GGGgATATCG GTGCTGGTCC 660

ArTGGTCCGG ATACGAAATC TTTTATGATC GCGGCGAAGA ATTTTTAGAT ATTCCAGAAG 720

ATGATCCAGA AAaTTATCCT GGTGGCGAAA aTGrACGCTA wTTAGrAATT TGGAmCTtAG 780

kGTtCyCyGG AATTTAATCA TACACCAGAA GATACGTACG AACCATTACC ACATAAAAAC 840

ATTGATACGG GCATGGGCTT GGAACGTGTT GTATCTATCA TTnCAAGATG GCACCAnCAA 900

ATTTTGGAAA CTGATTTATT TATGCCAATn CATTCA 936

(2) INFORMATION FOR SEQ ID NO: 606:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 890 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 606:

GCTGTTCCGG TTCTTCCTCT AGTTCCTTGG TTaTcAAAAa CGATATCTGC TCGTTCTTTT 60

TTCATTTCAA TCGGCCATkG GCTAGCAATT CGkTGTTGCG CTTCTTCTTC TGTTAATgGT 120

TTCGGGCCAT CAGACGCTCT TTnTGAATCT TTTCTGGTAC ATATACCACA GCAACTTGAT 180

CCATGATCGC CTCATAATGA GCTTCATAAA GCAGAGGAAT ATCAACAATA ACAAGGGCGG 240

CTTTCTTTTT AGCTTCTTCA ATTTGACGTA AAATTTCTTT TCTTAAAAAA GGTTTCAGCG 300

TTTCATTTAA CAACTCTCTT TTTTGAGGAC TAGCGAAAAT TAATTGTCCT AATTTTTTCC 360

GATCCAATTC ACCAGTTGTT AATACTATTT CTGGTCCGAA AGTTTCAACG ATTGCGGCTA 420

AGGCTGGTTG CCCTTTTGCA ACAATTTCGC GGGCGATAAT GTCGCCGTCA ACAATTGGAT 480

AACCTGCTTT TTTAAAAAGA GCAACAACCG TACTTTTACC TGTCGCAATC CCTCCTGTAA 540

TGCCTAAAAC TTTTGTCATT TCAATCGACG ACCTTTCAAT TGTTGGCATT GTGGACAATA 600

ATGCGTACCA CGTTGTGCCA CTTTTGTTTT AACAATTGGC GTACCACAAC GATTGCAAGG 660

TAAGCCTGTT TGCCCGTAGA CATTCAACGC TACTTGAAAA GTCCCTGCTT CTCCTAACGC 720

GTTTAGAtAG GTGCGGAtGG TGGTtCCACC TGCTTCAACA GCGCGTGCCA AGACGTCAAT 780

GATTGCTTGA TACAACGTAG CAACTTCCGC AGTTTTAAGG AATCTGCTGG TTGTTCTGGA 840

TGGATTTGCG CTTGCCAAAG TGCTTCATCT ACATAAATAT TCCCCAAACC 890

(2) INFORMATION FOR SEQ ID NO: 607:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1067 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 607:

GCCAACGTTA GGGAACCAAC CAAGGCGTTT nGGACCAACT TTGGAAAACT GACGGACTAA 60

AAGGGCCTCC nGATTATCTA AnGGCTAAGG CCnTAAGGGA ATCCGAATCC CCTGTCATTG 120

TGTAGCCGTT TAACGCACCA TATACAGAGA TTAAAATCCC AATAGTTACG ATTTTGCCGC 180

CAATTGGTCC AAAAATAGCC ACAGAAGCAT CCGAAGCTGC ATTCAAATTC CCAGCGATTT 240

GCTCAATTGG TAATGTTTTT AGAAACACAA AATTAATTAA GGCATAAATT AAAGTAATCA 300

ATAACAAACC AAAAATAATG GCTTTAGGCA AGTCTCTTTC TGGTCGCTTC ATTTCACCAG 360

CCACGTTACC CACACCCAGC CAGCCATCAT AAGCAAACAT GGTGGCGACT AAGGCGCCAC 420

TAAAAGCAAC TAAAAAGCCA GTATTCGcTG TTGTTTCAAC AGGAAAAAGT GAAACAGCTA 480

CTTGTCCGGG AGTAAATAAT CCTACTAAAG AAATcAAGGC AATTGGAATC AGTTTTACTA 540

CCAATGTCGT TGATTGGACC AAAGAAGCAA TTTTTGTTCC TAATAAATTA ATTATCGTGA 600

TACTGGTTCC CGCAAGAATG GCAATTGGAA TCAATAAATT TGCCGATAGA TGAAACAAAT 660

TGATTAGTTG TGTACTAAAA ATAATGGATA ACGCCGAAAT GTTGGCAGGA TAATAAATGA 720

TACTTTGTGC CCACCCTAAA AGAAAACCAG CTAATTTACC ATATGTATAT TCAATATACT 780

TCACTGCGCC ACCTGTTTCA GGGATTGCTG TAGCAAGTTC GGCGCTGGTT AATCCTGCAC 840

AAATGGTTAA gGCACCACCT AACACCCAAG CAAAGATAGC TArACTAGCT GACTGTGCGT 900

GTCCAACGAC ACTCGCAGCT TTAAAGAAAA CACCAGCGCC AATCACCGTA CCCATTACCG 960

TGGaTAACGc ACCAAACGTT GTTATCTCTC TTTTTAATTG CTTTTTCTCC ATrGcTCCCt 1020

CATTTCATTT AAACTTAAAA AGCGTAACAA AATTTCCACT TTTTTGG 1067

(2) INFORMATION FOR SEQ ID NO: 608:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 791 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 608:

TACCCGTCGC CAATGGCTGA CAATGGATAT GATATTtCAG ACTATTATGG TATTTCTAGC 60

GATTTTGGTA CAATGGCAGA CTTCGATGAA CTCATTGAAG AAGCAAAAAA ACGAAACATA 120

AAAGTTATTT TAGATTTGGT GGTAAACCAT ACATCTGATG AACATGCTTG GTTTCAAGAT 180

GTTTTAAAAA ATCCTCAAAG TCGTTTTCGA GACTTTTATA TTATAAAAGA AGGACGAGAA 240

GCACCTACAA ATTGGCGGTC TAACTTCGGT GGTAGTGTTT GGGAAAAATT ACCTGGGGAA 300

GATGCCTACT ACTTTCATGC CTTTCATAAA AAACAGCCGG ATTTAAATTG GGAAAATCCT 360

GAACTACGCA AAGAAATCTA TCAAATGATT CGTTTTTGGT TAAACAAAGG AATTGCTGGT 420

TTTCGTGTAG ATGCGATTAA TTTTATTAAG AAAGATTTGA CTTGGACAAA CTTACCTGCT 480

GaTGGGGcAG ATCAATTAGC CAAAGTAACA AAAGCTAGTA GAAACATGCC AGGAATGAGT 540

GACTTTCTGA ACGAATTAAA AGAAAAAGCT TTTGCTGGCT TTGaTAtTGT CACAGTTGCC 600

GwArCGGCCG GCGTCAATTA TCTAAATTTA TCAGAATTTA TTGGAGAAAC AGGGTATTTC 660

GACATGATTT TTGACTTCAA aTGGGCTGAT TTGGAwGtGA AATCAGGAAG CGAATGGTTT 720

TATAGAATCG ATTGGtCTTG GAATGATTTA CGCACATTAA TTTTCAAACA GCAAGAAGCC 780

ATGCAAGAnG C 791

(2) INFORMATION FOR SEQ ID NO: 609:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1243 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 609:

AAGACGACGT tCCAATGTGT AAwTGAACAC AACCTGTTTT ACAGTGGTGT TGCCTAATAC 60

GGATTGCTAC TTGCTCAGCm ATTTCTTTTA CGACCAGTyC GATTTCATTT CTTCTAGTAT 120

AATCTCGTGG TAACACTTGT GAATTGCCAT AAGATTTTTC CTTTGACGGT TCTAGTGGTA 180

TTGCTATGTC AGTTCGATCA ATACCATTTG CATGGAAGTA CAGCTGTAAC CCAATCACGC 240

cTAAACGGCT CTTAATGGTG TATGGATTCC AATTAGCTAA ATCTTTGATA CTCATGATCC 300

CCATTTGGTT GAGCCTTTTT TTCATTCGAG AACCGATACC CCAAAAGTCT GtCATTTCAG 360

GAATATTCCA TACTTTATCA GGAACATTTT GATACGTCCA TTCGGCAATA AATCCTTCmT 420

TGkTCTTCGC TtCGTTGkCT AArGCAAGTT tGGCTAACAA GGGATTATCT CCGACACCTA 480

CTGCAGCAAT CAATCCTAGC TCTTCTTTAA TACGTTCTTG GATCATTTGA GCGAGCTTCT 540

TTCTACGTTG GCTTCGTGTT CCTTCAGTCG TAAAAAGATT CAGTGATCGG GTCACTTTTA 600

AGATTGATTC ATCGATCGAG TAAATCAGTA GATCTTCATC AGCCACATAT CTTCTGAAAA 660

TATTATTTAC CTGCATATTT CGCTTGATAT ATAGTTTCAT ACGTGGTGGA ACAACATGTA 720

GTGTTTTAGG AAATGGTTGT GGTAAAtCAC GkGGTCTACT CACATTtGTA AtACCATACC 780

GCTTTTTkGC TTCAGGAGAA GAAGCTAATA TCAATCCTGA ACCAGTATTG TCACCTCGAC 840

TCATAACAAC AAGTTCTGTT GTTAATGGAT CTAAATTTCT TTCTATACAT TCGACACTCG 900

CATAAAAAGA CTTCATGTCG ATTAGAAAAT AATCATTTAC TGGTTCTTTC GAATAATCAA 960

ACATGAATAA CACCCCTTCT TTTTACATTA TACAAACGTA TGTTCGATAT TGTAAAGAAG 1020

AAATAGCTAA ACACACAAAA AAAGAAGCGT AGGTAAACTC TTACCCTGCT TCTTCTTTTG 1080

ATTTGTATGA TCAACTCATT GtATCACTTT ATTAAATAAT TTGAAAATcT AATaAAAACT 1140

CTCACTCTAA AAGAGTGAGA GCCTGCAAAT TAGAGTACGG AATTCTTTGA GAGTTCCGTT 1200

CCTTAATATT AACAAATGGT CTTAAAGCAT ACAAGGCGAA TTA 1243

(2) INFORMATION FOR SEQ ID NO: 610:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 586 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 610:

TAGATGnAAT nTGAAGTGAC TGGTGAAAAT CGACATATCA TTAATGGCCA TACACCAGTT 60

AAACGAACAA AAGGCGAATC GCCAATCAAA GCAAATGGCA CGCTTCTCGT AATTGACGGC 120

GGCTTTTCAA AATCTTATCA GACAATTACT GGGATTGCTG GTTATACATT GTTGTACAAC 180

TCGTTTGGTT TACAACTGAC CGCGCATAAA TCTTTTTCTA GTAAAGAAAC AGCCATTTTA 240

AATAATCAAG ACATTCATTC CATCAAACAA GTCATCGATC GCCCCTTACA GCGGCTATTG 300

GTGAAAGATA CCACTATCGG CAAAGAGCTA TTAAAGCAAT CGCAAGCACT CCAAAAACAA 360

ATGAAACAAA ACCAGTAGTG CCGTTAAAAA GACGCTTGCG CTCTTATCCG AAAAATGATT 420

AAATGGaGGA GTTGAGAAAA ATATCTATTA GAGAAACGGA AAGGGTTTtA CATTGAAGAa 480

CAGCmACTCT CACTTTACGG CACCATTTTC TTTCTTTGCT TTnTCCTAAT TGTGGGCTTC 540

TTCTnCTCAA GTCGTGGCGA AGATTATGCC AAAGATCTGA ACAAAA 586

(2) INFORMATION FOR SEO ID NO: 611:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 421 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 611:

GCACGCrCAA CTTTCAAGCA AAAACGAATT GGGAACTATA AAGAAACAGC CGATTGTTAT 60

TCTCTTCAAT CGGCTGTTTC TTTATTTATT TGCCTTGTCA TCAGTGTTTT CTTCTGCAGC 120

AATCCGCGCT TGAACTTGTT GACGTAAAGC TTCTAGTTCT TTTTCTTTCT TTTCTAATTC 180

AGCTAAATCT TCAGAAATTT CATGGGCTTC TTCTGCTGGC TTAGGTTGTA AAAAGACGTT 240

TAAACCAATC GCTATAAAGG TCCCAATGAA TGTATCCAAC ACACGTTCCA TTGCATATTG 300

aAAGGaATCG CTTTGCGGaA TACTTAAAGA AATCATTaAA AGTGTCGCAG TTGCaGAAAT 360

GATCCCAGCA TTGTTrTTAA TCCCATCAGA AACAACGaTG aCGaTAATGA CTAAAAGTGG 420

T 421

(2) INFORMATION FOR SEQ ID NO: 612:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 687 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 612:

TTTTTCAATA AATAgTTATA AAGATAATAT TACTAGAATT AACGAAGAAT GGAATCAATA 60

TCAGTCAAAA GAGACCATTA ATAAAATAGT AGAGTTGGTA GAGAGAATCT TGGGGGGATT 120

TTGTGAAAAG TAAGGTACGC ACTGAATATA TTGGGGGATT ACCAGTTGAC GTATTAACAC 180

ATCAAGAAAT TATGGATGAC TTAGAAAGTT ATCTCAAAGC TAAAGAGAAA ATGATTGTmA 240

CTAGCGTTAA TCCGCAGATT TCATTGTTAG CATTTGAACA TCCAGAAGTT AAGTTATATT 300

TGAAAAGAGC AACTCATCGG ATTGCAGATG GGATAGGTGT GGTCAAGGCA TCGAAGCTAA 360

TGGGTGGGTC TATTCGTGAA AGAGTTACAG GAATTGATGT CaTGGATAAT TTTTTAGAGT 420

ATGCAAATCa ACATaAGGnA CGAATCTTCT TGTmTGGTGC aAAGCAAGAA GTAGTTGAAG 480

CAGCAGCTAA AAATATTCCA CGACCATATC CAGGGATTGT TATTAGCGGT ATTTTACATG 540

GCTATACCAA GGAAAAACAG CAGGAATTGT GGACCAATTA ACCAAGCTGA ACCAGGTTGC 600

TTTGTAGCAT TGGTCnCCGA ACAAGAGTTT TTGGACAACA TTGTCnCCTG ACACTATGAT 660

TTAGACGAGC GGnCTTGCGT TACTGGn 687

(2) INFORMATION FOR SEQ ID NO: 613:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 889 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 613:

CGGCTTGAAT TTTTTATTCT TCTATAGTAG TTACAGCAAC TTTGGAAATC TTTTCTACAA 60

AGGCATCAAG TAAATCGACA ACTGCTTGTT GTGTTTCGAC TTGGTTAATC GCAACTTTCA 120

CTTTGGCTGC TCGCGAGATT CCTTTTAGAT AGTAAGAAGC ATGTTGGCGA AACTCACGGA 180

CCGCAATTTT TTCACCTTTT AAATCCACTA ATCGTTGTAA ATGAACTTTT GCAATGTTAA 240

TTTTTTCAGC TGGTGAAGGC TCAGGCATCA ATTCACCCGT TTCCAAGTAT TCTTTCGTCC 300

GTTGAATCAT CCAAGGATTG CCCAAAGCAG CTCGACCAAT CATCACACCA TCCGCTCCAA 360

CATATTCTAA CATGCGTTTA GCATCTTCAG GCGTTTTGAC ATCGCCATTT CCCATAAAAG 420

GAATGGTTAA ATGCTTTTTC ACTTCCTTTA AAACGTCCCA GTTGGCTGTC CCTTCaTACA 480

TTTGGACACG TGTCCGACCA TGCATAGCAA tTGCCGAAGC CCCTGCTTGT TCGGCAGCCA 540

AAGCGTTGTC TACAGCATAT AGATGATCTT CATCCCAGCC TGTTCGCATT TTTACAGTGA 600

CCGGCAAAGA AACTGCCGAA GAGACTGCGT GAACCATCTC ATGCACTTTA TTCGGATCCA 660

GCAACCATTT CGCACCAGCT TCAgCTTTAA TTATTTTATT GACAGGGCAG CCCATGTTAA 720

TATCAATAAT CGCCGCTTGG TATTTTCTTC CACAAACTTA GCAGCTTCCA CTAATGTTTC 780

GTTATCTCCA CCAAAAATTT GTACACTTAA AGGATATTCC CGCTTCATCA ATATATAGCA 840

TTTGCTAAGG TTTTTTGGTn TCTTnATTTT AAnGCCTTTG GTCACTAAT 889

(2) INFORMATION FOR SEQ ID NO: 614:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 995 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 614:

TGGTGATGAC TTTGTTTAGA ATACAAGTAG GATTTTGCTA CGACCGAAAT AAACATAGCC 60

ATAATCCCCA AGCCTACTGT AGCATAAATC ATGGTGAAAA TTTTTCCCAA GTCAGTAACT 120

GGATGCACAT CTCCATAACC AATCGTGGTT AACGTCATGA AACTTAAATA TAACGAATCT 180

AATGGCGACA ATTTTTCCAC TGTTGAATAA AAAATAGTCC CACTCAACAA CGTGGTTGCC 240

AGTAAAAAGA ACAAGGCACG AGTGTCTTCT TTTTTTAAGA TACCTAAAAA AGATGAAAAT 300

AATTTTTTGA CAGAAATAAG TAAAGATAGC AAAAAAATGA CCTCTTTCTA TTGTTAATCT 360

CTTCTTAACG TTAGCACTTT TTAGATGAAG TTCAACTTTT TAATTGGTGT GGTTAAAGAA 420

AATTTATTTT TTTGTTATTT ATAATACACT TCATCTAAAA ACAACCCTTG TGCAGGAGCC 480

GTTTCACCTG CGCCTTCTCG GACTTTTGTT GCGAACAATT CATCAATTAT CGCTAATTCT 540

TTGGTGCCTG CACaATTTCT AAAAGTGTCC CAACGAGGAT GCGGACCATT TTATGCAAGA 600

AACCATTGCC CACAAAAGTG AACTGTAACA TTTCGCCGTT TCTTTCAATC GTAATCTTTT 660

CAATCGTTCG AGTAGTCGAT TTTTTGGtTT TTTTCAATGC TGAAAAACCA ATaAAATCAT 720

GGGTCCCTAT TAATTTTTGG CAAGCTTGTT CCATTTTTTC TTGATCCAAC GCTTTGGGAA 780

AGTGAAAACT ATGATACCGT TCAAAGGcAT TAGGAATTgG GcTATTCCmA ACATAGTAGC 840

TGTATTGTTT GCCTGTGGCG TTATACCGTG CATGAAAACG TTCAGGCATT CTGGAACTTC 900

TTTAACAACA ATATCGCCAG GTAAATATCG ATTAAAAAGG ATAACATTCC GGTGTTGTAA 960

GTGACCAGnC GTTTAAAATC GCTACTGGAC CACGG 995

(2) INFORMATION FOR SEQ ID NO: 615:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 461 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 615:

TCAACCATTG TCGGCGCCAA TGCGTTTAGT TTATATTGGC TTTTGTTATC TGAAACAACA 60

GACTCAGAGG AGAATGGCTC GCTGTTTCAT GCAGATTTAA TTAGTTTGCT GGACTTGAGT 120

TGCCAACAAT TAGAAGAAGC TTGCTACAAA TTAGAAGGCA TTGGGTTGCT AGAAACGTAT 180

AAAAAAACAG ATCGGGAATT AGGAGATTGT TATTTATATT ATTTAAAAGC ACCTGAAACA 240

GCCGCTCGTT TTTTTAAAGA TGAAGTGTTA GCTTTAGTGT TATTTAATCG GGTTGGaCAA 300

CGGAAATTTG ATCAATTAGT GCAAAAGTTC CAACCGCATC CTACTAAAAC AGAGGGCTAT 360

CAAAATGTTT CCGCTAGCTT TCaAGArGTT akgCCTTarr GrAgACaAAT TGTTTcCGAA 420

GCGAAcCGrT TGACcACGAT ccAGrAACtT TTTCTCAAAA G 461

(2) INFORMATION FOR SEQ ID NO: 616:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 538 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 616:

CAAAATTTGT TAAAACGAGA TATTGCCAGT nTTTTAAGAC AAATGCACGG TTTAGATTAT 60

ACAGATATTA GTGAATGTAC TATTGATAAT AAACAAAATG TATTAGAAGA GTATATATTG 120

TTGCGTGAAA CTATTTATAA TGATTTAACT GATATAGAAA AAGATTATAT AGAAAGTTTT 180

ATGGAAAGAC TAAATGCAAC AACAGTTTTT GAGGGTAAAA AGTGTTTATG CCATAATGAT 240

TTTAGTTGTA ATCATCTATT GTTAGATGGC AATAATAGAT TAACTGGAAT AATTGATTTT 300

GGAGATTCTG GAATTATAGA TGAATATTGT GATTTTATAT ACTTACTTGA AGATAGTGAA 360

GAAGAAATAG GrACAAATTT TGGAGAAGAT ATATTAAGAA TGTATGGAAA TATAGATATT 420

GAGAAAGCmA AAGaATATCm AGATATAGTT GAAGAATATt ATCCTATTGa CtATTGTTTA 480

GGAATTAACA ATATTAAACC GGAATTTATC GAAATGGTAG AnAAGAAATT nTAAnAGG 538

(2) INFORMATION FOR SEQ ID NO: 617:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1164 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 617:

GCAATCAAAC GCGTACGTAC AAGTGCTAAC GCGAATGCGA AAAATTCATC TCAAACAAAC 60

GCAATGCtAC AGaATCAAGA AATTTGAAGA AGCTGTAGCT GCTGGAGCTG ATAACGTCGA 120

TGCTTTATAC AACGAAGCTG TTAAAGCTGT TGATATGGCT GCAACTAAAG GTTTAATCCA 180

TAAAAACAAA GCTAACCGCG ATAAAATTCG TTTAAGCAAA TTAGCTAAAT AATGTGAGAT 240

AGAAGTTAGC ATTCATGTTA GCTTCTTTTT TCATCCAAAA ATGTCCAAGT CTGTTTTCAC 300

AAAGACTCGG ACATTTTTTT ATAAAGGTGA ATCAATCCAC GGTAAATCCG TTGTGAGATG 360

TTGACTAAAT AATTTAGTCT CTTCTCGCAA AGCTAAACGT TGTTGTCGTC TCGTTTCATA 420

ATTGGAACAA ATCAACTTTT CTTCTGCAGT TTCTGGGACG ACTGCCGCTA ATTGGAAGTC 480

ACTTTCGGGA TTCATATGAG ACGGAATGGC TACAAAAGTT GTAAAACAAG TCGCCGCTAA 540

ATAACGCTCT CCAGTTGTTA AATCTTCTCC TACAACTTTA ACAAATACTT CCATCGATTT 600

ATGATGCACG CCAGAAACAA ACGTTTCAAC ACACACTGAG TGATTTTCTT TTAAAGGTTT 660

TAAAAAATTC AAGTTATCTA AAGATGCTGT AACTGCCCCT CTCCGACAAt GTCGTGAAAC 720

GGAAATGGAT GCGGCGTCAT CAATTAAAGA CATTAATTTC CCACCAAATA GTGCACCAAA 780

GGGATTTAAA TCAAAGGGrA ATACACGATG TGTTTGAATA ACTCGTGATT CTCGACAAAA 840

TTtTTCTTTC GTCATTAACG ATTCCTCTAC TTTCATTTAT AAACTTTTCC ACGCTTTACT 900

TTAGCTTACT TTACCATTgA TGCAATCCTT ACTTTTTAGG AAACATTTAA AAAATAACAA 960

ATTTGCAAGT CCCATTTATT CATTTTTAGG AAAGTTGCTG TATAATAATA ACGGTTAAAG 1020

AACTTTTAAG GAGATGGaTA CACaTGAAAA wTGCaATTGC TGGTGCCGGC GCTATGGGCA 1080

GCCGTTTAGG AaTTAkGTtG CATCAaGcGG rAATGakGTA ACTTTAATCG ATCAATGGCC 1140

CCGCGCACAT TGAAnCTATT CGTA 1164

(2) INFORMATION FOR SEQ ID NO: 618:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 989 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 618:

ATTnTCCCAT TCCTTGTTTC AATAAAGCAC TGGAATCTTT GACCGGACGT GTAGTTAACT 60

AAATATCTAT TACTAAACTC TTCTTTCGTC GAAGTATACC AATTTCTGTC AAATGTAGAG 120

TAAAATAAGC ATTCTCTATA TTTTTTCTGA GCTTTTAAAT AATAATCTGA GGTATCTTCT 180

ATCCCTATAT GAATATTCAT TTCCCTATAG TCTATATCAT CAGCAACATA AAACGACAAA 240

TCATTTTCCG AAAACATAAT AGCGGGTATA AAATCTCCTT CGCCAAAAGC TCCTAACAGT 300

CCAGGTTCCG ATAATTGCAT TTTAGCTTTC ATAAATCTAG GTAAAAATAT ATCATCAGTT 360

TCTACTTGTG AGATTGCTCT AAGTTCCATT GCGTTAAAGT AAGCCATTGG AATATTTCTA 420

TATATATTAC TTAATTTAAG TGAAATATAC ACACTTCCTC CTAAAGTATG AACTATTATA 480

ATTATGAACA ATTGATTATT CATAATATTA TGTTTATTTA GTTTTTCTAC TATTAATTCT 540

ACTCTGCTGG TTATTTTTTC TGTATAATCA ATATTTTCTT CTATACCACC AAAAATACCA 600

AAAGCTATAA GTAATTTATC ATTGGCAATA TTAAAAATTG CCTCCTTATA TTTTCTTTCA 660

TCCAGTAATT TTATAGCAAA ATTATTCTCA GCAATTTTTT CATGTTTCAA TCTCTTAAAA 720

GACTTTCTTA TATCATTCCA CATTTCTTCA AATACGTATA CTTCAGATAT TATTGTGGCA 780

ATTTTACTGG ATAATGCGTT AGCGACaGAT GTTATGTCTA ATATAAGTAC GCCTTCCCCA 840

GTATCCAGAT AAaGGCTTAG TTAAAAAGAG GATGTTGGTT TTGGGAAAAT GGTTTTTGCT 900

TTTCTAAATC AACATCTTCC TTCAnTATTG GATATTnCAT GTATTCCAGG TAATTnGTTA 960

ACTAAAAAGC ACCAAGTTTC CATAAAACC 989

(2) INFORMATION FOR SEQ ID NO: 619:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2013 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 619:

ATTGATCAGT GAATATTATG TTATGTAnCG nTTCTTTAAA CCAATTGCTC TTGATGAAAn 60

ACAAAATTTC TCTAGTGTTC AGCTATTCCT TAACCTTTTT ATATTTAATG CTTTATTCAT 120

GTTTGTTATG CTAACAAACT TTCATCAAGT TTTTGTAGGA AAAGAATTTA TTATTTTTGG 180

TAGTTTCATG GGTGGAATAT TTACGATTTG CTTAAAGGAA CAACTAAAAG AATACCAGAA 240

GCTAATGAAT CAATTTATAG ATATAAAGAA AGAAGTACAT TGATGAAAAA TAGAATTAGT 300

TAGAACAGCC ATGTATGGTG AAATATTGTA GTATATTACT TGTTAAGGAC GATGAACAGG 360

ATGAGTATCA AAAAATATTT ACTTTTTTAT ATTATTAGCA CATGTGCCAT TATTTTTGTT 420

TCTCTTACTT TTTTCTTTTT ACCTCTAAAA TCGATTGATT CTAGGAGTTA TGAAGTACTT 480

CATACGATAT CTCAATATGG AGTAATTTCT AATAAAGATG AGTGGAGGAA TATTTTAGAA 540

TTAACAAGAT ACGGTAGAAA AGTAACTAAT GTGGATAATT TAAATACATT AGTATATGGT 600

GTTAACAAAC ATAGTTCTGT AAAACAAATA AGTAGTGACA AAGATATGGA TAATATGGAA 660

ACTCTTAAGT TACCAAGTAT TTCAAAGTAT GAGGATCTTA CAATTATTAA CATTCCAAGT 720

GTCTATAGCA ATGATGATAA CTTCTCAATT AAATATGCAT CAAAGTTGAC AGATTTAATC 780

GAATATACTT CTGGTAACAT TGTTTTAAAT TTATCAAATA ATTATGGCGG GCTGAAAGAA 840

CCTATGATTA TTGGTGCGAG TTCTTTAATT CCTAATGGAA TGCTCTTTAG CAATATTAAT 900

AATAAAAAGG AAAAGTATCC TGTTTACTTA AAAAAATGGT AAATTATTTG GTGGCATTCC 960

AGGAACGATT AACGAACTGA ATCTTTATAA ATTATCATTA CACAAAAAGA AACTAGGGAA 1020

AAAAGTTGCA GTTGTAATAA ATAATAACAC AGCAAGTGCC GCTGAGAGTT TACTTTTAGC 1080

GTTAAAGAAT AATCCmAATG TAAAAGTTTT TGGTATGCCT TCTGCAGGAT ATACTTCGGT 1140

TAATyTAGGA CGTTTTTTTA CTAATCAAAA TTCCGATAAT TATTGGTGGt TTGTATATAC 1200

TGTTGGCTAT TATGAGACAA TAAAGCCCAT AAAGGGAAAA TGATTGTTTA ACAATCAACC 1260

TGTCCCGCCA GATTATTATG TAGATTTATC GATGTTAGAT AGAAATACTC AAGATATTAC 1320

ACAATTTTCT AATACCAAGA AACTATTCAA AGAAATAAAA AATTsGATTA ACTCTGATTG 1380

AGTTGAAGGT ACATTAAAGT AAATGAATTC TTAATGTATA GTATAACTTG TAAGCATAAA 1440

ATTTTTATAG GAGAAAGGCT TTTGCTAAAA TTATATTGTG GTAAAAGTCA CATGAAAGTG 1500

GGGGGGAGAA TCGATGACGA AAGGGAAAAA AATTTTGTTT AGTTTTATTT CTTTAGTTTT 1560

TGGAACTTTT TTATTTGCAG TATTTGTTCA TTATCAGTTC ATGTATTCAT TACCTATTTA 1620

TTTTATTAAG TTTGAAGAGT ATCCTTTTAT AGGGAATTAC ATTCAACTAT ATTTGTACTG 1680

GGGTAGTGTT ATAGGTATGG TCTTTATCCT TTTATTTATT TTAGGGATTG TTTTAAGTCC 1740

AGCAGAAGTA ACATCCGTGA AATTAAACGA TGACAAAGGA ATATTAGAAA TCAAAAAAAG 1800

TGCTATTGTA GGAATTGTTC AGTCGCAATT AGATCAGTCC AATTTATTAA GGGATTCAAA 1860

AGTAAATGTA AAAATGTATA AGAAAAAGAT AAAAGTACGT ATTACAGGAA ATACAAGTGA 1920

CAATCTAGAT ATTATAAATC AAACGAATCA ATTAGTAAAA AATATTGAGT TATATCTTAA 1980

AACTTTTATA GGATTAGATA CTTCAATAAA AGC 2013

(2) INFORMATION FOR SEQ ID NO: 620:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1676 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 620:

ATAAATTTCT GGTGTTnAAA AAGTTCCTTT AATTTCTTTA TAACCTAGTA TAGATAAnTC 60

ATCACTnATA TACGnATATT CAATATTAGG AATTTTTACA TTAGTTTCTA AATTTGTATT 120

TAAAAAATTA TATATTGCTT TTTCTTTTGC ATAACCTTTT TTCTTATTAG TACTAAATTT 180

TGTTTTAAAA ATGTATTCAT TATTAACTAA ATATGCCACA CTATCATAmC CACTACCGAT 240

TATTTCmATA CTATCTACTT TGAAATtATC aAAGTAATGC TCAATTAAAT ATTTCaTTGC 300

CTTAACATTT gTGGCATTAT CATCATATCT ATATTCCATT AAATAACAAT CTTCTTTTTT 360

GCCCTCGTGT AATTCATGTT CTGGCAAATC TTCAATAATT CTAAAACCAG ATTTTTGGTA 420

TGCCCTTATT GCTCTTGGAT TATTTTTATG AGGGTCTAAA ATAACTGCAT TAGCATTTCT 480

TTCTTTTTTC AAAAATTCAA AAATCAATTT AATATATCTT GTACCAATTC CTTTACTCCA 540

ATAATTTGGC TCTCCTATAA ATTGATCCAT ACCATAGACT ATCTCATCAG TTTTTGGATA 600

ATGATAATCA GTATATAACT CATCATACAT TTTATATATT TGTCCATATC CAATAGGAAC 660

ATTGTTATAT TCAATAATTA CTCTAAAAAC TTCATCTTCC CAAGGCTCTG TATAATGTTT 720

TTTTAATGAT TCTAATGTAT ATTTTTTATC TCTACCACCA TAAAATTCTA ATACTCTTTC 780

ATCAGTTAAC CATTTTAACA TCAAAGGaAA ATCATCATCT ATTAAAGTTC TTATACATAT 840

TTCATTTTCA ACTATATTCA TTTATTTATC ACCTTTTTCA TAATCATATA CATATACTAT 900

TTCATCTTTA TAATCATTTT TACCACCTAA TTTTTCATAT ACATGGCAAG CTCTAGGATT 960

ACCTTTATCA GTTATTAAAA ACATTTCAGA ACAACCAATC TCTTTAGAAT ATTCCTTAAT 1020

AAAAGATAAT AATTTTGAAC CATAACCTTT GTCTTGATAG TTAGGTAACA TTCCTATTGA 1080

GTGTAAATAA AACATTGTTT TTCCATCAGG nCTTAAAAGT GTATAGCAAT ATGCAAATCC 1140

TATAATTTTA TTATTTkCTT TAGCTATAAA CCCAAATGAA CTGGtATCAT TAAGAAATCC 1200

TTTTAAATTA TCAATATCAA AAACCATATT ATCATCAATT AAAACTTCTT CCATAAACTC 1260

AGTCAAGTCC AGACTCCTGT GTAAAATGCT ATACAATGTT TTTACCATTT CTACTTATCA 1320

AAATTGATGT ATTTTCTTGA AGAATAAATC CATTCATCAT GTAGGTCCAT AAGAACGGCT 1380

CCAATTAAGC GATTGGCTGA TGTTTGATTG GGGAAGATGC GAATAATCTT TTCTCTTCTG 1440

CGTACTTCTT GATTCAGTCG TTCAATTAGA TTGGTACTCT TTAGTCGATT GTGGGAATTT 1500

CCTTGTACGG tATATTGAAA GGCGTCTTCG AATCCATCAT CCAATGATGC GCAActTTTG 1560

AATATTTTGG kTGATCGATA TAATCATGAA TCAATCGATT TTAGCCTCAC GCGCnAAGTT 1620

AATACTGTGG ACTTAAAAAT CCCTAAnAGC GTCTCTGAAG AATTTGAATT GTTTTA 1676

(2) INFORMATION FOR SEQ ID NO: 621:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1635 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 621:

ATCCCCTTAG TATAAATACA AACTCAAAGG AGAGACAAAA CATGTCAATT ATTACTGATA 60

TTTACGCACG CGAAGTCTTA GACTCACGCG GTAACCCAAC AATCGAAGTA GAAGTTTACA 120

CTGAAAGCGG AGTTTCGGCC GTGGAATGGT TCCTTCAGGA GCTTCAACTs GTGAATACGA 180

AGCAGTTGAA TTACGTGATG GCGACAAAGC TCGTTACTTA GGTAAAGGTG TTACTAAAGC 240

TGTAGATAAC GTAAACAACA TCATTGCTGA AGCAATCATC GGTTACGATG TACGCGATCA 300

AATGGCTATC GATAAAGCAA TGATCGATTT AGATGGAACT CCTAACAAAG GTAAATTAGG 360

TGCTAACGCT ATTCTTGGTG TGTCTATCGC CGTAGCTCGT GCTGCTGCTG ATTACCTAGA 420

AGTGCCTTTA TATCACTACT TAGGCGGATT CAATACAAAA GTATTACCAA CTCCAATGAT 480

GAACATCATC AACGGtGGAT CACATGCTGA CAACAGTATT GACTTCCAAG AATTCATGAT 540

TATGCCTGTA GGTGCGCCTA CATTCAAAGA gCTTTACGTA TGGGGTGCAG AAGTATTCCA 600

CGCATTAGCT yCAaTCTTAA AAGGTCGCGG TTTAGCwACT TCAgTAGGTG ACGAAGGTGG 660

TTTCGCTCCT AACTTAGGTT CAAACGAAGA AGGTTTCGAA GTAATCATCG AAGCTATCGA 720

AAAAGCTGGC TATGTTCCTG GTAAAGACGT TGTTCTTGCT ATGGATGCTG CTTCTTCAGA 780

ATTCTACGAC AAAGAAAAAG GCGTTTACGT TTTAGCTGAC TCAGGCGAAG GCGAAAAAAC 840

AACTGAAGAA ATGATCGCGT TCTACGAAGA ATTAGTTTCT AAATACCCAA TCATTTCAAT 900

CGAAGACGGC TTAGACGAAA ACGACTGGGA CGGATTCAAA AAATTAACTG AAGTATTAGG 960

CGACAAAGTT CAATTAGTTG GTGACGATTT ATTCGTAACA AATACAACTA AATTAGCTGA 1020

AGGTATCGAA AAaGGTATCG CTAACTCAAT CCTAATCAAA GTTAaCCAAA TCGGTACTTT 1080

AACTGAAACA TTTGAAGCAA TCGAAATGGC TAAAGAAGCT GGCTACACTG CAGTTGTATC 1140

TCACCGTTCT GGTGAAACAG AAGATTCAAC AATCTCTGAT ATCGCTGTTG CAACAAACGC 1200

TGGCCAAATC AAAACTGGTT CATTAAGCCG TACTGACCGT ATTGCTAAAT ACAACCAATT 1260

ATTACGTATT GAAGATCAAT TAGGCGATGT AGCTGAATAC aAAGGCTTAA AATCTTTCTA 1320

CAACTTAAAA AACAAAaTAn TTCATTAAAT GgAaTAmCGC yTAAAGGGrC mATTGGTTAA 1380

TGAmCCAATT CTCCTTATAC TCGGACGTAA AAACcGCCTT TCTGCAAAGA AAGGCGTGnT 1440

nTTTTATATT TATAAAAATA TTGAAAAAAT AGATTAGAAA TAACTTCAAC CGTACATCGC 1500

TCTATAATCA CTTGTTTAAA AATAACAAAA CATTAACAAA ACAACATTAT ATAAAAATTT 1560

AATAAAACAA AATAATTGTT ATTTTATAAT TTTAGTATTG ACTGGTTTCA GATAACGCTG 1620

TACGCTTAAA GAAAT 1635

(2) INFORMATION FOR SEQ ID NO: 622:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8001 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 622:

AGAGGGTGCT TTGGTCCGTG TTCAAGCCTT TGCAACTATT AATGGAGAAG ATATTACCAC 60

TAGTTTAGCG GCGGACGCTT GAAATCTCTA AAATCAGTTG GTAGTAAAAA TCAATTGTCC 120

ATTTTACAAA TTCAAGAAGA AGTATCCAAA TATTATCATG TGCCTCTTAA AGATTTAAAA 180

GGGAAAAAAC GGGTGAAAAC AATTGTGGTT CCTCGTCAAA TCTCAATGTA TCTTGCTAGA 240

GAAATGACCG ACAATTCCTT GCCAAAAATT GGTGCCGAAT TTGGCGGCAA AGATCATACA 300

ACGGTGATTC ATGCGCACGA AAAGATTCAG CAATTATTGG AAAAAGATCC AGCGATTCAA 360

AAAGAAGTCA GTGAAATTAA AAATTTATTG AACTCTTAGG ATGTGGAAAA CTAAACAAAA 420

AAGAACAAAG TTATCCACAA CTTATGCACA AGTGGaTAAC TTAGTAAAAT CAATGCTCTT 480

TTGACTTTTC CACAGAATCA ACATGCCCTA TTACTTTTAT TATTTATTTA TATATAATAT 540

ACAAAACAGT GGCGTACAAA ATTgGAmCGT ATTGAAAATA CTTAAAAAGG AGTTTACATT 600

ATGaAATTAA CGGTCnAACG AAGTGtCTTT TTAcAAGAAT TACAAACTGT TCaACGAGCA 660

ATTtCTTCTA AAACAACGAT TCCcAATCTT AACAGGTGTT AAAATTGTGT TATCAGAAGA 720

TGGCTTATCA CTTACTGGGA GTAACGCGGA TATTTCAATT GAAAGTTTTT TAAGTAAAGA 780

TGATGAAAAA GCTCAAATGA CCATTGAACG CACAGGTAGC ATCGTTTTAC AATCTCGTTT 840

CTTTGGCGAA ATTATTCGTA AATTACCAGA AGATATGTTC ACAATGGAAG TTCTAGATAA 900

```
CAATCAAGTT GCAATTACTT CAGGAAAAGC TGATTTTACG GTTAATGGAT TAGATGCTGA    960
TAATTATCCG CACTTACCAG TAATTGATAC TCAAAACCAA ATGAAATTAC CTGTTCATTT   1020
ATTGACAAAA ATTATTAGTG AAACAGGTTT TGCTGTATCG ATGCACGAAA GTCGTCCAAT   1080
TTTAACTGGG GTTCACTTTA TTTTAGAAAA TCAAAAATTA CTTGCCGTTG CGACAGATTC   1140
ACATCGTTTA AGTCAACGTG TGATCCCGAC AGAACAAGCA GTAGAAGACT TTAACATTGT   1200
AATTCCAGGA AAAAGTTTAA CTGAACTTTC TCGTTCATTA ACCAATGAAG AAGAAATGGT   1260
TGAAATCAGC ATTATGGAAA ATCAAGTGCT ATTTAAAACA GAAACAATGT ACTTCTATTC   1320
TCGTTTGTTA GAAGGAAATT ATCCTGATAC CAACCGTTTA ATTCCAACTA GCCATAACAC   1380
GCAAATTGAA TTTTATGTAC CAGAATTGCT TTCAGCAATC GAACGTGCCT CTTTACTTTC   1440
TCATGAAGGA CGTAACAATA TTGTTCGCCT TTCAATTTCA CCAGATTCTG TTGTTTTATA   1500
TGGAAATTCA CCTGAAATTG GAAAAGTCGA AGAAGCTTTA AACTATGAAA ATGTTTCTGG   1560
TGAAGCGTTG GATATTTCTT TCAACCCAGA TTATATGAAA GATGCGTTGC GAGCGTTTGG   1620
CGATATGAAT ATTACCGTGA AATTCCTTTC TCCAATTCGT CCGTTTACAT TGGAGCCAAC   1680
CGAAACAGAA CTAGATTTCA TTCAACTAAT TACACCGGTT CGTACAAATT AATCATGTTT   1740
AAAAAAGACC TTCGGAAAAT CAATATGATC GCCGAAGTTC TTTTTTGTCG TCTGTTTAAA   1800
GCCGATAAAC GCTGCTAGAT AAGCTTTTTA TAACTAGAAT AGCTTTTACG CTAATATGAT   1860
TGTGAATGTT TGAACTATCG TTTGACGAAT AAATTCTGCC TTTTTTTCGT GGATAAAGAA   1920
AAAAGCCCTT TAATTTTTCG TTTAAGGGCT GTTCATTTTT TTATGAATAT TTCCTAGTGA   1980
AAGAAGTTTA CGTTGCTTAA AACGACGCTC AGACCTCTAC AAGCTACTTT TCTTCATTGC   2040
GCGCCCTCTT CGTTTGTCTT TGAGTTACCC CTAATTGTTT TTTAGTTAAA CGCTACTCGT   2100
CATTTCTGCC AAATCAGGCC CCTTTTTTCT CAAGAAAATA TCAATTTCCG TTTTAAAATG   2160
CTTTTTAAGA GTTTTTTTAT CCTAATAATA TTTTCGGTCA TTTAGAATTA AAATTGCTTA   2220
GAATGAAAAT TAGACAGATA AATTTGTTTT CTCTTATGAA AAAGGGTATA ATATATATGA   2280
GAACGTATTA GAAAAAAGAG GGATGCACAT TGAAAAAAAC GTTTGTTTTA GCAACTGAGT   2340
ATATAACGCT TGGTCAGTTT CTTAAAGAAA TCAACGTCAT TGGTAGTGGC GGCCAAGCCA   2400
AATGGTATTT GGCAGATAAT AGCGTGTTTG TTGATGGCGA GTTGGAAAAT CGTCGCGGAC   2460
GGAAACTTTA TGCGGGCATG ATGATTGAGA TACCTGAAGA AGGTACTTTT TTTATGGTGA   2520
AAAACGGAAA CGAGTCCGAC GATGCGGCTG AATGAGCTGA CACTTCAGCA TTATCGTAAT   2580
TATGAAACAG TCAGCTTAGA TTTTCCAAAA ACGTTAAATC TTTTTTTAGG CGAAAACGCC   2640
CAAGGAAAAA CCAACTTGTT AGAAAGCATT TATGTTCTAG CAATGACACG AAGTCATCGA   2700
ACCAGCAATG AAAAAGAATT AATTGGCTGG GAACAAGCAG CAGCTAAAAT TAGCGGTGTG   2760
GTTGAAAAAA AGACAGGCAC CGTGCCGTTG GAAATTTTAA TTTCCAACAA AGGACGTAAA   2820
ACCAAGGTTA ATCACATTGA GCAAAAAAGG CTCAGTGCGT ACATTGGGCA ACTGAATGTT   2880
```

ATTTTATTTG CCCCAGAAGA TTTATCTTTG GTGAAAGGCT CACCTCAAGT CCGCCGGAAA 2940

TTTATTGACA TGGAACTCGG ACAAGTGAGT CCTATTTATC TTTATGATCT TGTCCAATAT 3000

CAGTCTGTTT TAAAGCAACG GAACCAATAC CTAAAACAAC TAGCTGAAAA GAAACAAACA 3060

GATACTGTTT ATCTGGATAT TTTGACAGAG CAGCTAGCTG AATTTGGCGG GAAAGTTCTA 3120

TACGCTCGGT TAGGTTTTCT TAAAAAATTA GAACACTGGG CAAACTTACT CCATCAAAAA 3180

ATTAGTCATG GACGCGAAAC GCTCACCATT GACTACGCAT CAAGCATCCC AATTGACAAC 3240

ACGGATCTTT CCTTAGAAGC ACTTCAAAAT CAATTGCTCC AACAATTAAT GAACAATCGT 3300

AAGCGTGAAT TGTTTAAAGC GAATACTTTT TTAGGCCCTC ATCGTGACGA TTTGCTGTTT 3360

ATTGTCAATG GTCAAAATGT TCAAACGTAT GGCTCACAAG GCCAACAGCG AACGACAGCT 3420

TTAAGTATTA AATTGGCAGA GATTGATTTG ATGCATTCAG AAACAGGAGA ATATCCTGTG 3480

TTATTACTAG ATGATGTCAT GAGTGAATTA GATAATGAAC GTCAAATTCA CTTGTTAGAA 3540

ACGATTGAAG GCAAAGTTCA AACGTTCTTA ACGACAACTA GTTTAGATCA CATAAAAGAC 3600

AAGCTAACAG TTGAACCAGA TATTTTCTAT GTTCAACAAG GCAAGATAGA AAGGAATTCA 3660

GCGACATGAC AGAAGAAGAA AAAAACATGA GAGAACGTGC GCAAGAATAT GACGCTAGTC 3720

AGATTCAAGT TTTGGAGGGC TTAGAAGCTG TCCGGAAACG TCCTGGTATG TACATAGGTT 3780

CTACTAGTGG CGAAGGGTTA CACCACTTAG TATGGGAAAT CGTTGACAAC TCTATCGACG 3840

AAGCCTTAGC AGGATTTGCT AAATCAATTC AAGTAATTAT CGAACCAGAC GATAGTATTA 3900

CTGTTATCGA TGATGGTCGT GGAATTCCAG TCGGGATTCA AGCCAAAACA GGTCGCCCAG 3960

CTGTGGAAAC AGTGTTTACA GTGCTACATG CCGGCGGGAA ATTTGGCGGT GGCGGCTATA 4020

AAGTATCAGG TGGCTTGCAC GGTGTTGGTT CCTCTGTTGT TAATGCTTTA TCGACCAGTT 4080

TAGATGTTCG TGTTTATAAA GATGGCAAAG TGTATTACCA AGAATATCGT CGTGGTGCGG 4140

TTGTTGATGA CCTAAAAGTT ATCGAAGAAA CGGATCGTCA CGGAACCACT GTTCACTTTA 4200

TTCCAGATCC TGAAATTTTT ACAGAAACAA CCGTTTATGA TTTTGATAAA TTAGCAACAC 4260

GGGTTCGTGA ATTAGCCTTT TTAAATCGAG GCTTACACAT TTCAATTGAA GACCGTCGTG 4320

AAGGACAAGA AGATAAAAAA GAGTATCACT ATGAAGGCGG GATTAAGAGT TATGTTGAGC 4380

ATTTAAATGC CAATAAAGAC GTTATTTTCC CAGAGCCAAT CTTCATTGAA GGAGAGCAAC 4440

AAGATATTAC AGTGGAAGTG TCAATGCAGT ATACAGATGG CTACCACTCA AATATTTTAA 4500

GTTTTGCCAA CAATATTCAT ACTTATGAAG GCGGAACgCA TGaATCTGGT TTTAAAACTT 4560

CTTTAACACG TGTGATTAAT GACTATGCAC GTAAACAAAA ATTGATGAAA GAAAACGATG 4620

AAAAATTAAC TGGGGAAGAT GTTCGTGAAG GCTTAACCGC AGTGGTTTCC ATCAAACATC 4680

CCGATCCTCA ATTTGAAGGA CAAACGAAAA CGAAATTAGG GAACTCAGAA GTACGGACCG 4740

TTACGGATCG TCTATTCTCA GAATACTTTA CAAAATTTTT AATGGAAAAC CCAACAGTCG 4800

GCAAACAAAT TGTCGAAAAA GGCATGTTGG CTTCAAAAGC TCGTTTAGCC GCTAAAAGAG 4860

```
CCCGTGAAGT CACCCGTCGT AAAGGCGCTT TAGAAATTAG CAACTTGCCA GGGAAGCTGG    4920
CTGATTGCTC AAGTAAAGAT CCGGAAAAAT GCGAATTATT TATCGTCGAA GGAGACTCGG    4980
CCGGCGGTTC AGCAAAACAA GGACGTAGCC GTGAATTCCA AGCCATTTTG CCAATTCGTG    5040
GGAAAATCTT GAATGTTGAG AAAGCCAGCA TGGATAAAAT TTTAGCCAAT GAAGAAATTC    5100
GATCATTATT TACAGCTATG GGAACAGGCT TCGGCGAAGA TTTCGATGTT TCTAAAGCAC    5160
GCTATCACAA ACTAGTAATT ATGACCGATG CCGATGTCGA TGGTGCGCAT ATTCGAACGT    5220
TGTTATTAAC CTTGTTCTAT CGCTTCATGC GGCCAATTGT TGAAGCTGGT TATGTGTATA    5280
TTGCCCAACC ACCACTATAT GGTGTGAAAC AAGGGAAAAA CATCACCTAC GTTCAACCGG    5340
GCAAACATGC TGAAGAAGAA CTAGCAAAAG TGTTAGAAGA ATTACCTGCT TCACCAAAAC    5400
CAAGTGTTCA ACGTTACAAA GGGTTGGGTG AAATGGATGA TCACCAACTT TGGGAAACAA    5460
CAATGGACCC AGAAAAACGA TTAATGGCGC GAGTGAGTGT CGATGATGCC ATTGAAGCTG    5520
ACCAAATTTT TGAAATGCTA ATGGGCGATC GTGTGGAACC ACGTCGTGCG TTTATCGAAG    5580
AAAATGCCCA TTACGTGAAA AACTTGGATA TTTAATGTGT CGCAACGTTG TAAGAGACGC    5640
TAGAACATAT ATTTAGAGGG GAAAAAATTC ATGAGTGAAG AAATTAAAGA AAACATTCAA    5700
GACGTCAATC TGACCAGCGA AATGAAAGAA TCTTTCATTG ACTACGCAAT GAGTGTTATC    5760
GTAGCCCGCG CGTTACCTGA CGTTCGTGAC GGTTTAAAAC CTGTTCATCG CCGAATCTTA    5820
TATGGAATGA ACGAATTAGG GGTAACCCCT GATAAACCAC ACAAAAAATC AGCCCGGATT    5880
GTTGGGGATG TTATGGGTAA ATATCACCCC CATGGGGACA GTGCGATTTA CGAATCAATG    5940
GTGCGGATGG CACAACCTTT TAGTTATCGG GCTATGTTAG TTGACGGCCA CGGAAACTTC    6000
GGTTCCGTCG ATGGTGACGG CGCTGCCGCT ATGCGTTATA CCGAAGCACG TATGAGTAAA    6060
ATTGCTTTAG AAATGCTACG AGATATTAAC AAAAATACAG TCGATTTCCA AGGAAACTAT    6120
GATGATTCAG AACAAGAACC AGTAGTTTTA CCTGCTCGTT TTCCAAACTT ACTAGTTAAC    6180
GGAACAACGG GGATTGCGGT GGGGATGGCA ACAAATATTC CACCACACAA TTTAAGTGAA    6240
GTGATTGATG CAACAAGTTT ATTGATGGAC AATCCTGATG TAACGACGAA TGAATTGATG    6300
GAAGTGTTAC CTGGACCAGA TTTTCCAACA GGCGGTTTAG TGATGGGGAA ATCAGGGATT    6360
CGCCGAGCAT ATGAAACAGG GAAAGGTTCG ATTACTGTTC GTGCAAAAGT TGAATTGACT    6420
GAAATGCCGA ATGGAAAAGA ACGTATTTTA GTAACTGAAT TGCCTTATAT GGTGAATAAA    6480
GCCAAATTAA TCGAACGAAT TTCTGAATTA CACAGAGATA AACGAATTGA AGGAATTACT    6540
GATTTGCGGG ATGAATCTTC TCGTGAAGGC ATGCGGATTG TCATCGATGT TCGTCGAGAT    6600
GTGAGTGCCT CTGTTGTGTT AAACAACTTG TACAAAATGA CTGCCTTACA AACATCTTTT    6660
GGTTTTAACA TGTTAGCCAT CGAAAAAGGC GTACCGAAAA TTTTAAGCTT GAAACGTATT    6720
TTAGAAAACT ACGTCGAGCA CCAAAAAGAA GTCATTACTC GCCGGACGAT TTTTGACAAA    6780
AACAAAGCAG AAGCACGGGC GCATATTTTA GAAGGTCTAC GAATTGCCTT AGATCATATC    6840
```

GATGAAATCA TTGCCATTAT CCGTGGGTCA CAATCAGATG ACGAAGCAAA AGCAACGTTA 6900

ATTGAACGCT TTGAATTTTC AGATCGTCAA GCGCAAGCGA TTTTAGATAT GCGTTTACGC 6960

CGTTTAACAG GCTTAGAACG CGACAAAATT GAAAATGAAT ATCAAGAGCT ATTGAAATTC 7020

ATTGCGGACT TAGAGGACAT TTTAGCCCGC CCAGAACGCG TCATTGAAAT CATTAAAACA 7080

GAGTTAAATG ACGTTCGCAC AAAATTTGGT GATGCACGAC GCACAGAATT ATTAGTAGGT 7140

GAAGTCCTAA GTCTTGAAGA TGAAGATTTA ATCGAAGAAG AAGAAGTCGT GATTACATTA 7200

ACCAATAACG GCTACATTAA GCGGATGGCA AACTCCGAAT TCCGGGCGCA ACGCCGCGGC 7260

GGACGTGGTG TCCAAGGCAT GGGCGTTCAT GATGATGATT TCGTGAAAAA CCTAGTTTCT 7320

TGTTCAACAC ATGACACGTT ATTATTCTTT ACGAACACTG GAAAGTTTA CCGAGCAAAA 7380

GGTTACGAAA TCCCTGAATA CGGTAGAACA GCAAAAGGAA TTCCAGTGAT TAACTTACTG 7440

GGAATTGATT CTGCAGAAAA AATTCAAGCG ATTATTTCTG TTGAGGGCAA AGCGGAAGCA 7500

GGTAAATACT TGTTCTTCAC AACCTTAAAA GGAACCGTCA AACGGACAGC CGTAACAGCC 7560

TTTTCTAATA TCCGTAGTAA TGGATTAATC GCCATTAGCT TAAAAGAAGA TGATGAGTTA 7620

GTTAACGTAG TAACGACTAA TGGCAATCAG AAGATGATTA TCGGAACACA TGCAGGATAC 7680

TCTGTCACAT TTGATGAAAA TACTGTACGT GATATGGGCC GGACAGCATC AGGTGTTCGT 7740

GGAATCCGTC TCCGCGAAAA TGATTATGTG GTCGGCGCAG CGATTCTGGA TGAAAATAAA 7800

GAAGTCCTAG TCATTACTGA AAATGGTTAT GGTAAGCGTA CAAAAGCCTC TGAATATCCA 7860

GTTAAAGGAC GTGGCGGTAA AGGGATTAAG ACAGCAAATA TCACTGAGAA AAATGGTCCA 7920

TTAGCTGGTT TAACCACGGT CAATGGTGAT GAAGATATCT TATTGATTAC GAACAAAGGC 7980

GTCATTATCC GCTTTAACGT T 8001

(2) INFORMATION FOR SEQ ID NO: 623:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2508 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 623:

TTTATGATTT AACGATTATC GGCGGCGGTC CTGTTGGATT ATTTGCCGCT TTTTATGCGG 60

GAATTCGTAA AGCGAAAACG AAAATCATTG ATAGCCTGCC GCAATTAGGT GGTCAGCTAA 120

CAATGCTGTA TCCCGAAAAA TATATTTATG ATATTCCTGG ATTCCCTGCT ATTAAAGCCG 180

GTGAGTTAAT TGCCAATTTG GAAAAACAGA TGCAGCCATT TCAGCACGAT GTTTGCTTGG 240

AAGAAGAGGT CACACACCTT GCACAGGAAG CAGATGGACT TTTGCGCTTA GACACAACTA 300

AAGGCACACA TTATTCTAAA ACGGTTATCT TTGCGATTGG TAACGGCGCT TTCCAACCTC 360

GACGTTTAGC TATCGAAAtG TtGaAGCCTT TGaAGGTGAG TCcAATCCAT TATTATGTGA 420

```
CTGACATGAA AAAATTCGCT GGCAAAAAAG TTGCGATTGC TGGTGGCGGC GATTCCGCAA    480
TTGACTGGGC TTTAATGCTA GrAAACGTCG CAGAAGAAGT ATCTATTATC CATCGTCGCC    540
CACAATTTCG CGGCCACGAA CATAGCGTGG AACAACTGGA AAAATCTAGT GTTTCTATCA    600
GAACTCCCTA TATCATtAGT GATATTTtAA AGGAAAACGA AACGTTCACA GGCATCCAAT    660
TAACGGAGAC AAAAGGTGAC CAAACATTGG ATCTCCCTCT GGATGATTTA ATTATCAATT    720
ATGGCTTCAC TTCTTCTTTA ACACACCTCA AAGAATGGGG ATTAGACGTT TCTAGAAATG    780
CTATTAATGT TCATtCCGAT ATGTCAACTA ACATTCCTGG TGTATACGCT GTAGGGGATA    840
TCTGTTCCTA TGAGGGAAAA GTGAAATTAA TCGCTACAGG CTTTGGTGAG GCACCTACTG    900
CCGTAAATAA TGCTTTACAT TACTTACGGC CCGATGCTCG ACGTCAACCA GTTCATAGTA    960
CAAGTTTATT TGAAAACGGC GTACCTAAAT AAGGTTAGCA CTGGACTTTA GCAGTTTTTA   1020
AGCCAAGGAC AAAAATCTAA ATCGGATTTT TGTCCTTGGc TTTCTGCTTT ACTGAAAAGC   1080
CTTTCTTTTT CTCTTTTAAA TTTATTTTtG ATACAATAAA AAGTGACTnG GAGGAAGAAA   1140
GATGAACCAA CCAGTTATTT ATTTAGGTGA AGTATTCAAC GAAGAACAAC TGGAACAAGT   1200
GAAAGCTGTG GCACCCAATT ATCTAGTTAA AACATCTACC GATsATTTAT CATCTGCTGA   1260
AGAAGAGGCC ATCGAAATTA TGTTAGGCTG GCATAAAGAA ATTgGTCCGC GTTTaTTAGC   1320
ATCTGATACT AGTCACCTAA AAtGGATTCm GCTCATTTCT GCTGGTGCTG aTTATATGGA   1380
TTTTGATmAA TTAAGaGAAA AAGGCATTCT GTTGTCTAAT GGTAGCGGAA TCCATAGTGT   1440
TTCTATCTCA GAACATGTTT TAGGCGTTTT ATTAGCACAT ACCCGCGGAC TGCAAGAAAG   1500
TATCCAACAA CAAATGCaAC ACACATGGAA TCAAACAGCC CCTTCCTATC AACAACTTTC   1560
TGGGCAAAAA ATGTTAATCG TAGGCACCGG GCAAATCGGA CAACAATTAG CTAAATTTGC   1620
TAAAGGTTTA AATCTTCAAG TTTATGGCGT GAATACCTCT GGACATGTGA CGGAAGGCTT   1680
CATTGAATGC TATTCACAGA AGAATATGAG TAAAATTATC CATGAAATGT CCATTGTTGT   1740
GAACATTCTT CCTTTGACTG AAACAACAAA ACATTTATAC AATCAGGCAC TTTTTGAAAA   1800
AATGGCTCCT GAAACAATTT TCGTAAATGT TGGTCGAGGC GCCTCAGTAG CAACGAACGA   1860
TTTGATAACT GCCTTAAACA ATAAAACCAT TGCTTTTGCT GCCTTAGATG TCTTTGAAGA   1920
AGAACCTCTT CCCGAAGATA GCCCTTTATG GGAGATGCAG AATGTTTTAT TAACACCTCA   1980
TATTTCTGGC ATGACACCGA AGTTCAAATC AAAATTACTG GCGATTTTTA TTCCAAACTT   2040
AAATcAATTT GTGACGGATC AGACCCTCGT CAAAAACCAA GTTTCCTTAA AAAAAGGCTA   2100
TTAACTCAAA AAAGGCCTTC AACCATCTTT CAAGTAATTG TATGGTTGAA GGCTTTTTTT   2160
GCTATTTTTT ACCAGATAAC CTTGAAGTTA TCATTTGTCG ACGCTTGAAT GATTGGATGT   2220
TTTTTTAAAT ATTCTGCAAT CAGTTCCGTC ATATCAATTT GAATTTCTCG AATAATTTTT   2280
TCAGGTTGGA ACATTTGATA GTTTCCGCCA CCGACTGCTC GGTATTGATT TGTCACGATT   2340
TCCAAAACAT CTTCTGCTTG CACTGCTTTG CCATGATAAT TAAGTTTCGT CACTCTTTCA   2400
```

CCAAACGGTT GtCTTAAATC AAGCGTATAC TCAATGCCTT CcAwACcwwA yCATAAtnAT 2460

AGTAwTGTGG TTTTGGATCG ACATATTTCG GGTTGAAAAT TGGTTGGC 2508

(2) INFORMATION FOR SEQ ID NO: 624:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 763 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 624:

TTTAGGTTGT CTTTTCATGC CTAATTTTTA TATAGAATAC TTAGTAGAGG AGGAAGTTGA 60

TTAAATTATC AGAAAAAAAG TGTTTGGAAA AGTCGGATTG GTTGCTTAAA AATAGAAAGA 120

TATGGGGAAA AATATGATTA AAAAAATAAT AGTCGTTGTT GCTTTCATGC TTACAGGCTT 180

CTCGCTAACT GCGATGAGTG CATCTGCAGA AGAAATAACT GATTTATTTT TACAAAAAGA 240

AGTGACATAT TCTGGTGTAG AAGGAGGAAA AATTGGAGAG AATTGGAAAT ACCCTCAATT 300

TGTTGGCGAA AAAGCTGTCG ATGGCGATGA AACAACACGC TGGTCCGCTG ATAAgcAAGA 360

TGAACAATGG TTAATTGTTG ATTTGGGTGA AGTAAAAAAC ATTGGCGAAC TCGTTTTACA 420

GTTGCATGCA GAAAGTCCTG TGTATGAAAT CTTAGTTTCA ACAGATGGTG AAAGCTATCA 480

ATCTATCTTT AAAGAAGAGA ATGGGAAaGG TGGCCAACCT ACTAAAAAAT ATATTGATGG 540

CAaTAATGTm CAaGCACGTT TTGTAAAATA TCAGCAGATG AAAaTGTGsg CAACACACCA 600

AATAAGCAAT TTTACAGTTC AAGTATTATT TCGTTTGarG CATATGAnAA AAAACGACTG 660

GCCAGAAGCG ATTAAACTTT TTAACCGAGA ACCTGGACTA TTAGTGGAAA AAAGGAAAGC 720

CACCGCTAGC CTTTGGAAGT GTCCnCCAGC CGGGAGTAGG TTA 763

(2) INFORMATION FOR SEQ ID NO: 625:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 446 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 625:

AACCCACTCT GTGGTAATGC CAAAAATCCC AGAAATCGTT TCATTCTGGC CTGCGATGGA 60

TGCAGTGATC AATGATACGT ACAAAGGCAA TATCAAACCT GCGGATTATC AAGCAAAATT 120

AGATAAATTA GTCCAAGACA CATCAAAAGA AGCAAAAGAA TAAAAAGAGT AAAGAAATGG 180

GaCGAGCTTC AATCTGCTGT CCCATTTCCT CTATCTTTTT TATTTAAGTA GAAAGTGAGT 240

GAACGAAGAT GTTCAAGAAA AAGAAAGCCC AGACGTCATT TCGAGAGGTC TTTAAAAAAG 300

GCGACTTCGC TACAAAACTA TCTTTTTTAG TTATGGGTGC TGCAAACTTC GCCAACAAGC 360

AGTGGTTAaA GGGaATCmTC tTTTTAACaG cTGrAATTGG CtTATCyaTT GGGGTAATCC 420

GTAATnGGTT CCnCGCCATT AAAnGA 446

(2) INFORMATION FOR SEQ ID NO: 626:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1478 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 626:

AATTTTGGTT TGTTGGTGTT GAAACCAAGC ACTTTGTTCA AAAGGTGCTG ACAATGTTTG 60

TAAGTGATGC ACTTCGATCT GATTGGTTGC CTGcGGTTCT GTTGTGAAAT CATGGGCCCG 120

AATACCTATA GAATGACAAG AAATAGGAAC TTCTTGTGTA ACCATTAGTG TTTGTTGCCA 180

GCCAATCACT TGGACTCTGT GCGCATCTAT GCGTTTGACA GGCCAAATAT TTTTACAACC 240

AGTGAGTTTA GCTGCTTCAA TCGTTTGAGG TTGGTTAAAT AATTGAnTTG TTGCACCAAA 300

AAGTGTTTGG TTTTTTGTGA TTATCACTAA ACTCTGACAT AACTTATACA GTTCATCTaA 360

ACTATGACTC ACAATTAAGG CATGTTGATT TAAATTACTT AATCGCTGtT GAAGkTCaAT 420

TtGCaGCTCt TCCTTTAAGG GTGCATCTAA TGCGGAAAAT GGCTCATCTA ATAACAGGTA 480

AGCCGGCTGA GCCGCTAGCA TTCGCGCAAA AGCCACTCGC TGTCTTTGTC CTCCAGAAAG 540

TTGACTAGGA TATTGTTGTT GGACTTTAGT TAAGTGAAAA GATGCAAGCA ATTGAGTGAC 600

CAATTGTAAA TCTTTTGTCA CGCACGTCAG ATTTTCTACT ACCGACAAAT GAGGAAACAA 660

CGCATACTGT TGAAACAAAA GACCAATTTT TCGTTGTTGC GGCGACAAAT TGATCTTTTT 720

TTCAGTATCG AATAAGACTT GATCATTTAA AACAATCTGA CCTTTATCCG CAGTCTCAAC 780

ACCCGCAATA CATTTTAACA ACATACTCTT GCCACAACCA GAAGCACCTA mAATTCCTGT 840

TGTTTGAGkC TTTGTTTCAA ACTGAATTTG CAAAGTATGA CTTTTCAATT CTTTTTGAAT 900

ATCAACAATT AATCTCATTC CGATCTTCTC CAAGCGCGCT GaCTAAAAAA TTCCGTAAAA 960

AAtAAAATAA TCAGGCmAAG TAATAACATr ATTACAACAT ATTGATTGGC TAAACGCCAA 1020

TCCCCACTTG CTACAGCAGA GTAGATAGCT AAAGGAAGCG TTCTTGTTTT ACCAGCAATA 1080

TTCCCTGCTA ACATTGTCGT CGCTCCAAAT TCGCCTAACC CACGGGCAAA GGCTAACACG 1140

CCACCGGsCA AGAGACCGTT CATCGACAAA GGTAACGCAA TTTTTCTAAA ATATCGACTT 1200

TCCGAAAATC CGAGTGTCTG GGCCACAGCC AACAATTCAC TATCCATTTG TTCAAACGCC 1260

GTTAAGGCTG AACGtACtTA AAGGAAATGA AACAGCAACG GCAGCAATAA CCGTAGCTGG 1320

CCACGAAAAA ACmACTTGAa TGGCmAAAAA ATCmAGTwAA AACTGGCCCA CTGGTTGCTG 1380

AACACCmAAA aTAGTTAAGr GaAAAAAACC AAAAACAGTC GGCGGTAAAA CTAATGGCAA 1440

CGTAAATAAA CTATTCAAAA GAATTTTTAA TTTACGAT 1478

(2) INFORMATION FOR SEQ ID NO: 627:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1661 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 627:

CCTCTTTCGT AGAACGTACA GGAATTTCTT TAAATATTAC CGATGTTGGT TGGGCACCTT 60

TGGCCACCAT CACTTGGGGA TCTCCTTACA CCCTTTACTT TCTTTTAATT ATGGTGATTG 120

TCAATGTCAT TATGTTAATT CTAAATAAAA CAAACACATT AGATGTGGAT ATTTTCGACA 180

TTTGGCATCT ATCGATTGTT GGTTTATTTG CTATTTTTTG TGGCGCAAAC TTAATAATTG 240

CAACAATCTT AGTTATTTTC ATAGGTGTTT TAAAAATTAT CAATTCTGAC TTAATGAAAC 300

CAACGTTTAA CGaTTTACTT AACGCCCCAG ATACCAATCC TATGACTACC ACACATATGA 360

ACTATATGAT GAACCCAATC ATTATGGTTT TTGATAAAAT TTTTGATAAG CTTTTTCCTT 420

GGTTAGACAA ATACGATTTT GACGCTGCCA AATTGAATAG TAAAATTGGC TTTTGGGGAT 480

CAAAATTTGC CATTGGGATT TACTTAGGGA TTTTTGTGGG TCTTTTAGCT GGCCAAACAC 540

CCACGCAAAT TTTTTCTCTA TCCTTTACGG CTGCAGTCTG TTTGGAACTA TTCTCTTTAA 600

TTGGAACTTG GTTTATCGCT GCAGTTGAGC CTTTATCCCA AGGGATCACT GATTTTGCTT 660

CGAATAAATT AAAAGGACGT ACACTTAATA TTGGTTTAGA TTGGCCTTTC TTAGCTGGTC 720

GTGCTGAAAT CtGGGCAGCA GCTAACGTAT TGGCTCCAAT TATGCTATTA GAAGCCATTG 780

TCCTACCTGG CAATAAACTT TTACCACTTG GCGGAATTAT TGCAATGGGT GTAACCCCTG 840

CTCTACTAGT AGTAACAAGA GGTAAACTCA TCCGAATGAT TATCATTGGA GCAATCGAAT 900

TACCTnTGkT CtTATGGTCT GGTACCCTAA TTGCACCGTT CGTTACCGAA ACCGCAAAAA 960

AAGTTGGCGC TTTCCCAGCG GGTCTAAATA ATAATACTTT AATTTCACAT ACTACCATGG 1020

AAGGGCCAAT GGAAAAATTC TTAGGATACT TAGTCGGCAA TGCATCACAA GGTCAACTCG 1080

AATTTATTTT ATACGCATGT CTGGCGCTAG TCGCCTATCT ACTGATTTTC ATTTGGTATG 1140

CAAAACAAAT GAAGAAGCGA AATATTATTT ATGCTCAAAA AGCTCAATAA GGGGArcTAA 1200

AAAGAAGAGG TGCATCTGCA CCTCTTCTTT TTAGTAGATC ACCCTGACAA CTTTTTCCAA 1260

TACTTTAGAT GTGTCAAATT TATCTTTATT TGTAACTACC GCTGTCACTT CTTCTAATTG 1320

ATAAAAAGAA TAAAAATCCT GTTTCTCTAC TTTAGAAGTA TCCATTAACA AATATTTTTC 1380

TATTGAGTTA TCTAATGCAA TTGCTTGTGT TTGTCCTTCT TCAATAGTGG CTGTCATAAT 1440

TTTATTATCT TTTAACGCAT TGCAACTAAA AAATGCTTTA TGGAAATGCA TTTCTTCTAA 1500

TGCTTTATTA GCGATTTCTC CAAAAAAAGA TTTAGTTAAA AGTCGCATTT CTCCACCAAC 1560

CAAATAAATG GTCAACGCAT TCTTTTTTTT ACTTAACGTT TCAAACACAG GCAAACAGTT 1620

TGTTACTACT CTTAAATGTT TAAATTCCaT TGcCTCGGCC A 1661

(2) INFORMATION FOR SEQ ID NO: 628:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 628:

TCAGTTCCAG TTGGGTTTCG TTATGGCATC TTGAACAAAA GTAACACACC ACAAACTATC 60

CATTTAAATG CGAGAAATTA TCAAACTCAC ACGAACGTTG TCACGGATGG TTTCTATACA 120

CGTTTGAGGG ATGATCGCAC AAAaGATAAC GGTTGGAAAC TAACTGCGCA GTTATCTGAT 180

TTTAGCGATG AGACAAATTC AACGGTGTTA ACGAATTCTG GGATAGCTTT AAAAATGGAA 240

AATATGAAAA TTGAATCTAT AAAAAATCGA GATACTCCAC AAGAGAGTAT TGATCAAAAT 300

CCAACAGGAA CACCTTCAAC TGTTaGCACs AATGAAACGC TGATATCAGG ACAAGCAGCr 360

AArACATTGA TyAATGCACA GGCAAATGAr GGyCACGGTA CGTGGCAACT AAGGATTCCT 420

TTTGATAAAG TGTCCTTGAC TGTACCAGCG AATACAGGGG AAATAAATAA AAATTATACA 480

GCGACTTTGA CATGGTCGTT AGATGATACA CCATAGGAGG AGAGTTCTAT TGATACAAGT 540

ATGGATTGGA TACCTTTGTT TTCTTGCAGC AAGTTCTTTT ATAGGTTTTC TCTATTTCTT 600

TTTGAAAGAT AGTCGATTGC GTGTGGACAA CGGTAAAAAG TCAACGGAAT TACTAGTAGA 660

TTGTGGTTTG ATGTTTATTT CTTTATTATC AATAGTTGCA TCTGTTCTAT TATATCTCAA 720

CTGGCAAGAA CAGTTACAGT ATTTTCTAAA CTGATAAAAA GAAGTAACTT CAATGACGAA 780

GGAAAGATGT GTCATTTGAA GATAGTCAAC TTACAATTAT CTAATGGCAA AAAAGCTATT 840

GCATCATGGT TTCAAATTAA TCAATTAAAA GCTCTTTAAG AATTATTGAT TTAAAGGAAT 900

ATACATCTTA AAGGGGCATA TCAAGTTGAT TATCGTATAG TGTTACTTCA TAAGAGAAAC 960

TAATATTGAT CTTAATACAG AATAGTCTTT GATATTCAAA TATTTAGTAA CTGTGTTGTA 1020

TTTATAGTAA ATACTCATTA GGAGCTGCAC CAGTTGCTTA ATTAGTTGGC AATGAGGAAC 1080

TATTAAGAAT TGTTTAACGA ATTTAACTTA GTTAAAAGGG GAAAATAGTT AATGAAAAAG 1140

CGCGTAGTTA TGACTATTGC GGGGATTATG GGAGGAACCT TTACCTCGAG TGCTGTTGCA 1200

TTAGCACAAA CAAAAACAAG TGTGGAAGAA AGGGGTCCAG TACAAGAGAG TACGATTTCA 1260

ATACCAGACA GTGAGATATT GGCGGAGGAA AAGGAGGTAG CAGCAACAAT AGAGTCTACT 1320

ACCACGACGT CCACCATAAA GACAACCAGC GATAGCCTCG AAAAAGCCCT TGAGGCAACT 1380

ACCTCAGAGA CTTCTTCGAC AGAATCGGGG AGCGCAAGTC AGTCAAGCGC TTCACAACAT 1440

AGGAAAGAAA TAGAAACAGA GGAAACTTCT CTAAAAGCAC AAACTACATC AACGGTTTTA 1500

ACAGGGATGA TAGGAACGAG TCAATGGAGT TTAGATCCTA CATCTGGAGA ATTAACTATA 1560

GGCCCAGGAG TTTTTAGTGA TGATGTATCA TCTGCGTTCG ACTCCATATC AAATGTTAAA 1620

AAGGTAACAT TTTCTGGACC GGTGAAAGCA GGAAAAAGTT TAAAAAGTTT ATTCAGTAGC 1680

TGGTTTGCGC TATTGTCCAT TGAACGTCTG GATTATCTGG ACACGAGTCA AGTGACAGAT 1740

ATGAGTTATA TGTTTACTAA TTGTACTGGA TTACAAACGT TAGATGTATC AAATTTTGAT 1800

ACGAGTCAAG TGACTAATAT GAGTATGATG TTTTATAATT GTCGTGGATT ACAAACATTA 1860

GATGTATCAA ATTTTGAAAC GAGTCAAGTG ACAGATATGA GTTATATGTT TGATGCTTGT 1920

AGCGGATTAC AAACATTAGA TGTATCAAAT TTTGAAACGA GTCAAGTGAC AGATATGAGT 1980

AGTATGTTTG ATGCTTGTAG CGGATTACAA ACGTTAGATG TGTCGAATTT TGATACGAGT 2040

CAAGTGACTA ATATGAGGGG TATGTTTTTT CTTTGTACTG GATTACAAAC GTTAGATGTA 2100

TCAAATTTTG ACACCAGTCA AGTGACAGAT ATGAATAATA TGTTTGCTGG GTGTAAGGCT 2160

TTACAGACAC TAGATTTGTC TAATTTTAAT ACAAATCAAG TGACAGATAT GGCTGGGATG 2220

TTTTCGTCTT GTGATAAAAT CCAAACAGTA AATGTGTCAA ATTTTGACAC CAGTCAAGTG 2280

ACAAATATGg GGGTATGTtT TCTCATtGTt GGGAaTTACA AGCATTAGaT GTAcTAAGTT 2340

gtACATCCAA AGGAnGTTG 2359

(2) INFORMATION FOR SEQ ID NO: 629:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1130 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 629:

CTCCACAAGA CTGACATTGG TACTGATGGA AAACTGCTTT TTTCCGTTCT ACTAAAGGCG 60

GTGCATAACG GGCACCGCCA GTCTTAGCCA ATAACTCTTT AAAGTCCTTA TCTTTATGCT 120

GATAGCCTTT GCCAGCTAAA TGTAAATGAT AGTGACATAA CTCATGCTTA ATGACTTTTT 180

CCAGTTCCTC TTGTCCATAC TTTAAAAAGA CAGTTGGATT AAAATCCAGG TGATGCGAAn 240

TAAATGATAG CGACCGCCCG TAGTTTTTAA CCGCCGATTA AATGTAGCTT GATGCAGAAA 300

AGGkTTCTGA AAAAATACGA TTGAAATTTT TTCCACTAGC GTTTGAAGCG CTTGATCTGT 360

CATCAAGACT GGTTGAGACT CTTTCATTGT TACTCTTTTC TATCAGCCCG AGGCAACATG 420

GTTAAACTAA TACGCCCCTT TTTCAAGTCA ACATCTTCTA CCCAAACTGT GACTACATCC 480

CCCACCGCAA CAACATCCGT TGGGTGCTTA ACAAATTTAG TACTTAACTT AGAAATGTGG 540

ACTAAGCCAT CTTGCTTCAC ACCAATATCG ACAAACGCAC CAAAGTCAAT TACGTTTCGA 600

ACAGTCCCTT GCATTTCCAT TCCTGGTTTC AAATCTTCCA TCGTCAACAC GTCTTTACGT 660

AAAAGTGGTG CTGACATCTC ATCGCGCATA TCACGACCTG GTTGGATTAA ACTAGCAATA 720

ATATCCGTGA GTGTTTCTGA ACCTACTTCT AACTCAGTAG CTAATTGATT CACTGGkAAC 780

TGtTTCAATT TTTCGATAGC TTCTGCTGTC CCTAATTCTT TAGAAGCAAT CTCTGCTTTT 840

TCTAAAATTG CTTGTGCTAC TGGATAACTT TCTGGGTGAA TGCCAGTGTT ATCTAACACA 900

TTTTTGCCAT TAGGAATTCG TAAAAAGCCA ATTGCTGTTC AAATGCTTTG GnCCTAAACG 960

AGTACTTTTT GACTGATTCG TGCTGTAAGA CACCATTTCA CACGATAAGG ACAAGTTGCG 1020

CGTGGTTTAT TAAGCAGAAT AGTGGTAAAG TGAGACCGCG GATCACATAC ACCACTGGGT 1080

ACGAGTTCAC ACAAATCCAT GCCGCAGCGT TGCGAACGCA GCGAATGACG 1130

(2) INFORMATION FOR SEQ ID NO: 630:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2132 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 630:

CACTACAAAT GAAnAnGAAA GAAAAACAnG CCAATTAAAG AAAAACAAGA AGAATCTACA 60

AmAAATATCT ACGCAAAATC CAaGCGGATT ATCTACCAAG GAmACAAAAA TATGAAACCT 120

ATAACCGATT TTTTCAAGAG CGAAATAGTT TTACTAAATC AGATACAGAT GCCACATTTA 180

TGCGAATGAA AGATGATTAT ATGCGCAATA GTCAACTAAA GCCAGGCTAC AATTTACAAA 240

TCGCAACAGA AAATCAATAT GTCCTAGCGT ATGATCTGTT CCCAAATCCA ACAGATATAA 300

AAAACTTACA GACATTTTTA GACAGCTTTT TGGAACAACA CAACGAATTA CCCGAGTATA 360

TTGTAGCGGA TACCGGCTAT GGAAGTGAAG AAAATTATAC GTATATCAAC GATGTCTTAC 420

ATAAAACGCC ATTAATCACT TATGCTAGTT ATCGTAAAGA AAATAAAAAG ACCTATAGAA 480

ACAGTCCTTT TATGGTTGAT AACTGGCGCT ATTTGGAAGA AGATACATAT ATTTTTCCTA 540

CCCATCGGGC AGTTCCTTTC AAGAGATATA GTCGAGGAAA AGAGAAAGGC GGATTTGTAC 600

GTGACTTCAA AATTTATGAA TGTGaAAATT GTCGAGATTG TCCTGTTCGT AGTCAATGTA 660

CGAAaGCAAA AAGTGaGCAA AACCGACAAA TTTTAGTGAa TAtACATGAC GCTAtTTCAA 720

GCAGAATGTA AAAAGAAGCT TTTAGAGGAA AAGACAGGTT CGATTTATCG GAAACGAAAA 780

ATCAATATCG AACCAGTGTT TGGCCATCTA AAGGCTCATT TGGTGTTCCA ACATTTCCAC 840

TTACGAGGGA AACAAGGAGC CGAGATTGAC ATTGGATTGG CACTCATGGA ATTAAATTTA 900

AGGAAATTAG GGAAATAATT TCTGAGTATT GTACCGGATT AGATGCCCAT CAAAATAATG 960

TTGTGTTTTG AAAGGTCCAT TTATTCAAAT AAATAGAATG TTGGAGTATA ACTATATTGC 1020

TTATGTAATA TAATTTATTT TTATTGTATA TTTGTTCCCT TTTCGTACTT TTCAACTTTT 1080

ACTACCrGrG TwTmCTAAAa TATGCTATAT TTTTACTACA GTAGTTATAA aTTTGAGaTA 1140

AAAAAGGAGT TTTTTTTTGG ATAAAGGAGT CGCTGATATT GTAAAAATTA AGCAAGTGTT 1200

GAAACAAGAA TCAGTAAAAA GCTTAGTAGA AGGAACTGGG TTAAGCAAGA GTACTATTTC 1260

AAGTTTAAAA TCAGGTACTC GAAAAGTAGA AAAACTAAAC CTTTCTGCAG CTATTAAACT 1320

GACAGAATAT TCAGATCAAG TTTTTAGACC TATTATTGAA ATCTGGGGAG AAAAaCCGAA 1380

AAAaTAACCa CAGTTATTTT ATGAAACTAA TAGGATTAAA AATAGAAGAA CTAGAAGTTT 1440

TTTAGATGAA TGCTTGTTTA GCTCTATAAT GGATAGAACT GAAAGATAAT ATTCTTCAGT 1500

CCTACCATTA TAGAGCTATA ATTTTATTCA ATTTCAAGAA GTCTTTTTTT TACCAATAAT 1560

TTAAGATATT TCATAAGATC TTCTTCAAAT TCTATAGATT CCCTATCTTG TTGGTACAAA 1620

GATAAATAAA GAAATATTTT TTTTCGATAG AATACGTTAA AATATGAATA GATATAGATA 1680

GTAATTATAT TATCTATAAA TAGTAGAGTA TAACGATCTT TTATTTTTGG ATTTTCTATA 1740

AATTTTAAGT AGTAAGAAAA TCTTTTTCGT TCAAACTTTT CTATAATCTC TAAATTTTTA 1800

ATTTGAACAG AATTAGTTGA AATAAGCATA TAAAAATTTA ATAGTAATTG CTCCTTATCA 1860

GATTTTAGAC GTACTCTTTC AATTATATTC ATGATATATT CATCGATGGT AGAGCTTTTA 1920

TCAGCAATTT TTTCTAATTC AGAGTTTATT ATATCCAAAT TATACACAAT CACTGCCTCA 1980

TATAAATCAT GTTTTGTTTT AAAAAAGCTA TATACGGTAG TAGTGCTGGT TTTAGCTTCA 2040

TTAGCAATAT CTAAAAGTTT TGTTTTTyCA TAACCAAATT TAGaAAAgTG GTTCCATTGC 2100

GGGTAAATAT AATTTCTnCC nTTTTnAATA TC 2132

(2) INFORMATION FOR SEQ ID NO: 631:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1610 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 631:

nTnAGGTGGA AGGCCCAACG GATTACTGGT TAGCCCTGTA CATGGTTAAT ACTAATAGAT 60

ATGCTTGATT GTAAATTTGT AGTGAAGTAC AATTTATTTT AAGAGTATAT TTGTAATATT 120

CTGTATAAGA AATTATTTAA GGAAGGAACA AATTTAATGA AGAAGAAAAA AATGATGAAG 180

AACTTAAATT TTGTAATCTT ACCTTTGTTA TTAGGGGTTT GTATTTCATT AATATGGAAT 240

ATACCATTTT TTGCkGTTTC kGGrGCTATT TACTTTGTAC TTTTAGTTTT TTTGGTTCCG 300

TCAGTTGACT TTGCCTTTAC AGATTTCAAT ACTATACGAA TAAaTCCGCA TTACAGAGGA 360

AGAAGAAAGA TAATTTTATC TAATGATACA TTGACATTAG TATTAGTTTT AATAGCACTA 420

ATAGTATCTG TGGTACTATC ATATTTTTAT TATAGATAAG AAGTATAATG GTAAGAmGGG 480

ATGAAAAAAA TCAAATATAA AACCATAAGG ATAGATAAAT ATTTATArGG TCAAAAACTT 540

TTAAGTTTTT GACCTCTTTA TTTTTTCGTG CTATACTTGG AATTATCAAG TGTCAAAAAA 600

TTATGTTCAG AAAAGGTGAG GAAATTTTGG CCATTTATGG TTATTCAAGA GTGAGTACTA 660

GTAATCAAGA TTACAAAACC CAGATTCAAA AACTAGAAAA AGCCGGAGCA GAAAAAATAT 720

TCAGTGAAAA ATATACGGGA ATAAAAAAAG AAGGAAGAAA AGAATTAGAA GAACTACTAT 780

CTATTGTAAA ACCTGGCGAT AAAGTTTTAG TGACGAAAAT TGATCGGCTA GCTAGAAGTA 840

TCGTTGATTT AAATTCGATT ATCTCTTCTT TAAATAATTC AGGAGTGACT ATCmCCTTCT 900

TAGwTAaTGC CCTAACCTTT GAcCAAATAA AaGTGATtCT ATGCAAACGT TAATGATGAA 960

TATGCTGGGT AGCTTTGCGC AATTCGAAAG AGATTTAATC GTGACGAGAA CGCAAGAAGG 1020

GAAACAGTGG CATCGAGCGA ACAATAAAAA TTATCGAGAA GGTCGTCCGA AAAGAGTTCT 1080

AAATGATAAA TACAAACATG CACTAGAGTT AATGGAAACA AACmGCmTGC gAGaAGTGGg 1140

AAAGGcAAAC AGGTATTTCT TTATCTACTC TCAAACGAAT CaAGAAACag CCAAgGAAGA 1200

ACmGTTACTT AGTGrGrAAT AATTCCAGGA GTACTTATAA TGGAAAATAG AGAGAAAATT 1260

ATTCArTTGT TCAAgAATCC TTTAGTAACA GGTTATGGGA TTGAGATAAT GTCAAACGGG 1320

CGACTTTATT CAGCAAACTT CCAAAGATAT AAGAATCGAG yAAAGAAAGA AGAGAACCCr 1380

TTGATTATTT TTGAGAGCAT GACAGAAAAA GTTGAGCAGG TATTTTTAGA ATTAGCTGAA 1440

GAAGTCATAC GAACGAACCC TAAAACAAAA CAAGAATTCA AkGAAATGAT TArAGAATAT 1500

AGTTATAAGG AGrATArCAA ATGrTAGyTC GAAAAAmATA TGATCATTTT GGTATTGAAA 1560

TTGGTATGTG GAACCGAGAC AATGTAGTAA ATAAAATAGA GTGTGATTGT 1610

(2) INFORMATION FOR SEQ ID NO: 632:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 626 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 632:

TCAACAAATT ATTGATTATT TACAAAATGA AATTGATTCA GGGCGATTGC CTGTTGGAGC 60

GCAAGTGCCT ACCGAAAAAG AATTATCGGC TCAGTTTAAC GTCAGTAGAA TTACGTCCAA 120

ACGAGCGCTA ACTGAGTTGG AAACGCAAGG CGTAATCCAA CGTCACCAAG GAAAAGGAAG 180

TTTTGTTCAA GCACCAAAGC ATCCTCTTCA TACCTCTTTA AATAAAGTTC TTTTTTTATT 240

GCCTTTTGCA GACGATTTAT CAGTAGGAAA TTTCTATAGC GGTTTGGCCC CTACTATTCA 300

AGCAGCTGGT TTAGAAGTCT TTATGACCTC GCCTAATTTT TTGAGAGAGA AAAATGCTGC 360

GGACATTGTG AATGAGTTTG CCGGGCTTGT CTATTATGCT GCTACTAGTA ATGACTATCT 420

CGACTTACTG TTCGAGTTGG CTTkGATGAA TTTTCCAGTC GTTGyCTtGG ATAAAAAGAT 480

CCATGATTTA CCTTTTGCAA CCATTCAATC AGACAACTTT GCTGGTGGTA AAACAGCAAC 540

CGAACGCTTA ATTGGCTTAG GACATACAAA AATTGGATAT ATCCTTTCAG GAAGTCCAGC 600

CCTACAATCC GTTCGGCAAC GGTATT 626

(2) INFORMATION FOR SEQ ID NO: 633:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5678 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 633:

ACTCAGCACG TCAAAACATG CACAAATTTA GTGGCCAATG GTCTGTGTTT GCCAaTTCTA 60

ATTTTGTTTT AGCAGmCGGA aCCAAAGTGA ATaAAGGGGm CGACATTTCG AAaCTATTCA 120

CGATGACCTT TGAACAAGGG GTAGTTAAAA TCACGGCCAG TCAAGCGTTT TTGGATGCGA 180

TGAATCTAAA AGAAAACAAA CACGTTGCGC ACTCATGGAA AGCGTTCATT GGTGTAGAAC 240

GAATTGCGGC AGGAGACGTT TACAACACAA TCGAAGAATC TTTCAACAAT GAGAAGATTA 300

AAACTAATAC GGTAGTCACA CATACGCCAG AAAAACCACA AACACCACCA GAAAAAACAG 360

TGATTGTACC ACCAACACCA AAAACACCAC AAGCACCAGT AGAGCCATTA GTGGTAGAAA 420

AAGCAAGTGT GGCACCCGAA CTACCTCATA CAGGTGAAAA AGAAAACACC CTATTATCTG 480

TACTAGGTGC CGGAATGCTA GTTGGTCTGG CTTGGTTTGG TTTGAAAAAA CGTGAAGTGA 540

AATAATGAGT AAAAAGAAAA AGCACTGTGA AAAATCAGTG CTTTTTGTTT ATTTTAGAAA 600

AAGTGTGGTG TTGTCAAGTG TGGATAATAA TTTTTGCGTT AAGTTGTATC GTAGTATCTA 660

TAATAATTGG AGCATTTGTC AGTGCAACGG TCGATGAACG ATTGGGAGAC AATATTTTTT 720

TGCTACTTCT AGTTATTTTA TTTTTTTGGT TTATTTTGAT TTTTAATTGT TAAGTATATT 780

TTTTTATATG CTATTTTAAT ATATTTAATA AGACTAGGTG AAAGTATTAT TAATTCAGAG 840

ATAGCTGCTA TTGTAGCTCC GAAATAAATA TTGGTCATAC TTTTTGAAAT AAACGACATT 900

ATTATAGAAG GTAGCATTAT AAATATTGTT GGCAAAAAGG CAAAATAATA AGTTTCTTTG 960

GTGATAGATT CGTTAAATTC TTTGATTTTT TGAAAGAAAG ATAGTTGTTT ATGATAGAAT 1020

ATCTTAATTT TCTGTATCTC GTTTAAAGAA TAAATTAAGC GCACAAATTC TTTATTTAGG 1080

TAATAAGTTA GAAGAACGGT ATCTTTGTTA ATCATTGAAA CTAAGAAAAC TTTTGTATTA 1140

CATGGGATAT CATTTAAAGA AAGTTCTTGG GACAGTACGT AAAAGTTACT ATGCTTACTT 1200

CTTGAATTAG AAGGTATTAA GGTTCTATAA ATATCAGGGA GCTTTGCTAT TGAGAAAATA 1260

AATACAATAA AATAAATGAT AAAGTAAGTT AAGTAAGCTT TACCCTCTTT AGTAAGGTTA 1320

TTTATATCTG AAACAAAATC GGTGAATGCT GTGTCTGTAG TTATTAATGC ATAAGCAAAA 1380

ATACCAGATG AAAGTAGTAA AGAGGATATT AGATTAATAA AACCTTTAAA GATTTTATAT 1440

AATCCGATTT CAGTGGGAGA GGAAAAAAGG TTTTTTTTAA ATGAATCACT TATTACGTAT 1500

CCATAGATGG AGGAATAAAT TGTGAAAAAG CCTATAACTA TAGGTAAAAG TATACCCACT 1560

GGTTTTATTA ATTGGTTAAA AAAATCAATG TTCATATAAA TAATTACTCC TTATAAACTT 1620

TTTATAATAA TTATATCAAA AGGACGTGAA TATATGTTAA TAAAATTAGG TACAGAATTG 1680

ACGAAAGAAG AATATGTTAC AAGGTATATG CGAAACTTTC AGAAATTATT GTTGCTCGGC 1740

GATCGGCCAA AAGTATTAAC GAATCGTGAA GAACAACTCT TGCAGTATGA AAAAGAATTA 1800

TGTGTGCTGT TTTATGAACA ATTTATAAAG AAACATCATA GAGCGCCTGA TGAAGCCACG 1860

CTAGACGATC AAGTTAAAGC GAATTTCATA GAGCGCTCTA AAATTTTCGC TCGATCTCCC 1920

TTAGTAATGG ACGAAGGAAA TTTTACGCAA GCACATATTG GTCAACTAAA ACGGTTGCGT 1980

GAATTGCGCA TGGAAGACTA TTTGCCAGAC AATTATACGC ATATTTTACA ACGAGAAGAA 2040

GAATTGGCAC GTAACTATTT TCGAAAACAT GATGATTATC CATTTGGGTA TGAGTGTTTA 2100

TGTATTAGTC GGTCACGAGA AGTGGTGAAC CAAGGGCTAG AAAAGCTCTT AGAGGGGTTT 2160

TATGACAGTT ATCAGGTTTA TTATCGTCGG TATCGTAAAA ATGGTTAGAA ATGGAAGTAG 2220

GAATAAAAAG CGCAGGAGGT GTAGCGAGCA ATGGCAGAAG AATCCAAATA TCCGCCAGTG 2280

AAAGAAAAAT CTGTGTTTTA TCAGCGTTAT AGTCGTGTTT CGACGATTTC TATGTTGATG 2340

GATTTAGGTT TTAGTGTGAG TGAAGCCCAG ATTGATAATG AAGATATTGT TTATATGGAA 2400

CAGTATAAAG AGTTTTTATT AAATAAAGGA ATTGTGGAAG AAAAAGTTCC TTTTAAACGA 2460

GAATTTGTAG CGGAGTATCC ATTAAGCAAA ATGCGACGAC TTTTTGTAGA AGGCTATTAT 2520

TTACGAAAAC CGAGTGGTGA AAAAATGTAT GGTAGCTACT ATTACGATTT TTATTTTAAA 2580

GAGTATTCAG GAATGGGGAT TTATTGTACG AAAGCAACAA AATCTCTTTT TTTACATGTT 2640

TTTGAACAAA GAATTAGAAA AATAAAAAAA TTAAAACATA GATTGGAAGA ATGAAAATGA 2700

AAAAAATGAT TATTATTGCC TTATTCAGTA CAAGCCTTTT AGCAGGGGGA AGCAGTGTTT 2760

CTGCTTATGC GCAAGAATCA GAAGGAAATC TTGGTGAAAC AACAGGGAGT GTTTTACCAG 2820

ATGAACCGAA TGTACCAACT GACCCAATAA CGCCAAGTGA GCCAGAGCAA CCAACAGAGC 2880

CAAGTACACC AGAGCAACCA TCGGAACCGT CAACACCAAC CGAACCTAGT GAGCCTTCAA 2940

AACCGACGGA TCCTTCGTTA CCAGACGAAC CGAGCGTACC AACAGAGCCA ACAACGCCAA 3000

GTAAGCCAGA GCAACCAACA GAGCCAACAA CGCCAAGTGT ACCAGAGCAA CCAACAGAGC 3060

CAAGTGTACC AGAAAAACCA GTAGAACCAA ATAAACCAAC CGAGCCAGAA AAGCCTGTGC 3120

CAGTTGTTCC TGAAAAACCA GTTGTACCAC AACAACCAGA GCAACCAACA GATGTGGTGG 3180

TAAAGCCAAA TGGAGAAATT GCAACAGGAG AATCTACACA ACAGCCAACT GTTCCAATTG 3240

AAACGAATAA CCTTTCAGAA GTAACACATG TCCCAACTGT GACGACACCG ATTGAAACAG 3300

CAAGCGGAGA AGCAATTGTC GCAGTGGATA AGGGCGTTCC TTTAACACAA ACGGCTGATG 3360

GATTAAAACC GATTAAAAGT GAATATAAAG TATTACCAAG TGGCAATGTA CAAGTGAAAA 3420

GTGCTGACGG AAAAATGAAA GTACTTCCTT ACACTGGTGA AAAAATGGGC ATAATTGGGT 3480

CAATCGCTGG TGTATGTTTG ACTGTTTTAT CAGGAATCTT AATTTATAAA AAACGTAAAG 3540

TGTAGTGAGT GATGTATGAA ATTTTTAAAA AAAATTCAAA CAAAGGAAAA AAGCAAAAAG 3600

AGCCCAAGAG TTCGTTTACT ATCACAAAAA AAAGCGAATC GATTAGTATT AATTGTGGGG 3660

TGTGCTCTTT TTGGATTGTC ATTGATTGGT GCTATTCGAG CAAATGTGAT GGCAGGCAAC 3720

GTGATCCGCC TCAGTCAACA ATTAGAAACA ATAAATAAAG ATGTCAAAAA AATAAAAGAA 3780

AAAAAGAGAG TGTATGATTC GTCGGCACTC TCTTTTTATG TCAGAAATTT TGTGAATGAG 3840

TATATCAATT TTAATGGAAA AGCAACAGAT GAAGAGAAAA AAGCACGTGA TGAGAAATTA 3900

GCGAGTTATT TTTCAGGAGA TTTAGCGTTA GATAAAGGAA CAGACAATGC TGCAAAAGAC 3960

TTAATTCGTA AATTATTACA AGTCTCTGTT ACTCGTGTAG AAGAATCTGA AGAGCTATTA 4020

CTTGTGCATG CGAGTATATC TTATAAAGTG CAACAAAAAG GGCAGGATTT ACTAGATACA 4080

CAAGAAATTG TTTTACCTAT TCAAGAAAAA AATGGGTTGT TTTCTATTGT TGGTCGTCCC 4140

TATTTTTTAG CAATGAATGT TCCGAAAGGA AAACAGGAAC CATTGCAATC AGTGAAAGAA 4200

CCGCTTGATG TTGAAAAAAA AGAACGACAA GGGATTGAAA AATTTTTACA ATTATTTTTT 4260

GAAAAATATG CGAAGGGGAA TACAGAAGAA TTGTCTGTTT TAATGAAAGA ACCAGTTAAA 4320

ACAAGTGGTC AAGCAACTGT GTCGTCTATT GAAGAAAAAG AACTACGTTT TTTTGAAACG 4380

AAACAAAAAG GTGTTACAGG CGTTCAGGTG TCAGTGAACT TTAAAGACAA AGCAATAAAT 4440

TTCCTTTACA CGGAAGATTT TTCGTTGTGG TTAACGAAAA CAGAAAATTC GTATTTTGTA 4500

CAAGAGTTAA AACATTATTA TACAGAAAAG GCAGGAGATA AATAGTGGAT ACAAGTAAAT 4560

TAAGTCAATT TATGAGTAGT GGATTGATTA CACTTATTGC GTTTGGTGTG ATCATTTTAA 4620

TTTTAAAGCA TTGGAAAGAA GCTGCGTGGT TAAAGATTGG TTCAGTTATT GTCATTGCAT 4680

TGATTTTAAA TGACTTTGCG ACAAATAATG GGCAGAACAT TTTCAAAGTT GTGAAGTGGG 4740

TCTTAGGATT ATTTGGGATT CAGCTGTAGA GGAGAGAAAT AATGAAACAT TTTGATTATT 4800

CGAGGGGCTT AAAAGCCCCT TACTCATTAC AAGTTATTAA GTCCCCAAAA GGGAAAATTG 4860

TTTGGTATTT TGCACAACCC TTAAGTTTAG CTTATTTAGT GATGTTATTT TTAGGGATTG 4920

TACTAACAGG CATTTTTTGG AAGTTTGTAC CATTGCCGTT AATTTTTGGT ATTAATCTTA 4980

ACCTAATGAT TATGTTGTAT TTTCCAAATA AAGTGGCTCG ATGGTATACA GAAACAGAGT 5040

TTGAAGGGAA AACAGGATTA GCTTTTTTAA AAGATGGTTT TGTTTATGTA AAAAATTATG 5100

TGTTAGACAA TAGATCAATT ATATCCTTTG AACGAGTAAA GGAAATAGAA GAATTTTCAT 5160

TTAAACGTTA GTAATAAGAA AGAGTGTATT ATGGTGAAAA CATGAAAAAA AGAAGTGTGT 5220

TGTAATAACT TCTTTTTTCA TGTTTAAATT ATGAATTTCT ATATTTTAAA TAAAGCATAA 5280

CTGCACAGAC TATTATTCCA ATAACTAAAA GTAGCACAGT AACAATACTT TCATTTTTAT 5340

CTATTTTTCT TTCAGGTCTA TAGCTAGGAT TAATTTGTTT GGCGTTAAAA TCAGTAATTC 5400

CAAAGTCTAA AGTAGGAAGT AAAAAGAGTA ATAAAATAAT ATATATAATT CCCGCTATAA 5460

GGAAAATAGG GGCGCTAGTG AAAAGTGAAA TAATGGTACC TATGATTATA GGAATTAAAA 5520

TGTATAAAAA GGATTTGATA ACAGATTCTT TTCTCATTTT AATTACCAAA TTTTGTTATT 5580

CATGTTTCCC ATATTTCCTA AAGGAACACA ATAACCACCG CTCATATTTT TCCCACCTGT 5640

AGACCAGCCA CCTTTTTTAA AATAAAAACC TCCAGGAC 5678

(2) INFORMATION FOR SEQ ID NO: 634:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1068 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 634:

```
CCTTTTTATT GGTAACAAGG TACGTATGAT CGAAGGGTAA TTnCCACTTT ATACGGnCCA      60
ATTTTTTCAT AACTTGCCGT TGAATTTCCA CACCCAGTTC AAATTCAGAT AATTTGGATT     120
CTGTAGGCAA AnCCAAGTCG AACAAAATCT TCGGCTCTTC GCCCATGGGT AACTAAGCGG     180
ATATCATGGA CTTTTAAAGA CGCGTCAAAA GAACGAATAA TCTTTTTGAT TGTTTGATGA     240
ACAAATTGTT GTGTTGGGTC ATACAAATTC ACTGGnATCA ATATGACAAA CAAGATTAAC     300
ATCCAACTCT TCTTTAAACT TCGCTTCAAT CGCATCGATT GTCTGATGGG CTTTATTTAA     360
ATCCCAGCGA TCATCAATTT CAATGTGAAC AGACGCAAAC GTTTGGCTGG GACCATAATT     420
ATGAATCAGT AAATCGTGAT ACCCTACAAT TGTTTCCATT TTTGATAAGA CATTCTTCAT     480
TTCATCAATT TCTGTTTGGC TCGGACGCAT GCCCATCAAC TCATTAATAA ATTCGCGAAT     540
TAACTGCAAG CCACTATAAA TAATATATGC AGCAATTAAA AAACCAATAT ATCCATCAAT     600
TCGCAAACCA GTGACGCCTT CAATAAATGC AGAAATCAAA ACAGCTAATG TAGTATAAAC     660
ATCATTAAAG CTATCCTTAG CAGAAGCAAC TAATGCTTGA GAATCAATTT TTGCTGATAC     720
TTTTTTGTAA AACAGCCCTT GCCAAATTTT AATCCCAATT GATAAAGCTA AAACAGCAAA     780
TAAAATGGGT GtCAsCTTGA TTGATTCTGG ATGCAGGATG CGATCCACGG AGGTCGTTAG     840
AAACTCAAAC CCAATAAACG TAATCACTAA TGATACAAGC ATTCCACTAA TATATTCAAA     900
ACGTTCATGG CCATAGGGAT GTTCTTTATC CGCTGGCTTA CCAGCAATAT AAAAACCGAC     960
TAATGTTAAA ACTGAAGAAA TAGTATCGGA CAAGCTATTA ATGGCATCTG CCATGATTGA    1020
AACACTACCT GnAAGCAAGC CAATCATAAA TTTACTAACG GATAGAAT                 1068
```

(2) INFORMATION FOR SEQ ID NO: 635:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 995 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 635:

```
TCGGAACGGC AATTACATAC ATCTCTTTTT CATGAATTGT TACATCACTC ATGACTTTCA      60
TTCCACTTAC AGAAATGATG CCACAAACAA CCACAAAAAC ACCGCCGATA ACTGGTGCCG     120
GAATGGCCGA AATCAAGGTG GAAAGTTTTC CAGATAAGCC AAATAAGACA AACCACGCAC     180
CTGCAGCCAC AAAAACCTTG CGACTAGCAA CTCCAGTGAT TGAAATAATC CCAGCATTTG     240
TAGAGTAACT AGTTACTGGC GTGCCGCCAA TGAATGAAGA AACTAAGCAA CCTAAGCCTT     300
CACCAATTAC GCCGCGATTA ATTTTTTCTT CGGTCAAAGG TTCAGCGACA ACATGACTCA     360
CAGCAAACCA AGTGcCTGTT GTTTCCGcCA tTAAAACCAG ATwAATaACT AACATCGTGG     420
CAATAGCTGA CCAATcAAAT TGGaAAGAAA AGTTAACTAA TGGTAACTGA GGCAAACTAA     480
```

ACCAAGCCGC TTTACTAACG GCTGATAAAT CAAGtACCCC TAATAATTGT GCCGCGATGC 540

TCCCAAATAA TAATGCCAAA ATGaCAGAAG AAACGCGGAA AATCCGCCCT TTTTTCAAAG 600

CAGAACCTAA GAGCGCAAAG AAGATCAACG TACCCGCTGC TATAAATGCT AATAAAATAT 660

TTTGTCCTAA TTCCCCTTTT GCGTGGTAAA TATTATCCgA TAAAGCGACT GGCATTAACG 720

AAAGCCCCAC CACAAAAATA ATTGTTCCTC CCACAATTGG TGGAACAAAT TTAGTGATTA 780

AACGATTAAA AATACCAGTA AATCCTAATA TAATAACAAC AATTGCCCCC ACTAAACTAG 840

CACCTAAAAt GCGcCCCAGC CATTTCCTTG tTGGTTATTG GcAAAATAAA TTCCTGcAAT 900

TGCCCCGATT GGaATAAATG AAGGGcCTTG CGcAACTGGt AAtCnCCATA CCAAAAATGG 960

GGAnTGGAAT TAATCGTGGG CAAGTnCCTG CTGGC 995

(2) INFORMATION FOR SEQ ID NO: 636:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1038 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 636:

CGnCTATCCA AGATGGCTGG AAAAATCCnG CACGTTCCAA AATTCCATGG AAGGAATTAA 60

AAnAAACCGG GAAAGnGGGC ATAAGTCGTT CCGACTTATG CCCCCCCCCT TCGTAATCTT 120

AAGATTATCa GsTAAATCaA AGTGAGTTCT TTTTAAAGAA ATACTTTTAT TTTAGCCTTT 180

TATAGATCTA AATAATTTTC TAAGTAATCT GCTAAACCAT CTTGATCATT TGTTTTTTCA 240

GTGACATCGT TGGCTACGGA TTTAATTTTA TCCGTGGCAT TGTTCATAGC AACACCCCAA 300

CCAGCATAAC TTAACATTTC TTCGTCATTA TGTTCATCGC CAAAAGCAAT GATATCAGCT 360

GGTTTGACAG ATAAAAAGTT CGCCACTTGA TCAACGCCAT GGGCCTTTTG AATTCCTTTT 420

GCGACCATTT CTAAAATaGG CaTTGGACCG CCCCAAGTGC GGACATCAAT ATAATCACCA 480

TATtGTTTtG TCAATGTATC AGCAACTTTT TcAGCTTGGT TTGGTGTGGT CCGAACCATC 540

ATTGAGGTAG GATTTGTTCG TAAATTTTTA GCAGTAAGTA AATTATCTAA GGTTGCTTTG 600

GAAGCAAAAA ATTTCGGATC GAACCCATCT AACGTATCGA TATAAAATGT TTCTTTGTTT 660

TCAGCAGCAA CAAAATCTAA TTGTAATTCT TTTTTCTGTG CCAAAATGTC AAAGACTAGG 720

TCACGCTCAA TATTGGcTTC GCTTTCCAAG TCCCATTTTT TCTCTGGAAT GTGGACTAAA 780

GCACCGTTAA AGTTAACCAT AGGGGTCGTT AGGCCAAGTT GTTGGTAAAA TTGATGACTC 840

ATTCTATAGG GACGTCCAGT GGCAATACTG ACGTAATGAC CGTTGTCAAT CGCTTTTTTT 900

AGTGTTtGTT CGGkTTTTGG aCTGATTAAA GACTGTGaTT aAtGtGTGcC GTCtAAAtCA 960

ATAGCAATTA ATTTTTGCAA GTAAAAATCC TCCTAAAATG CCTTTnACTC AnTCTAGCAT 1020

AAGACCGAAG ATTATCCn 1038

(2) INFORMATION FOR SEQ ID NO: 637:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 551 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 637:

```
AAATATACCA TAGGCCAGTT TTATGGTTTT TTTAAGTACC CATCCCAAGT TATTTGGGAG        60

TAATTGTATT TATCCTTTAA TGCAATTGGG ATCCTACAGT AACTTCTAAT TCCCTTGGTC       120

CTGTCTTTGa ATTGAAGAAA ACAAGCAATT CCTATAGAAA CTGCTGTAAA CCCTATAATT       180

AACAATATAA CAGTCTTTTT ATTCATTCCT ACTTCACTCC TTTATCAAAG ATTCTTCCAG       240

TAATTCATAT GAAGATACTC CTTCTAATCG TTTATACACT TTTCCTTCTT TGATAATAAT       300

TACAGTAGGA GTCTGTGTTA TATTTAATTG ATCTAAAAGT AAATTACTTT CTTTATTCTT       360

CTCTTCTAgT CTACGCCATA TACTAATAAA TTTTTGTCTT TAATTACCTT ATTCAAGTCT       420

TTTTTTTGTA CATTACATGG TTGGCAACCT TTTTTAAAAA AATAGATATT CAaGTCTTTA       480

TTTTGAGAAA CCTTGTnCCT CAATACTCTC AGAAATCGGT nATTTTATAT TnCAATCTCT       540

TCTTTATTTA A                                                            551
```

(2) INFORMATION FOR SEQ ID NO: 638:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1636 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 638:

```
nCTCTTAATG GTTTTACnTT TTCGTAAGTT TTCATACCTA TATCTGnCCT TATCTAATGT        60

TTATTTTAGT TTAACATATA TTTAAAATAG ATTTTTAAAA AAGTTTATTT CAGATGTTGA       120

CATGTTTGTT TTGGAATGAT ATACTCTGTG TATAAAGTCT GAAACAGATA ATAGACATTC       180

TGTTTTGGAT AAACGAAATT AAATAAACGA AAAAGCGTTC ACGAGAAGAA AATGAATCTC       240

CATTTCGTAA ACGCCGCATA ATAAAACCTA GACACTTTTA TTGTACGGCA TTTGACGAGA       300

TAATACAAGA GTGAAGTTTT CATTCAAGGT TATCAGTCTT TTTGTGGTGT GGTATAGTCA       360

GCCAATGAAT GACTTGCAAG GTGGCAGTAG TATCTTTAAT TAGAGCCATA TTGATGAAGA       420

TTTTAGGTGG AAGGTTTCTA GCTTTCATAC GAGGACTAGG TTCGCATACT GTGTATTTGT       480

AAAAGATTAA CAGTTAGAAC AATGGAGTAA TTTCCGCTTG GTGATTGAAG GGTGTTCATT       540

TAAAAGAATC AGATTATTTC CGATAACCAT GTCGTGATGA CACGTTAGAA ATAATTGCCG       600

AATTAGATTG AATAAAAAAA TCtATGCAAG CTCAATAAAA GGAGGAGAGC TGTACACCGT       660

ATGAGGTAAA TCTATTAATG TGCCGTAAtG TAATTTTTAT tACTAGCCGT TGAACgATAT       720

yCyATGATTT ACTTTGAATG TAAGAATCCT AAATTTTAAT GCAATtAAGG TTGTTATTTA       780
```

CACGAGTGGA AAAGATTTTT TCTAAAAGAA TAACCAcCTT AACTGGAAAC GTTAAAAATA 840

GGAGCTAATT ACTTAATTCC tATTTATATA rATATTTTAG AAGTGAATAA TCTTCACTTT 900

TAAAACTTTT TCTAAAACTT TGTCGGTTGC ATATAGAAAT GCAATTGCAC AAAGTCCTAA 960

GAAAAGCAAT CTATTTACGA TTGGTCCGTA AATAAGAGTT GTCGCTAAAA CAGCAGCGAT 1020

AACAAAGATT ACTTCAAAAA GATATTTGAA GTTACGCATA GATACCACCT CCTTTTATTT 1080

AGTTATTTTA TTTTAACATA ATAAGGGAGG TTGGTACTTA TGATGTAGCG ATTGTAAGTG 1140

AAGCATAACT AGAATTCTCT GGTTGTGTTT TTTTGTGTTT AAAAAAGAAA AAATCGATAA 1200

ATAAATTACC AAGAGGAGGG AAGTCATGAT AGAAGAATTA TTACTTTTCG TTGAGAAGTT 1260

TGTTGAACGT ATGAAACGTC AGAAAAAAGC ATTCTCTATT TCAGATATAG AAAAGAGCTA 1320

TAACTTAGAA CGTAAGAAGT TAGGTAAGAG TGCAGTTAAG TTAACAAATA TGGACCGTCT 1380

AACAATTGAA AGTCAATTGT TAAAATATCA GATTTTACAG CGGACATATA AAATGACAGG 1440

TTATCATAAA CCTTGTCAAG TTGTGTTTTT TAGTTAGTAA TAAGAATTGC ATAAAGAGAG 1500

AAAGGCAGTT AGTTGTTGAC TGTtTATTTT TGtATAAAAA TTAAGTgAAA GAAGGAAATT 1560

ATTTgGCmAA TAGAATTCAG GAAATTGCAT TACCAGAAAA AATCTAGGAG TAAATTAAAG 1620

CTATTTGTAn AnCAGA 1636

(2) INFORMATION FOR SEQ ID NO: 639:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 474 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 639:

TGGTATGATT ATGACAACTC CAGACCTTGG TGGAGCCACT GAATATTGGA ATATTTTTGG 60

ATACCATGTT GCACAAACAA ACTATGCCTA CCAAGTAATT CCGGTATTAG CTTCCGTATA 120

TCTTTTGTCA ATATTGGAAA AATATTTTCA TAAAAAACTT CCTTCATCTA TTGATTTTAC 180

ATTCACACCA TTATTATCCG TAATAATTAC TGGATTCTTA ACTTTCACAG TTATTGGTCC 240

CACAATGTTA TTACTATCTA ATGGAATCAC AGATGCTATT GTTTGGTTAT ATAATGCAAC 300

TGGGkTCATA GGAATGGGGA TCTTTGGTGG AACATATTCA CTCATTGTTA TGACTGGTCT 360

TCATCAGTCA TTTCCTGCAA TCGAAACACA GTTACTTTCA rCAtGGaCmA ATGGTATCGG 420

ACATGGGGGT TTwayCTTgT AGTTGCTTCC ATGGGCAAAT GTAGCnCAAG GTGG 474

(2) INFORMATION FOR SEQ ID NO: 640:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1094 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 640:

TGTGTGACTT CTTTGCTTTA TGAACGTATC CgTCCAGAGT TTCATCtTGC ACGCTGGATA 60
TATTACGAAA AAGCACGATA TGAGTTAAAA GGGGTAGAAC TTGAAAGTGC AAAAATATTT 120
TTCAACGGGC TAAAGAACCT ATCCGAATCA GATAAAAAGA TTTTAATTGA TGTTTATTAT 180
CGCTCAAAAG ACTATTATAA ATTTAACCGA CAAACAGGAT TATATCAGTC TGTCAGACCA 240
ATTTCCGACG ATGCAATAGC AGAACAATAC GGAATTACTA AAAAAGAAGT AACGAAAGTA 300
AGACGGCAAG CGATAGATCA TTTAGCGGAA GAAATGCGGA AAATTATATT AGCTATAAGT 360
ACAGCCTTTC ATTTGAAGAT AGGTAAAGAC TTATATCTAG TTAGGCTTAT AAACGAAGGA 420
ACATACAAAG AACAGTTTGT TCTTGGAAAT AAAAGAGAAG CTAAAGTGTT TAGCGCAGAA 480
AAGGAAGACA CAATTAGAAA GTTTATGCAG CTAGGTTTTG AAAGAGAGCC AGCATGaATA 540
AGGCTTAACC TTGAGAGGTA TTCTTATGAC AAAAAATAAG TTATTTCATT TAACAGATAA 600
GTCCAATATC AATAGCATTA GAAAACACGG TTTAGTAGGA GTAAAAAAAG CAAAGGATCT 660
TTTAAGAAGA GAGACAGGAA CAGACCATAC GCTCTTTAAT CAGCAAACCT ATTCTATTCT 720
AAATACTTAC GGGTGGAAAG GCTACGATCT TAGATGTGCT ACTTTTATGT TTGAAGAAGA 780
TAGTTGTCTT GGCTATGAAT TGCTGCAGCT AATGAATGAC GATCCATATA TTTTAGAAAT 840
AGAGATAGAC CGTTTAAATA AAGATAAATT ATTTGTGTTT AATACGGAAA TTGCTAGCCA 900
TTTGCTAAAT TATTCGAAAC AAGATCAACA CAGGcTAGCT AAGTTTTATT GGAATACAGC 960
AATTCCATAC AACACTTATG TACCAAAACC AAGGACAAAG TTAAATGGAC GCTTTTCAAT 1020
TTAGGCAATT ATCCAGGTAA TCnAAGCAGG AATAAGGTnn AATTTCGGTG GAnATTTCCA 1080
CCCAAAATTA TTAT 1094

(2) INFORMATION FOR SEQ ID NO: 641:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 692 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 641:

ACAACAAAAA AAGCATAGAA CAATCCCGTT AAACATGTTG CGTAACCATT GTACACCACA 60
TTGTATTACG AAATTCTTCC ATGCCTAGTA TTATCTAGTC ACATGAAAAC GCTATTCATT 120
TGTTAACAGT GCTTATTCTA ACGAGTATCA TTATTTTTGT CAACCTTTTT CAATTATTTT 180
TTTATTAATT GTTGACTATT AACAAAAAAG AAGCTCTTAT CGGAAAAAAA GTACCTTCAG 240
CTAAAGGTAC TTTTGTTCAC TAATTTATTC TTTATAGGCA TACTTATATG TGCTCGTTGC 300
ACCGAAGTTA TGATACTGAA TACCTTTGAC ACCTGGTCGT ACAAGCTGAC TACGGGCCTC 360
TTGGTATAAA GGAATTAAAG CGGCATCTTC TTCCACCAAT ACTTTTTCAG CTTTAATTAA 420

TGTTTGCCAA CGAGTTTCAG GCTCATTTGC GTATTTAGTT GTTGCTTCTT CAATTAGTTG 480

ATCAAACGTT GGGCTAGAAT AGTTCATACG GTTACCACTA ACTGAGCTTT CTAAGAAAGT 540

CATTGGATCC TGATAATCGG GTCCCCACAT GCCAACAGAC AATTCATAAT CCCCTTTTCT 600

CATTAATTCA ATtGCATTAT TTAAAGGAcT GTTTTGACAT TAATAGTGAG ACCAGGCACG 660

TTCTTGTACT GGCCTGGATA ATnCAnCACT TT 692

(2) INFORMATION FOR SEQ ID NO: 642:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1485 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 642:

AGTACAACCA TATGGCAGAT GGCTGAAGGT ATTATTAATC ATTGTTAATC TTGGCGCTTT 60

GTTTAGTATT CTGCGTCTAT TTATTCATTT GTTGATTTGA AGGGGTCGAA AGGTGAGAAA 120

ATGAGAAAGT TAATTTTAGT TTTTGTGGGT TGTGCCATGG CCTTTTGTTT ATGGCCTAAT 180

CAAGCGGCGG CTGTAGCTGG TGGCTATGGC GGCGGGCGTG TAGAAAGTCC TAGCGGAGGT 240

AATCCCTCGA ATGGTTTTTC GCCAAGCGGT GATGGATCAA GAAGTTATGT TCATCATAAC 300

GAGACCGGGA CCAGCAACAC AACGCCACTG GAAGGCTTAA TTTCAACAGT GTTTTTTGTT 360

CTTTTTGGGG GCACGATGAT TTGGCAAAGA CGCCGTAGAA ATCAACCAGT CGATGAAGGA 420

GCTTTTCAAG AAATGTGGGC AkCATTACCT GGAACATCAA AAGAAAAAAA GGCACTCTTA 480

AAAGAAATAC AACAAACGTT TTTAGAAATA CAACAAGCGT GGGATCAAGA AAACTTGTCT 540

GTGGTCAAAG ATAACTATAC AGAAAAACTT TATCAAAAAC ATCTTGCAGT ATTACAAGCA 600

AATCAACAAG TGGGCATCAG AAATCATGTG AAAAAAGTGA AAGTAGCGCC AGCTTCGAAT 660

TATCATCATA TTAGTGAAAC GAGTTTTTCA CTGATGCTTC ATTTkTCTTG TATTGACTAC 720

GAGGAAGATA CTCAAAGCGG TCGTGTTTTA TCGGGATACA AGCATCAAAA ACAATATTTT 780

TCACAACTTT GGTATTTTGA TTACAATCCA TCTCTTGGAA AATGGCAAGC AGACTTCATT 840

CAACCTAAAG GATAGTGCTA AAAAGCGCTT GGGGTTTAAA TCCCAAGCGC TTTTTTATTC 900

TTCGATGcGT TTTACAAGCT GTATTTTTTT ATTTAATACG TAAATCAGGA TACCGCCGAC 960

GGTCATTGAA AGAAGTTCAC CAAGTCCAAC GGTCAACCAT GTGTACCAAA AAGGTAACTG 1020

ATAAAAGAAC GTTAAGGGTC CGGCGACAGT GAACATGGAC AAAGCAAAAA gAATGGCCGT 1080

CAGCATTAGT TTAATCAATA AATTTTGCTG ATTTTTTGTC GCAAAATAGG CAATCCATAA 1140

GACTGCGCAC GTTCCAGCAC TACCGATGAC TATATCGATG ATACCAAGAG GCGAAGCGGT 1200

GTTGGCAATC GCAACCCCTA ACGTAATGGA CCAGATATAT CGCTTGTTAA AAACGCCTAA 1260

ATAATTAAAT AATTCGGCAA GACGAAGCTG TATCGGACCA AAACTAAGCG GAGCGAGAAG 1320

TACAGTGACA ACGACATACA AAGCGGTAAC CAGCGCCATT TTAGCCAACG CCGTAGCTGA 1380

CCACGTACTG GTTTGTGGAT TTGATGAATG ATTCATTTAA GATTCTCCTT TATAAAACAA 1440

CGACGTGTTT TAGTGATAAC CATAACCAAA GGAAGGGCAG TAAAG 1485

(2) INFORMATION FOR SEQ ID NO: 643:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 951 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 643:

AGAAATATCA TTTCAAGTTC TTTTTAAAAA TATTTTTAAA TGAAATTACA TTTCTTATAT 60

AATTTTTGTG CTATACTAAA AGAAAACGGT TTAGGAGTGA TTATGTGcGG GTTTTGCTGA 120

TTTATTTCAT AACTATTTTT CTATCCAATA CAGTTGGGGC GCTCTCTGGG ATGGGTGGAG 180

GCGTCATTAT TAAGCCTGTT TTGGATTTTC TTGGGTTCCA TTCGTTAAAT AGTATTGCGT 240

TCTATTCAAG TGTCGCTGTT TTTGTCATGT CCATTTCTTC GACTTATAAA CAGTATCAAA 300

ATGGTGTCCA AATTGAATGG AAAAAAGCCG CGAGTATTTC GTTTGGTTCC TTAGTCGGCG 360

GCATGTTAGG TGATTTACTT CTTAATCAGG CGATTGCACT AGCGCCAAAT GAAGAAAAAG 420

TACAACTAAT CCAATATATC ATCATGTTGC TGACCTTGGT TTTGGTTCTT CTCTACAATC 480

AATTTTCCAA CTGGCACTTA CACCTCAACG GATTGAGCAT TTTTCTGATT GTCGGATTAG 540

GCTTAGGGAT TCTATCAACC TTTTTAGGCA TCGGTGGCGG TCCCATTAAT GTAGCTTGCC 600

TTATTCTCTT TTTTGGAATG GATATCAAAT CAGCAACTGT CTATTCCATC ATTACTATTT 660

TTTTCTCCCA ACTCGCCAAA CTTGGCAATA TCGGCTTGAC CACAGGGTTT GCGGTTTTTG 720

ATTTGACAAT GCTTTGGGCT ATTATTCCTG CCGCTTTATT CGGTGGTTAT GTCGGAGGTC 780

TTTTCAGTAA AAAATTATCC CAACAACGGG TCGCCCAAAT TTATAGTTTA GTTGTTTTTT 840

TAGTTATTTT GTTAAATATG TATAATTTGT GGACTGTnAT CnAAAAAAGA AGCGCTTTGw 900

CGTATAGACA AAACGCTTCT TTTTTAATTA AATCACTGTn AATTGAACAT C 951

(2) INFORMATION FOR SEQ ID NO: 644:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 647 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 644:

TCCAGAGAGT TGCCCTCCAG CTGAGGCTAT CCCATTTGAT GGAGTTTTTT TCAGACTATG 60

TAAGGGTGAA AACTTTTGCs AAGAAGATTT TCTTACTCAT TTTGAAGCTG GGAAAAGTTT 120

CCCAGATAAA AAACTGTGTG AAGCTATGGC ATTATCATTT TTTGATAGCT ATGATCATGC 180

AGAAAAAATTA AAAAAAAGAT TTAGAAAAAA ATTnGCAACA TGTTCTATCA AACCTGTTGC 240

AATTATTGAA GATTATGGAG TAGGAATTTT TGAAGAGAAT TCAGGGCATC TTAATTTATG 300

GGAGCACAGG AACGTTGATA TTTTTGCTGA TTTAATGAAG GAGGATAAAG GAAAAGATGA 360

ATAATCTAAT ATTTTCAGAC ATGGgAATAT ATTTTCGkTT TTGACTATTA TGaTACTCCt 420

TTATTTTTCa TTtCTCGAAG TCGGAAAAAT AATAATTACy AaTTAwTCTA CTCTATAAAT 480

GATGACGTTT ATTTTTACAG TGAGCTATCT GCTTCGGATA TTAATTACTT ATTTAGTAAT 540

CCAAATGGAT ACGACATTCT TCATTACCTT AAAAAGAAGT CTAAATTAAA TTTTGTTATG 600

GTAAATAATA ATTCTTTTGA TATCTATAGC TTAGCGGAAT ATACAAA 647

(2) INFORMATION FOR SEQ ID NO: 645:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1258 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 645:

GACGAATGGC CCAACAGGGC ACTTACAGAA ACAATATCTG AAGTACTCTC TAAGCAACGT 60

GTCGCAAAAG TGTGTCGTAA TTGATGGAAG TGTACTTTGT CCAATTGGAC TTTATCGCGT 120

AATTTGTGGA AGTAATACGT AAGTAAGCGA GGCTCACAAG GCTTATTATT ACGAGAAAAA 180

ACAAATTTTC CTTGGCTGTG TTCTTTTTGA CACTTCAATG CATTTTTTAC TGTATTACTC 240

ATAGGAATTA TTCGCTCAGA ATTTATCGTT TTTGTACAAT CGTAGTAAAG TTGAGTAGAT 300

TCTCCGAACG AGCTTACAGA TACTCTTTGA AAAGTATGCT TTACATGAAT AGTATCCTTC 360

TTAAAGTTGA TGTCCTCCCA TCTGAGCGCA GTAATTCTCC TATACGTAAG CCGGTATGTA 420

AAGCGmCAAT AACTGaTAAT CCyTGTTGCT GAGGATTTTt ATAGGCAACA TTCtCCaATT 480

tCTTTkGcyC TTCAATATCT AGTGCACGAA TATGTCGACG CTTTGGTTTA GGAAGCGAAA 540

TAGTTGAACA GGGATTGCTT TTCAAATGCT TTGCTTTTAC TGCATGATTC AAGGCTCTTG 600

AAAGGATACC GAATATGACA CAAATCGTAC TTTTACTCAA TCCTTGTGCG ATCAAGTCCT 660

CTACTAATTT ATTCAATCTC AAATCATCTA TTTCATTTAA TGCAGAGTCT CCCATTGTTT 720

TTAAGACATA GTTGCTTAAC TTGTAACTAT AACTAGCGAA AGTAGAAAGT TTTACTTCCT 780

TCTTTACGAT GACTAACCAC TGTTCTATCC ATTCATTAAG TGTAAAAGCG CCTTGACCTC 840

GAGATTGATA GATAGTTTGT TGATTTGCTT TTGCCACATA GAGTTTTTGT CGTACCTCTT 900

GATAAGTAGA ACCATAGATA TATCCATACT TTAATTTTCC ATCTATCTTC CTTCCTCGTG 960

GGAAGCGACC TTCCCATCGT CCATCTTTTC TTTTATATAT ATTTTCGCCA CGTTTACTCA 1020

TTTCAATTCC TCCGTTTAAC TACTATGTTC TGACGGCTAT ATTGACGGCT TATAGAAAGG 1080

CTTATTTATA GCTTTAAAGA ACATTAGTTA TAATATAAAT CGTTATTATT GCTAAAAAAT 1140

ATCATTATTT TACTTTTTGT TGACACATAT TTTGGAAGTA GATTATAATT CTnCTGTACG 1200

GGGATGAATG ATGnCCTACA TTGGTCTATA TATGAnTCCC CACAGGGGTT TCAATAAC 1258

(2) INFORMATION FOR SEQ ID NO: 646:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 555 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 646:

ACTGTAACAC TTGACGGCCA AGACCGCGAC CTCGTTGCGT TTTTGTCACA ACAAAGCCAT 60

AAATCCCCCA TTCGTycTTC AAAATGATCC AAACGTAGCG TGGCTAAAAG TTGATTGTTT 120

TCTTTATAAA TAAGTGTTCG TTGAAGGTCC TCCGGCAAAG GCGTACCTTC GAGTAACCTA 180

GCCAAGGAAT CCAAATCAGC CACTTGTGGT CGGAGAAGTG TCAGCTTCGC CAACGCCTGT 240

GAGTAAGTTT CTCTTTGAAA AACCAGATAC TGTTCTGAAA AAGCTGGCAC AATTCCTTGT 300

CTATTAAAAT AATTCAGGCC AACAGATTGC TTGTCGTCCA TCACGAGAAA GACTTCTTTC 360

ATTCCCCACA AAGCTGCTTG CTTTTCGAGT TCTTGAACAA GTTCTTCAAA AAATGGTCCT 420

ACATTTGGCG CAACAATCGT CGCTTCCAAG GTTTCCCCAT CAAAACAATA CCCTACGATA 480

TATGCTTGAA TGGTTgTTCC TTCCTGtATA AACCATAATA AGTATAAGCT GATtCCTCAn 540

GAAATGTTGT CCAAT 555

(2) INFORMATION FOR SEQ ID NO: 647:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1030 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 647:

AAGGCAAGCG TCGATTTTCG ACAATATCCA ATAATGCATG GCTTTGTCCG TCCATCATGA 60

TAAAACTCAT TGCACTGGTC ACATGTCGAA CCGATTTGAA TTCATCAAAG CATAAGACCT 120

TTGGCAAAGT CTGTTTGAAA GGATTCAATG GGCGGTAAAA TCTCTTCATG ACACGATAGA 180

CCGAAGAAGT GGAGATATTT TTCTGACGGG CGATTTCCGT CATGGAAGTA TTCTCTTTTA 240

AGCGAGCAAT AACAGACCAA CGAACTTTTT GACTAATGAA ACAGTATTTT TCAGCGACAG 300

AGGTATCCGC TACGAATGTT CTTTGGCAAG ACTTACATTT AAATCGTTGT TTCTTCAACT 360

CCAGATAAGT GCGATACTCA TATATTCTTT TTCAATCAGC CAATCTTCAT CAAAAAATAA 420

ATGTTTATCT GTTAATCCGA GTAATTTTCT AGTATGATTA TCCATAGAGC GACTCCTTTG 480

AtaggGTGAT GTTTTGGtCG ACTAAATTTT ACCAAAAGTG AGTCTCTTTT TGAGTTGAAA 540

CGAAGaAAAA GGGTGTTGGa AAATCGCCTT CGATTTTCCA ACaACACAAA ATATTATAGA 600

ACCCTTTATT CTAAAGGaTT TCAAGGTGTC TGTTTGcaTT tATTAAAAAA AAgAAAAAGT 660

ATTGGCAACc AGAAGGCGTC ACCaACaCTA GATATTATAG AGCCcATAAT ATGCTGTATA 720

AtATCTAGTG TATAGGCTTA AGTGGAGGTA TATATGAAAA TATTnACTGT TGAAGATGAT 780

AATATGATAA GAGAGGGAAT AAGTGAATAT CTTTCAGAAT TTGGATATAC TGTTATTCAA 840

GCTAAAGATG GAAGAGAAGC CTTGTCAAAA TTTAATAGCG ATATAAATTT GGTTATCTTA 900

GACATTCAGA TACCTTTTAT AAATGGTTTA GAAGTGTTGA AAGAAATTAG AAAGAAAAGC 960

AATTTACCAA TTCTAATCTT GACTGCATTT AGTGATGAAG AATATAAAAT TGATGCATTT 1020

ACTAATTTAG 1030

(2) INFORMATION FOR SEQ ID NO: 648:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1023 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 648:

AAATATTTCT TCTnTCTTTT ACATAATAAG TAGCTCCTTA GTTTGACAAT TGGTTGTACC 60

CAAGGGTAGG CTTTCTTTTT GCAAGCGGAG CAACATGGTA CCATGTTGCG ATATTTTGAA 120

TGACCCCTAC CAAAAATACT GGTAGGAGTC GTTCAAATTT TTGGGGTTCC CCCAATCCCC 180

ATCGCTACAC TTTGACATTT ATTGACGTTT CGACTTACCA TTTTTGACCT TTATTTTTTC 240

TTTCTTATTG ATAAGGAGCT AATAACTCAT CCAGAAAGAC TAAATACGCG TCATTTCCTT 300

TCAACTTTCG TAAATAAGTA GCTGCTTCTT CGTAAGGAAA TAAATCAATT TCTAATTCTT 360

TTTCTTTAAA ATCATGAATC AACGTATCCA GTTCTTTTAG CACCGTTCCT TTTGATTTTT 420

CTTTCTTTTT CTGGTGGACT CTTGGTTGTT CACTAAGCGG GTCTGCCAAG AAAAATTGTG 480

TCAGAATAAA TTTACTTTCG TCCGTTAATT GGTCCATTTT GTTTTCTCCT TTTCTCTTAT 540

TATTTGTTTT TAAAATCGAG TGGGATTtGC TATTTGAGCA GAAATTTTTG GTTCATTTTG 600

GTGCCTTTTT TGCTCGTTAG CTTGGGACTT TCCCACTTCA GTTTCTTTTT TCTnCGCATT 660

CAATAAACGT TCTTCAAATT TATTTTGGGG gAAGTtGAAT GCCcTTTTTC TATGGATAAA 720

GCATGGTCGG CTTTTTCAAT TAGCGTGCGC GCCcTCTTTC GTGATCGTTT CTAACGATtG 780

CGTAAaTtCT tCTAACTkTT CTCCTTGCTT CGTCCAAGAA GAAAGATAAC CAAAAGAATA 840

CGAGGACGTA TCCAGCCCTA AGTGACGGGA GACAATATAG GCAACTGATT CCGCTTCAAA 900

TTCTTGTCGA GAATAAACGT CATCTCCAAA TATCGCTTGC GAATCTGGGT GAAGCATGGC 960

ATGCGTCATT TCATGAAGTA AGACTTTTAA AGTCATCACT TCTCCTAAAT GTTGTTGGAG 1020

GAC 1023

(2) INFORMATION FOR SEQ ID NO: 649:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 829 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 649:

TATTTTCGAT TCATTCTTTG GCGGCGGCGG TGGCCGTTCT GTTGATCCAA ATGCACCAAG 60
ACAAGGCGCC GATTTACAAT ATACAATTCA ATTGAAATTT GAAGAAGcTA TTTTCGGTGT 120
CGAAAAAGAG rTTAAGTACA ATCGAGAAGr TACTTGTGCm ACCTGTGGcG GTAATgGCGC 180
CAAACCAGGC ACACACCCAG AAACTTGTCA TAAATGTCAT GGTCTGGAAC AATCCATGTT 240
GACGTCAAAC GCCATTAGGT CGTATGATGA GTCGTCAAAC TTGTGATGTG TGTCATGGCA 300
CAGGGAAAGA AATCAAAGAA CCATGTCCAA CTTGTCATGG CACAGGTCAT GAGAAAAAAG 360
CACATACTGT CAAAGTAAAT GTGCCTGCTG GAGTAGAAGA CGGTCAACAA ATGCGCTTAG 420
CTAACCAAGG AGAAGCGGGT ACCAATGGTG GACCTTATGG TGATTTATAT GTAGTCTTTC 480
GTGTTGAAGA TAGTGATATT TTCGATCGAG ATGGCGCAGA AATCTACTAT GATTTACCAG 540
TAAGTTTCGT CCAAGCGGCT CTAGGGGACG AAtAACTGTC CCAACTGTTC ATGGTGACGT 600
GAAATTAAAA ATTCCGGCTG GAACACAAAC TGGcACAAAC TTCCGCTTAA GAGGCAAAGG 660
TGCGccACGT TTaCGTGGTG GGGGCAATGG TGATCAACAC GTGAAAGTGa AATTAATAAC 720
ACCAAAAAAT CtTAATGaAG aACaAAAGGa TGCCstACgA gCnTTTTGCC AAAAAGCTGG 780
GCCGGGACCn AAAAATGGTT AACCTnGGAA CCCCAACCCA AGGGAAGGA 829

(2) INFORMATION FOR SEQ ID NO: 650:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1374 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 650:

TGACACTGTC GGTAATTTCT TAGAATCAAT GCCCAGCTTG GTTAnCTGAT GGTATGGCTA 60
TCGGTGGTGG TATGGTCGTT GCCGTAGttA TGCATTAGTA ATCAACATGA TGGCAACAAA 120
AGAAGTATGG CCATTCTTCA TTATCGGTTT CGTAGTCGCA GCGATTTCTC AATTAACATT 180
AATCGCATTa GGTGCTTTAG GTGTCGCTTT AGCTCTTATk TACTTGAACT TATCTAAAAT 240
GGGCGGTTCT TCTAACTCAA ATGGCGGAGG CAACTCTGGC GACCCATTAG GCGATATTTT 300
AAACGACTAC TAaTTCTGAA GGAGGAGCAA AGAAAATGGC AGAAAAAATT CAATTAACTA 360
AAAAAGATCG TTTAGCTGTA GCTTGGCGTT CAACATTTAT CCaAGGTTCT TGGAACTATG 420
AACGTATGCA AAACGGCGGC TGGGCATTCT CAATGATTCC AGCAATCAAA AAATTATATA 480
AAACGAAAGA AGAACGTTCA GCGGCgTTaA AACGTCACTT AGAGTTCTTT AATACACACC 540
CATACATTGC CtCACCAATT CTTGGTGTAA CATTAGCTTT GGAAGAAGAA CGTGCCAATG 600
GCGCACCAGT TGATGACGTA GCGATTCAAG GGGTTAAAGT TGGGATGATG GGTCCTTTAG 660

CCGGTGTCGG TGACCCAGTT TTCTGGTTTA CTGTTCGTCC AATGTTAGGC GCTTTAGGTG 720

CTTCCTTAGC AATGGGTGGT AACATCTTAG GCCCAATCTT ATTCTTCGTT GCTTGGAACT 780

TAATTCGtTG GTCATTCATG TGGTATACAC AAGAGTTTGG TTACAAAGCA GGTTCTAAAA 840

TTACTGATGA CtTTCAGGTG GTTTACTACA AGATATTACA AAAGGTGCTT CTATTTTAGG 900

GATGTTCGTA CTGGCCGCAT TAGTTCAACG TTGGGTATCG ATTAAGTTCT TACCAATTGT 960

CTCACAAGTT AAGTTAGACA AAGGTGCTTA TATCGAGTGG GATAAACTTC CTGCTGGCGG 1020

TGAAGGAATG CATAAAGCCT TTGAACAAGT GAACCAAGGC TTGGCGCTTT CACCAACGAA 1080

AGTTACTACT TTACAAGATA ACTTAGATCA ATTGATTCCA GGTCTAGCAC CATTACTACT 1140

TACATTCTTA TGTATGTGGT TATTGAAGAA AAAAGTTAGC CCAATCGTTA TCATTCTTGG 1200

TTTATTCGTA GTCGGCGTTG TCGGCCACTT AATCGGCTTA TTATAAGAAT GACGAGGTAA 1260

GTTTATTACT TCTAATCAAA AAAAGCGGAA GAGAAGTCGA GAAACTTTTC TTCCGCTTTT 1320

TTAAACAAAG ATGTAATTTT GTTTGTTTCG ATCAGGAAAG AAACGTATAA TAAA 1374

(2) INFORMATION FOR SEQ ID NO: 651:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8911 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 651:

GGAAAAATTC CTGGCGCCTT CATTAAACGT GAAGGTCGTC CAAGTGAaCG TGCGACATTA 60

mCCGCTCGTT TAATCGaCCG TCCTATTCGT CCAATGTTCT CAGAaGGCTt CCGTAmCGAA 120

GTCCAAATCA CGAmCATCGT TATGAGTGTG GAGCAAGACT GTACACCAGA AATGGCTGCA 180

ATGTTTGGGT CTTCCTTAGC GTTAgCAATT TCAGATATTC CGTTTGATGG ACCAATTGCT 240

GGTGTTGACG TTGGTCGTAT TAATGGCGAA TATGTTTTAA ATCCAACGGT AGAACAAGCG 300

GAACAAACAG ATATTGAATT AACTGTTGCT GGAACAAAAG AAGCCATCAA TATGGTTGAA 360

AGTGGTGCCA AAGAAGTATC AGAAGAAGAT ATGTTAGGTG CCTTACTTTT CGGTTTTGAT 420

GCCATTAAAG AACTAGTGGC GTTCCAAGAA GAAATCGTTG CCGCTGTAGG TAAACCAAAG 480

ATGGACGTGG ACTTATTACA AGTGGACGCT GATCTTAAAA AAGAAATTTT TGATGCGTAC 540

TACAATACAA TGAAAACAGC CGTAATGACG GAAGAAAAAT TAGCACGTGA AGTTGAAATT 600

GACAAAGTAA AAGATACAGT CAAAGAAGTT TACGCTGAAA AATTTTCGGA ACACGAAGAA 660

GAAGCACAAT TACTAAAAGA AGTGAAACAA ATTGCGGAAG ATCTTGAAAA AGATGTCGTA 720

CGTGAGTTAA TCACGATCGA TAAGATTCGC CCGGATGGCC GTAAATTAGA CGAAATTCGT 780

CATTTATCAT CAGAAGTAAG TATTTTACCA CGTGTGCATG GTTCTGGCTT ATTTACTCGG 840

GGTCAAACAC AAGCTTTATC CGTTTGTACA TTAGCGCCTT TAGGGGAACA TCAAATTATC 900

GATGGGTTGG GTGTCCAAGA TAGCAAACGT TTCATCCATC ATTATAACTT CCCGCAATTC 960

TCTGTTGGTT CAACAGGACG CGCGGGCTCA CCAGGTCGTC GTGAAATCGG TCACGGTGCG 1020

TTAGGTGAAC GTGCCTTAGC ACAAATTATT CCATCAGAAG AAGACTTCCC TTACACAATC 1080

CGCTTAGTAG CAGAAGTGTT AGAATCAAAT GGTTCTTCTT CACAAGCAAG TATTTGCGCG 1140

GGAACATTAG CATTAATGGA TGCGGGTGTG CCAATTAAAG CGCCAGTTGC AGGTATTGCA 1200

ATGGGTCTTG TTAGTGATGG CGAAAACTAT ACAATCTTAA CAGATATTCA AGGTTTAGAA 1260

GACCACCTTG GCGATATGGA CTTTAAAGTA GCAGGAACGA AAGACGGAAT CACTGCTTTA 1320

CAAATGGACA TCAAAATCCA AGGAATTACA GAACAAATTT TAACGGAAGC ATTAGACCAA 1380

GCGAAGAAAG CACGGATGGA AATTCTAGAA GAATTAACAA CAACGATTGC TGCTCCTCGT 1440

GAAGAATTAA GTCAATATGC ACCGAAAATT GAAATGATTC AAATTAAACC AGCGAAAATT 1500

AAAGATGTCA TTGGTAAAGG TGGCGAAACA ATCAACAGCA TTATTGATGA AACCGGCGTT 1560

AAGATTGATA TCGATCAAGA CGGTAACGTA AGTATTGCTT CATCAGATGC TGAAATGATT 1620

AAAAAAGCCA TCAAGATTAT TGAAGAATTA ACCAAAGAAG TTGAAGTCGG TCAAGTTTAC 1680

TTAGCAAAAG TTGTCCGCAT TGAAAAATTT GGTGCGTTTG TTAACCTAAT TAAAGGCAAA 1740

GATGGCTTGA TCCATATTTC TCAATTAGCA AACGAACGTG TGAACAACGT TGAAGATGTC 1800

GTGAAATTAG GCGACGAAGT CCTTGTCAAA GTAACTGAAA TTGACAAACA AGGCCGTGTC 1860

AACGTGTCAA GAAAAGCGTT ATTAAACGAA GAAAACAAAG AAAAATAATT TCTTTTTAAA 1920

CCAACGAAAA CCAAGGAAGC TTCCTTTCTT TTCGTTGGTT TTTTCGTTGA AAACAAGCCC 1980

TAAAAAAGAA ATGATAAAAA AAGCTAAAGA CGATCGCCAA CTCTTTTGTT TATAGTCAAT 2040

TAGTGGTAAA ATTAATACTA TCAGTCTGAA GAATATCTGG GGTGTCTTTG TGAAAAAATT 2100

TAATCCAAAT AAAAATATAA TCATTACCTT AATTTAGTC ATTATTTTAG TAACCATTAT 2160

TAGTTTAACT GctGCGCAAC GTTCTGAAAA AGGCAAAACC AATGTCGTTC AAGGTGGCGT 2220

CAATAATAGC GTTAGTATGG TGGATCGTGT GATTTCTTTT CCAGCTCGTG TGGTGGAAAA 2280

CGGCTTGTCT TCAATGGGCA ATTTAATTAA CACGTTTAAA GAAAACGAAC GTCTAAAAGA 2340

AAAAATTGAT AGCTATAATG AATTAGCTGT CCAAAACAAT AACTATAAAC GTGAAATTGA 2400

TAGTTTAAAA CAAGAATTGA ATCTGAACGA AACGTTAGCC AACTATGAAA AAGTAACTGC 2460

GAATGTCATT ACGCGTTCGC CGGATACTTG GCAAGACTTA TTAATTATCG ATAAAGGAAC 2520

CAATGACGGC ATTGAAGCGG GCATGGCGGT CATGGCGCAA AAAGGGtTAG TGGGGCGTGT 2580

TATTGAAGTC AATGCAACGA CTGCAAAAGT GGAACTACTA ACGTCTAAAA ATGTTAACTC 2640

GAACCATTTC CCTGTGCGAG TAACTTCCGC AAATGGCGAA TCGTTTGGTT TACTTAAAAA 2700

TTACGACAAT AAGACGAATG CATTAATTGT GTCACAATTA ACTGGCGATG CCACTCTGAA 2760

AGAAGGCGAT GTCGTTCAGA CTTCTGGCTT AGGTGGTAAT TCCCCAGCCG ATTTAGCGGT 2820

AGGGACAGTT ATTAAAGTGA AACCAGATAG TTTTGGTTTA GATCGTGAAG TGTATGTGAA 2880

ACCTTATGCG GATGTCTATG GTATCTCCGT TGTGACGGTC GTGAAAAGAT CGGCAGGGGA 2940

AAGCGAATGA TTCGGAAAGA GTATTTAAAA TACGTCACAC CTATTCTCTT GTTTTTCTTA 3000

ATGTTAATTG ATGCTCATTT AACAAAGTCT TTTGAAACGT GGACAAGCAA TGTTTACTTC 3060

GCCAGTTCTC ATCTGTTATT ACTTGCGTTT ATGTTTGCCG TTCCTAATTT TTCTAAACGT 3120

TATTTATTAA TTACTTCGTT AATTTTAGGC TTTATTTGCG ATAGCTACTA CATTGGTATT 3180

ATCGGGATTT ATACAGTTGC ACTGACGATT TCGGTTATGT TGATGTATAC ATTTAAAGAA 3240

GTAATTCAAA CAAATTTACT AACTGGTTTT TTTGGGATCG TCATTTTTAC CACGTTGTAT 3300

GAACTGATTG CGGTACTGAT TCAAGTGGCC TTTCATTTAG CAGATGTCAG CCCGATTTTA 3360

TTTATTACGC GTGTGTTGGG CCCTACATTG CTATTGAACA TGGTGTGGTA TGTTATTTTA 3420

TCTTATCCAC TGAAAAAACT GTTTTCCGTA AAATAAAAAT TGAAAATTTT GTTCGTTTGC 3480

GAAAAATGCT CATTTGTTAG CATTTTTCGC TtTTTTTATT TTCTTTAAAT AACAAAACAA 3540

TGAGAAGCTT ACCCTGGAGA CTTGAAGAAA GTTGTGAAAA AAGTGAAGCT TTTTTTTGTC 3600

ATTTACAAGA AGAAAATTGA AATATTACTG TAACATCCAT ATTATCTCAG TGACATACAG 3660

CTATGCTACA ATTGCTCTAT CGTTGAAAAA GAGAAATTTA TTCAAATTTT AATAAATAAT 3720

AAAAAAGCTT AACAAAAAAG GATTTTCGGA GGACAACAGA TTGAAGAAAA GTGTATTATC 3780

GGCTCTAATG GTATGTTCCA TTACATTAAC AAGCGTAscG TTGCCATCCG CAGCATTTGC 3840

AGATGAATAC GATACAAAGA TTCAACAACA AGATCAAAAA ATTAATGCGT TAACTAGCCA 3900

AATGTCAGAT GCAGAAGCAA AAGTTGCCGC GATTGAAAAT GATATGGTTG AAACGGCCAA 3960

ACAAATCGAT ACATTAACAG CTAAAAAGAA CAAGCTATCA TCAGAAGTAT CTAAATTATA 4020

TAGTGAAATT TCTGATTTGA ATGTCCGTAT TCAAAAACGT GAAGTACAAA TGACAAAACA 4080

AGCACGCGAT GTCCAAGTGA ATGGTCAAAG TGATTCAATT ATTGATGCTG TCTTAGATGC 4140

AGATTCAGTA GCAGATGCAA TTGGTCGCGT TCAAGCGGTC TCAACAATGA TGAGCGCCAA 4200

TAATGAATTA CTAGAACAAC AAAAAGAAGA CAAAGCGACT GTTGAAAAGA AACAAAGAA 4260

TGTTGAAAAA CAAATTGCTG AATTAGAAGC AGCAACAAAA GAATTAAATG ATAAAACAGA 4320

ATCATTAAAA ACATTGAAGA TTCAACAAGA AGTGGCTAAA AATGATTTAG AAGCACAACG 4380

TTCTGAAGAA CAAGGGAAAA AAGACGGCTT CATTAAACAG AAAAAAGAAG CGGAAAAACG 4440

TTTAGCAGAA GAACAAGCAC GTCAACGTGC AGCTGCTAAA AAAGCAGAAG AACAAGCGGC 4500

AGCGCaAGCA CAAGCCGCAG CACAAAAAGC GGCAGCAGAG CAAGCGAAAG CAACAAAAGC 4560

AGCCAATGAG GCAGCAGCAT CCGCCGCTGA AGAAAAAGCC GCAACACCAG TAGTAGAATC 4620

ATCAACAACT ACTGAAAGTA CCACAACGCA AGAAACAACA ACTTCAAGTA CTGAAACAGA 4680

AAGTGTTGTA ACAACACCTG TGGCAGCAGC ACCTGAAAAA GAAAAAGAAG TTCCTGTAAC 4740

TAACCCAACG ACTCCTGAAA AAGGCAATGA AGCAAAACCA GGGAATGGTG GCGTGACTTC 4800

AGGAAAACAA GCCGCAATTA ATGCAGCTTT AGCAGATGTT GGTAATTCTT ATGCAACAGG 4860

TTGGAATCAA CCAGGCGAAT GTTTAGTATC TGTTCGTCGC TGGTTAGCGG CAGGCGGTAT 4920

TAACTTCGGT TATGGTGGTC CAAACAGTGG CTATGTTGCT TCTGGCGCAA CACAAGTTAG 4980

CTGGAGTAAC GTGCAACCAG GCGATGTTGT TCAATATGAA AGTGCCTATA GTCCAGATAG 5040

CTGGATTGGT GGCGTGCATA CTGTCTTGGT TACAGGTGTA AGTGGTGGAA GTGTTCAAAT 5100

CGTTGAAGCG AACAACCCAG GAGGTTCTGG TTATGTTTCT TCTAACTCTA ACTGGTCACC 5160

AGCACCACCA GCAGGATTCC GCGCTGTTGT ATGGCGTTTT CCAGGCTAAA TAGTCATCTA 5220

TTAATAATTG ATAACAAGGA GTGGTAACTA ACAAAAATCG TTAGTTGCCG CTTTTTTTAT 5280

TCATTTAATT TTAAATATAT GCATAAAATG TATCTGAAAA TATTGTAATC TGGCCGTTTT 5340

TTTGCAATTA TAGAGGTATA TACATATTAG GAGCGGAGGA ATCAGCATGA GCGGAACGCA 5400

CGATACAAAA CAAGCAATCG CTTCAGCTTT AATAGAATTA TGTGAACAAA AGGATTTTCG 5460

TAAAATAAGT GTCCAAGATA TTACCAAAAA AGTGGGCTTG AATCGTCAAA CATTTTATTA 5520

TCACTTTACA GATAAATATG ATTTATTACG TTGGATTTAT CGTCACGATG CATTGATTTA 5580

CTTAGAAACA GATATTTGTT TGGAAAACTG GGAGGAACAA GCCCtGAAGT TATTAAAAGC 5640

GATTAAAGAG AAGAGTCATT TTTACTATAC AACGGTGACC TCAGATTCAG AAGTTCTTCT 5700

TAATGTTTTT TCCGCTTCGA CCAATCGCTT ATTTATTTCA TTATTTGAAC AAGTCGATGT 5760

TGAGAACCAT TTAACAGATA AAGACAAACA ATTTTATGCG AATTTCTTTT CCTATGGATG 5820

TAGCGGCGTC TTAACGAAAT GGATTCTTGA GGAATATCCT CAAACTCCTT TGGAAATGGC 5880

GACGCAATTA TTCCGTTTAG CGAAAGATAC AGAATTTATG GCTTATCATT TGTATGAACA 5940

AGAAGCCAAT GAGTAATGTG TGAAGGATAA AAAGTTTTAT GAAGGAGTGA AACGtTGTGG 6000

TAACTGTGTT ATTTGTTTGT TTAGGCAATA TTTGTCGCTC GCCAATGGCA GAAGCGATTT 6060

TTCGTCAGAA AGTCCAACAA GCCGGTTTAG AATCGACTAT TCAAGTGTTT TCAGCAGCTA 6120

CTAGCCACTG GGAAGTTGGC AGTCAGCCAC ATAAGGGAAC CAGAAAAATT TTGGAGCAAC 6180

AAGGAATTTC CTATCAAGGA ATGCGGGCGA CACAAATCCA GCCAAGTGAT TTTAAAAAAT 6240

ATGACTATAT CATTGGGATG GATACGAATA ATGTGGCCGA TTTAAAAGCA TTAGCTCCTC 6300

AAGAAGAGCA ACCGCGGATT CATTTATTTA TGGAAGTTGT GGCTGGTAAA GAAACCATGG 6360

ACGTACCTGA TCCTTATTAT ACTGGAGATT TTGAGGAAAC TTACCGCTTA GTGGAAGCAG 6420

GTACCTCAGA ATGGTTGAAA AAAATAAAAG CACAGCTGAA AGATTAGGAC TATAGTCCGA 6480

TAGTACTTTC TGTTTCCCCT TGGTACAATA GTGTTAACAA ATTGTTCAGG GAAAAAGAGA 6540

CGGGAGGGAA GTAAGATGAA CAGTAGTTAT TTATCCTATG TTTTTGAATT GAGTTTGTAT 6600

TATTTATTGT TGATAATGAG TCTGCCGCTA GTGTATGCTG TCACGTATCA TTTATCTTTT 6660

AGTAGTATGT ATACAAGTGA ATGGTTAATA ATCTCTGTTT TCTTATCTCC GTTAGTTTTA 6720

TTATTTGCAG GTATTCGTTA TGGTTTTGCT CGTTTGAAAC AGCAAGAACG CCAAGTGATG 6780

AAATAAAGCA AAAAATCCTT GAAAAGTGTC AAAACTTTTC AAGGATTTTT TTATTAAGCC 6840

```
ACCCAATCGG CTAAAGTTTT TTGCTCCAAT TGGAAAGCCT TCAACATATC GTCGTTGGCG    6900
TCTTTAATAG GCGTCGTTGC TGCTAATGTT CCTTGAACAG CAATTCGCGT CGTTGCTTGT    6960
AAATCGCCAC AGGTAATTAA GGTAATCATA TTTTGACCAG GAACGTCATC AATTAACTCA    7020
ACACGGGTTG GTTCGATTTT TTCTACAGAA GTTATTTTGT ATGTATAAAC AGTAGATAAA    7080
TCAGTGATAT AAATGAGTTC GTCTTTTTTG GTTCTTTCTA AAGGTGAAAA TAAGGAAACG    7140
CCATCTTCCG TTCGATGACT AGCCAAGGCA TAATTGTTTT TCCCCATGAC TTGATCTTCT    7200
TTCATGGTCC CGGCACCAGT TAATAAAGCG ACATTGGACA ATCCTTTAAA AATGGGcAAA    7260
TTAATTTCGA CACTTGGTAT CGCAATGGCA CCAATCACAG GTAAGTTTTT GTTTTCAAAT    7320
TGGGCTTTCA TCACCGCTTC TGTGCTCAAG GACTCAACTG AATCAAAGTC AAACGTTGTT    7380
TCACGAGCCA TATTTTTCTT CACATCAGCT GGTTTCAACT TGCTAACGGC GTACGAGCGG    7440
CTATTTTGTT GAACCACCCA ACTACGTATC TGATTGTTAA AAATTAAGGC TAAGCCAATG    7500
ATAAATAGTA AAACTAATAA ACTGTTGATT AACCAATTTT TTCCTCTTTT TTTCTCTTTT    7560
GGGCGCATAT TTTCCCTCCT TTTAATGTAT CAAGTTTACC ACAATTTATT TGAAAAACCT    7620
ATTTATTTTC TTTGGGGAAT AGGTAAGCTG GCCGGACCTT AATTTAATAA AAAACAACTG    7680
AATCGATTGA TTTTTCAAAA AAATCCTCAT TTTGTAATGG AATTTTAAAG TAAAAGAAAC    7740
AAGCTTTTCT TTTGAAAATA TTGTAAAATG GTAAGGAGGC TAGTAGAGGG GAGATGGGAA    7800
TTTGCTTGAT TTAATTTATG TACATATAGA TGTAACCAGC AACGCGGTTT TATCTAAAGG    7860
AATTACCCAT TCCGATTTTG TTCGTTCAAT TGTGCATCAT CCACAGAACT TACTATTATT    7920
AGATCCTTCT GCCGAAGCTG GCGAATACGA TATGCATTCA GGCTTGAAAA TTATTCGTGG    7980
AGAAGAAAGC GTCAATCGCT ATTTTCAATC GCTACAACGT CGTTGGATGA CTGAAGAGAT    8040
TAAATGGATT GACTTTTCCG ATGTGACGAT GCTAAAAGAG CTAACGCCAT TAGAAATTTC    8100
GGAACTTCTT TATTTTGGAC ACATGAAGAA TAGTTTGCAT TCGCCTTTCT TTTATAAGCT    8160
ACAAAATGAT TTTGTCTTTT TTGAATTTGC AGATCAGATG ACGCGTGTCT ATTATCGTTA    8220
TATTGATGAA TTTTATCGAA TCTTTGCTGA TAAAATTACA CGTGTTGTTT TAGAGAAAGT    8280
CAACCAGAAA AAGGCCTTCT TTAAACGAAA TACGCCTGTC GAAAAATTAA GTCCTGAATT    8340
ACTTAAAAAT ATGAAAGGTA TCTTGCAAGA AGGTGTTATT TTTTGCTTCC AACAAACGGA    8400
AGTGGATAAA GAAGAATATC GGATTCCTAT TTACTTGGTG GAAGATTCAC TTAGCAAGAT    8460
GAAAAATACC TCTTACCAAA AAGAAGAAAT TTTGGCAACG CTTGTCTATA ATAGTGAAAA    8520
ACGCATTTGG CACATTGAAC AAGAGAATGA TTTCGGTGAA ATTTTTATGT TGGCCCATGA    8580
AGGCTGAGAG AGCTCTAACT TTATCAACTA AAGTAGGGCT CTTTTTGTAG CTATGATTTG    8640
AATGAGAAAT CAAGTATAAT AAAATGATAA AAAGGAAAGA AGGTTGTTTA ATGAAAAAAA    8700
TACTCTATGG CGCTCTGACT GGTTTTGCCC TCTTTTTGGT AGCTGGCTGT ACACCCAAAA    8760
CGCCTACTGC TGTTTCATCA ACAAATGAAC AAAAAGAAAC aGCAATATCm ACTAGTGAAA    8820
```

CmAGTTCATC CGTACAACaA GAaGCAaCGA cGACATCAGA GAGCCAGTGC GTAnTGGTCC 8880

CTTAGAAAAA GTACGCAGCT TTGGGAAAAT T 8911

(2) INFORMATION FOR SEQ ID NO: 652:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2533 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 652:

AGTACGCCAT GAGCAGCCAC GTGAAAAAGG CATACAACCA CTGATACCTT TTAAAACTTT 60

TACTTCGGTT ACTTGACCCG AGCCATCCTC GCTAGCCCGA AGCCAAGCAT AAAGCCCTGT 120

AATGTCCATA ATAATATAGA AGCTTTGTAA AAGAACTTCC GCATATACCC GTGTCACTAA 180

CCCAAAATAG CCAATAATCA CTGCACTGAT TAACCCAAAA ATGTAATTGC TGACTTTTCC 240

TTGGGCAATT AAATTCACAC AAATGATGCC GGCTAAACCA CCTACTAACG TAATCCAATT 300

CGATGGATTG ATAGCGAACG CAACCAACTG TACCCCCACC ATCAAAGCTA AAAACAAGTA 360

ACTTCTCGTT TCCCAACCTT TAAAATCATT CACCAAAAAA TTCTTATTCA CGACAAATTT 420

CCGACCTTCC TAAAACCAAA CTTTTCTCTG TTGACGTCTC AATAAGAGTA CTCCGTTTTA 480

AAAAACAAGT CAATCAGATT TCGGAATAAA AAGAAGATTT TAATACTATA TATTGATTCG 540

AAAAATATCr AAACACTACA TATAGTTATT AAGCAAGCTT TCTGATATCC TTTATTGCCA 600

CAATATTTTC TTCATCTTGT TTTAATCTTC TATGTATTCA TGGATTTGTT TCCATCTAAT 660

AATTCCTGCA GCAGAACCCT TTACGACGGT AGCGCTTCAA ATTAATATTG CGAAAGCGGA 720

AGCAACCGCA CATATTTATA CkTCAAACAC AGCCATCGTT CACTCTTTTT ATTGAAATGC 780

TTAAAAATTG aTGGCCTTAT TTCTCGAATC TGGCCACTTA TTAACTTATT ACATGAAAAC 840

ATTCGCCTAA AAAAGCAAAA AAAAACGACT CACCTAAATA AtCGACCTAC AAACAATTCC 900

TkTCATTTGT AGGTCGATTA TCTTTATTAG GTTTGATACG GATGTTTCGC CATTCCTAGT 960

TTCCCTTTAC TGCTCATCTC TATTTATTTT TTTACTGTTT GCCAACAGCT CGTTGTCTGA 1020

AAGTACTATC CGTCCTTTTG TACTTtGTTC AAAGCTGTAA GCTkGCkCAA ATAACGTTTG 1080

cwCTTtATTG GcAGTTTGTG TAAAAGTCAG CCCATAAGGT TGTCCTTCAC TGTTTTTCGC 1140

CAAAGGAACA GTCAATTCAG GATATCCTGC CACAGCACTT AAACCCGTAC CTTCAGAATC 1200

AATAAAGGCA TATCCGTCTA ATtGTTGTTT TTTtAGCAAC GCATCAAACG TTTGCTGTGC 1260

TTTTTTAATG GTTGTTTGAa TGATTtCTTT ATCTGGTTGT TTCTTTTTAA CATCTGCTTC 1320

TAATAAATCT TGCCCATACT gGCACGAACC TTTTTATCTT GTTGATTATA CGCAATTAGT 1380

TCCTCCAGTG TTTTAAATGG TAAGTCATAA CGTTTCGCAA ATTCTTCCAT GCAAACTTAA 1440

ATTCATTGGA AATAACTTTC CCACCATCAA TCCCTTCCAT GTCTGGCGTT ACCGGAATCA 1500

CTTCTGCTCC TGCTTTTTGG AGGGCCTTTT TAGCAGCCGT AAATTCTTCT TCAGCACTTG 1560

GCATTTCTAG CAACCCAATG CGTTGACCTT GTAAGTTATT TTTAGTAAAT TTCGTTGAAT 1620

CAATGtCTCG GAACCGTGTC TTTCTTAAAC GCATTATAGC CTTgCGCAGC ATCAACGACA 1680

CTTCGGGCAA TCGGGCCAAC TGTATCTAAT GACGGTGCCA ATGGCAAAAC ACCTTCGCCA 1740

CTTACTGAAC TGTGGGTCGG TTTAAAACCA ACCACTGATT GATGAGAAGC GGGGGAAACA 1800

ATCGATCCCA TCGTCTCAGT GCCTAAAGAG AAGGCGCCGA TATTCATCGT CACACTAGCC 1860

CCACTCCCAG AACTTGAACC TGAAGGTGTT ATTTTGAGCG GCCCATAAGG ATTAAGTGTT 1920

TGTCCATGCT TGGAACTATA GCCACTTGGT GCTTTCATTG ACATATAATT CGCTAGTTCA 1980

GATAAGTTTA CTTTTCCTAA AATCAAAGCT TGCGCCTCTG TTAACTTCTT AACTACTTCC 2040

GCATCTTCCT TAGGCTTGAA CGTGCGTAAG GCATAAGCTC CTGCACTCGA AATCATGTTG 2100

GTCGTATTTA TATtTTCnTT CAAAGTCACA GGAATACCAT ACAAAGGCTT CTTCGGTGTT 2160

TGAGAGGCTT GCTGGTcAAA TGCCTTGGCT TCCTTAATTG CTTGCGGATT GATTTCAGAG 2220

ATGCTGTTCA TGCCaTTATC TATTTGATCG AACTGTAAAA TACGGTCCAA ATAAAACGCC 2280

GTAAGTTCTT CATAGGTCAA TGTTCCGTTG GCAATCGCTT GCTGTAATTC AGTGATATTT 2340

TTTTCAATAA TCAACGGTGC TTTGGTCCGA GCTTTTTGTA CATCTGTTTC TTTTAATTGG 2400

TCTTCAATCG TTTTAACGAT TCTCTCTTGA TTGTACATTG CTGGGGTTTC TGCCTCAGGA 2460

GTGAACTTTT TGAGTATAAA CCAATACGCT GCAGCGCCTA TTATGACTAA CAAAACAGCA 2520

ATTATTCCGA TTA 2533

(2) INFORMATION FOR SEQ ID NO: 653:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1447 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 653:

AAACACTGTG GGTGAAAAAT GGCCATCACG TGTTCTGGGT ACTTTTAATT CTAGCGTGCC 60

TACACGTGTC GTAAAGCTGC GCTCATAATA GCCATTTCGT TGACTTTGTC GGTTTTCTGT 120

TCGTTCATAT TCTTTTGCTT GAATATATTC TGTTCGTTGA TTTTCCATTA GTTGATTAAA 180

TACCGTTGTT AAAATATTTT TAGAAACGTC ATCCTTTACm GAATATTCAA TAATGCTTTG 240

AATCTCTTCG CTTTTCAGTG TAAAATGTAc TTGGGTCATG TAAAAGTCCT CCTGGGTATG 300

TTTTTGTCGT TAAAAACATT GTACCGTAAA AGGACTGTTA TATGGCCTTT TTACTTTtAC 360

ACAATTATAC GGACTTTATC AATATTTGCC TAATATAAAT ATTTAAGTTC GCTCTAAACA 420

CTAATTCTCC AGTTGATCTA TTACTAAAAT ACATTAGTGG CATACTAAAA AGTTTATCAA 480

TATAAGAAAA CATTAATTTA AAAtCAAaTT CATACTGAAa TTCTGCTACT ACAGAGCTTT 540

TTAAGTATTG AACTACATAT AAAACAGAAA AAAAACTAGT TAATATTAAT AACAGAACAT 600

AGGTAGGAAT TTCTTGAAAA GATCTAATAT TATCTATCGA ATACTTTGTT GCAATAGGAA 660

TTAAGAGTAA TAAAAGTTGT GAAACAAATG ATAATATTAT TAAACTAAAG AAATATCTTT 720

TGAATTTTGT AAAAATAAAT GACTTTAAAA AAAACTGTTT CCTTTTGCCT TCTTTTTTAG 780

TCTTTTTCTT ATGTGCGTAT ATTAATATAT TTGAAAAATT TTTTTTGAAT TCTGAAATAT 840

CAATCCAGCG TTTATTACTT GCAGGATCTA AGATTAATAC TTTCTTCTTT TTTATTTTCT 900

CTATGACCAC AAAATGTTGA TTATTCCAAT AACTTATTAC AGGAGTCGGA AGATCTAAAT 960

AATTTGAAAA ACTTGATTTA AATGTCGATA CATCAAATCC ATATTCGTCA AAGACAGTAC 1020

GAATATTCTT GATAGTTAGT CCTCCTTTGG GCACCCCATA TTTTTCCCTT AGTTCTACTA 1080

GTGTACTTTG ATTACCATAA TAATTAAGTA GCATAGTGAT ACATGCCAAA GCACATTCAC 1140

TATGCTCTCC TTGCGCAACG TACTTCAATC TTTTCATAAT TACACCTCTT ATAACTCAAA 1200

AAGTAATGCA TTCGGAATTT CAGAATCTAG GTTTCTCAAT AATTCATAAC CCACACCACT 1260

GATTCCTACA AAGAATCCTA GTGATTCTAA ATACTCACTT cCGGCAACCT TTAGTGTATT 1320

AtTTTTTTCa AAATCaTTTA gCATATAGCT AATAAGTTTA TTTTTTTTAT ACTGATAAGT 1380

TTCAGGATCT TTTTTAGCTA ATTGAATTAG CCCTTCTAAA GTACCTGCAT TTCCATGGCA 1440

CAGACAA 1447

(2) INFORMATION FOR SEQ ID NO: 654:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 581 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 654:

TATGAAACAC GAACATTATA TCGAnTATAT ATATAATAGT TAATATTTTG TTGTAAAAAG 60

ATAGCGATTC TCTTATAATC AAGAGTATTA TTCGAAAACT GGATTTTAAG GGAGAGAAGT 120

GCAATTATGA ATACGTTTTT CAAATTAAaG GAAAACAaGA CAACGGTATC AACTGAAATA 180

ATGGCGGGGG TCACGACCTT TTTTGCAATG AGTTATATTT TATTTGTTAA TCCGTCCATT 240

CTTTCAGAAA CAGGGATGCC TTTTCAGGCT GTTTTTTTAG CGACCATTAT TGCTTCAATC 300

ATCGGTACAT TAATCATGGG GCTGTTTGCT AATGTACCTT ACGCTCAAGC ACCAGGGATG 360

GGACTCAATG CTTTCTTCAC GTTTACGGTC GTTTTTGGGA TGGGGTATAC ATGGCAACAA 420

GCATTAGCAA TGGTTTTTAT TTGTGGTTTG ATTAATATTT TGATTACAGT GACAAAAATT 480

CGCAAAATGA TTATTAAAGC AATTCCAGAG AGTTTACAAC ATGCTATTGG TGGTGGGATT 540

GGAATTTTTG TCnCTTATGT AGGGTTGAAA AATGCTGGCT T 581

(2) INFORMATION FOR SEQ ID NO: 655:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 533 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 655:

TTTTATCATA GATCGGAGGT TATGAGGATG AAGTTAGATA TGGTCGGTAT TATCGTCAAC 60

GATATGACAC AAGCACTCTC TTTTTACCAA ACGCTAGGTT TTAACGTGCT GGAAGAAGCT 120

TCGCCTGATT ACACAGAATT ACAAAACAAT GGCGTCAGGA TTTCTTTAAA TTCACTACAA 180

ATGATTACTG GTGTCTATGG GTTTGCGCCT AGCCAGCAAG GCGATAAAAT CGAACTTGCT 240

TTTCTTTGTG AAcGGCCGCA GAAGTTGATA ACACCGTGAA AAAAATGAGC CAAGCTGGTT 300

ACATTGTTTT TAAAGAACCT TGGAATGCAC CTTGGGGCCA ACGATACGCA ATCATTAAAG 360

ACGTGGATGC AAATTTAATT TCACTCTTTG CCCCTCTATA ATATTACTAA nAmATGCGCC 420

ACTTTCGTTT CTTTTACTCT CTTTCGgAAA GATTAGTGCA GAATGGGtCT TAAGTAGGaT 480

GTCyTTTCtG tCyCtTTTTT TTGGTATAAT GAACCTTATC CAATGGATGG GGn 533

(2) INFORMATION FOR SEQ ID NO: 656:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1061 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 656:

GCACAAGCAA ATGACATTTG AAGAAGTATT GCCACAGATT AAAGCAGGAA AAAAAGCCGT 60

ACGTAAAGGC TGGAGCGGTT TTGAACTTTT TATTGAATTA CGTGATGAAA TTGGCACCTC 120

CGAAnGTGAA TTTTTACAAG TTACCCCGTA TTTTTTAATC AAAACTTCTG ATGAAGGCTA 180

CAGCATGTTT TCACCAACAC CCTGTGATGT CTTAGCAGAA GATTGGGTGA TTGTTGAAGC 240

ATAACTACAC GCATAAGGTC TGGCAGATTA CGGCGATTGC TCGTTCTTCC AGACCTTAAT 300

TTTATAAATG TTTAAAGGAG TTTTTCTTAT GCCAAAAGAA TTTGAAACAA AACGGGTGCT 360

TGTAACCGGT GCAGCTTCAG GGaTTGGCCA AGCACAAGCA ATTGCCTTTG CTGAGCAAGG 420

TGCTGAAGTT ATTGGCATCG ACCTAGACGA AACGGGGTTA AAGCAGACAG CCGCACTGGT 480

TGACCCAGAT TCTGCTAAGT CGTTTACTTA TTTTGTCGGT GATGTGTCTT CTCCCTCATT 540

TGTGCAAGCC ACGATGAAAC AAATTGTGAA AAACAACGGC CAAATTGATA TTTTATTAAA 600

TACGGCAGGT ATTTTAGATG ATTATCGCCC TTCTCTAGAA ACTTCAGAAG CTTTATGGGA 660

TCAAATTTTA GCAACCAATT TAAAAAGTGT CTTTTTAGTG ACCAATGCCA TATTACCTTA 720

TTTCCTCCAA CAAAAAAAAG GAGTAATCGT TAATATGGCA TCTATCGCTG GCTTAGTAGC 780

TGGGGGCGGC GGCGCAsGTA gcACTGCCTC CAAACACGCA ATCaTCGGGT ATACAAAACA 840

ACTTTCCTAC GATTATGCCA AATTAGGCAT TCGAGCAAAT GCGATTGCGC CAGGTGCCAT 900

CCArAcACCC ATGAACGCAG CTGATTTTGg CAGGAGAAGG TGAAATGGCT GCyTGGGtAG 960

CACGTGaAcA CCCGCGGGCC GTTGGGCACA GCCACAAGAG GTAGCnAAAC TTTCnTTATT 1020

TCTAGCTAGT GnTGACGCTG ATTATATCCA TGGCACAGTT A 1061

(2) INFORMATION FOR SEQ ID NO: 657:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 488 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 657:

GAACAGCCGC AATGTGTCAA GACACATAAA AGAACTTTCT GTTATTTCTC ATAAATAACA 60

GAAAGTTCTT TTTTATACTG GTTTATTTTT TTCTAAAGCA TTACTTAAAC GAGCGAATTC 120

TTCTAAAGAC AATGTTTCAC CACGTCTAGA AGGATCAATT TCTGCTTCAG CTAGGCTCTC 180

TTTTAGCCAA GCCACCGTTT GTTCATCTTT CCCATAAAAA TGCGTTAAGT TATTCCACAA 240

TGTTTTACGT CTTAATTGGA AACTGGCTTT GGTTAATTTG AAAAATTCTT TTTCATTGGT 300

CACTGTCACT GCTGGTGTAG CACGGCGCGT CAATTTAATA ATCGCCGAGT CAACATTGGG 360

cTGCGGCACA AAAACTGTTT TTGgGCACAA TAAAGGCGAC ACTTGcTTCC ATAAAATATT 420

GGACAGCAAT TGaTAATGAG CCATAAGCTT TCGTTCCTGG TTtAGCAGCA ATTCGATCAG 480

CGACCTCT 488

(2) INFORMATION FOR SEQ ID NO: 658:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1906 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 658:

TAAAGGnGTT ATTTTAACGG ATACTTTTTA CAATTATTCG GAGCAAGTGT AGCCAATATT 60

CCnGGAAGGC TTTCAACAAA TGnCnCTGTC ATTACCAGAT ACCGTCTTGG CCTTTTTGCC 120

CAGCTTTGAT TGTTTGGTCT ACCTTTAAAG CATTTAATGC GACGCCTGTA ATCGGGATTG 180

TTATTGGTTT AATGATGGTA TCACCCATTT TACCAAACGC TTATGCAGTT GCAACACCAG 240

ATTCTGGCGT GAAAGCTATT ATGGCTTTTG GTTTTATCCC AGTTGTCGGT GCACAAGGGT 300

CCGTTTTAAG TGCAATTGCT GCTGGTATAA TTGGCGCAAA GATTGAACTT TTTTTCCGAA 360

AAAAGATGCC GAATATTCTT GATCAAATTT TTACACCCTT TATGACTATG TTAATTACTT 420

TTTTAATCAT GATTCTCGGA ATTGGCCCAA TTTTGCATAC CGTTGAGTTA GGCATGGTAG 480

ATGTTGTCCA ATGGTTGATT GGGTTGCCCT TAGGATTAGG TGGTTTCGTG ATTGGTGCCT 540

CATATCCTTT AATGGTTTTG ATTGGTATTC ATCATACACT AACTATGGTA GAAACATCAT 600

TATTAGCAAA TACAGGATTT AATGCATTGA TTACAATATG TGCCATGTAt GGGTTTGCCA 660

ACGTTGGCAG TTGTTTAGCC TTTGCGAAAA AAgCGCAAGA TAGCAAAGTG AAGTCGACAG 720

CCATTGGTTC TATGTTGTCT CAATTATTTG GAGTTAGTGA GCCTGTGTTA TTTGGATTAC 780

TTATTCGTTG GAACTTAAAA CCGCTGCTAT GTGTCTTGTT TACATCAGGT TTAGGCGGAG 840

CAATTTTAgC TATTTTTCAT ATTCAATCAA ATTCTTAkGG CTTAGCCGTT ATTCCTTCTk 900

TCTTAATGTA TmTCTATAGT GCACATCAAT TAGTGATCTA TTTATTAGTC GCACTTTTAT 960

CTGTCGGTGT ATGTTATGCA TTAACCAGTC TATTTGCAAT TCCGCAAGAA GTTTTGATTT 1020

CAGATaAAGT AATCGAAGAG GAAGAACGCG AAGTCTTTGA AATGCAACAC AATACCTTAG 1080

ATGAACAACT TTTTTCTCCG GTAACTGGTT ACGCCATGAA TCTTACTGCT GTTAATGACC 1140

CAGTATTTTC AAGCGAGATG ATGGGCAAAG GTTTAGCAAT CATGCCAACA GCGAACAAGG 1200

TTTACGCGCC TGCAGATGGT TTATTAAACT TGGTTGCAGA AACTGGGCAC GCTTATGGTA 1260

TCCAAACGGA CGCAGGAGCT GAAGTATTAA TTCATATAGG AATTGATACT GTCACATTGG 1320

GACAGGAGGT CTTTCAGACA CAAGTAACCC AAGGTCATCG TGtTAAAAAA GGAGATCTTT 1380

TAGGAACGTT CGATCGAAAA gCCATTAAAG AAGCTGGGCT TGACTCAACT GTCATGGTTA 1440

TCATTACAAA TACTTCAAGT TATTTAAGCG TAGAGCCAAT GATGTCAGAT CATAATGAAA 1500

TTACGCCGGA ACAAATCATT TTAAATCTAA ATACACCTAA CTAAACAAAA GArCAGCAAG 1560

AAGTAGTTAT CACTTCTTGC TGTTCTTTTT CTTAATAAAA AACCAATGAT TTTCGGTTGT 1620

ATTTGTCTAT CCGTTTGTAG TATGATTGAG TCAGTTTTAA AGGAAACATG CCAAGTATGG 1680

AAGACCATAC TTGACACGCA TTGACTAATT AAACATGGGG AACGTTGCCA TAAGTTGATG 1740

TTCCAGTCAA TAATGCAAGT AAATCmAACc ATAGCATGTA CCTCCTTTAA TTGATAATCA 1800

AAGTATAACC ATTTTGnAAT GGAATTTACC ATAGATAATA CCAAGTGGAG GTCTGTnCGT 1860

GGAAGGAGnC AGGAGGCCAG AAATTAAACC AAGCCGGGGA AAnCTT 1906

(2) INFORMATION FOR SEQ ID NO: 659:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1150 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 659:

GGCTGAAAAG GGGTGGCTTC AnCCCAAAAA CTTTTTTGGG CTTCCGCnAA CGACCGCTGG 60

TGGGTCAAGT CACnGCTGGT GGACTTGayC CAAgCACCTT GtTTtGAtCG ACCACTGCTt 120

GgTTGGTCTT tAAaCAGATt GgAtCAATGG TGtCTTTTtC TTTTTTCGCA GTGTCAGCAA 180

CTTGTTGTTT TTGATCGACA ATTGCTTGTT TCTCTGTTAC TTGTTGCTGT TTTTCAGTAA 240

TGGCTTGCTC CGTTGTTTCT GTAGCTTTCA CTGCTGGTTG TTCTGTAGAA CTGTTTTCAG 300

GTGTTTTTGG TTGCGCTTGT TCTGCTgCCt GTACTTCATT TCCCCCAATA CCTGCTACCC 360

CCGTTGCGAC TAAAATACTT AACCCTGCAC TTGCGATCTT TTTCATTTTA TTTTCCTCAC 420

TTCCGATTAT TTTTGTTGTA AATGGGATAC CCATtGTCTT TACGCCATtG TTGTAACGCT 480

GTAGTTGTTA TATGGAACTC TTCCGCTATT TCTTTAAACT GTAAGCCCAT TTTTTGATAG 540

GCTTCAAATC TTTCCTTGCT AATATGCGAA GAAAATTGTC TGGTGTTTGC TGTTTCTATC 600

TTTCTTGCAA GTTCTTGTAT TCGTTTAAAA CGCTCGGTAT CTTCATTCCT ATACCAGTCT 660

GCGTCTTCAT TCATTAATGC CAACATTTCT TTACGCCATA TTTTTATTAA TTTTTTAGTC 720

ATACCCCTTT CACCTCAAAT TAACAATAAC AAAATAATAT TTATTATGAA AATAGTAACC 780

TTTTAGAGAA AATAAAAAAT GTCTTTTACA TGACAATTAT TAAATGCCAT AAACAACATT 840

ATCTTTAATT TCAAATTCAA ACCATTCTGA AAAATCTTGT TTAAATCCAT TGACGCTATA 900

TTTTTCTTGT AAAACTACTT CTTCCGCATT CGCAGGGTCT AAATAGGTCG GTTTATCTTT 960

TGGACGCGCA TAAACACCTA CCCAACTTGT AAAATCCCCC TCAAAAGCAC CATCAATTTC 1020

ATAACCAACT TCTTCTAGCT TCTGACGAAC TTGTTTCTTT ATATCAATAA TATCTTTAGG 1080

aATATAATTT TGaTACTTTT CTTGAAATTC CTTAAAAGCT AATACTTCTt TACCATAGAA 1140

CCTTTTTATn 1150

(2) INFORMATION FOR SEQ ID NO: 660:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1175 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 660:

TCnGATTnTC TTATTATTAA CGTAGTGGCT TATTCTAnTA TCCCCAAAAT CCCCACTGTG 60

TATTTAATCT AATCTtAAAC TGATTTTCCC ACGAAAGTTT TATAAaCCCA AATATTAATG 120

ACTCTGGTCT AcTTTATTAA ATAATCCAAA AGCTAAATGA CCTAACTCAT GTACAAAGAT 180

AGACAATAAT AAGGCAAAGT ATGCTATAGA TACTAAATTT AATTCAAAAA CAAGAGTAAA 240

GATTACGAGA AGTAACGTAA AGATCATCGG AGATATAAAT TTCATCTATT TTATTCTCAC 300

CTCTTTGTAT TTAAGCATGC TATTTTTATC TATAAAAGAA ACAGCATTTT ACAATATTTC 360

TATATGATCT AGTTTCAACT TCTATATTTG TTTCTAATGC TGTAAAACCA ATCCCAATCA 420

TGCCTCAAAT AACTAAACTT TTAACTAACT TCTTTTTCAT TAGTATTCAT CTCGAAAATA 480

AAAATTAGAA AACGTCTTCC ACGCAAGAAT TGTAGCTCCA TTGATCTAGT TTTGTAAACA 540

TTGAGACAAA GGTTTCTATC TAACAATTTT GTAAGTTGGT GCCCTTCGAA TTTGATTAAA 600

TTAGATTTTA TTCCGTTTTA TTTCAAGGTA AGTATTAATA CAAGTTTTAA TTTTTATAAA 660

AAAATGAATC TTACCATAGA GATTTACTAA CTAATTATTT ATACTTTTTA TAATAGATTG 720

GAGGTTACTA TATGAGGGAA AAGTTATTAA TAGGGAGTTT TTGGTTATCA TTCTTTAATA 780

TTGTTTGTAA AATTCTCGGA TTCATTTATC TAATACCTtG GCTAAAATTT ATGGGrACTA 840

TCCACAATCA ACAAACTGCT CAAGCTATTT ATAATGTTGC ATACTTACCT TATGCTTTAT 900

TCTTATCACT TGGTACAGCT GGATTTCCTA GTGGCATTGC TAAAAAAATT GCTGAATTAA 960

ATATTAATGG AAATAAAAAC CAAATAAAAG AATTGTTCAA AAGTGGCTTG ATAGTTATGG 1020

AAATAATAGG GATTCTCTCT GCTTTAnTAA TGTTTATTTT TGCTCCTACA TTAAGTAAAA 1080

TAAGTCCAAT AGTTGATCAT ACTGCGGGTA TCACAGCAAT CCGTAGCCTT TGTTTTTCTn 1140

TATTAATTAT TCCTAnACTT AGCGCACTAA GAGGG 1175

(2) INFORMATION FOR SEQ ID NO: 661:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 965 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 661:

TAGTCCAAAC CCCCTTTTAT ATAGGGCCAG GCGCCTTAGG CTATCCACTA ACnAnAAAAT 60

CGGATnTCTG CCCCCTTAAT CATTGGTTTT CCCCCCCGTC AAAGCTACTA ACTGTTGATT 120

AATTTTnTTA ATCTTTTGCT CCTCCCATTT GAATTAAATG TCGCGATTGA TCTCGTTTTC 180

TTAACAAAAT AGTTTCCaAA AAAGAAGCAT CCTTAAGTTT TCTCAAAGTT TCCGCCATCT 240

CTTCAACTTT AAACCCTGAA TCTCTTAACA GTAAAATTTG TTTCAATAAA GGAATTTGAT 300

GTGCGCTATA ATAGCGATAT CCTGTTTTTT TATCAATTTC ATTTGGAACA AAAATTCCTG 360

CTTTTTCATA ATGACGCAGC ATCCGTGGGG AAATATCCAT TAACTTGGCA AATTGCCCAA 420

TTTGATACAT AGAGTCTCCT TTTCATTCAA TTGATATTCT TATCATAAAC CTTGACACCA 480

TGTGAAAGTC AATGAAAAAG TTTCTATTCA TACATTTCTA TTTTATTTGA ATATTCCCGT 540

GATCTTTCTC TAGTTATTCA GAAGTATATT GAGGGGTGAA ATTTTGGAAT TAAAAAAAAC 600

GATTTCCAAC AAAATAATGA TTATTTTAAT TGCAATTATC GTTGCAATCT TTGCAATGGG 660

ATGGATTCTA CCTATTGGCA TCGATAAAGT AACTAGGTTA AGTTACCGAG AATACTTATT 720

TAGTACATAT ACTGTATTTA CACAATTTGG CTTTCTAATG TTTTCGTTTC TAGTCTCTTT 780

CTTTATTAAT AAAGAGTACT CAGGAAAAAC AATTTTATTT TATCGTATGA TGAATACCAA 840

TAGCCTTACT TTTTATATAA AAAAAGTTTT GACACTTACG GTAGAGACCC TAGGCAGTAT 900

TCTAGTTCTG TTATTTATCG TTTCCTTTAT TTTTATGGAT TTTTCTGTAA TcTGCAAATG 960

TTTTT 965

(2) INFORMATION FOR SEQ ID NO: 662:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1610 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 662:

ATGTAACAGT TTACAACGAA GAATAAGAAG ACTATnATTA TGTTAACTAT TTTTAGGTGT 60

GCCTAAAAGT GAAATGTGTG AACAAGATAG ATTGAGGGAT GTTGGATGAA AGAACAGAAA 120

ACGGaAAAAA TGCTTAATTA TGCAAATGGA CCTAGTTTAG AAGAGATAAA TAATACCGTC 180

GAGGTACCTA AAAACGCAAG CTTTTGGCGA ACTTTGTTAG CCTATAGCGG ACCTGGCGCT 240

TTAGTGGCTG TTGGTTATAT GGATCCAGGA AATTGGATTA CGTCAATTGC TGGCGGAgCG 300

GAATATAAGT ACACGTTACT TAGTGTGATT TTGCTTTCTA GTTTAGTTnC CnGGGGAnTA 360

CAAAGTATGG CTGCGCGCTT AGGAATCGTG ACAGGAATGG ATTTAGCTCA AGCGACTAGA 420

GAGCATACAA GTAAACGAAC GGGCGTTGCC TTGTGGGTGG TGACGGAGTT AGCCATAATG 480

GCTACAGATA TCGCTGAGGT AATTGGTGGT GCCGTTGCTT TGCAATTATT ATTTGGTTTT 540

CCATTATTAA TTGGTGTGTT GATAACAACG TTTGATGTTT TATTACTGnT GCTACTGACA 600

AAGTTAgGCT TTCsCaAAAT CGAAGCAAwT GtTTCTTGTT TAATTGCAGT CATCTTTTTT 660

GTTTTTGCTT ATGAAGTGGC ATTAGCAGAT CCAAATGTTG GTGAAGTATT ACGAGGTTTT 720

ATTCCAGACA CAAAAATAGC GACAGATAAA TCCATGTTAT TTTTAGCCTT GGGGATCGTT 780

GGAGCGACAG TCATGCCCCA TAACTTATAT TTGCATTCTT CCATTGCGCA ACACGGAAAa 840

TTTGATCGTA ACGATGATGT TGAGAAAGCC AAAGCAATTC GTTTCACTAC TTGGGATTCA 900

AATATTCAAT TAACTGTTGC TTTCGTCGTA AATTGTTTGT TGTTAATTTT AGGAGGAGCA 960

TTATTTTATG GAACCAACAG TGAATTAGGT AAATTTGTTG ATTTATTTGA TGCTCTGAAA 1020

AATCCAGATA TTGTTGGTAA TATTGCTAGC CCAGTGTTAA GTATTCTTTT TGCAATTGCT 1080

TTACTGGCTT CGGGTCAAAA TTCAACGATT ACAGGTACGC TGTCAGGACA AATTGTAATG 1140

GAAGGCTTtA TCCATTTAAA AATGCCGTTG TGGGCACGAC GAGTGTTAAC TCGTTTGTTA 1200

GCAATTGTGC CAGTCATTAT CTGTGTCATT ATTTATGGCG GTAGCGAAAC GGCTGTTGAA 1260

GACtTGTTAC TTTATACGCA AGTTTTCCTA AGTATTGCTT TgCCAGTGTC GATTATTCCT 1320

TTAACGATGT ATACAAGTGA TAAAAAATTA ATGGGCCAAT TTGCGAATCC AGCGTGGGTG 1380

AAATTTTTAG CTTGGATTAT TGCTATTGCC TTGACGTTAT TGAATTTATT TTTAATTTAT 1440

GGCACATTAA CTGGTTTAAC GAATTAATAA AAACAAAGgC TGACTAGAAA AGTTAGCCTT 1500

TgTTTTTTCA TAGAATTTGC GCGTAAGAAA TGGTAGAATA GGAAAGATGA AATACAAGGA 1560

AGAGAGGCTT TCACGTGACA GATTATTTAA ATGTAGGAAA AATTGTGAAC 1610

(2) INFORMATION FOR SEQ ID NO: 663:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1029 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 663:

AAATTTAGGT GGTGGCGGTC ACTTTACTAA TGCGGCAGaC AATTATCAAA CGTAmCAGTA 60

GCAGAAGTAA AAGAGCAACT ACTTGATGTA ATTCGTCAAA ATATTAATGA AATGTATGAA 120

CAGGAGTGAA GGAAATGAAA GTCATTTTTT TACAAGACGT TAAAGGAAAA GGAAAAAAAG 180

GGGACGTAAA AGAAGTGCCA ACTGGTTACG CCCAAAACTT TTTAATTAAA AATGGTTATG 240

CGAAAGAAGC CAATAAAGGG AGCATGAGTG CGTTAgcGGG TCAAAAGAAA GCCCAAGAAA 300

AACATGAAGC AGAAGTTTTA GCACAAGCAA AAGAAATGCA AGCCTTTTTA GAAGATGAAA 360

AGACTGTTGT AGAAATCAAG GCAAAAGCTG GAGAAGACAG CCGTTTATTT GGCTCAATTC 420

CTTCAAAACA AATTGCAGAA GCCTTAAACA AACAATACAA TGTAAAATTG GATAAACGCA 480

AAATCGAATT AGCCAACCCA ATTCGTTCTC TAGGTTACAC TAACGTACCA GTCAAACTAC 540

ATCATGAAGT TACAGCGAAA ATTAAGGTTC ATGTCGTAGC GGAATAATAC AATCGATAAA 600

AAAGAAGCCG ATTTCTCCGG CTTCTTTTTT ATGCAAAATA AATTACTAGA ATTATTTTTT 660

ATAATATTTA AAAATCAAAT ATACAACAAT TTAAATTAAC GAAAGTATTA TATTTTTATT 720

AATTGTATAT TTTTAAAAAT TGGAAAGTGA ACTAAAAATA GAGAAATCAG CAGAGAAAAA 780

ATGTACTTTT TTAATGAAAT TGTGTTAAAA TGGATAGACT AAACGAACCA GAAAGTAAGA 840

CATCGTGGGA GAGAaCAAAA TGAATGAATT ATTGCAAGAT AGGGTACCGC CACAAAATAT 900

TGAAGCGGAA CAAGCCGTTT TAGGTTCTAT TTTCTTAGAT GCAGATGTGC TAATAGAAGC 960

GATGGAATAT GTGGAACCTA AAGATTTCTA TCGTCGTAGT CATCAGTTGA TTTTCCAAAC 1020

GATGATGAC 1029

(2) INFORMATION FOR SEQ ID NO: 664:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 567 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 664:

CTTTCTCAGA AAGAGAAATA TGATAAAAGG AGTTGTTCAT ATGGACTATG CTAAACTAGC 60

TTCCGACATA ATTCTAGCAG TTGGAAAAGA TAATCTTATC GCTGCTGCAC ACTGTGCAAC 120

ACGTCTTCGT CTCGTACTAA AGGATAATAC AAAAGTTAAT CAAAAAGCTT TGGATGAAAA 180

CCCAAATGTA AAAGGTACCT TTAAAATTGA TGGTCAGTAT CAGGTTATTA TTGGTGCTGG 240

TGACGTTAAT TTTGTCTATG ATGAACTTAT AAAAAAAACT GGCCTTTCAG AGCTTTCAAC 300

AGACGACTTG AAACAAATAG TTGATAAAAA TAAAAAATTC AATCCAATTA TGGcATTAAT 360

TAAACTACTA TCAGAAATTT TCGTACCAAT TATTCCTGcA CTTGTTGCaG GTGGtTTATT 420

AAkGGCTCTA CGAAATTTTT TAACTGCCGC GGGtCTTTTk GGACCAAAAT CaATCGrAGa 480

AAgTATCCTG cAATTAAAGG AATTTCCTCT ATGATTCAGC TTATGTCAGC TGCACCATTT 540

ATGGTTTTTA CCATTCTAGT TGGTATT 567

(2) INFORMATION FOR SEQ ID NO: 665:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2442 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 665:

TTTCCGTTTT CTGCAGTGAT GAAGCCAAAG CCTTTGTCTG AGTTAAACCA TTTTaCTGTA 60

CCTGTTTCCA TAATAAATAT CCTCCTAGAT GTGTGTTATA ACACATTCAT ATTTTGCAAT 120

TACTTGAAAG GTGAAAAATG TGGAAGATGT TTATTTGAAA CAGCTACCCA TATTACCGAA 180

CAAAAACTAC ATCTATTACT ATAGCATACT AACTGGTTTT TGTATAGCGA TAACCTTATT 240

TTCTTTTTTA TCGGATGTTT TTTAAAAACT AGAAAGTGAA GATAAGGCTG TTATAGTGCT 300

GTTTCAATAA GGCTCAGAAA AAAAGCCCAT TGTAAGGCGC TTTCTTACAA TGGGCTTTTT 360

GCCTGTTTTT AATAACGGAG GATAGCCGCT AAAGATTTTT CATCTGGTGC AGCTTTTTGG 420

TCTAAAACAA AAACTTGACC GCCGTTTTGA ATGACTTGGT AAGCAATGGT GTTTAAAAGT 480

TTGCGACGAT CATACTCTTC AGATGACATT TCATTGTTTT CAGTAACAAA GAGTGACGTT 540

GCAATAAAAA GATGGGAAAT TTTGCCGTCA GCTGCCGCTG GCGAAATATC TGTAAATTGA 600

TCTAAGAATT TACGATCTAG TAATTTGTTA TAAGAAGCTG TTTCTTTTTC TGTTAACTCT 660

GCAGCTATTT TTTCACTAGC GCtGCGGATA TCTkGAATGG TTGCTTGTGC AGGTGAAGCA 720

GCAACTGAAA TTGAGCAATC ATAATAAGGA TTTTTAGCAA TCTTTTTAAA GAGTGTTTGG 780

TTTTCTGGTA AAgCATAAAG ATAAAGCGGT AATTTTTCAG GATTATCCAG TTGATCTTTT 840

AAGAAAGTAT CGACTGCTTG ATAGTAGTTC ACCCAATCAA TTTCAACTTC TTTATCTTTA 900

GTATTGATGC TATGATAAGC GACACCTTCT TTAGAACCGT TAGAACTACC TTGAGTTGAA 960

AAGTTTAAAT TACCACCCGT TAATTCATCA CCTAAAGCTG TGACCACGTC TGTTGGCGCG 1020

TCTTTAGGCA ATTCAACAGG AGTCACAGTC GTATGATCAA CCTTGTAAAG TTTCATTGAA 1080

TCGCGATTTA GAGCAAGTAG GTAATAATGG TAGTTAAATT GTTTGTCTTT GACTAGTCCA 1140

AGTAAATAAG GAATCGTGTC CACGTAATAT TGATCGTCAA CGGGCACATC TAATCGTTGA 1200

ACAAAGGTGT TTTCTGGAGA TAAGATGATT GAAACACTTG CTGTGCCGTT GCGCCAGAAG 1260

GAAGCATCAG CTAGAAGTGC GTCAATTTTT TCTTGGAAAA CAGACCATTT TTCCTCAGGA 1320

AATTTTTTTT CAAAGCGTGT TTTTGCCGCT TTTGCAAAAT TTTTCAAAAC AAGCGAATCT 1380

TTTTCTACGT TTTgATGAGC GACATGGGTG TTAAACAAGA AAGTAACAAA CGGTCCATGA 1440

aCTTCTTCAG AAAATAATGT TGATAAGCTA TCTTTTTTTT GATTTGCCAT AATAAAATCC 1500

CCCTTCTAAA AGTAAGTGTA GCACAGGAAA TCGAATGTTT AAAACAATAT GTTCAATAAA 1560

AGArAACGCT TTAAATAATC TAAATAAAGG ACTGTTAATG GTGTTCTAAT AAAAGTTTTA 1620

TAAAAAATAT AGCTGAAACT CATGATAAAA AAGGCACAAA CCACTATAtT ACGTTATGCC 1680

AATTATCCCG AATTGGTTTT TTCATACTTA CTCCTACCAA GAGTAAATAA CAACCTTCAT 1740

TTCCACCACA GACGACGGTC GTGGTGGTTT TTTTTGTTTA GGTGCTAGTT TTTTTGCTAT 1800

CAATTGACTA GGAGAAGTGA ACGTTCTTTA AGAATTTGTC ATTTCAAATT TACGGAACCG 1860

CAAGAAACGA TTTGCTTCTT CCTCTAACTG AGTAATTTCT TCTTCTGTTA ACTCTTCCGT 1920

AGTAATGTGA ATATGTAACT CTTTTTGTGT GCTTTTCATA GGTGCCTGAT GATTAATCTC 1980

TGTTTGTTGG GTGCGGCCCA TTAAATAGTC CAAGGAGACA GAAAAATAGT CTGCTATTTT 2040

TAAAAGATTC TCTACTTTAG GCGAAGAGGT ATCCCAACGA CGAATTTGTC CATTGGAAAT 2100

GCCGACGTTT CTTTCTAATT GGGCGAGGGT TAGTGCTTTG ACTGAGCTAA TTGTTTAATG 2160

CGAGAAATTA AGGTCATTTC AAGTCCTCCT TGTGTTTTCC GTTTGCTAAT TTTAAATTAG 2220

CGTAAAAGCG AATGnTTGTC AATTGCGAGC GTGTAATAAG AAGGTGAACA TCTCATGAGT 2280

GATTTTTTAG AAAATTTCCG ATGAAAAAAA GAAGAAATAG AGTTATTATT AGTAAGTAAA 2340

CTGTTTTGGA GCGAGAGGGG ATGAATCAAT GGAGAAAATT AAAAGTTACC AAGAGGATTC 2400

TGAAGTGCAG AAAAATCGTT GGTGGATCTT AGTTTCTGTG GC 2442

(2) INFORMATION FOR SEQ ID NO: 666:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2413 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 666:

CCAATTCGGn CCAGCCTnTT CGACCAATGT TGCACCAAAn ATAAGTCnTT TAGACAATTA 60

TTCATCTGCC GAGGGAAaGC CTAGCACGAG TGATCGGGAT TCATTTTTTG TAAGAGGAAC 120

GTCGAGCGCA ACAATTCGCG CACCAAAATC TGTTGCTTTT AAAGTCACAC CATTTTTATT 180

CGTCAGTGTG TAAAGAAGCA TTCCTTGTCC AAATGATTCA GTTGAAATAG TCAAAATAAT 240

CAATCCTTCC TTTAAGTAAA TATTTAGTAA ATTTAACTTT GTTGTTTAGA TTATTTAAAT 300

AGTCAAAAAG AAATGCtAAG AAAATAAATC TTTATTTAAT TGTATTGTAC ACGTTTTCAA 360

TTGAAATGAC AATTAAGAAA ATATTAAGTA AACATTTGCA AAAGTTTAGT AAATATTGTA 420

CAATTTGTTT ACTAGTGAAA GGAGCGATTA GTGATGGAAG AACTGTTAAT GAATCGATTT 480

GTAGAAATAT TTGGTGAGAA AGGGACAGCC TGTTATTTTG CCCCTGGAAG AATTAATTTG 540

ATTGGTGAAC ATACAGATTA CAATGGCGGT CATGTTTTCC CAGCGGCTAT TACTTTAGGG 600

ACATTTGGTG TGGCAAGAAA AAGAGACGAC CAAAAAATTC GGATGTTTTC TGATAATTTT 660

AAAGAGGTTG GTCTCATCGA ATTTTCATTA GAAGATTTAA CCTATTCAGA TTCAGATGAT 720

TGGGCAAATT ATCCTAAAGG CGTTTTAAAT TACTTGATTG AAAGCGGACA CAATATTGAT 780

TCTGGcTtAG ATGTGctCTT CTACGGTACG ATTCCTAATG GTGCTGGCTT GTCTTCTTCT 840

GCTTCTATTG AACTTTTAAT GGGAACGATC TGTAATGATT TATATGCTTT ACACTGTCCA 900

ATGCTTGAAT TGGTGCAAAT CGGTAAAAAA GTGGAGAATG AATTTATTGG CGTTAATTCA 960

GGAATTATGG ACCAATTTGC CGTAGGGATG GGAGAAAAAG ATCAAGCTAT TCTACTAGAT 1020

ACCAATAATA TGCACTATGA AATGGTTCCC GCCAAATTAG GTGAGTATAC GATTGTCATT 1080

ATGAATACTA ACAAACGCCG TGAATTAGCA GATTCAAAAT ATAATGAACG TCGTGCAGAA 1140

TGCGAAGAAG CCGTTCGTTT GCTTCAAAAA GAGTTATCTA TTGAATTTTT AGGTGAATTA 1200

GACAGCGAGA CATTTGAACA ATATCAAGCA TTGATTGGAG ATCCAGCGTT GATTAAACGA 1260

GCGCGTCATG CCGTCACAGA AAATGAACGA ACTTTATTGG CAAAACAAGC GCTAACTGAA 1320

GGGGATTTGG AAGAATTTGG CTTATTATTA AATGCCTCAC ATCGTTCGCT AAAAGAAGAT 1380

TACGAAGTAA CTGGCATTGA GTTGGATACC TTAGTGGCTT GTGCGCAAGA ACAACCAGGC 1440

GTTTTGGGCG CACGGATGAC TGGTGCAGGG TTTGGCGGTT GTAGTATTGC TTTGGTACCA 1500

AAGCAAAATA TAGATGCATT TATTGAAGCT GTTGGCCAGA GTTATCAAGA TAAAATTGGC 1560

TATGCTGCTG ATTTTTATCC TGCATCAATT GATGATGGTG CCAGAAAGTT ATTTTAGAAA 1620

GGAAGGCTTT AAATGTCTAT TTTAGTATTA GGTGGCGCAG GCTACATCGG CTCACATGCC 1680

GTCGATCAAT TAATTTCAAA AGGCTATGCC GTTGTTGTTG TTGATAATTT ATTAACGGGG 1740

CATCGTTCAG CAGTTCATGA ACAAGCGACT TTCTATGAAG GCGATATTCG TGATAAAGCT 1800

TTTTTACGTA GTGTCTTTGA AAAGGAATCA ATTGAAGGGG TGTTGCACTT TGCGGCCAAT 1860

TCTTTAGTAG GAGAATCCGT GGAGAAACCG TTAATGTATT TCAATAACAA TGTTCACGGC 1920

ACTCAAATTG CGTTGGAAGT CATGCAAGAA TTCGGGGTGA AACACATTGT TTTTTCTTCC 1980

ACAGCGGCCA CTTATGGCGA ACCAAAAGCA ATGCCTATTA CAGAAGAAAC CCCAACGAAT 2040

CCTAAAAATC CGTATGGGGA AAGTAAATTA ATGATGGAAA AAATCATGAA ATGGTGCGAC 2100

AACGCTTATG AAATGAAATA TGTTGCTTTG CGTTAtTTTA ATGTtGCAGG AGCaAAAAAA 2160

GATGCCTCaA TTGGTGAGGA TCACACGCCA GAAACGCATA TTGTGcCAAT TATTTtACaA 2220

GTGGcATTAG GCCAACGAGC AGAGCTAAGT ATTTTtGGGG rTGATtATGA TACACCAGAT 2280

GGCACGTGCA wTCGAGATtA TGTTTACATT GAAGACTTGA TTGCAGCACA TATTTTGGCT 2340

TTAGAATACT TGAAAAATGG TGGCGAAAGT GACGTCTTTA ACCTGGGTAG CAACAACGGC 2400

TATTCTGTTA AAG 2413

(2) INFORMATION FOR SEQ ID NO: 667:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 548 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 667:

```
GCCGnACCCT TAAATCCGGC CAAATGGTAT AATTGGTCCn CTGGrTTmCC CAatArCCcT      60
cTTCtCyTCt GTCaCAGtGC TTCCgGCTCC aCTTGGgCAT cgCTTaGCTA ATTTTTTCCC     120
aTAAGCAGTT TGACTTTTAG CATTTTTCTT TGCGGCATTC AAACGGGAAC GCGAACGTAA     180
ATCTAACGCA TCTTCGCCAA AACGTTCTAA GTCTGGTTTG ATTTGTTGAT ATTGTCGTCT     240
AGTTAAAGGA ATCTCAGGTT CTGAGGTTGG AATATTAAAT TTCCGCACCG TTGGATCTTC     300
TTCCGTATCT TTCACTTGAA AAGCCGTTGG TTCATCATCA AGAATGACAT ATGAACGACG     360
ATCTTTTTTC ATTGCTTTTA GCAACTCTTC TTCAGTAAAA CTAGGTTCCT GTTCCTCAGG     420
AATCACAGAA GCCGGGaTAT ATTTAGGaAC AAAATAAGAA GGtGtCGTTT TTTGCACTAA     480
TGGTGATGCG GCAGGtCTAC TTGGTTTCGC TaCGaCTTCt TTTtTTTTAt AcGTTTTGGT     540
GGTCCCAG                                                              548
```

(2) INFORMATION FOR SEQ ID NO: 668:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2119 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 668:

```
TATGATGTGT GGACTTAGTG GTGGGTATGC TTCGGTAGAA GCCAAAATTG ACTTTTkGAG      60
AATaCGATTT AAGACATTrG ATGTTCGGAC AGTTGTTGAA AAGTTGTTGC ATATGGATAT     120
GAACTGGTTT ACTTATGaAC CTAGAGGGTT TTATCATTAT ACTGAAACAT TTTCGTATAG     180
CTCTATTCGG ATTTTTCGGA ATCCaGAAAA CGTGAATATG GGAATTATGT TGGACTTGTC     240
TGGCGAAGGT TGTCGTCAAT TAGAAAAAAT TTTTGAAGAA GACAACAAGC GATCGTGGAC     300
AGAATTTTTT CGCTCACTCT ATGATGATGA TATTTTCGGG CAAGGTGTTT TAGTTGATAC     360
AAAAATAACT CGAATTGATA TTGCTCTTGA TGAATTGATT GTGAAAGAAC AAGAAAATTT     420
TGATTTATAT GTTTTGAAAG AAAAAATGGA ACAAGGGCTT GTGGATACCA CTTTTAAGAA     480
TTTTGATTTT AGCGGGGGCT TTGCTTATGA AAATAAAAAA ATGGTAAATA AAGGTTTGTC     540
TCTTTATTTT GGTAGTCGTC AATCGCCATT GTATTTTAAC TTTTACCAAA AAGATTATGA     600
GTTAGCACGA AAAGAAAGTA TAAGTGTTGA GGAAGCAAGA GAAAAACATG AAATTAAAAA     660
TCGGTACGAA ATTCGCCTTT CTGATGAAAA GGCGTTTTTG TTTGTAGAAT ATTTTCTTTC     720
TAGTGGTGAA ACGTTGGATT GGCTAGTCAA GGAAATTATT AATCAATCTT TGACTGTCTA     780
TGACATTGAA GATGATATGA AAGTCTATTG CAAAGCTTGG TATGATGTTG TTGATAAATT     840
GGAAGGTTTA AAATTGTCTG TTCAAGGTGA AAAACCGTCT ATTGAGAAAA CACTTCGGTG     900
GTTATCAAAT TATTTAGCCC CTTCATTAAA GATGATTAAA GAAATTGATC ATTTGTTAGG     960
```

AACAAATGAG CTGATGGAAC GTATCGATTT AGCAGAATTA AAAGAGAAAC ATGAAGAAAT 1020

AATTGAAATG GTCTGTGTTG ATGCGAAAGA TTTATTGTTT ACCACTAATC AAAATAGCAG 1080

TGTTCGATCT TATATGGAAC AAGAATTTGA TTTGGAGGAA GTATATCCAT TTTAAAGATT 1140

AAGGGGGATA TGTATACACT TGATTCGTTA TTTTAAATAG ATATGAGTTT AGTTGAAAGG 1200

AATGTATGAA TTATGAGTAG AGAATTTATT GAACGCAATA CTAAAGTAGC AATTTCTATT 1260

ACTGAAAAAA TGAAAAAAGG TAAAAATGAT TTACAAAAAA CAAAAGAAAA AATTGTTCAA 1320

TTAGATGAAC AAGGGGAATT GaCTaTTcCt TATTTAAAGA TTACTTTTGA AAAGTTTTCA 1380

GAATCAAATG AAGAGCTTTT AAAAGAAATT TCAAGATATG AATATACTTA TGTGGTTCAT 1440

GAAGCAGAAA TGGCGGTTAA AGAGAAAGCA ATTTGGGAAG AATTTTTCAG TATAAAAAAA 1500

TTGTATGATA AAGAGCTTTC AGAATTTGCC AGTTTTAAAG AGAAGTATAA ATATTTTGAG 1560

CCTAAAAATA GTGAAGAATT AAAAAAACAA GCTCGAGTGT TGCTAGAAAA GAAAGGGTAT 1620

ATTGTTGACA GTCCATTTGA AGGAGATTTT GAmaGATGGA TTGGCGTTTA TGCTAGACCT 1680

AAAGATAAAC CAACTTATTT AGATCCAACG GATGGAGAAG AAGCTGGATT ACAAGAATTG 1740

TACAGTGTAG ATGGATTTAA GCAAGACTTT GCAGAATGGT TTGAGTTTGA AGTTGTGGAA 1800

GGGaAATTAA AGGAGGACAT TTTGTAGTGA AAAAGATTTG GCAATTGGTA TGTTTAGAAA 1860

ATAATTATGA TTCACACAGC TCTTATTTTC GGGrTATGGT AAGATATTTT TCCTATACAC 1920

TGTTAGCACT ACTTTTATyC yTTTtAGGTT kGwATTTtAm CATAGGTTGG GTAAGTGGTT 1980

TCTGTTTTGT AGGTTTGATA CTTTTACCAT ATATGGGCTT TTTATCTCGT GATAATTATA 2040

GAGAATTTGC ACCGAGAGTA TGGAAAATAA GAGATGTTCG AGAGTTAGAA GAATTGTAAT 2100

CCTGTAGTTG TTGATAGAG 2119

(2) INFORMATION FOR SEQ ID NO: 669:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1587 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 669:

AAAGATTGAC ATGTGCAACA ACCAATTGTA ATTGTTCTTT TTCTGCCACC TGTTCCATTA 60

CGTCAAAAGG GCCATGGAGT CTACACCACC AGAAACGGCG ACCACGATTG TTTGATGCGG 120

TTGCCAAAAT CcTTGTTCTk TTCCTCGTTT ATAAAACTGC CAAAACATTG TTCGCCCTTC 180

TTTCTTCCTT AGAATAAAAG AGGAACCGAG AAATAACCCA TTCGTTATCC CCAGTCCCCC 240

TATTTGTTTG TGCTTAGCCG CGACGTCCGC CACGTCCGCC ACGTTTGCCT TCAGTATTAC 300

GTTTTAATGA AGATAAACGG TCATCACTAT CTTTCAAGAA AGAGCTCATC AATGAGTCAA 360

AATCTTGTTT ATTGTTATTG TTATTTGCTT GTGGTTTAtT GAAAGGCTTT TTGCCCCCTG 420

CGCGTGGAnT ACGGTCACGA TTCTCATAAT TGTTTTCACG ACGTTGAAAT TCACGTTTTG 480

GTTGTTGCCC TTCTGCTGGC TCTTGTGCTT TACGGATCGA TAAGCCAATT TTGCCGTCGT 540

CACCAACTGA GGTTACTTTC ACTGTtAmTT CGTCACCTAC TGTTAATACA TCATGAATAT 600

CTTTAACGAA TCCATTAGAC ACTTCACTGA TGTGTACTAA CCCAGTTTTT CCTTCTCCTA 660

AATCAATAAA TGCACCAAAG TTAGTAATCC CTGACACTTT TCCTGGCAAT TTAGCTCCTA 720

CTTCGATTGA CATAAAAAAA TTGTTCCTCC TATTTTTTCC CTAATTATAA TTGCAGTTTA 780

TTCAGCTGAA GACTTAGTTG TTTCAGTTGA AGACTGTGAT TGACTATTTT TTTCTGAACT 840

GCTCTTGCTT TCTTCCGCTC CAGAAGTCTG TGTTTTTTTC ATTTGTTCTG GTGTGGGATA 900

AATCTGTTCG CCTTCTTTTG AGAGAAGGAA ACGGCTACGT GCTAATTTTG CCACATAATC 960

ATCATCTTTT AACAAAGCAA CGTCTTTTTT TAGATCCTTG ACCTTCTTAT CAACTGCTAC 1020

GGATTCAGCA ATTGCATCTG CTCGTTCTTG TTTGAAAGCA CCTAAGCGAT GATAATCTTT 1080

CATCAATTGG ATGCCAGAAA AGGCAAAAAT AATGAAGGCC ACGAGAAAAA TCGCTGCCAA 1140

GCGCTGATGA ATCCCCTAAT GATTTTTATC AAAATCATTA AGTTAAGGTA GATACACATC 1200

TTGTCATATG ATCAAATGGT TTCGCCAAAA ATCAATAATC AGACAACAAA ATGTGCGAAC 1260

TCGATATTTT ACACGACTCT CTTTACCAAT TCTGCCCCGA ATTACACTTA AAACGACTCA 1320

ACAGCTTAAC GTTGGCTTGC CACGCmTTAC TTGACTGTAA AACTCTCACT CTTACCGAAC 1380

TTGGCCGTAA CCTGCCAACC AAAGCGAGAA CAAAACATAA CATCAAACGA ATCGACCGAT 1440

TGTTAGGTAA TCGTCACCTC CACAAAGAGC GACTCGCTGT ATACCGTTGG CATGCTAGCT 1500

TTATCTGTTC GGGCAATACG ATGCcCATTG TACTgTTGAC TGGnCTGAAA TTCGTGAGCA 1560

AAAACGACTT AAGGTATTGC GAGCTCn 1587

(2) INFORMATION FOR SEQ ID NO: 670:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 620 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 670:

CAATATGATA ATCATACAAA AGAAGaACTT AGTGAAATGA ATACAGAAGT CAGCGTTGCt 60

GgACGTATGA TGACAAAACG TGGTAAAGGA AAAGCTGGTT TCGCCCATTT GCAAGACCGT 120

GAAkGCCAAA TTCAAATTTA TGTTCGGAAG GACCAAGTTG GAGACGAAGC CTATGAACTA 180

TTTAAACATG CTGACTTAGG CGACTTCTTT GGTGTGACAG GTCAAGTAAT GAAAACCAAT 240

ACGGGGGAAG TTACTGTTAA AGCTCAAACA ATTACGTTAT TAACTAAAGC TTTGCGTCCA 300

TTGCCAGATA AATACCATGG ATTAACGAAT GTGGAACAAC GCTATCGTCA ACGCTATTTG 360

GATTTAATCA GTAACAAAGA AAGCTTTGAT CGTTTCATGA AACGCAGTCA AATTATTAGC 420

GAAATCCGTC GTTATCTAGa TGGCAATGGC TATGTCGAAG TAGAAACGCC TGTCTTGCAT 480

AACGAAGCAG GAGGAGCTGC TGCTCGtCCA TTTATCACAC ACCATAATGC GTTAGATATG 540

GACTTATACT TACGCATTGC CCTAGAGTTA CATTTAAAAC GkTTAATTGT CGGtGGCATG 600

GrAAAAGtAT ATGAAATnGG 620

(2) INFORMATION FOR SEQ ID NO: 671:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 841 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 671:

CTCTCTCGTC ACTGCGTCTn TTGGATCGAA TAAGTCGTGC GTGTATTTCG TGCGAGCTCC 60

CnGnCATTCT CACACTAACG CCGTAGTACT GTGTAAGTGT ATGATGACGA GCGCCGATTC 120

TCCTCTGTAG GAGATCTACC GAATCCACTT AGGCTTCATC GTCTGGTGAA CAAGGTTCAA 180

AATAGTCATT TTTCAATAAT AAATAATTAA TCAAAAGATT CACCCCGAAT CAACATTTTT 240

TGTTCTAAAT CCATTATAAT AGAAGTATGA CACAATACGA GAAGAAAAGC TTATAAAAAA 300

CATTCAAAAA TAAGTGGCCG CGATAGGTGA GGCCTCGTTT TGAAAATGGT TTGATAAGTA 360

ATCCGTTTTT TATATGAAAA AGCAAGCAAA AAGGGCTTTT AGCACATATA TTTAAACGAA 420

ATGAAACATC TGTGTACGTT GATGAAGTAT AGTTATAGCA GGTGGAATGA AAGGATTTGT 480

TGAGAGAATG GTGTCGTTTT CGGCAATAAT GTTATTTTGG TATATTTTTC CGGTAATTGT 540

GCTCTTTGCG TGCAATTTTC TCGTGTCCAC TTTTTCCCTA ACGGAACTTG GGAAAATAAA 600

AGCACCGGAT TTGGCCATTC CGTTTTTATT CATAGGGATC CATGAGTTAT CCAAAGATAG 660

TTATGATGAA TCGATTATGC CGTATTTTGT GATTTCTATT TTGTTATTAG GAGTCGCTGT 720

TGCCCTGTTT CAAGCTTATT ATTATGGAGA GATTTTATAT AGACGATATT TCCAGAAGTT 780

TTGGCGGTGA CGTTTTTACn AACGTTTGGG TGGAAGGGGT GCCAACCTGC CAAAAACAGn 840

A 841

(2) INFORMATION FOR SEQ ID NO: 672:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1377 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 672:

ACnAAATGTC TTAAAAAAAT GCTTTTTTCn TAAAAAAATT ATTACAAAAA AAAaCAAGAA 60

GTCAGCCAAG TATTGGCTGA CTTCTTGTTT TTTTTATTTA ATTATTCAGC TAAACTAGCT 120

TTTAACGCTT CAACTTTATC TGTTTGTTCC CAAGGTAAAT CAATATCTGT TCGACCAAAA 180

TGACCATAAG CCGCAGTTTG TTTATAAATT GGGCGACGTA AATCTAGCAT TTCAATAATG 240

CCTGCTGGAC GTAAATCAAA GTTTTTACGA ACTGCTTCGA TTAATTTACT TTCTGGAAGA 300

TCACTCGTAC CAAACGTATT AATAGAAATA GACACTGGTT GCGCCACACC AATTGCATAA 360

GCTAATTGGA CTTCTACTTT TTTCGCCAAA CCAGCAGcAA CAaTGtTTTT CGCGaTATAA 420

CGCGCAnATT AACTAGCAGA GCGGtCAACT TTGGTTGCAT CTTTTCCAGA AAAAGCACCG 480

CCGCCATGAC GAGCATAACC GCCATAAGTA TCAACGATAA TTTTACGACC AGTTAAGCCG 540

GCATCTCCTT GAGGGCCACC GATTACAAAG CGACCAGTTG GGTTGATAAA GTATTTTGTT 600

TCATCATCTA AAAGTTCATG AGGAATCACC TCATTGATTA CTTGTTCTTT AATATCTTTT 660

TCGATTTGTT CTAAGGTTGT TTCATCATCA TGTTGTGTAC TGATGACGAT CGTATCCACA 720

CGTTGTGGTT GACCTTGATC ATCATATTCC ACTGTTACTT GAGATTTTGC ATCAGGACGT 780

AAGTAAGGTA AGACTTTTTC CTTCCGCAAT TCTGCTAAAC GGCGAACCAA ACGATGACTT 840

AACGCGATCG GTAATGGCAT TAATTCCGGT GTTTCGTCAA CAGCAAAACC AAACATTAAG 900

CCTTGGTCAC CAGCGCCAAT ATCGTCTTTT TCATCTTGAT CACGCACTTC TAAGGCTGCG 960

TCAACCCCTT GAGCGATATC AGGTGATgCT CATCGATtGC TACAAGAACT GCgACAGTCT 1020

CTCCATCAAA TCCATATTTT GCTCGTGTAT AACCAATATC TTTCACAGTT TGACGAACAA 1080

TTTTTTGGAT ATCTACATAA GCTGTTGTAG AAATTTCTCC AAATACTAGG ACTAAACCAG 1140

TTGTTACGGa TGTTtCACAT GCCACGCGGG CCATCGGATC TTTCTCCAAA ATCGCATCTA 1200

AAATTGCATC ACTGACTTGG TCAGCTACCT TATCTGGATG TCCTTCTGAA ACGGATTCTG 1260

ATGTAAAtAA ATGTCTnTCT GTCATAATGT TATTCCCCCT AAATAGTTTC GGTTACAAGG 1320

CATCTTTAGT AAGTAACGAC AATTACTTGC TAAATCCTTT CGGGAACTTG TAACAAA 1377

(2) INFORMATION FOR SEQ ID NO: 673:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 536 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 673:

TGTTAACGTG aACTATTATT CATCGTTTTA AAACGTGTTT CCATTAACTC CACTGTATTT 60

TCCAAGCATT GaGTAATATC TTTTTTTCCA GTATGAACTT TTCCTGCAAA TAGCGGTAAC 120

TTGCCTAAAG CCATCAAGTC AGAATTTTTC GGGTCACAGA TTyCmAsmTs ACCcACTCGA 180

CAAAGCGCAT AGATCAGAGA CAAAATTGTA AAAGTTTTAC CTCCACCaGT GCCTCCACCT 240

AATAACAAAT GTGGATCTTT ATCaAAGTGC CAATACAACC CTTTCATCAA CTGAACTTTG 300

CCTtCGTCTG CTTGCATATC CTTAATCCAA ATACGCTTAT TGTACACATC TGTTGCTAAT 360

TCATAGGTAA CAAAACCTTT TTCATCAATC TTATTTAAAT TATCTGCATG AAACGkTGkT 420

TCCAAAAAAC CACCAATGGT TTCAAATTTT TCCTGAAATT TATTACCCGC TGTTTCAAAT 480

GAAACAAAAA TACTATTTTT TGCAGACTTA TAATAAATTT TTGGAAACTT GATTTT 536

(2) INFORMATION FOR SEQ ID NO: 674:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1079 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 674:

CGATTGACTT TCTAACACTA CTTCCATTAT AATTAGCTTT GTTGTGTTAC AACTATTGTC 60

ATAGTAGCTC AGCAGGATAG AGCATCCGCC TTCTAAGCGG ACGGTCGGGG GTTCGAATCC 120

CTCCTGTGAC GTAACGATAA CAAAAGCGCC TCTTAGAAAC CTTGGTTTCT AAGAGGCGCT 180

TTTGTTATCC CAAGTTCATC CTAACCTAGT TTAACTTTCT ACCAACTTGA AGTACCTCCA 240

CCTCCGCCAA AACTCCCACC TCCGAAACTG TCACCGGAGC CATCATCAGA AGAACCCTCG 300

TTGCTTGTAA CCCACCAATC ATGACGATTC GCATACATAC TGGTCTTTTT TAGTTGCCTG 360

TCAGAACGTG AACTTGCTAG AAATTTAGCA ATAGACACAT TTCCATAAGC AAAACGCATA 420

TCATAGTATT CAGGATACGT AnTCTTAATA AATCTTATCG GGTAGAAGGA TTTTGAAGCT 480

TTTTTGTTGG TAATCTTTTT TtCTCTAATT TTTTtCCGCA AtAATTTTTT TTCAATTTwA 540

TATkGCtGGT TGTCtAAATC GTTTGCyAAT TGTTTtAATG ATTTTGATTT TTCTGAACTT 600

ACTTGCTGGA TTTCTTTTAA AAACCTACGG ATCATAGAAT CCAGGTGCAA TCTATTTTGT 660

AAGCCAACAC CTATTGATTT TATATAACCA AATAAGAAAA AGACCGGAAT CAAAAAATAG 720

ACCAGACTAA CCATGGGTAC TAAAGAGTCT TCGTCTTCTA AAAATGTCAG AATCTCTGTG 780

ATTAGTGTGA AAACTGTCGA AGGaAACGAC TGCyCTGATA AACTAACTTG AACATAAATA 840

ATCAGAGTCA CGGATACCAA AAAAGGATGT TTTAATAAAA TCGAAAGAGG AAACAAAATA 900

ATTTTTAGGT AAGCAGATAT TAAGTTATCA TTAATTCCAC TAATCCGACA CGCGGGCAGT 960

TTCTTTTCTG GGAACTTTTT CCCAATATAC TTCACCTATC CAAAAACGCG GGTAAGTAGG 1020

ATCCATGGGT TnGATTCACC CTTCCCAAAA AnnAGGCACA CATAGAAGAG GCTTCnTTA 1079

(2) INFORMATION FOR SEQ ID NO: 675:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1090 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 675:

GTATGAATTG GAAGACTTAA TGAAACGGAT TCCACCAATG TATGAAGAAT TAAATGAAAC 60

TTTCCCAGAT CAATTAAAAG AAATCGAAGA AGGCTACAAC CAATTATTAG CTGACGATTA 120

TGTGTTCCCT GAACAAAACT TTGCTGAAGA AATTCAACAT GCGAAAAAAC GTGTCGAGAA 180

TTCAATGGCT GATTTAGAGA AAACAGAGAT TGCTGCAGTA GAGGTCGCAA ATCGCGACAC 240

AGCAACGGCG ATTGATGCAT TATATGAAGT GATGGAACGT GAAATTGAAG CGAAGAAATA 300

TGTCGTGACC AaTCAAAAAA TTATTGACGA TTATATTTCT CATTCCTTAA AAAATAACCG 360

TCAATTAATG ATTGAATTGG aCCATGTATC ACAAAGCTAC ACCTTAAACA ACAACGAGCT 420

AGGACGTAGT CGTGGCTTCC AAACGGAAAT TGAAGAAATT ATCCGCCGTC AAAAAGACTT 480

GGAACCTCGA ATGAAAGAAC ATACCGTTCC TTACTCTGAA ATTCAAGCGT TCTATAAAGA 540

ATGCTATAAG ATTTTAGATG ATATTGAAAA TCAACAATTA GAAATTGATG CTTCATTGAA 600

AGAACTAAGA AAAGGCGAAA AGGTCGCCCA AGAAAAAGTC GACGAATACG AATTCCGTTT 660

ACGGAGTATC AAACGATATG TTGAAAAGCA ACGATTACCA GGTTTGTCTG CTGATTACTT 720

AGAATTTTTC TATGTGGCAA CGGATCGAAT TGAAGATTTA AGCCGAGCCT TAAATAAAAT 780

GCGCATTAAC ATGGATGAAA TCAATCGTTT GTGTGACTTA TGCGAAGATG ATTTAGAGTT 840

ATTAGATAAG AAAACAAAAG ATTTAGTCmA TGCAGCAGCT CTGACAGAAC AAATGATGCA 900

ATATGCAAAC CGCTmtCGTC ATACTCACGA GAACATCCGA GCGGCTTTAG ATAAAAGCAT 960

GTATTTATTC TCGACAGrAT TCCGTTATCA AGATGCATTA GATGAAATTG GAACGGCATT 1020

GGAAGCTGTT GAGCCAGGGG CGTTCAAACG AATCGAAGaT TTTTATTTTA AAAATAnCAA 1080

TAACCCTAAT 1090

(2) INFORMATION FOR SEQ ID NO: 676:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 527 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 676:

TATCTTCTTT TTCCGTTATT GAAATAATTA ATAGCGTTAA TAAAAATAAA GGAATAAAGA 60

GGAAGTGCAT AACCATTTGA GTAACAACTA TCGATACAGT AATCGTGTGC TTGTCAAATA 120

GCTGAACTAG TATTGGATTA AACCAAACAA TGGCGACTAT TACTAACGAA AAATATTTAG 180

TCACGCTTTC CAGCCTTATT ACATACAACA AGCCAAAGAA AATTAGTAAA CTAATAATGA 240

TATCTTTATA AGTAAAAAAA TGTTCATAAT TCATAATTTC TTTCGTTTCA TCTCCACTCA 300

ACTTGCTTAA CTATTCAAGG TTTAGGTAAA TCACGTTTCC CATTTTCTCA CATGATCGTT 360

CAAAGAAACA GAACGTTGTT TGGTGTTTTA AAACGTAAAA GTTGAGCAAC ATATCCTGTC 420

TAAACATTTA CTTGGCAGTT TATTTATAGA AAAATTGTTA CTTTCCAAAT TTTTGTATGC 480

TTTTTAGAAA TGAAACCAAG CTTATGTTTT AATGCTTCTT ATGCAAA 527

(2) INFORMATION FOR SEQ ID NO: 677:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1493 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 677:

ACACGGATnA TTTATTGATC CCTAGGTCAT TnGGCGGGAT TCAAATTCCT TCGCATAAGC 60

CAGGCCaTTA AAtTmCTCGt GGtCaTAwGa cyCCtTaCaA TATwCyTTTA AAAAaGcTTT 120

TATCaTCTGG TGTCATAATk GTGtTATTTC atATCTCGAC TTAATAGkTT CtCAAAgCAG 180

ATCGCTCTTC aTCAGTGACA CTAATATTgG TCTTAAAACA GTAGTATTAT TACGATTTGA 240

CATATTACCA ACATTGATtC TTTTTATCTC TAAATCATTT TCAATAAATG GTAATAAATC 300

TACGGGTGAT TTGACAATCA TCAGAACATT TTGTCCTTCA TATTTGCCAT TTTTAATATT 360

ACTAATAGCT GTTTCTTCAG TTATGATAGA CGAAGAAATT CCAGCAGGAA CTGCCATTCT 420

TAAAACCTGC TTTTGTAAAT CATTTTTAGC AACGTTGTCG TTCACTACCA TTAAACGTGT 480

TACTCCCAAT TTATTACTCC ACATTGTAGC TACTTGACCA TGGATCAAGC GATCATCAAT 540

TCTAATATTG TTAATACCTT TCATTTCCAT CGAAAATGCC TCCCATAAAT TATTTTATTT 600

TAAAATACCT AGAGCCGCTA AAACAATTGA CAATACCATA ACACCAAGAA TTAAGTAAAT 660

AGGTTTAACG TTCTTCTTTA ATGCCCAGAA AACTAAACCA ACTAATAATG CAGGCAATAA 720

TCCAGGCATT ATTTGATCCA AAATATCTtG ACCTTTCATT GAAAAATTAC CCTGTTTAAA 780

GTTTAAAGTA ATATTAGCTT TAACGACTGA AGGTATTAAT GCCCCCACGA CCATTAATCC 840

TAAAATCGAa GTTGACTCAG TTAATAATTT CAAACTTTTT TCTATAGTAG CAATTAATTT 900

TGTTCCTGAG GCATATCCTG CATTAAATAA TGGAATACTA ACTAATTTTa GCGCTACACT 960

TACAACAATC CATAACAAAA TTCCAACTGG ACTACCATCA ATGGCCATAT TTGCTGCAAT 1020

AGCACCAAAT ATAGTCCaAG GAATAACTAC AAATAAACTA TCACCCAAAC CAGCTACCGG 1080

GCCCATAAGG CCTGTCTTAA TAGATGTTAC AGCCTCTTCT GTGTTTTCGT CAGCATTCTC 1140

TTCCATAGCT ACATTAATAC CTAAAATGTA AGGTGCCACC CAAGGATTAG TATTAAAAAA 1200

TTGTAAATGA TTTAATACCG CTTTTTGTAA TGCTTCTGGT TTATTTTTAT ATGCTATTTT 1260

TAAAAATGGT AATACTGCAT AACAATATGC TAGCGCTTGC ATCTTTTCAT AATTAGAAGA 1320

AACTGAGCTT GTAAAAAACC ATCTCCACAA AaCACTTCGT TTAGTTTTTG TATCAATGCG 1380

TAATTTTAAT TGTTCGATAT TATTCATCGT CATCCATCCC CCCTGCTATA TTAATCACAG 1440

TATTCTTTAC TAAATCTTGA CTATTTTTTT TATAAATTAT TAATGCAATA GCT 1493

(2) INFORMATION FOR SEQ ID NO: 678:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 582 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 678:

CCAATGGATA ATCCTCCTTT AATTTATATA AACTTATTTT TCATTTCCTG TTAGTAGGTC 60

TATCATATCT CTTTTGTACC AAAAAAGCAA GAAAAAACCG AACAAACATT CGCATATTTA 120

TTTTTTATTA ATGATTCGGC GCCGCAAAAG ATCTAATCCT CGCATAACCG CACTCTGTCT 180

AATATAATTT CGGTCACGGT TAAAATGGAA ACATTCTGCA ACGGTTGGTT GGCCTTCTTC 240

CGCCAGCCCA ATCCAAACAG TGCCTGTCGG TTGACCTTCG AGTGGGTCAC CAGCGACCCC 300

AGTAAATGAT AAACCATAGT TAGATTTAGC TAACTGACGC GCTTTCTCTG CCATTTCTTT 360

GGCACACGCT TCACTTACCG TGCCATGTTC TTCTAAAAGT TCATGAGAAA TTCCTAAAAA 420

GTTCTCTTTT GTTTCTTGCG AATACGTCAC GAAGCCTCCC TTAAAAATCT TCGAAGCACC 480

CGCAATTTCC CCCAGTGTAC TTTGGGAtAG GCCAGCGGTG AGGCTCTCAG CAGCCGTAAC 540

GGGTTTGnCC CGTTTTGGTA nTAACAAATC nACGGTAAAC nT 582

(2) INFORMATION FOR SEQ ID NO: 679:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1192 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 679:

CTGGTTCGCC ATCTTTCATT GTCATAGTGG CACCAATTCC TTCAAAATCA CCAGATAAGC 60

TTTCGTTTAA ATCATTGGCA GCAGATTCAT TTAGATAAGT TGAGTATGGG TCTCCAATGG 120

CTTCAGACAT GCCTTTTAAG GCGCCTTCAA CTAATTCATT TTTATCTACT TCACCTACAT 180

AATTGGTACT AATTTCATTA TATAAATCCT GCACTTTGCT TAAATCAGCA TTGGTAATTG 240

CCCCTTCTTG ACTCATCTTT TTAACACGAT GATCAAAATA AATATAACtA CTTCCTCCAG 300

CTAAAAATGC GACACAGAGG AGCGAAATAA TATATTGATA AAAAGGAACA GTTCGTTTGT 360

TCTTCATAGa TTTCATCTCA TTTCTAACAA ATTCGTCACT ACTTTTGTTT ACTATAACAT 420

GAATGAAAAA AGAGGAAAAG CTTTGTTTGC TTTTCCTCTG GATTTATTTA TTGTTTTCAA 480

GGTACTTTTC CCATAACTCA TCAAAGATGT CCATGTTGCT TAAATAGTCA GCATTCATTT 540

CTAAATAAGA AGAAAGTTCA TGATAATCCT CGGTTTGTTT TGGAAACTGA ATATCTTTGG 600

CCGCTTCATT GGCAAAATCA GTTTCAGAAT CTTTAGCCGG ACCTTTTAAT GTCATTAAGT 660

AATGGTAAAA GCTTCGTCTC ATGTAAATTA ATCACTCCAT TTTTCTTCAA TAAATTCACT 720

TCTACGTTGA TGGGACAATC CATAACGCAA AGCGTCTTTT TTATAAAAGT CTTGATGGTA 780

TTCTTCAGCA GGATAAAATG GTGCTGCTGG TTCGATGGTG GTAACAATTG GTTCTGTAAA 840

ACGGCCGCTA TTAGCTAGAC GTTCTTTACT TTTCTCAGCA ATTTCTTTTT GTTCTTGAGA 900

ACGATAAAAA ATAACTGGCC GATAATTATC GCCGCGGTCT TGAAACTGGC CGAAAGCATC 960

AGTAGGATCC GTTTGTTGCC AATAmATATC CACTAATTGT TCGTAAGAAA TAATCGCTGG 1020

ATCAAAAGTG ATTTCTACTG CTTCCGTATG GCCTGTTGTG TGTGTTAGAA CCTGTTCATA 1080

TGTTGGATTA GGCACGTGCC CACCTGTATA ACCCGAAACA ACTGAGATAA TCCCTGGTTG 1140

TGTATCAAAA GGCTGCACCA TACACCAGAA ACAGCCGCCT GCAAAAATTG CT 1192

(2) INFORMATION FOR SEQ ID NO: 680:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1442 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 680:

GTTATCCACG AGCGCCTTAA AATATGTCCn GGGAATGTGC CCGATTCCCT AGGCATTACA 60

AAAACCAAGC CTGGGAAATG TGCTTTAGCT GCTTATTCCC AAACAGTTTC CTGGAAGGAC 120

TTGTATACGG CGATTATAAA CTGTTTTTTT AAACtAACGG TACATTAGGG ATCGTTGAAT 180

GGTTTTAAAA CAAAGGAGCT ATATTATGAG AAAAAGTATT TTATTCATGT TGGGAATATC 240

TTTGTTCTCA ATGACTGCTT GTTCTAATGG GGATAAAAAA GAAGTAAAAT CAAAAGCATC 300

CTCsTCTACT GTTATCTCTG TCAACAAAGA ATCATCTGGT AaTACGAAGA AAAATGAATC 360

AAATTCAATA ACTAAAACGT CTAACGTTGT TTCTTCTACA TCTGTGGAAA AAACACgtGG 420

AAGTAAGTTT AGCgATTTTA TkGGTtGGGG TATtCmAAGT GGAAaCCTTT TTTTATTAAT 480

GAAGATGGGA CATATTCAAA TGGTCAAGTA GACCATTCTT CATTAATTGc CTTAAATTTA 540

GTATTTTAGC AGAcGGAAGA AAATCAATGA GCTCAAATTT AGGAACTCTA ATTAAAGAAT 600

CTGATGGCAC CTTAACAGAT GGTGAACTAG TCTTTCAACC ACTTGAGTTT AGTAACAAAG 660

AAGATTTCCT TGCAGATAAA CAAAAAGAAT ACAAAAATTC TCCCGAAACT AATGCACCTG 720

AGAACGAaGT TGAAATTACT AATACTATAG rAGATATCAC TAACGGAAGT ATCTCTACTG 780

ACACAAATAC GCTCACAGGA TTCTTAAATG TGTATGGTAT GAGTCCTGCA GCTTATAAAG 840

TTACTATTGA AGGAATGAGT GAAGAAGAAG CTTTAAGAAA CACACCTAAA GAAATGAAAA 900

CTTCTGCTGA AATACAATTA GGTATCTCTA AATATGGTAT TCAATAATTT GTCTACGATT 960

TATAATTAAC ATTAGCCATA AAAAATACAC ATAGGAsrCT AAAAAATGAG TGCATTTTTA 1020

GGTCTATCAT CTTTATTAGG ATTATTCGTA ACTTCTGTAT ATTTTGTTTA CTGTATTTTT 1080

AAGAAGAGAA ATAATTTATC CTATTTTGTA TGGATTTATA TGATTCTTAT TGGATTATTT 1140

TTCmCTGCGT TAGAGTTATT TATTCCAGGT ACTATTTTTC TTACTATTTC ATTGTATAAA 1200

ATATTAAAAG AkCCAAAAAC AAAGGaATTG cTACATGaGC aAGaAATAAA AcgAACmGaA 1260

AAAAtAATAG aAAAACAAAA CCCTATTCAA GAATTCACAG AAAGTAATTC AAACGAAAAA 1320

GATCTAAATC AGAAAGTTGC TTCAACATAT TCTAACGATT ATTATCCCTC GGATAAAAAA 1380

GcTGGGCTAT TCCATCTTGG CGTCTATTGT CCTTATTGTA GAAGCCTTAA CGTCCAATAT 1440

AT 1442

(2) INFORMATION FOR SEQ ID NO: 681:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3919 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 681:

TTGCTTTACG CGCCGGCATG GACGGGCGGC TTGCTCCGTG CGGAAGCnAG CGACGACTTT 60

TCCGGCGTGG CCGCAATCTA TATTGACGGC GAGAAATACn ACCGATTTTA ACCAACGGCA 120

CACTGGACAT ACCGCTGCGG GATTTGCCGG ATGATTACGA ACAGCTTTCC GTGCAGGCGG 180

TGGACGCTGC GGGAAATAAA TCAAAAATCG TGCAGGTCAA GAACCCCAAC TATCAGGCGG 240

AGGATAACAA GCAGAAGCCC TCCACCGGCC AGACACAGCC CCCGGCGACC ACCGCGCCGG 300

TTACATCCAC CACGCCTGCG CCAACGACAC CGGCCACCAG CGCAACCACA CCCCCGACTA 360

CACCGACCAC CAATCCGGCA ACAACGACTG CCACAAAACC GGCTGCCGGC ACGGGCAGCG 420

CAAAGCCCTC CGCTTCGGCA AGCGCAgCCC gGACGAAGAA ACGGACACCA CGCCCCGCGA 480

TCCGATCCCT TTGACCCCGG ACGGCCAGGC TACCGTTCTG GACAACGCCA CCGGTGAGGA 540

CGGTAAGGAG TTTTATACCA TCCAGACACC GGATGAAAAC GTATTCTACC TGATTATCGA 600

CAACCAGCGC GATACGGAAA ACGTGTATTT CCTCAATGCT GTCACCGAAG CCGACCTGAT 660

GGCTCTTGCA GTAAAAGAGG ACGATACGCC GCAGACAGAC GCAATCCCTG ACCCGGAGCC 720

GGCCTGCATC TGTACGGAAC AGTGCGCCGC CGGCGAGGTC AACACCGCCT GCCCGGTCTG 780

CACCCTGACC CGGAAAGACT GTACCGGAAA AGCCCCCGTG ACCGACACGG AAACCGAACC 840

CGACGAAAAG CCGGAAAAAA CGAAAAGCGG CGGTTCCGGC ACACTGATTT TAGTGCTGCT 900

TGTGGCGCTG GCGGCTGGCG GCGCTGGCTA CTACTTCAAA ATTTACAAGC CCAAAAAGGA 960

TTTGGACGAT GCAGAGGACT TTGACGAACT GACCGGCGGG GACGAGGAGG AAACGGTCAA 1020

CGAGGACGAG GATGTGGACG CGCAGGAAGA AACCCGGCAG ACGGGCGATG AGGAACCGGC 1080

AGCCTATGAC GAGCCGGAGG AACCGGACTA TCCTGAAAAC TATGGCGGGG AGGACGACTG 1140

ATGCGCTTTA CGGACAGCCC TTTTGAGCGG ATGATGACCC AAAGGCCGGC CCCCAGCCGC 1200

GCCGCGCCTG CCTCCCCCGT CCATCCGCCC GGACACCCTT GCCACCGCTG TCCcTACGGG 1260

CTAAATGCGC CCTGCGTGGG GATATGCTAC AAAAATCTGA AAAAGGAGCG TGAGAAAAAT 1320

GCAGTTAGTG ATCGCGGAAA AACCGTCGGT CGGAATGGCG CTGTCTAAGG TATTAGGCGC 1380

GAAAAGCCGT GGCGATGGTT ACATGGAGGG CGGCGGCTGG CTGGTAAGCT GGTGCATCGG 1440

GCATCTGGTG GAACTTGCCC CCGCCGATGC CTATGACCCC CGCTATTCCA AATGGAACTA 1500

TGCCGACCTG CCGATTTTAC CAAAACCTTG GCAGTTCCAG GTGCTGCCGG ACACCAAAAA 1560

ACAGTTTGAA GTCTTAAAAA CCCTGATACA CCGGGCCGAT GTCACCGCCT TAATCTGTGC 1620

AACAGACGCA GGACGCGAGG GGGAACTCAT TTTCCGCCTG ACTTACCATC TGTGCGGCTG 1680

CACCAAACCC GTCAAGCGGC TGTGGATTTC CTCAATGGAG GAAGCCGCAA TCCAGAAAGG 1740

CTTCGACAAC CTGCTGGACG GGAAGAACTA CGACAATCTC TACGCCGCCG CTCTCTGCCG 1800

CGCCAAGGCG GACTGGCTGA TCGGCATCAA CGGTACGCGC CTGTTTACCA CACTGTACCA 1860

GGGCAAGTCC CTGAACGTGG GCCGGGTGCA GACCCCGACG CTGGCACTGC TGGCAGAACG 1920

GGAAGCCGCT ATTTCCGGCT TTCAGAAAGA GAAATTTTAC ACGGTGGAGT TGGAACTGGA 1980

GGGCTTTCGC GCCGCCAGTG AACGCTTTGC ATCCAAAACG GACGCAGGCA GGCTGCGTAC 2040

CGGCTGTGTG GGCAAGCCTG CAACGGTACA GACGGTTTCC CAAAAGGAGA AAACCGAGCG 2100

CCCGCCCAAG CTGTACGATT TGACTACCCT CCAACGGGAG GCCAACCGGC TGTTTGACTA 2160

CACCGCCCAG CAGACGCTTG ATTATCTGCA AGCCCTCTAT GAAAAGCGGC TGGCGACCTA 2220

CCCCCGCACG GACAGCCGGT ACTTGACCGA GGATATGGCG GCGGGGCTTC CCGCGCTGTG 2280

CCAGTCCGTC GCTGCGGCGC TGCCCTTTGC GGCAGATTTG TCCTTGCCGG TAAATGCCGC 2340

GCAGGTaATC GACAACAGCA AAGTCACCGA CCATCATGCC ATCCTGCCCA CGACAGAAGC 2400

GGCAAAAACC GACCTGTCCG CGCTGCCCTC CGGGGAGAAA AACATCCTCT TTCTGGTTGC 2460

CGCCCGTCTG CTGTGCGCGG TAGGCGAACC CCATACCTAT GCGGAAACCG CTGTCACACT 2520

GGAATGTGGC GGCGCTGTCT TTTCCGCAAA AGGCAAAACG GAAACCGCCG CAGGCTGGAC 2580

GGCTATGGAG CAGGCTTTCC ATGCCACACT GAAAAAGAAA CCGAAGCAGC CCGAACCCTC 2640

CCCGCCACTT CCTGCACTCA CCAAGGGCCA GAAGCTGACC GCCGGGGGTG CATCCATCAA 2700

AGAGGGCGCG ACCTCCCCAC CGGCCCATTT TACCGAGGAC ACGCTTCTCT CCGCAATGGA 2760

ACACGCCAGT GCAAAGGAGT TTGCGAAACT GGAAGATGTG GAACACAAAG GACTTGGCAC 2820

TCCGGCCAcC CGTGCCGGCG TGATTGAAAA GCTGGTGCGC TCCGGCTTTG TGGAACGGAA 2880

GAAAAAGCAG CTTTTCCCCA CGGAAAAAGG CACGGAACTT ATCAAGGTAA TGCCTGATAC 2940

GCTGAAATCC GTCAAGCTGA CCGCTGAATG GGAGGAACGG CTGGGCGCGG TGGAGCGCGG 3000

GGAACTCTCC CCGGAGGACT TTATGGCGGG CATTACCGCC ATGCTCACCG AACTGGTAAA 3060

AAGCTATCAG GGCGTGACCG TGGCATCTGC CGCCCTGTCT GCATCCGGGC GCACGGTAAT 3120

TGGTATCTGC CCCCGCTGCG GCAAAAACGT GGTGGAGGGC AAAAAGAGCT TTTTCTGCGA 3180

GGGCTGGaAC gTGCGCCCGC CTGCGACTTT GCCCTGTGGA AGAACGACCG CTTTTTTACC 3240

AGCaAGCACA AGGAACTGAC CCGcAAgGCT GCCGCCGCCC TGCTGAAAAA AGGGCGCGTT 3300

GCCATGACCG GCCTGTTTTC CGAGAAAAA GGCGTCCTCT ATGATGCAAT TGTTGTGATG 3360

GAGGACACCG GAGAGAAATT TGTTCGCTTT AAGCTGGAAT TTGACAATAA CACACCTAAA 3420

AAGAAAGGAA ACTGACGCTT ATGGAAAAGA ACATGAACGC ATTTAACCGC AACCTGCACA 3480

ACCTCCTGAC CATCAGCGGC ACGGATGCCG CTGAACTGGC GCAGACCCTT AACGTGGAAC 3540

TGCCGGAGAT CAACCGCTGG CTgmAacsTC CGCCGCGCCG GATGTGTACC AGCTTCAGGA 3600

AATCGCCAAG CGGTTCGGGA TGCCCTACTC CTACTTTTTT GAGAGTGAGC AGAGCCTGCC 3660

CAATGTGCCG AAAGTAGCGG CATGGCTGGG CTTGTCGGAG GAAACGGTGG AAACCCTGCT 3720

GGCAATGGCC GATACCGAGC CGGAGGATGT CATGGACGCG CTGGACGATG CCATTTATGC 3780

AATGGCTACC GCTGTCAGTG CTGCGGGGGA GGTGGATACG GAATGAAAAA AGAAGAAGCG 3840

AAAGACATCC CTATCGGCGG AGAAATCGCT GTCCGGCAAA CTGGCGTCCC CTATGGCTGG 3900

CTCTACCGGC GCACGGAAC 3919

(2) INFORMATION FOR SEQ ID NO: 682:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1338 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 682:

GGTCCGTCCT TGATATTGAA CGGTTAACGC TTGTAACAGC TAAGTTAAAG CTTTTTCTCA 60

TGAAATTCCT CCTTTTTTAG GTTTGCCTTA AAATCAGTAT AAGGGATAGT TTAAGAAAAT 120

GCAAATAATT TTTAGGCGCA CCTAAATACA AGCGCTTTTT TCACAAAGAA AACCGTTTCT 180

TTCAATAAAT TAAGGAGGAT CCCATGATTT TTATGACAAA AAATGGTAGA ATATGCATGA 240

AAACGACTAT TATTTTATAT TGACGGAGGA TATCATGACC GAAAACTATC AAGATGGACA 300

CCTTACGCAT CTTCAGTTTA AATGCAAATC GACCACTTGG CTGAAAAAAA TCGCGnCATC 360

AGTAGGaACA GAATTAGGAA AAAGCACCGT GcGCCaATTT AGTGaTGGcG AAaTTCaAAT 420

TAAtATTGAA GAAAGTATTC GTGGCGATCA TGTCTACATC ATCCAAGCAA CGAATGCTCC 480

TGTAAATGAT CATTTGATGG AATTGCTTAT TTTAATTGAT GCATTAAAAC GTGCTAGCGC 540

AAAAACAGTT AATGTTATTT TGCCTTACTA TGGTTATGCT CGCCAAGACC GCACAGCGAA 600

ACCAAGAGAA CCCATTACTG CGAAATTAGT TGCTAATATG TTGGTAGAGG CAGGCGCAAC 660

ACGTTTGTTG ACATTAGACT TGCATACTGT TCAAGTCCAA GGATTTTTTG ATATCCCAGT 720

AGATAATTTA TTTACAATGC CTTTGTTTGC TCATTATTAT CGTCAACAAG CATTAGTAGG 780

GGAAGAAATT GTGATTGTTT CGCCTAAAAA CAGTGGCGTT CAACGCGCAC GTAGTTTATC 840

GGAGTACTTA GATGCGACCT TAGCAATTGT CGACCATGAA GAGATCGATG GGGTGCGCCA 900

AGAAGGTTAC GTTATTGGGA ATGTCGCTGG GAAAAAATGT ATTCTTGTTG ATGACATTTT 960

GAATACTGGT CAAACGTTAG CCACAGCGGC GGAAGTTTTA ATGAAAAATG GTGCGCAAGA 1020

AGTGTATGCT TGCGCCTCGC ATGGTTTGTT ATCAGAAGGC GCAAAAGCAA CCTTAGAAAA 1080

TGCGCCAATT AAGGAAATTA GTATTACTGA CTCTGTTTAT ACGACAGCAs ACCGCCAACC 1140

AGCAACCTTG AATATTATTT CTTGCGCTGA GTTAATGGGG GAAGCGTTAC TTCGTATCCA 1200

TGAAAATAAA CCAATGAGTC CATTATTCCG TTTAGAACCA AAAGGAGAAT AAGTGATTCC 1260

TTAAACTAGA AATAGGTCTA ACAACTTTAA TAGTGTTGGG CCTATTTTTC TTATCTCTTT 1320

ATAAATnGAA GAGAAAGT 1338

(2) INFORMATION FOR SEQ ID NO: 683:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 894 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 683:

TTCAATGAAT AGAAGGCTGC ACCTAATGTG TAAGACACAC CCCCTGCGAC CAATAACGCT 60

GTTCCTGTAA ACCCAAGAGA TTCATATAAC GGACGAGCAG CTATGATACA GAGCCAACCT 120

AAAACAATAT AAATAATCGT TGAAATGTTT TTCACTGTTT CTTGTTTATG CAACGTGAGG 180

GATTTATAGA CAATACCGAC GATGGCGATT AACCAGACTA AACTAAATAA GGTCCAACCT 240

AACCAACCGC CAATACTAAT TAAGCAAAAC GGCGTATAGC TGCCAGCAAT CAATAAGAAA 300

ATAGAACTAT GATCGAAGAC CTGAAAGACC TTTTTGGCTC TTGTAAAAAT TAAACTATGA 360

AATAACGTTG AAGATAAAAA AAGTAGAATC AACATGGAGC CATAAATAGC ATAAGATACC 420

ACGTGAATCG GTGAGCCTAA ACGAGCTCCT TTAACAAGTA AAATGACTAA GCCAGCAATA 480

CTTAAACCTG CGCCTATGCC ATGTGTAACT GCATTCAAGA CTTCGTTCAC GATTAGGTAT 540

TTTTTTGAGA AaTGCGTTTT TTCCAATAAA TTCACTTCCT TTATTTTAAT TTGATGAAAA 600

ATtCGGCAGT GTCTGTTATA CTATAACTAG ATAATTAACC GTTTAATGGT ATTCTACCAT 660

GAAAACAGGT TCCTTTTAAA GAAAGAGTGT AAATTTATGA CAAACGTAAA ATCGTACGGT 720

TCTCATGACG TGGAAAAAGC TAGGGTGATT AATATCATAT GTGCTTATCA TTATGGTGTG 780

GGTGTTATCA GTGTGTCATT GCAGCGAAAT TATGGCAGTG GCAATGGAGG CTACCCAACG 840

nGCACCCCAT TGGGATGnGA TAnTGTCCTG GGAGTGCGGA GATTCGTCCA GCAA 894

(2) INFORMATION FOR SEQ ID NO: 684:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 644 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 684:

GTAACGATCG TATTGACAAT TGACCTTAAT TACAGCCATT AGCACCGCTG CCCCAAAAAC 60

TTGAGAAGCC AAAGTTCATT TCCCCCACAA AAaACGAGGT CATTTGTAAA ATCGCTGCAT 120

TGACAACAAA GCTAAATAGA CCAAACGTTA ATAAAGTAAA TGGGAGCGAC AAGATTGTTA 180

AGACTGGTTT CACCAACATA TTTAGGATTG AGAGAACAAA GGCGGCAATA ATTGCAGGCA 240

TTATGCCTCT CACATGAATC ATGCCAGGAA AAACAACTGA TAAAGAAATG AATGTCAGTG 300

TATTAACAAC CAATCGCTGA AAATATGTCA TTAAAAGTCA CTCCATTCGT CATCATCAAT 360

TTTTTCCGCT TCTTTACGTT GCTTTTTATT GGCAGCATAA GGATTTTCAT TGCGAACATT 420

TCTATTATAA GGATTTTGAT GGCCGTAGCC GCGATCGCTT CCTGTTCTTC CAGAAGGGAT 480

CAAGACTGCT AAAGCAATGT ATAAGAGAAT GCCTGGAAAT CCTGcTGAGA AAAAGCTGAG 540

CmAGACATAA ACGACACGAA TAATTGTTGG GGTCAATTCC TAGCCAnTCA GCAATCCCAG 600

CTAAGCTTCC TGnTAAAACC ACGTTATTCG GAGATTTTGG TAGT 644

(2) INFORMATION FOR SEQ ID NO: 685:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 501 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 685:

GCATTTnnTA GTGTCCGATT CGnCTTAGAA gTATTTTcAA TcGTAAaGCT AATAAAaTTt 60

CGTtGGCTTt CCGTTtAAaG GTCGTTyCAT CAGTTAAAaT CTGkTGcTtG rCCAATTGCk 120

GAGTCAaTGT TGAmCCmCCr GAgAcCGCsG ATACCTGTCG CATCAGAAtC AACGCTCGAA 180

CTAGCGCTTT TGGTACCACC CaTGATGTTC yTCmAAATAT caTCTTCTGT GGAAATAATG 240

GCTTTTTTTA GCAtGGAGAC ATTTGATCGC CATTAATTCG TGTCCGAATT AAATCTGATT 300

TGATGTTAGC AATCGGAGTA CCATCTGCGT AAGTCATTTT TGACACTTCG GTAATATCAC 360

TGATTTCTTT TTGCAATTCT TCCTTCGTTG GGGGCTGTGT ATCTTCAACT AGAAAAGCAA 420

AaTAGCCCAT TCCAATACCT AAGCCAAGAG AGCCACCTAA GACAATCAGA tAAcCCCAAA 480

ACCACTAATG CTGAAAACGC G 501

(2) INFORMATION FOR SEQ ID NO: 686:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 829 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 686:

CTTTGTAATT GATGTACGCC AAAATCCTGG TGGACTGCTA GATCAAGCAG AGCGAATGGC 60

CAkTATGTTT TTAAAAAATG GCGAAACAAT TGTTCAATTT GAAGACAAAA AAGGACGTAC 120

GATGAAAGAA GTTGCTTCCA AAGAATTAGA CGGCGGCTTT aAAGTAAAAG AACCTGTTGC 180

AGTTATCATT GATGGCAATA GTGCCAGTGC ATCCGAAATT TTTGCTGCAG CACTGCATGA 240

ATCTGCGAAT GTCCCATTAA TTGGAACGAA AACATTTGGC AAAGGAACGG TCCAAACGGT 300

GAAAGATTTA AATGACCAAA CCGAAATAAA ATTGACTGTT TTAAaATGGT TAACACCAAA 360

AGGCGAATGG ATCArCGAAA AAGGAATTGA ACCAACCATC AAAGCTGATT ACCCCGAGTA 420

TGCCTATTTA AAATTGATTC CTCGCGATAA AACGCTAAAA GAAGGAGATC AATCTGAAGA 480

TATTCAGAAC TTAAATGCTA TTTTAGCTGT ATTAGCGTAT CCTGTTGATG AAAACAATGC 540

AAACTACACC GCAGAAACAA AAGCAGCGGT wAGCGATCTT CAACaGAAAA ATGGCTTGCC 600

TGTAACTGGA GAAATCGACA ATGAAACAGC CACAAAAATT GAAGCCACTT TAGGTAAATT 660

AATTCTTGAA AATGACGCTG CTTACGATAC AGCAGTCAAA GAAATTCAAA AAAACTAGAA 720

AAArGTGGCA CGAaGTTGAA GaAGAAcGTG ATTATGTTGG gTATTtAATG kACTTyCTGA 780

AAATnTTAGt ACCAGCAtCg TAgCCGTTTT tATTTtAAGr GGtTTTCCC 829

(2) INFORMATION FOR SEQ ID NO: 687:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 485 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 687:

TAGATTCTAT AATGGTGACA GGGCTAGTAC CATATTTAGC GATGAAAGAT AAGTCAGATT 60

CGAAGGATGA AGGGGTTAAC CTTAGTGCGT TAAACTTTAA AGAGCAAAGC GTACCGTTCA 120

AGATATTAAA AATTAAAAAT ACATTTACTG ATGAAATGCG TTTTGAATAT CAAACACACA 180

TTATGGATAC TGCAAAAAAT ACTCCAATTA TGAATAATGA AAAAATTAGT TTTATCAGCT 240

ATGAAAAATA TATAGTTACT GGTATGAAAA GTATTTTAAT GAAAGCAAAG GATTCTAAGA 300

AAAAAATATT GGCATATATT AATAATAATT TACAAAATCT TATAGTTAGA AATGTAATCA 360

GACCAACTCA AAGATATGCT GATATGTTGG AATTTTCATA CCATCCTAAT TGCTTTTCTA 420

ATGCAATAGA GAGAGAAAAA GTACTTCATA ATAkGTGGGG CCTATCCCTT ATAAGGAnTA 480

AAAAG 485

(2) INFORMATION FOR SEQ ID NO: 688:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 970 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 688:

CTATACGCCA CAAACAATCT CCAACATGAG TAAATTGGTG GCTGAAGATG TTTTAGCTTT 60

TAAAGAmAGA ACGTTAGAAG CCAATTATTC CGTTATATTT ATGGATGCGA CGyaTATCCC 120

TGTGAAGCGA CAGACTGTTT CGAAGGAAGC AGTGTATATT ACGATTGGTA TTCGTTTGGA 180

TGGAACCAAA GAAGTTCTTG GGTTTACTAT TGCCCCAACT GAGTCTGCTT ATATTTGGAA 240

AGAAGTTCTT CAAGATCTTA GAAAACGTGG GTTAGAAGAA GTTTTATTAG TAGTGACAGA 300

CGGATTAAGC GGTATTGAAG AAAGTATCCA TAGTGTGTAT CCGAATGCCC AATTTCAACA 360

ATGTTGTGTG CATGTATCTA GAAATATCGC TCATAAAGTT CGTGTTCGAG ATCGAAAAGA 420

AATTTGTGAG GATTTCAAAT TGGTTTACCA AGCGAATTCA AAAGAAGAGG CATTGGATCA 480

CATCGACTTT ATGATTAGGA AATGGAAAAA GCAGTATCCA AGAGTCGTCA ATTTACTCTT 540

GAATCCTGCC CTATTAACCT TTTATAATTT CCCTCACGCC ATCAGACGAA CAATTTATTC 600

GACGAACCTG ATTGAAGGCT TTAATAAGCA GCTAAAACGA TATACTCGAA GAAAAGAACA 660

ATTCCCTAAT GAAGAATCTC TAGAGAGATT CTTCATTTCT CAATTTAATC AATATAACCA 720

AAAATTTTTA GGTAGAATTC ACAAGGGATT TAAAGAAATT CAGGATACAT TAGAGTCGAT 780

GATTTAACTG TAATTAACAG AATGGATTTT CCATTTACAC ATAATTCTTG ACGCTACCAT 840

AATCTAAATA ACCATCATAT TCATCTACTA CTTATATAAT TGAGTTTTAT TAAACTTTTT 900

AAGTTTCTAT TCTATCAAAT ATTTAAACAA TCTCTGCATC CGAAAAGACT ACCTCGTATA 960

GTTTTACTAA 970

(2) INFORMATION FOR SEQ ID NO: 689:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 383 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 689:

ATAAACATAT TGGTAAAAAT TATGTCTCTT CAAAAAATCC ATCTGkTTTT GAAAATTTTC 60

TTTATCCGTT ATATCCATAT TAGCATCTCC cTTTTTTAAC GATTACTATT GCTTATTCTA 120

TCGACTTTTT CCCTATTATT TACTTTGAAT AAATCCTCTT TTTATTAGTG TTGTTAATAT 180

AGACGAAGCA CACCGCAAAA CAGCGCTTTC ATAAAGGACA ACTATAATAA GAAATAATAG 240

TTAGTTTTAT TAAATCATAA ATACAGCTAA AATAAATCAT ACAAAAGTCC TAAATCAAGT 300

TTTATTTTAA GTkTCTAACT TTGATTGCkT TAGTTAAAAA TGACACmACC aTCTTATATC 360

nTATTCCATA CCATTTTCnT AAT 383

(2) INFORMATION FOR SEQ ID NO: 690:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 850 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 690:

GnCTATGTTG AGAAnAAAGT CGGTAAATTA GAnAGATATT TCCAGCGATT CACCTGAAGC 60

AACGGTTCAC GTGAATTTGA AAGTGTACAC TGAAAAaAaC AGCCAAAGTC GAGGTTACTA 120

TTCCTCTTCC TTATCTAGTT TTACGAGCTG AAGAAACGTC ACCTGATTTA TACGCAAGTA 180

TTGATTTGGT TGTCGATAAA TTAGAACGAC wAATTCGTAA ATTTAAAACA AAAATCAACA 240

GAAAAtCTCG TGAAACAGGA CGTAACAATA CAAAAGCCGC TGTCTTCTTA GTTGGTGAAG 300

AAACAGAAGA AACACCTTCA GAATTAGACA TCGTTCGTAC AAAACGTCTA TCATTAAAAC 360

CGATGGACAG TGAAGAAGCT GTTTTACAAA TGAACATGTT AGGACACAAC TTCTTCATTT 420

TTGAAGATGC TGAAACAAAT GGAACAAGCA TTGTTTACCG TCGTAAAGAT GGCAAATATG 480

GCTTGATTGA AACAGACTAA AGCTTAAAAT ArGAGCTGAA ATGCCTTCGT GCATTTCAGC 540

TCTTTAAACA TCTTTAAAGC ACGGAGTAAA GAAAAAATAA AGAAATACAT AAAATAACTT 600

TAAAAGTTCC TCTTTTAAAG GAAAATGTCT GTAAAATAGA GAGAAATGGT TTTCTTTCTC 660

GTCAAGTAAT GATAAAATAA AACGGATAGA TTGTGCTTCT TGTACAAAAA GGATACAAAA 720

ATAAAAAGAG GTAATTGAAA GGACACAGTA CCATGGCAAT TTTTTGAAAA AATGTCGAAA 780

CGTAAAAAGA TTAGCGTTGA AAAACGCGTA ATGTCCTGCA GGCATGACAC TACGTGACAT 840

TCGGAAACGT 850

(2) INFORMATION FOR SEQ ID NO: 691:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1426 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 691:

ATACTCACCT ACATCTTAAT TCCAACTAAA CTTAATACCC ATTTAACAAC TGCAAAAaTA 60

TTTTTTCCTT GATTCGTCGC AAAGTCCTGT AAAATTAAGG CAATGACAAT TACTGAACCA 120

ATTTTTGCCC ATTCGGCACC TTTCCAATGT TTTAAAATCA AGATAATCGC ACCAGCCGCT 180

AAAATTGTAA TAAAACCTGT ACTTAAAAAA CTACTCAAAT TGCCTGTATT CATTAGTCAT 240

TTCCTTCTTT CTCTGTAAAG TAATGTTTCA TCGTATTAAT AAAATAAGAA TTCTCCGTTT 300

TAGATAACCA AAGCGAAAAA TCCTCAGTAT GCGTTAATTT TGTTGTTTTA TCTGAAAAAG 360

TGACCGATAC TTGAACCCCT GTTACACCTT TCTGCTTTGT ATCAAAAAAT TTCATATCAC 420

CAGAATCTAT TGAAACAAAT GAGTCTTGAC CTGCTGTCGT TGTTGGATTT TTCATTAACA 480

ACATTAATTC TTGTTTGTTC CCAACTGAAT ATTTTGTAAA AAACATCTTT AAAAATTTCT 540

CTATAGATTG TTTTTCTTTC CCATCAACAT CTATTGAATT TTCTGTCTGT TTTAGAGATA 600

CTGTTTTTCC TTGTGGTAAA CTACTTGCTA CAAAATAAGG ACGTCCTACA ATTGAAAACA 660

AATTATCTTT TACTTGAATT GGTAAAACCA TATTCTGTGT CACATTTTCA GTCTTTTTAT 720

CTTTCTCCAA TTCATAATTC ACATTTAGAT AAACCAAATA ACAATCTTCC ATTTTCTCTA 780

TACGTGATAA AGACGTATTT TTAAACGTAC GTAAGACATT CCCTTGTAGA TTTTCTGTTT 840

CATCAACATC TTTTACATCT AGTGATAAAT ATTTAGATAA TCGATCAATA CGAGCCTTTT 900

TAGCTTCATC ACTTGCTTTA GAGTTGTAAT TTAAATACTC TTTCGCAAAA TTTTCGACAT 960

AAAAAGAGAG GGCTGAATAA TCAAACTCCT CTCTTACGAC GTTATTCTTT TTTGATTGTT 1020

CGCTTAATGT ATCCACTTTT GAATATAAAC TATCTACATT ACTTGCCATC ACATTGGCCC 1080

GAATCGCACC AACTCCTGAC ATCACCAGTA AAACACTAGC GCCAATTAAC ACAACCCGAT 1140

TCGCTTTCTT TTGAGAAAAT GTCTTAACCT TTGGTGTTTT TTCTTTTTTC TCTTTCTTTT 1200

TAATTTTCTT TAAATTCATA TTCCCCTCCT GCAATCAATA AACCTAAACA ACTTTCAAAT 1260

GATTTTAAAA AGGTAGATTT AATAACTTTT TCAACATAAA CTCTTTCTTT AACACGACCT 1320

TTATATGTTt GAAAGCCTtC TTTATACAAA TAAAAATAAT ATGAATTGTA CATTTTATCT 1380

CCATTAGGTT TTCTCAAGTA TTGACCTTCT GCTATTAAAC TAGTTA 1426

(2) INFORMATION FOR SEQ ID NO: 692:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 540 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 692:

AATAAAGAAA ATGAAATtCT AGGAAaTAAG rAAATTACTA CTAAGcAGAC CAsATGTTGA 60

GGGGAAAGAT GATATTTCAn ATAGTCGATT TGCTCTTTGG CAAAGTGCGA TnGATTTAAT 120

GAAATTAAAT CCTATATTTG GCACnTCTCC AAAAGGTTTT GCAGATATAG CTAAAGATAA 180

ATTGCCTCAA TCTCATATTG CTAAGACAGC TCAAACACCG CATAGTTTTT TCTTCTATTT 240

ATTAGCAGCT ACAGGGATTT CAGGTACAAT TGTATTTCTA ATTTTCTTGC TATCTAAAAT 300

GTTTAACTCT GCTAAATTGC TGTTTAGCGT TCGAATTAAA AATTATATGG ATTTTGTGTT 360

AGATAACCAA ATTGTGTTAG TCATTTTAAT TTCTGGATTA CTGATAACTG AAGTCATTTT 420

AACAAGACGT TTTGCAACAG TTATCTTTTG GTTATACTTA GGTAAAATTC AGTATCGCAC 480

TGACATGGAA AATGATTTGA TTAGAGGAAA TGAGCaATGA TGCTCATTTT tTCTACTATA 540

(2) INFORMATION FOR SEQ ID NO: 693:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 483 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 693:

GAAAAACAAA AAGGTGTCAT GACATCTTTA GGAGGCAGTT TTAATGAAAA CAAACTCAGT 60

AAAACGTTTA ACAATTTCAG CACTTTTAAT TGCAATGGGG ATTATCATTC CAATGGTTAT 120

GCCAAGAATT ACCATTGGTC CAGCCAGTTT TACATTAGCA AGTCACGTCC CAGTTTTTAT 180

TGCCATGTTT ATTTCTCCAG TGGTGGCAAT TGCGGTTAGT CTAGGAACAG GTTTTGGTTT 240

TTTCTTATCC GCAACACCGA TTATTGCTTT GCGGGCCTTA TCTCATTTGA TTTTTGCAGT 300

CATTGGCGCA tGaTTTTACA AAAGCATCCA GAAATTTTAA TTAATAAAGA GGGGAAATTC 360

ACTTTATTAA ATGGAAAATT ACAGTTGTTT AACGTGGGAA TTGGAGTGAT TCATTCCGCt 420

GCCGAGCTAG TGGTTGTTct GTaTTtACAC AATGGGCAAC TTGcCAGGAA CCTATnATAC 480

AGC 483

(2) INFORMATION FOR SEQ ID NO: 694:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 416 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 694:

AAATGAAAAA TTGTCTGACG GGGAACCAGC TCAATTTGAT GCAAAAGAAG AAACGTTTTG 60

GAAAAAAATT TCGCAAAAAG TTATTTTATA AAAGTAATAC ATACGAATAA GAGCGTACAC 120

ATAGAAAGAA GCGTAAgCAG TTTAACACTG CTTACGTTTC TTTTTTAATG AATAGGAATA 180

ACTGTTTCAC GCTTAATTAA TTTGTGAGGA ATAACGAGAC GGACACCATT TGTCAATTCT 240

TCGTGGATTA ATTCATATAA TTTTTTCATC GCCATTCGAC CTAATTCTGA CGTGTCGATG 300

TCAATGCTTG TTAGATAAGG ATGTGTCAAT GTTGgAAAAG ATTGAGTtGT TAAAACTAAT 360

AATGGATAAC GTTTCCrGGr ACyTGaTAAC CAwAnGATTT GCGCaAGTTg CATCgG 416

(2) INFORMATION FOR SEQ ID NO: 695:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1442 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 695:

GCCTGATACA TACATTTCCA ACATTGAAGC CATGAGGGCT TTTTCGTTtC GTTGATAACG 60

TTCAAACACT GTGGGTGAAA AATGGCCATC ACGTGTTCTG GGTACTTTTA ATTCTAGCGT 120

GCCTACACGT GTCGTAAAGc TGCGCTCATA ATAGCCATTT CGTTGACTTT GTCGGTTTTC 180

TGTTCGTTCA TATTCTTTTG CTTGAATATA TTCTGTTCGT TGATTTTCCA TTAGTTGATT 240

AAATACCGTT GTTAAAATAT TTTTAGAAAC GTCATCCTTT ACAGAATATT CAATAATGCT 300

TTGAATCTCT TCGCTTTTCA GTGTAAAATG TACTTGGGTC ATGTAAAAGT CCTCCTGGGT 360

ATGTTTTTGT CGTTAAAAAC ATTGTACCGT AAAAGGACTG TTATATGGCC TTTTTACTTT 420

TACACAATTA TACGGACTTT ATCATAATTT AAGTAATACT AATCAATGAA ATTTGAATTT 480

TGAAATAGTT ATTTATTTTC ATTTTAAATA CGTCTTTTAA TTAACATTAA CCTATGATTG 540

ACTATAATAA AAAATAGGCT CTCTTTGTAT AGAAAGTCTA TTAAATTTTA ATAATTTATT 600

TAGTTGACAA TAAGAATATA CCTTCAGCAA TTTTATTTTT AGTAATAGGA ATTACTTCTG 660

AATAACCTTT AGAGTTTGTT ATAATAACCA TTATTGTATC ATAAAGGCCT GAATTTTTGA 720

TCATTGGGTT GTCCAATGTT CCAAGAACCT CCCCCTTTTT ACTTTTTGAT TAGCAACAAA 780

TTTTTTAGCG AATCATTTAC TACTTAATGA TACTGTATCA ATACCAATAT TAATCAGTAA 840

CATGTCCAGT TTCAAATAGA ACTTGAACAA TACCACATCT ACTGGAGAAT AAATAGTGTT 900

TCTATTTGGC TTAATAGGAA CTATTTTCTT ATTGTTGATC CGTCTTGTGT TTGTATTGAA 960

AAAGAAGAAA TAATGAGGGA TTTAAAATTA GCAAAAAAAT TGGTAAAATG AACTAAGGTA 1020

ACGATCGCAC CCTCAATATA ATAGTGTCAT GAATTTAAAC CTCGATTTTT TTGCGGTTGT 1080

TACCTTTTTA TTTTAGGGTG GAAAATAAAT GTcTCTTTAT TTCTTTTTGT TATTTCGCAA 1140

GTGATAAACT CCTGTTAAGA TAGTGGCTAT ATAAAAAAGG CAGATACTTA AGAAGAGTAT 1200

ATTCATGATT TCCTCCTAGT TTAAATGAAA GTCATGAATG ATTGTACCAA GATTTTTTAG 1260

AGAAGTTAAG ATTGAATAGA AAAGTAAATA CGATArAATT TTTATAGACT GTTACGTTGA 1320

AATGTTCTTC AGAGACTATT TCTTCCAATA CCnCCTACTA GATAGGTGGT TTTAATCATT 1380

ATTTTCGGAG GTTGTGTTTT GGCCAAAAAG TAAAGATTTA ATGGATGTAG GTTCATTTGT 1440

GA 1442

(2) INFORMATION FOR SEQ ID NO: 696:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1627 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 696:

CTATGATTGG CTTGGAGAGC tTTCGTTAAT TGATTTCTTA CGTGATGTTG GAAAAAACTT 60

TACAATTAAC TATATGATGA GTAAAGAAAG TGTGAAACGT CGAATCGAAA CAGGAATTTC 120

TTATACTGAA TTTGCGTATC AATTATTACA AGCGTATGAT TTCTTGAAAT TGTACGAAAC 180

AGAAGGTTGT TTATTGCAAT TAGGCGGCAG TGATCAATGG GGAAACATTA CTTCAGGGAT 240

TGAATTATTA CGCCGAGAAC AAGAAGTTCA AGGATTTGGC TTAACTATGC CTTTAATTAC 300

AAAAGCAGAT GGTACCAAAT TTGGTAAAAC AGAAGGAAAC GCAtTTGGCT AGATGCGGAA 360

AAAACTTCGC CATATGAATT TTATCAATTC TGGATTAACA CAGATGATCG TGATGTCGTG 420

AAATTCTTGA AATACTTTAC TwTCTTAACA TTGGATGnAA TTGCAACAAT TGrAGaAGAA 480

TTTACAGCGA ACCCAGGTCA ACGTGCTGCG CAAAAAGCGT TACCAAAGAA GTGACTACTT 540

TAGTCCACGG TGAAGCAGCC TATCATCAAG CGGTTAAAAT TTCAGAAGCT TTATTCAGCG 600

GGGATATCCA AAGCTTAACA GCAGAAGAAA TTAAACAAGG CTTTAAAGAT GTACCGACGT 660

ATGAGGTGCA ACCAGAGGAT CAATTAAGTT TAGTCGACTT ATTAGTTACC TCArAAATAG 720

AcCATCCAAA CGTCAAGCCC GTGAaGATGT TCAAAATGGA GCGATTtACG TCAaTGGAGA 780

aCGTaGACAA GATTTAGCAG CTGAATTGAC AGAAACTGAT AAAaTCGAAG GACAATTTAC 840

TGTGATTCGT CGCGGTAAGA AAAAATATTT CTTGTTAAAA TACTArAAAA ATGaAAGCTT 900

GGGACGaAGG TCGTTTaGAC AACAGGTTCC AAGCTTTCGC TTTTAAAAAA gATAAATTGG 960

GaTTAATTCT GTTACAAAAA ACTCTTTTTT aCTATTGGGc mTATTCaATT aCAGATAACT 1020

TGTGCTATGC TAGAATAAAA AGTGCAGTCG CTATAAGGAG TAGGTTATGA AATCAATTGT 1080

CGAAAAGTTA AACTTAACGA AGTATCAGCA ACTAGTTATT TTAAATAGAC CACAGGGAGA 1140

GTATCTATCT GAATTTGCTT CAGCCGCGCA GCAATTACCC TCAGAACCAG TTGAGCTGAT 1200

TTTTGCATTT GTGAAAACAA TGGCGGAATT TCAAGAGATT GTTTCAATGG TGATTGAAGG 1260

TCAACATTTA GTAGAAAATG GGTTGTTATA TGTGGCTTAT CCGAAAAAAG GCAATAAAAT 1320

GTATCCAACC TTTGTCCACC GAGATGAAAT TTTTCCAACA TTACAAGTAA ACGAGGCAGA 1380

TGGGTATATT AAAGGCTCAA CCTTAAAATT TAATCGCATG GTTAGTTTAG ATGAGACTTT 1440

TACGGTTGTC GGCATGAAGA ATGTTCTTAA AAAGCTAGCG AAGAATCAGA CCAGTTCTGC 1500

TGTTGGGGAT ATGTTCATAT TTACCGGAAA TTGAAGCGTG GGACAACAGA CCACTGCTGC 1560

AGCGACATCG TGCATACGCn GGTATCCAAA GATGGCCCGn ATGTTTAGGC AACGACGnGC 1620

CCGnAAC 1627

(2) INFORMATION FOR SEQ ID NO: 697:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 869 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 697:

TCAAACGTTT ATCATAAATA AATGTAAATT ATAATATTCT TTTGATTTAA ATAGATTTAT 60

ATATGTnATA CGTTTGACTT TTTTCTGATT TGTTTATACT ATTAGACTAA CTACTATTTT 120

CTAATAAAGG AGACATTACT TATGTCAAGC ATCATGAACC AATGGACGGA TGAATTACGT 180

TATGCGCCTT ATTCTTCTTG GACATCTGCT CACCTCGAAA ATCTAACTTC TATTATCGCG 240

CAATCTAGTT GGCGTTTTAA GTATCATATT CAACCACAGA CAGGACTACT AAATGATCCC 300

AACGGTTTTT CGTATTTCAA TAACCAGTGG CATTTATTTT ATCAAGCGTT TCCTTTcGGG 360

AGTGTTCACG GACTAAAAAG TTGGGCCCAC TTAACTTCCT CCGACTTAAT TCACTGGGAT 420

TATGAAGGAA TTGCCCTTTA TCCCGACTCT GAATATGATT CTCATGGCGT CTATTCAGGC 480

TCAGCTTTAA CAATAGATAA CCAACTATGT TTATTTTATA CAGGAAATGT TCGTGATCAA 540

ACTTGGCAAC GATTTGCATA TCAAAATATT GCATGGCTGA ATTCTTTAGG TGCGATCACA 600

AAGGArTCAA CACCATTCCT ACCTATTGAC CCCAATTATT CTTCCCATTT TCGTGATCCG 660

ATGGTATTTC CTTATCAAGA AGGACTTGTT TtATTAATTG GTGCTAGTGA TTTAAATGGA 720

CAAGGAAAAA TTGTGGTCTA TTTTTCTAAA GATCGAAATG TACACAATTT TCATCAACTT 780

GGCGAATTGA CGTTCACCAA CCAAGAATTA GGCTACATGG TTGAATGCCC CAATTTGGTA 840

TTTATTGATG GCCAGCCTGT CTTATTATT 869

(2) INFORMATION FOR SEQ ID NO: 698:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 774 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 698:

TGTCAGATTG CTCGTTTATC GTTTCAATGC AGCATTAAGG AGTAACGAAA TGCGATTCGC 60

TACTCCTCAA AGACTGATTT CTATTCTTTT TCTGAACCTT CTGTTGCATG ATAGTTCACG 120

CTATTTGAGG TACCTTTTTT AGAAAAGTTT TCAATTAGCT CTTTCACGTC GAGGCCCGTT 180

GTTTCTTTTA GTGTTTCTTG TGTACCAGCC AACAAATTTG TTGCATAATT TGTAATACGG 240

TTGGCGCCAG AATTTTCGCC ACCTGCGCCT GTATCTACCA CTGAAATCTT GTCGATGTTG 300

CCCAATGGTT GTGCTGCTTC TTTCATTAAT TGTGGCAACA TTTCCATCAC CATACTTAAC 360

ACAGCTGCTT CACCATATTC TTTGAAGGCA TCAGCAATCT TTTGTTTGGC TTCGGCTTCC 420

GCTGCACCAC GAGCTAAAAT AGCTTCGGCT TGCGCTTGAC CTGTTAAACG CGTTTTGTTG 480

GCTTCCGCTT CGGCTAATGC TTCAACTTTG AAGCGTTCTG CTTCGGCTTC TGsCACTTCG 540

CGTGcCTTTT GAGCAAGAGC TTCTTGTTCG CGTGCGTAAC GATCTGCATC GGCTTTTTTC 600

TTCACTTCTG AGTCGTATTG TTTTTCACGA CGCGTAATTT CTTTTTCTTC TAACTCAATT 660

TGTTTTTGAC GTTCAACGAC TTTGACTTCC ATTTCTTGTT CCACCACGTG TTGTTGTGCA 720

CGCGCGCTTT CCAAGtTATA GGCTTGGTCT GCTTTGGCTt TGGCAATATC TTGT 774

(2) INFORMATION FOR SEQ ID NO: 699:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 992 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 699:

ATCTGTAAAA ATAATCAGCn TCTCGTTCTT GTGCGTGTTC AATTTGTGTC AAAATTTCAT 60

CATTGATAGC AGACCCAATC GTTTCAATCG TATCCGCTGC AACTACCTCT TGAATGCGGC 120

GTGTGTCGTC TTTTCCTTCA ACGACAATGA TTTCTGCAAC TTTCATTTTT TCTTTCATTT 180

ATCTATCCGA AACAAGCGAT GAGCGkTGAC TCTCGTTTGT TCTGCCACTT CTTCAAAGGT 240

TAGTTGTCGT aATTCAGCAA TTTTTTCCAC TACGTATCGT GTATAGCCTG GTTCATTCCG 300

TTTACCACGA TAAGGAACAG GGGCTAaATA GGGTGCATCT GTTTcCACCA ACATTcTTTs 360

tAGTGGTACC GCCATCGCCG CTTCTTGAAC ATCCAAtGCC TTTTTAAACG TTACGACTCC 420

ACTAAAaGaa TATaCmAACC TAAATCTAAA AATTTCTTCA TCCACTCGGC ATCACCACTG 480

AAACTATGCA TGATTCCACC AATATCTGAA ATTTTTTCAT CTTTTAAAAT TCGATAGGTA 540

TCTTCTAATG CCTCTCTCGT ATGCACACTA AACGGTAAAT TCATTTCCCG AGCAATCGCA 600

ATTTGACGGC GAAAGACTTT TTCTTGCACC TCTTTAGGAT CCTCCATCCA GTAATAATCA 660

AGrCCGATTT CACCTAaTGC TACAATTTTT GGCAACGTTA ATtGTTCTTG TAAtCGTCGT 720

TCAACCTCTG GCGTATAGCT ACCCGyTTCT GTTGGATGCC AACCAATGAT ACTATAAATA 780

TTGGCATATT CCTGACTAAG TAAAAGAGAT TTTTCAATCG TTGGTTCATC AAAACCGACC 840

ACGGCCATCT CTGTAACACC CAATTCTTGG GCATGTGCAA TCGTTTCTGG AATATCATCA 900

TTAAATTGTT CTGCATTCAA ATGTGTATGT GTATCAAAAA TCATTATGAT TGCCTCCTTT 960

ATCTCCTCTC TACTTTATCA TATTTCTTGA AA 992

(2) INFORMATION FOR SEQ ID NO: 700:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 748 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 700:

GCCAGATTAT GAATTGTTAA TTAAAATCCT TAAAAACGAT TCTAACTATC AATTACTTTC 60

TTTTGAAACA nGAACATTTT CGTTAnCTTT TGCAArGTtA GTkGAyCCTA aTACTGTAGC 120

TCATTCCAAA CAATATTTTG GTGGAGAAGA TAAGGACTTA GGAGTTTTTG TAGATATTTT 180

TCCAATTGAT GGTATCCCTG ATACTCAAAA GGAGAGAAAT GAACTAAGAA AGATAACGGA 240

ACAGTATCGG TTAAACTTAA TGGATACTTT AGGTTTAAGC TATGCACGAA gTTTTCATCT 300

TAGTAAGGCA TTAATAAAAT TAATAACTCG TTATCCTAAT CATTTTAAAG TGAAAAAAGT 360

AGGaAACAAT CAATATTGGC GTGATTTATA TCAAACAGAG GCAAAAAAAA CCCCATTTGG 420

AGCAACAAAA CATTGTGGTT ATCTAGAATG GATTCATATA GATTGGGGTG TATTTCCTAA 480

TGAATGGTTT TTGGAATATG AGGATGTAGA GTTTGAAGGA CATAAATTTA TGGCCATTAA 540

AAACCGAAAa CAATTTTTAA CTTTACGTTA TGGAGATTAT ATGGaAATGC CTCCAGTAGA 600

AGAAAGAaTC ACTCATCATC CCwAwGAwww wtACCAAAAA TAmAGATTAT TTATAGGTTT 660

TTTAGAAAGA AGGCGTACTT AAATGAAGCA GCAATATACA GCAACTGTTG TTATGGCTAC 720

CTATAATGGA GAAAAAAATA TCTTAGAA 748

(2) INFORMATION FOR SEQ ID NO: 701:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1002 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 701:

TGGTGCCGAC TTTCGTCATC GCAnCGCACG ACCnTTTCGA AAnGATATGG GTTCCATGAC 60

GTTGTCTTTG nGCTGGTAAG TTTGTGTTAA CTCCCATTCG ACTCTAGGTA TTCCAAGTTA 120

ATCACCATAT TCGGTTTGAT TGACGAAGTA AGAACCGTTT CTCGAATTAC CTGATTAAAA 180

CCAGTGCGAT CATATTTTTC GTAATCTAAA AAGCCGTTAA TTCAGTCCGT AGGAGTGTGT 240

TTACGCTTTT TCTAAGTGCG AACGGAATAA TTCATTTAAA TCGCCTTTAG TGACTAGAGT 300

TTGCACAATT TCkGTAGTAA AATCATTCAT AGGGAAGTCC TCTTTTCTGT GAATTGGTTG 360

TCGTTAACTT TATTCTACAG AAGCGACTTC CTTTTTTGTA TGGATTTTTT CATyTACACA 420

AAATATTTTA CACTCTCAAG TTAwCGTACG ATAACTTTAA wwCATATGTT GTACAAGTTT 480

TTTTGTTAAT TGTTTTTAGT TATATTATCA TTAAATTAAG TGAACTGAGA CAGTTTTATT 540

GGCTTwAAAA TCTATATTAA ATAAAACAAC AAAGACCTTA AGTTGGAGTT TAAGGTCTTT 600

GTTGTTTTAT TTAAGATTCA TAATCACAAT GCCGACAATA ATAACCACAG CACCAATAAG 660

GGCATTCGTT GTGATTGTCT CTTTTAGAAT AAGATTTGAA GCAATTAAAA TAGCGATGTT 720

TGTACAGGCA ACGCCTAAAG GAACAATGTA AGAAACATCG GAACGACTwA TAATTAACAT 780

CCATAAAATA AAACTAATTA AATAACAAAA TAAACCTAAT AAAGATATAA GTGAGATATT 840

CATTGaAAAG ATTGTTTGTT GGaGTTGGAT GTTCmAACTG CTAGAACCTA GTTTAAATAA 900

GACCmGACCA GCmGAAGACm GAAATACGTA AwTAATAAAA rGAaCCmTAA TTTCCCTTCT 960

CTCAAAAATA ATTTCCGTGC AAAAAAATGC AAGACCTCCA TT 1002

(2) INFORMATION FOR SEQ ID NO: 702:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 855 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 702:

CTTGAATAAA CGAAATCAAC TTGTCATTGA GAAAATGCTT GTAAGGAAAT CTTTGATTAG 60

TGCATAAAAA ATGCTCACTC TTATCTGGGA GCAAGACTyT CGTTAATGAC TGTTTAATTC 120

TGATAAAAAT AAAAGACCGT AGCTTATAGT AGGCTAGGGT CTTAAATACG TTATTGAGTA 180

AAAGAATTTA ATGTATACTT TATTTTCGTT TCGTTGGTAT CGATATCAGT AATTTGCTCT 240

TTAGTAATAT AAGCATGAAT ACTTTTTGAA ATCGTAATTT GTACTTCAGA GGTCTTTATT 300

TGTTTAAATA GCGTTCTTTC AAATGTACCT GaGATATCTC CATGTGCTAA AGACCTAATC 360

TTAAAATAGT TCTTACTTTT ATTTGCAGCC AAGACCTCTC CTTCTATCTC TATTTCTTcA 420

GTTAACACGt TTCnTTCATT TATTtGATCA TTCACACTAG ACAACATAAC ATCTAAATTT 480

CTTTTGATCT TTTCTGATTT TtTTATATTA CTCAACTTAA CTTGATTGGA AATTAAATCA 540

ATTGATAAAT gCATTTTCTT CAATAACTTT ATACAGCTTA GCAGTTTCTT CAATGAGCTT 600

TCTTTCATTA AACAGTTCTT CTTCATCGTA CATTGGTATA TAGGCTAAAT TATCTACGAA 660

GnCAACAGAA AAAGAGGCCT CATCAGTAAA TAATCCCGTT TGnCGATCCT TTTCTAATTC 720

GAAATCAAGT TTTAGCGACC CTCTTGAAGG TGCGGACATC AAAAGTTTCT TCTGAGTTCC 780

TAATTCTGCA TGCCTTTGAC TTAAATCTTG CCATGTACTT CTTATAAATT CTAAAGCACT 840

ATAAACAATA TCAAT 855

(2) INFORMATION FOR SEQ ID NO: 703:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 565 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 703:

ACAGAAACAG TTGATGTTTC ATCAATGATT GATGAATTAG CAACAAAAGC GAATGTCGCT 60

TTAAAAGCAA TGGAAGATTT TACACAAGAA CAAGTAGATC ACATTGTCCA TCAAATGGCG 120

ATGGCGGCGT TAkACCAACA TATGCCTTTA GCAAAAATGG CTGTCGAAGA AACTGGCCGT 180

GGaATCTAkG mmGATAAAGC AATCAAAAAT ATGTATGCAT CTGAATACAT TkGGAATAAT 240

ATTAAACATG ATAAAACAGT GGGTGTAATT AACAAAGACG AACAAACCGG CTTAATGGAA 300

ATTGCTGAGC CAGTTGGCGT GGTATGTGGG GTCACACCAA CAACGAACCC AACGTCAACA 360

ACTATTTTTA AAyCcTAATT GCCTTAAAAA CmAGaAATCC mATTGTTTTg CTTTCCATCC 420

TAGCGcACCA AAAATGTTCG GCAGAaGCGG cACGTATTGT ACGGGATGCG GCGATTGCAG 480

CAGGTGCACC AGAAAATTGT ATTCAATGGA TTGAACAACC ATCGATTGAT GCAACATCTG 540

CTTTGATGAA TCATCCAGGG ATTGn 565

(2) INFORMATION FOR SEQ ID NO: 704:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 673 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 704:

ACGTATGTTC GCTATAGATn ATATGTATTA TTAATAAAAG GAGAaCTACA TATGATTACA 60

aAAAmaGAAA ATGGTTACTT TGTTGACGTT AGCTTGGGGA TTGATCCAAT ATCaGGCAAA 120

CAACGTAGAA AGAGACTTAC AGTTTCAACT AAAAAAGAAG CTCAACAAAT TGAATCAAAG 180

TATTTAAGGT TATACCACGA AAATAAATTA ACAACTAATT CAGACTTAAC TTTAAAAGAT 240

ATTTACGAaG TTTACGAAGA AAAATACACG CAAAACCTTA AGCCATCTTA CAAGCAAACA 300

CAAGAAAGAA TTTTTAAAAA CTATATAGAA CCTTATTTTA AAAATACTCA AATCAAACTA 360

ATAAAAAAAC AACAAATCTA TGATTTCCAA CAGTTTCTAT TAACTAGTAA ACCTAAGAGA 420

AAAGAGACTT TATCAAATAA GACAATCAAT ATGATTATTA TTCATCTACA GAAATTATTT 480

AACGTAGCTA TGAAAGAAGG ATTATCCTAC GAAAATCCTT GTAACCAAAT AGACAAGTTA 540

AAAGTGCAAA AAAAAGAAAT TGATTTTTGG ACATTAGACG AATTTACAAC CTTTATTTCT 600

CATATTGATA AAAATAAACC ATTTTTAAAA GTATTCTACC AGTTTGCGTT TTTTACAGGA 660

ATGCGTGCGG AGA 673

(2) INFORMATION FOR SEQ ID NO: 705:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 814 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 705:

ACTGnCGAAT TGCTTATCAG CAAGTATTAG ATGTCGTTTC GCAAAATTTG CCAGCAGGAT 60

TTGAATTAAC AATGCAATAT GGCATGCCAA CCTTTGTTGT GCCGCTTTCA GTATTTCCAG 120

AAGGCTATCT TAACCGTGGT GATGAACCGT TACCATTTAT CAGCTTAGGC GCTACTAAAA 180

GGCAAGTAGC TTTATATCAC ATGGGATTAA TGGGAAATGA AGCCGTCCGA AGTTGGTTTC 240

AAGAAGCCTA TCAGGAACAA GTGTCAACAA AGCTAAACAT GGGGAAAAGT TGTCTTCGCT 300

TCACTAATCC TAAGCGAATA CCCTATGAAC TAATTGGCGA ATTAGTAACA AAAATCACGC 360

TAGAAGAATG GCTGCAACAG TATCAACACT ACCAAAAGAA AAAGAAGAAT TATTAGGAGG 420

AACCACATGT TATTAGGATC aCACGTTAGT ATGAGTGGCA AGAAAATGCT ATTAGGCTCT 480

GCCGAGGAAG CGGCTAGTTA TGGCTCAACA ACCTTTATGA TTTATACAGG TGCGCCACAA 540

AACACGCGTC GGAAACCGAT TGAAGAAATG AATATTGAAG CTGGGCAAGC TTTTATGAAG 600

GAACACAATT TAAGTAATAT CGTCGTTCAT GCCCCGTATA TTATTAATTT AGGTAATACA 660

ATTAAAACGG AAAATTTCGG TTTTGCCGTA GACTTTTTGC GTCAAGAAAT TGAGCGAGCA 720

CAAGCGTTGG GAGCAACGCA AATCACGTTA CACCCAGGTG CGCATGTTGG TGCAGGTCCA 780

GAAGCCGGCA TTAAACAAAT TGTTAAAGGT CTAA 814

(2) INFORMATION FOR SEQ ID NO: 706:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 590 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 706:

TATAAATAGC GAAAATTTTT GACCAATAAA ATCTTTTTCT CTATAAAAAA ACACCTTTGT 60

CGAACTCGTA TAAAAATAAG TCCAACAAAG ATGTCAATAT ATATATTTAA CCTTTCAAAT 120

ATTCTCTTTA CTTTAATAAT CAATAAGGTT GTTTATTGGC GGATAGCTTA ACCCACTTTT 180

GCTACATACC ACTTATTAGC AGTTAAATAC CCTTGTTTCG TCTTTAACCT TGGTATACCA 240

TTGTTTGTAT ATTCGATTGC TTCAACTTCA ACAATCGTAT TCTTCTTAAC AGCCTGTCCT 300

TTTTTAGTAA ACTCAGTATT ATTGTAGAAA CAATCATCAA TTAAGAGCTT AACTTTTTTT 360

GGATTAGTGA TAAAGTAATT GTTATTGGTA TTTCCAGCAG CTGTCACATA GTTTTTATTT 420

GcAGTAAGAT ATCCTCCTTC TGtCTTCAAG CGAGGaACAC CTTTAGAGGA ATACyCaATG 480

CCCTTAAaCT GGaACmACaG TCCCaGATGa TAAGGTTCTG CTACCATTAC GGAAACTCCA 540

ACATCnGTAT AAAAGTTTAT CCnGGACACC CGCCAnAAAA ATTTGGGnGG 590

(2) INFORMATION FOR SEQ ID NO: 707:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1283 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 707:

CCCTAnTGTA AAAATATTAT CCCTAAAAAT AGTCCCAAAC ATCCAAAAAT CAAAGGATAT 60

TTaTaTTTkG ATATGAATTC TaCCaTTGkt TTCTCCtATA AAACTTTTGA TTTAGTCTTT 120

GTATTTTTTT CATGATAATC CTTAAAAATT ATTTCGGTTA ATATTATTTG CTCTGCACCA 180

AATAATTCAT TTAAATTAGT ATTTACCTTT CTTATAAAGG CTGTTTGTCG CTCTCCCGtT 240

TGaAAAGCTT TGCGTAtTTT TCCTTGTATA TAGATTTTGA TTTTCTTTTT CTTAATTACA 300

GATTTTACTG TTGGATCAAC AATAAAAGGT TCTTCTTTTG CAGCAATTAA TACAAAGTTT 360

TCAATTGCTT TTTTTTGGAC TGCTATTTTT CCACTATCAG TCTTTTTTTC TATACAATTA 420

TGTGCTTTAG GATAAAACAA TACAATAATT AGAATAATAA ACAATAATGT TAGCAAACAT 480

CCACTAATCC AAAACAAAAA CTGTGCCATA CTATTACCAA TGTACGGATA GTTTGTAAAT 540

GGAAGTAGGT AAAATGGTAA ACTAACAACA CCTTGCATTG TTATTAATGT GTGTGCTAAA 600

GGAAAAAATA AAAAAAATAA AACAAGGATT AATAATAATT TTAATCTTTT TCGCATAATA 660

CACCTCTTTG TCTACGTTGT GAATATATTC AATTATCTAG TCATTCCTTG AAAATTCACA 720

ATTCAAATAA AATATATACT ATTAAAAATT TCTAAAGATT TTGTCACTTT GACTyCTwTA 780

wCyTTCTCTC CTTTCAACTG TATAACTTAT TTAGGTGAAC TGATAAAATA AATTTTATCA 840

GTTACTATTT CTAAAATTCT TAAAGTAATC AGACCTCaTT TCAGAAAAAT AATGAATGTT 900

TATTTTATAA AATAAAATAT TTGATGAAAC AATATTAACA AGAAAAATCT ACTAAGTCAA 960

ACAAAATGAA CGATCATATT TTTAGGGAAG AAAAATAATT TAAATTTTAC AATGAATGTA 1020

TGAGcAAATT TGTTGAGCTA ATATTTGGAT ATTTTTATAA TAAGTTTAGG GGTATCCCAA 1080

TATCACTCTC TAGGGGTGAC TCAGAGcAGG GGAAGGAGTT ACCCCTAAAA AGTATGTTAG 1140

AATGATTTTA TCGCATCATA TCCTTTTTTA TGTAATTAGT TGATAAAATC TAtCTGTACT 1200

TATATTTAAC GACyGaAATC GGTCACTTAC TGAAAAAACA TCTAAGGTAG AAATGAATCT 1260

CACCTAAGAT GTTTTTTGAT TnT 1283

(2) INFORMATION FOR SEQ ID NO: 708:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3738 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 708:

GGCCCCTGGG TTTCCAATCC AACCCCAnGG ACAATCCCCT AAnTTTTTTT TTAAACnTCC 60

TCTTTTCCGG TAAACTGGGG CTAGGCTTCC TTGGGTGGGA ATTctTTCtC tAtTCtTCtC 120

aTGgAAACtC aCTCCCaATT AATGgaATTT AtCtTTTTCt AGtAATCGtT TCAAtTCaCT 180

TTcAAAtTCt TCAGCTGaAA AAACATGCAA TCGATCAAAC ACTTCTTkGT ACTCGTCCCC 240

TTCTTTGTCA ATAAAATATT CTTTAAACAT ATCAAGTTCT TCTTTAATGA TTGAACTAGG 300

TGTTACATAA ATTAAGCGTG CATCTTCATA ATTTTTTAAA TGTTTAATGT ATTTTGGAAA 360

CACACCCTGA CTATAACGTC GTTTTGTCTT TAGTGTTAAT TCTATTTCAA ATTGCCACTC 420

TACTTCTCGT CCAATACTAT CCAAAAACAT AAAAATGAAG TCAGGAACAA AGTATTTTAC 480

GCCTTTTTGA TTTTTCTTCT CGTCTTTTAC TTCATATACT TTACGGCGTT CTGATACTAC 540

CGCTTTAAGG TGCGGGTGGT TATTACAAAG TTCCAACATT TTTATCAGAT AATCATTAAT 600

TTGTAAGTGG TGTTGTAAAT TGTATTCAGG AACTTTTGGC AAGGTGTAGT ACCCACCAAT 660

GTAATTATGG CCTTCTTTTG TTAGGTAATA AAATTTTGTT AATCCGATTT GGGAACATTC 720

CACTAGATTT TTTTCCTTCA GTTTTTTTAG ATTACCGTAT ACACTACTTT TTTTTGCCAT 780

ATTATCAAAT TGACAAAAGT CAATAAGTTG TTTTACCGAT ACAACTCTCA ATTCATGTAG 840

TAAACACAAA AGAATCAAGT GTATCTTATC TTCTTTTGAA TAATCAATAA AGTAATTGTA 900

AAGTAATGGT TTCAAAGTAT CACCGCATTT CTATAAAATG TACAAATAGG CAAAGCCTAT 960

TTGTTTCATT GAAAAAGAGT TTCGGAACAA CGTGTAGAAA TACTTTTTCA GCTTTTAAGT 1020

CCTCCAAGTT TTGGTGGAGG ATTTAGGAGG TTGTTGCTTT GGGCGATAGT CCCTTCAAGG 1080

TTTGGTGAAG GGGCGTCTAT CGCAAGGGGT AGTTGTTAAT CTTTTGAACT TTATCTTTTG 1140

AACTTTTTTC TTTAATTTTT ATTACTTCTA TCACCTTACT CATTCTATTT TTTTACTCTA 1200

TTCTGTTtGT CCTCGGCGCT TGGTATTTTT TTCGTAGTkG CAAAATTCCT GACCATTTTC 1260

GCTTAATGGG GGTTTTAACC CACGTTGAGC GAAAATGGTC AGGAATTTTG CAGGCGGAGA 1320

AAGCCTTTTA TACCAAGCGC CGAGGACACG CCACTACAAT TTATAATTCC ACACAATTTC 1380

CAGTGGTTTT CCTAAGTATT TCCATGATGT TTCCAAAATT TGACACACTA TATTAATCTA 1440

TTTTGAATTA CTTTTTTGTA ACCATTCGAG TGCGTTTCGC ACATAAACTT TTTGTGGTGG 1500

GAGGTTTACA TTTTCTGCTG TACGATAAAT AATAAACTCC CCTGTACGCA ATTCTTTTAT 1560

CTTGTTTGGA CTAATTTTGA AGCGATCTAC ATTACGTTTG GTTCCTCTAT CTCCTGACCA 1620

ATCAGATTTT CCAACATCTG AATAACCATC TTCTTGTTCC GTAACTGAAG TTACATCAAT 1680

GTCTTGATAC GTTCCCATTG TTTCACTCCA ATACTCGGCT TCACCTTTTT CGTTGGTCTT 1740

TCCAATGAAC AAAGTATTCA CATTTCCGAC AAGTTGTTCC ATTAATTTAT CGCCTAATTT 1800

TGCAATATCT GCCATCGTTT GAGGGGTGAA AATACCGTAA ACACCTGCAC CTCTACCTTT 1860

ATTTACCATA TCAATAAACG TTTCATTTAG ATAAGAAGCA GGTTCATCAA AAATCACAAG 1920

GACTTCACGT TTAACATTTT TTCGATACAT TTCAGAAGCA AAATAATTTA CGTCTCCGAC 1980

TAACATTTGC GCTAAAGTGC GGATATATTC TTTGTAAATC AATCCATTGA ACGATACGTA 2040

AACTAACTTT CGTTGATTGA CAATTTCTTG CACATCGAGA ACATTTTTTG CGTTTTTAGT 2100

ATCAAATAAT TTTCCTAACT CACTGTCTAA TAACATGTTA ATGTTTGTCT GTAATGCTTC 2160

TGATGTGGCA TACAAATAAA ACGGTGAATC TTTATGTTCa TAACGAATAA ATAATCGTTC 2220

GAACAATTTT TGTTTGATAT TTTCTCCTGT TTCTTTGTCT GTTGGCAAAT ATCTTAGGTT 2280

TCGTTTGAGT AATAAATAGT ACGAGTCTAA ATCTAATGTT TCTGGATTAA GGACAATCGT 2340

TTCCGTTTCT TTTTGTGCTA TTCCTAGTTG GGAAATGTTT CTAAATTTAC TGTTCTTAAT 2400

TTGTGAATCT TCCTCTTTTT CTTCTTTATC AAGGTCTGAA TCAGGCAGTA TCTCGCTTCC 2460

TTCTTTCACT GATTTTTTCT TAGGTTTTTG TATTTTCTTT TCTACTTCAA TCTCAAATAG 2520

TTTGGGATTG TtTGGTAAAA TTGCATCCGC AAATAAGTGT AAAACATTTC TCGGAAGTAA 2580

AAATCGTTGA ACAAATGGTA GACTTCGTTC AACGGTTTCT GTCGTACGAG TGTCTCCAGA 2640

TAAAGTAACC TTGGCTCCTT TAAACTCGTC TAACAATTGA ATTGTTACTT GCAAAAGAGC 2700

TTTnCGctGC GCCTGAATAA AAGACACTTT CTGTTTCGGC TAAAGACACT AACTTATCCC 2760

GAATAGATAT GTCATTTCCA TGCTCTACTG GGTTGTATCT CATATCGCCA TCATCCGTAA 2820

AAGCATGCAC TTCTTTGTCA TAAAATTTGG CAATTTCTCG CATACTTTCT AATGTATCTC 2880

GAGCGCCTTT TCCATCAATC AATATCACCG GAAAATTAAA TTTGACTGCG TGTTGGACGA 2940

GTAATTGAAG TAACGTTGTT TTCCCTGAAC CTGTCGTTGC GGGAACAAGA ACATGCTGAT 3000

TGAACTCTGA AAAGTTGATG ATATATGACA ACCCGAACTC ACTCATACCA AGTAATAATT 3060

GGTTGGATAG TTCTTTATAC TTGTCAATAT CAACACTTTC TACTGCTTCT TGGTATTCTT 3120

TTTCTTGTTC ATCAGTAAAG CGTTGTGACT CTTCAAACTT ATTTTTTTGC CTATTTTTGT 3180

ATTCTTTAGA TTCCGTCACT TTCTTCCGTT CGTCTTCCTT TGAAATAACT TTTCGTTTTG 3240

CCACAAATGA AGAAAATGCC AAAAAAATGG CCCCCATTCC ACTTGATCCA ATAAGCAAAG 3300

CGACTATTGT CTCCCAATTA AAACGAAGTG CTGTATGAAG AAGATTTAAG GGGAAAAGAA 3360

GATACGCAGA CAGATACCCT CTGGGAATCC AGAGGATAAA TGCAAAGACA TTTATTACAA 3420

ATAACACAGC TGAAAAACTA AGAAAGACAA TACTATTTCG TTTCAAAAAA CTCTCATAAT 3480

CCATATCAAA TACATTTGCA GCCTTGTCTC TTTTATCAAC CGCTAGACCA ATTAGTAATG 3540

GAATACTGCC TAAAATCGCA AAAGGATATA AAAGAACTCC GACTAATCCA AAGACTAGTC 3600

GGAGTTGGTG GTAGCGAGTC ATTTTTTGTT GTTCTTGCCC TATTGGCGGT TTCTTTGGTT 3660

GTAGCAACCC ATTGTTAAAC AAATTATTTC CTTGCACGAG GTTGTCTCAC TCCTTTAGGT 3720

TTGTTTGTTA TCTTCTGT 3738

(2) INFORMATION FOR SEQ ID NO: 709:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 555 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 709:

GCGAGCTTTT ATTCAAATTG TTTTACCCAA TGTGTCTTCA GGAATTTTTG CAGCCTTTAT 60

GTTAGCATTT GTCAATTCGT TTAACAATGT TCCTGTATCG ATGTTTTTAT CAGGTCCCGG 120

TGTGACGATG CTGCCTACTT CTTTATTAAG TTATATGGAA TmCaATTATG aTCCaACTGT 180

TTCmGcAATT TCTGTATTGT kGATGCTACT AACaATGGGC TTAATGTTTT TAATTGAAAA 240

AACGTTGGGA CTAGCTTCCA TTGCCTAAAA CATGGAAGAA GCACGAGATA GAAAGAGGAG 300

AAAATTTTGA CTTTTGTGGA TTTAAAAGAT ATTCGCGTCA cTATGACGGA AAACAAGATA 360

TTTTAAAAGA TTTAAATATT TCAAgGAAAA AGGAGAATTA GTTtCTTTAT TAGGACCGAG 420

TGGTTGTGGT AAAACGACTA CGCTAAGAGT GATTGCTGGa TTAATTawGC CAAaCGACGG 480

ACAGTTTytC GTTGACCAAG AAGAACTAAC AAAGGTTCCT GTCCAnAAAA GAAACTTCGG 540

GATGGTTTTT CAAAG 555

(2) INFORMATION FOR SEQ ID NO: 710:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 700 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 710:

TAAACCAAAC TTTTTTAATT TTATCTAATC ATACTAnAAA CATTCAATTT GGAGTCACTA 60

ATCCnGGTTn CTCTTTTGGT TCATCATAAT ACTCTTttCA aCCCcTCTCt ATACTCTtCT 120

CGAaTCTCAA CCaTTGATTT tGGTtCTTGT ACATTGCCTT tCTCTATAAA TATAACGTCA 180

TCAAGAAaTG TTTCCATATC TCTCAATATA TGACTACTAA TTATCAACAC TCTATTTTCG 240

GACAAGGTCT TAACATTCTC AATAAATAAC TCTACACCAT CAGGATCTAA ACCGTTAAAT 300

GGCTCATCTA ATAATACTAA TGAAGGGTTG TGAGCTGTAG CAAGACCTAG CAAAAACCTT 360

TGTTTCATTC CTAAAGACAT CGTATTAAAC TTTCTTTTTT TATTTACATT TAGACCAACA 420

TCCATTAATA GAGTCTCTAT CATTAACAAA TTTTTTAACT TGTAATAAAA AGCATGGTCA 480

TTGTATATAT CTTCAAGCAT AAATCCATTT AAACTATCTG AGTTAACAAA GTTAGTACCC 540

ACCTGcAATA ATATTTTAGG GTCTGCTATT TCTACTTGCC GATTGTCTAT TTTAAGTATT 600

CCACTATCAG GTCTAAnGCA AACTACATAG ACATTTCCAT CAAAGTTGGT TTTTACCTGC 660

ACCATTCGGA CCCAACTAAT CCCATAAAAT TTTnCCCTAC 700

(2) INFORMATION FOR SEQ ID NO: 711:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1226 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 711:

CGTAnATTGG GTGGnGCGGA GCTGTATTCC AAATGACATt CaCGTATTTC AAACTTTTTC 60

AAaGCTAAAT ATAACCTAAT CATACACTCT TGATAACAAT CTTCATCAAA ACTCCCATTT 120

CTACAAGATT CCTTAATTAG TAACGGTTaA ATTkGTyCAC AATCTGAGCG AACGCTTCTy 180

CATCATCCTT TTGAACTAAA ATAACAAGTt CTACCAATGA TTTCATTTCG TATTTCCTCC 240

AAACACTTAT CACTTTTTTT CAGTGAAAGC TCTGGTGGAT CACGATTCTT ATATATTTTT 300

TTAACTAACG TAgTAAACGT TCTCGTATTT TCTTTAGAAT ACGTTGACGC TTTTTTGTTA 360

CACCTTGTCT ACTTAACGAT AATTGAATTG AAATTTCATA ATCCGAATAT CCAAGAATAT 420

ATTTTCGCAT TAAAAAGGAT TTCTCTTCCC TATTTAATTT TTCAAATTCA GTAGCTAAAT 480

CTAAATTGTC GAATTCAAAT GATAAATTAT CTGtTTCAAC yCTGATCATT TCyTTTACmA 540

AGrAAGTCtT ATCTTCAATT AAGAAATCAA GAAACTTTTG TCCAGAAATA ACCTCTGCTT 600

CTTCTATTCT GGAATTTTTT ACAAAGTAAT TTATCGCaGT ATTATGAACG GTTTttTTAA 660

TATAATTTAT AATTTGAGCA TCATATTTTG TAACATCAAA GTTCTGTTTC ATCGCAGCCA 720

CTTCTTCCTT TGGAAATTCG AAAGGAAAAA GATATACGGA CAAGTTAGGG TGTGTTTCTA 780

TACTGCCAAT ATCCTCTCAT TGCACGCTCC ACAGAGATTG AATCATCTCG ATTTATTTTT 840

GGAAGAATAT TATTCACGAA ATTTCTATAA TCGGGAAGAA TATCATTATT TTTAGCTAAT 900

TCCtGATATA TTTGTTTACC TGACAAAGGT TTTCCTGACT CTTTTAAAAT CGATGCAATA 960

ATAAACCCAT AATCTTTGTA ATTTCGTTGA TGAGTGCCAA TCATTTTTCG CTGAATTAAA 1020

TAATCCTGAA GCACTTTTTC ACTTTTGAAA TTTGnAAATT nAGTTTCTTG AGATAGTTCA 1080

TCACTAGAAT CCGAAACAAT CTTGTCTTCT TCTAAATGAA TCAAGTACTC TTTTAAATCA 1140

TAATACTGTC TTGAAAGTTC CCTCCTTTCC TCCATTAACA GTTTCAGCTG TTCATACACA 1200

TGACATTTTT CATCAGGAAT ATCCAG 1226

(2) INFORMATION FOR SEQ ID NO: 712:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 503 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 712:

AGTTGTTCAA TTTTCTACAG AAAATTTTTG GTTTCGCTTT GTTAGTGGAT TAGTGGCTGT 60

TGGACTTGGT GCCTATTTCG TCTATCTATT TGTTGCTTTT TTGAAAAAGA AAAAGCTCAT 120

AATTTGGACT TTTTTAGGAA TCAAAGAAAC CTTGCCAGTG GTCTTGTTAG GTTTGGTGAT 180

AATCTTTTGT GTGGGAATAT TTGGTGATTT CTTATTACGT CTGGAAGGAT TAGAGACGAC 240

CTCTAACCAA CAAAGTGTTG AAGAAATTGT CCAATCTATT CCTAAAGGGC TTATACTTAT 300

TTTTGCAGGT GTGATTGGGC CAATTTTCGA AGAGGGAATT TTTCGTGTCG GGATTCAAGC 360

GTTTTTCAAG CCAGCTTCAA AAATAGGCAT ACTCGTTTCT TCwAwTTTAt TTGCACTTGn 420

GCACACGCCG ACAGATATAG GkAgTTTtGT tATTnATgGG GCAnGGGnTT AGCTTaAGTT 480

GGGnCyAwCw ACGCACGGGC GCT 503

(2) INFORMATION FOR SEQ ID NO: 713:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 714 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 713:

AGTCAAAGAA TTACAAGCAT TGGCGGCAGC AAAAAATCTG CGCCTGATGG TTGGCTTCAA 60

TCGTCGTTTT GCCCCTTATA CGGAAGTATT GAAAAATATT CCTGAAAAAC AAACAATTTT 120

TATTwAAAAG rACCGCATCm ATaCAACGAG AGAAACAAGT TTTATGATGT ATGATTTATT 180

TTTACATGTC GTTGATACCG CAGTTTATTT AGCGGAAGGA CCGCTCCACG TGGTGCAGTC 240

AAAATTAGTG GAAGAAAACG GTCATTTAAA ACGAGCAATT TTACAATTAG AAACCGAACA 300

GACAACGATT GTTTGTTCCA TGGATTTACA CTCTGGGGCC AATACAGAAA CTTTTGAAGT 360

CACCAGCCCC ACAGGTACTT ATCGCTTGGA AAATTTAACG CATTTGACCA TCCAAACAGA 420

AGAAGAATAC CAAGTAAAAG AGATGGGAGA TTGGACGCCA ACATTGGAAA AAAGAGGTTT 480

TTACCAAATG GTTACTGCTT TCATTCAGGC CATTCAAAAA CCGCAAGAGC AAGAGTTAAA 540

GCAAGAAAAA GTCTATGAAA GTCATGCGTT ATGTGAAGAA ATGCTACGCC AACAACAACG 600

GCATGTTCTT TAAAGGGGAC TAGCAACTTT CTTTCTATCA TGCTAGGGAG GTGCGACAAG 660

TCCCTGnTAT GnGTCATGTT TGTCATCTGG AGCCATAGnA CAGGGTnCAT CATG 714

(2) INFORMATION FOR SEQ ID NO: 714:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1190 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 714:

TTAACTTTTC CCGGTnGCCC CAAAnAcCAg GCyTACACCa ACTAATACTG ATGCAACACC 60

CACACTTAAT TtTTTTATGC TATACTTTAA TACTCTTTTT CTcCCTTTTT TTAAAATCAT 120

ATGCTTATTG TTTTTTCCAA ACAAACTAAA TTTGCTCCCT TCCTTATTTA AATTTAAAAT 180

TACATTTTTA TTATAAATAA TACAAAATAG GAATTCTTGT AAATATCACA AATATTTGTA 240

AAAAAATCAA AAAAATAAAA TATTATAAAA AAAACCGACC TTTACGGGGT CACTGGAACA 300

TCCCTAGAAA TCAAGTAAAA AAAATCCAGA GGATTTTCTC CTCTGGATTs ACTATTACTC 360

TACATTTTTT AGTTATTTTA TTATTCAATT ATTCTCATTA AAAATTTTTC TTTCTATATT 420

ACTTAACACG AGATTCCGTT AATTCTTTTG CTTTAGAACT TCCTTTAGAA GCTAGcTCTT 480

TTGcTTTCTC TGTTTGCTCA CTGGcAAAAT TACTTACTGA TTCTGTTGCC TCACCAACCT 540

TATCTTGTAA ACTAACGGAA TCTTCTTCAT GCTGTTCTTC TGTCTTAATA TCAACAACAT 600

TTACATTTAC TTCAATAACT TCTAACCCAG TCATATTTTT TACTTCTTTT GTAATAATTT 660

CTTTCATTTC ATTGTAAATT TTATAAATAT CTTTTCCATA TTCTGCAACA ATATCTAAAT 720

CTACTGCTAC CTGTTTTTCT CCTACTTCAA CCACTACACC TGATGTAACA TCATCGGTAT 780

TCACTAACTT ACCAGCTAAA TTTGAGAAAA ATCCACCATC AACCGTAAGC AATCCTTCTA 840

CTTTTTCTAA AGAAATGCCA ATAATTTTTT GAATCACTTT ATCTTCATAC GTTAATTTTC 900

CTTGTACATT TTCCTTCTTT GTTTCTTCCT TTACACTTAC TTTTGTTTTC TCCATTTGTA 960

TCCACTCCTA AATATTCTTT ATTTTATTTA TTTTGGAAAA ATTTTCCAAA TAAATCCTCA 1020

TTTTTAATAT AACTCCCTAT AAATACACCA ATCCCAACCA TAACTAATAA GAGAATTGTT 1080

TTAAAAAAAC CAATAGTGAA AAAGAAAAGT GCTAATACCA ATCCTATTAG TCCTCCAATA 1140

ATTGGATACT TATATTCTAG GAACAATTCT TTCATATTCT CCCCTCTTTC 1190

(2) INFORMATION FOR SEQ ID NO: 715:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 491 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 715:

TTTAAAACAG ATACAGGATC GATTCTGATT AAACGTGTGA TTGGTTTACC AGGAGAAGCT 60

GTGCGTTACG AAAACGATCA ATTATATGTC AATAATCAGC CAATCGCTGA ACCGTATTTA 120

ACTAAAAACA GAAAAAAAGA TCATGAAACG ATGCCTTACA CTACGAATTT TGATTCAAAA 180

GAATTGTTAA TGCAAGAAAA ATTACCTAAA GATAGCTATT TTGTGCTTGG TGATAATCGC 240

CGTATGTCCA AAGACAGCCG TTCTTTTGGT GCAATACATG CAGATCAAAt CTTAGGGAAA 300

GCACAATTTG TTTATTACCC ACTCACTCAT ATGAAGATCA TTCCTAAATA AGTAAGACCA 360 mTAGTTCGTT TATTtAAATA ACGACTAGTG GtCTTTCtAA TCTGaAAAAT TGTyCTCTTA 420

GCTGnAAAAT GGTAAGGGTA AAGTAATGGT AGATGTAGGA GGnCATGCCA ATGATTAGAA 480

AAACAATTTT n 491

(2) INFORMATION FOR SEQ ID NO: 716:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 434 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 716:

TAGTACTAGT AATCGTTATT ATTGCTGCGG TGCTATACTT AATTGGTTAT TTTATGCGCA 60

AAAAAAATCA AGAACAATTA GATGAATTAG AAGTTCGTAA AGAAGCCCTG TTTGACTTGC 120

CAGTCTTTGA AGAAATCGAT GACATTAAAA AmATGCATCT AGTTGGACAG AGTCAAAATA 180

GTTTCCGTGA ATGGAATCAA CGTTGGGTAG AACTATCGAC GCGTTCTTTC GCCGAACTAG 240

AGAGTCAAAT TTATGAAGTC GAAAATCAAA ATGAAATTTT CCGTTTCATG AAAGCGAAGA 300

AAGCGGTCGT GGAAGCGAAT GAAACGATGA CCGAAaATGGA AGCCGAAGTG GAAGTCaTcG 360

TAATGGCTTG AaAGAATtAC GTGaAAGTGA AGAGCGCaAT TCATTAGAAG TTCAAAAAGC 420

TTnGGAGGnC ATGA 434

(2) INFORMATION FOR SEQ ID NO: 717:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1278 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 717:

AGTGCAGAGA TTGCATTATT TTTTATCGCG AAAAAAGAAC CACCTGGATG TGTGGAACTA 60

GAAGCGAATA AGGAAGGAGG AGAACAAGGA ATGCCTACAA TTAACCAATT AGTACGTAAA 120

CCTCGTAAAT CTAAAGTAGA AAAATCTGAT TCACCAGCCT TAAACAAAGG ATATAACAGT 180

TTTAAGAAAA CTCAAACAAA CGTTAACTCT CCGCAAAAAC GTGGAGTTTG TACACGTGTG 240

GGTACGATGA CACCTAAAAA ACCGAACTCA GCTTTACGTA AATATGCCCG TGTTCGTTTG 300

TCAAACTTAA TCGAAGTTAC AGCTTATATC CCAGGTATCG GTCACAACTT ACAAGAACAC 360

AGCGTGGTAT TATTACGTGG TGGACGTGTA AAAGACTTAC CAGGGGTACG TTACCATATC 420

GTTCGTGGTG CGCTTGATAC TGCCGGTGTT AACGACCGTA AACAAAGCCG CTCTAAATAC 480

GGTACAAAAC GTCCTAAAGC TTAATTATAA TAAGACAATA ACAACTTGAA AAATCATATT 540

TCGGAAGGAG GAGTTACGGA TGCCTCGTAA AGGTCCTGTT GCAAAACGTG ATGTTTTACC 600

AGATCCAATT TATAACTCAA AATTAGTAAC TCGCTTAATC AACCGTGTAA TGGTTGACGG 660

TAAACGCGGG ATTGCTGCTA ACATTATCTA TAATyCATTT GATATCATCA AAGAATCTAC 720

AGGTAACGAT CCATTGGAAG TTTTTGAACA AGCAATGaAA AACGTTATGC CTGTTCTTGA 780

AGTTAAAGCA CGTCGTGTTG GGGGTtCTAA CTACCAAGTA CCAGTTGAAG TTCGTCCAGA 840

ACGTCGTACA ACTTTAGGTT TACGTTGGGT AGTTAACTAC GCACGTCTAC GTGGTGAACA 900

CACAATGGAA GAACGTCTAG CGAAAGAAAT CATGGATGCT GCTAACAACA CTGGCGCTTC 960

TGTTAAAAAA CGCGAaGACA CACACAAAAT GGCAGACGCA AACCGCGCAT TTGCTCATTa 1020

TCGTTGGTAA GATCCTCTCT GTCTGTCGCG TAAAGTGGCA GCAGAAaGTC AATTTACmAT 1080

CGATTTaAAT AAAGAGAGGA GTAAACATAC AAAATGGCAA GAGAATTTTC ACTTGAAAAA 1140

ACTCGTAACA TTGGTATCAT GGCCCACGTT GATGCTGGTA AAACAACAAC AACTGAGCGT 1200

ATCTTGTACT ACACTGGTAA AATCCATAAA ATCGGTGAAC CCATGAAGGT GCTTCACAAA 1260

TGGACTGGAT GGACAAGA 1278

(2) INFORMATION FOR SEQ ID NO: 718:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1463 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 718:

CGTTnCCACC AGGAACAACC ATTAACnGAT GTTACnTCCT GGTGGTACAA TCGATACTAA 60

TACACCAGGT AATTACGAAG GTGTCATTGA AGTAACGTAT CCAGATGGTA CAAAAGATAC 120

AGTAAAAGTT CCaGTAGAAG TAACAGACAA TCGCTCTGAC GCTGATAAAT ATACACCTAA 180

AGGTCAAAAA GTAACTACTG AATTAAATAA AGAACCAGAA GCATCTGATG GAATTAAAAA 240

TAAATCGGAT CTACCAAAAG GAACTATGTA TTTTTGGAAA GAAAAAGTAG ATGTGGGTAT 300

ACCTGGAAAT AAAAAAGCTA CTGTTGTAGT AATATATCCA GATGGATCTA AAGAAGAGGT 360

AGAGGTTGTT ATTTCTGTAG TAGATAAAAA AGCGCCGAAT AAACCTCaAG TTGATCCTAT 420

TACAGATGGT GACAAAATTG TTACTGGTAA AACTGaACCm AATGCaGATG TaACaGTwAC 480

GTTACCTGAT GGAAGTCAAT ACCATGGCmC AGCTGATAAG AGTGGTTACT TTAAAGTTAA 540

CGTTCCTAAA TTAGAGGCAG GGACAAAAGT AAAAGTAACT TCAACTGATG AATCTGGCAA 600

TACTAGTGAA CCTACTGATG TAGTTGTTTC GTCAAATGAA TTAAACGGTG GAAAAGGTAA 660

TGGAACAGAC AGTAAGACTA ACAACAATCA AGATAAAAAA CAATTCTTGA AAACTTATCC 720

TAAAAcAGGA gAgGTTgAtA GTAATATTTA TACTATTGCT GGTGGTTTAA TATTGTTAGG 780

AACTTTAGGG TTATTAGGGT ATGAAAAATG GAAAAAAGAG GATGAATAGA AGTTTAAATT 840

ATTAAATATT CTCTGAAAAG AGGAAGAAGT CTCTTCCTCT TTTTGTTTTA tTGTTGTTTA 900

tTTAAATAAT AtTGTGGATT TTATTATtTT TTATTAAAAT AATGTTGATA TTAAAATGAG 960

TAAATAATTA CTAGATAATA TTTATTTACT TATAGTAAAA GTTCATATTT AAGTAAAGGA 1020

GTTTTTTTTT TGAATACGGA TAAAGAAGTT TTATTAAAGT CATTACATGT ATGCTTAGGG 1080

ATACCTATAT TAGTATTTAA TGAAGATTAC ACTCTTGTAG AAGAATATAG ATCTGATAGG 1140

ACTATATCAT TATTTTATGA TTTCCCAACT TTTTTAAAAA AAGTTGGGAA AGACAAGCTT 1200

AAATTTGATT ATATAACTGG TAATTATAAT GAGTTGTTTT TATTGTACGC ATACAATAAA 1260

AAACTGTTTT TGTTTGGTCC ATTCAGATGT AATACCATAG AAAAAGATAA ATTCTATTCT 1320

ATGGGATAAA GTCCGTATAA TTGTGTAAAA GTAAAArGGC CATATAACAG TCCTTTTACG 1380

GTACAATGTT TTTAACGACA AAAACATACC CAGGAGGACT TTaCmTGACC CAAGTACATT 1440

TTTACACGGG AAAGCGAnGA GGA 1463

(2) INFORMATION FOR SEQ ID NO: 719:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1064 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 719:

CTAGAGTATT GAACCCGCAA GATAAGCAGG TGGGTTCCTC GATCGCAGTA TTGAACTACC 60

AAAAGAAAAG GCATGTTACC AACTTTGATA CAAGCAGGAT CCCTAGATAT CAATAAAAAT 120

GTTCGGTCAA CACAGTGAAA GGACAACGCG TACATCACAG ATTTACCATC TTTATAATAG 180

CATACTCTCT CArAATCGGC AACGAGATTG TTTCTTTTTT TGAAAGCGTA TTATTTTGTT 240

ATTTTTTCTG TTTCCATTCT TGGATACGCT TGAGTTGCTG ATGTAAGGCT TGTTCATATT 300

TACCAATATC AATCGGTTCG TAATAACGTG CCTCTTTTAA TTTGTCAGGT AAGTATTGCT 360

GATTGACCCA ACCGTTTTCA AAATTATGTG GGTATTGATA ACCAATGCCA CGGTTTAATT 420

CTTTGGCTCC AGCATAGTGG GCATCACGTA AATGATCAGG AACATCTCCC ACAGAGCCTT 480

GTCGAATATC TGCCAAAGCA GCATCAATCG CTGTGACTGC AGAATTTGAT TTTGGCGATA 540

AACAkAAATC AtGACGACAC TAGCTAGAGG AaTCCTAGCT TCTGGAAACC CTAGCTTTTC 600

GGCGGCTTGA ACAGCGGTGA CCGTCCGAGC GGCGGCCGCA GGATTCCCAA GACCGATATC 660

CTCGTAAGCA ATGACCATTA AACGACGACA AATAATCGGT AAATCACCTG CTTCAACTAA 720

GCGACCCAAA TAATGAAGTG CAGCATCTAC ATCGCTTCCT CGAATCGACT TTTGGAAAGC 780

GGAAATCACG TCATAATGGG CATCGCCATC TTTATCATGC GTTAAGGCTT TTCGTTGCAC 840

GCACTCTTCA ATAATCGGTA AGGTGATCTT AATTTCGCCT TGTTCATTTT TGGGtGTTGA 900

TTTAACAGCT AATTCTAGTC CATTTArCGC ACTACGTAAG TCACCGTTCg TTGCGCGkGT 960

CAAATGTgTA ATGCTTTTTC TTCTAAGTGG ACAGGATAAT CGCCTAAGCC ACGTGTGCTA 1020

AACGGnTAAn GCTCTTGnAT CGCTTGCTGA ATGCTGTTCC GAAA 1064

(2) INFORMATION FOR SEQ ID NO: 720:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 684 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 720:

GCTATTAnAA TTATATTTTA ATTCnAAAAT GnATAAAAAA GTTAATTCAG AAATATrCAT 60

TGAAgCAAAA TAAAATAATT ATTTCTGAGA cTGCTACATC TATAGTTTTA AAAGAAAATA 120

ACAATATTAA TAAACACGCA TATATAAAAA AATACATTGA GTGTTTAGTT GAATTGTATA 180

AAGATGACAA CTATAAAAAT ATTTCAAATA AAATAAGATT TTATCATGAT TCAATAAATA 240

AATTGTATGA AGATAATTTT GAATATTGGA TGTCACATGA GCAATCAGAA TTAGAGAAAA 300

TTAAGAACAC ACCACCAGAA AAATTAGAAG GTGAAGATTA TGATTATGTG GCTGCAGAGA 360

AGAATATTGA GCATTATCAA AGAAAAATAA AAGATATAGA AGTTAGTTTA ACAAACTATC 420

ATAAAGCTAT TGATTTCTTT ACAACAGTTA TATCCTTATA TTTAAAAATA GArTGGGATA 480

AAGCCAAAGA AGGCCAGTAA CAGTATTGTA AaTTACTCAA TTTAATTAGG CATTTATAGG 540

ATAGTTGAAA GCAGTTAAGC TAATATATGG TACTGTAAGT TAAATTAAAG TTTGCTGAGA 600

TACTTTTTTT TCGATTTGTC ATACGCTATA CTTAAATTCA TAACGTTGAT TTTAATAACC 660

AAATTTATGA ACGGAGATGT CGTT 684

(2) INFORMATION FOR SEQ ID NO: 721:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1120 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 721:

TGTCTATTCT ACTCCCATTA TAGAACAGCA AATCAAGCTA AACAATATAT AAACACTAAC 60

CCATTGCTTA ACTTTATTTC CTGAAACAAC CTATTAAATT GTTACTACCG CTTCTAATAA 120

TAGCTTTACT TGTTGTATAT CTTCACTATT TTTTTCACTA GGTGTCTCAA TCATAAGATC 180

TAGTTTACCA ATCGCTTCTT CAATATTATC CGAACCAATA CAAACTTTTC GTTCAAAATC 240

TTCATGTATT TTACACGCAT AAGACCTTAC AAGTTGTTCT AAAAATCCTT CAAATAACTT 300

TACTACTGCT TCTTCATTTT CAAAAAATAT ATTATCCAAT AATAAAAATT GAACTAAATA 360

TTCACTCGTT TCTTCATATA AATTCAAATA CAATTGATGA TAATATGGAA AACTGCACCA 420

ACAACGATAG GCACTAAATT TACCATTTAA AAAACGTCTA TACTTTTGTC GTTTTAATTC 480

TAATtCTTCT GATAACATTC TTCTTTTAAT ACACCTCCTT CGTTTGAAAT TTCCTTCAAA 540

TTACCACTTC CTTTATGAAG AAAAACAATA CTGCTCTTCA ATCTTCTCTC AACTAGCCAA 600

TTTAATGGAT TCAATCTACG ACTTTCCAAT AATTCTTTTT CTTTTCTAAC TGGACGTTTC 660

TCTTGTTTCA TTCCAATCAC TCCCATTTTT TTATTAACTT TACCCACTCT TTTTGAACTC 720

ACTCAAAATA CTTTGAACTA AATCACTACT AATAAAACCA GCAAAATATA GTTGCATTAA 780

GTATCTCCGA GCATAGTCAT ACCTGCTCAT TGGGAAAAAT GGGTTCGTTT CAAAAGTAAC 840

AAATAATTCC CTTAATGAAT CAATCGCTTC ATTtGATAAG TtGAATTTTT TtGGATtCTT 900

AATTAATTCC TCTAATCGAT tCATCkTTtA ACTAACCACC TCATTTTTTA TTATTAATGT 960

CTATTAACAA ACCTTGTTGA CACAAATTAA TAATAAAAAG GGCACTCGTT AGAATGCCCT 1020

TTACAGCTGT TAAACGGCGC TACCGTTCGT CCGAATACAA TATTTATTTA CTTCTATTGT 1080

TTATTTTCTT GTTTTTCCTT TGGTTGAGGT GCCTGTGTTT 1120

(2) INFORMATION FOR SEQ ID NO: 722:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 539 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 722:

AAnGAnCCTC TACATACAGT AGGCCCATCT AACATGTCTG GTnCTTTTCC TAACATCTTA 60

GCTAGTAGCC TTAGGTnTCT TTCACACTTC TCTTTCCCTA AGTATTTCTT TGTAACCGTC 120

ATGGAAAAAA GTAATCTCGC TTTCAACGTT AAAAATTGGA TCTAATCCTT TGTTGTTCCT 180

CAACTCTTCA TGTATTAGAC CATAATACTT AATTTTATTA TTGTTTCTAA ATAATCGACT 240

TTGGCCATGT GTGATAACAT TATCTAATTG ATATATTGCA CAACTAACTC TTAATGAATC 300

TTTAAATTTA AACCTGTCTA ACATATCTAT TAAGAGCCTT AATTTGGTAA GCTCATCTGA 360

AGATACATAT TCATCTGCAT CAATAATTAA TATCCATTCT TTACTTGCTT TTTGTATTGA 420

ATAATTTCTC GCATCTGAAA AATCATCTTG CCAAGTATAA TCAAATATTT TTACTTTTTC 480

GCTGCnATTT CTTTTACTAT TTTTTTTGTC AAATCAATGG ATCCTGTATC AATAAATAG 539

(2) INFORMATION FOR SEQ ID NO: 723:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 633 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 723:

CAGTTGAAAC AATTGCTACT GCCAATTTAT GGCTACGAAC AGCAGATCGG ATTAAAATCG 60

TCGTTGGCGA ATTTAACGCC TATAGTTTTG ATGAATTATT TGAAAAAGTT AAAGCACTTC 120

CTTGGGAAGA TTATTTGCCT TTAGATGCAG AATTTCCAGT TGCAGGTAAA TCAATTAAGT 180

CAAAATTGTA TAGTGTTCCT GATTGCCAAG CCATCACCAA AAAAGCCATT GTTAACCGTT 240

TAAGAGAAGT TTACCATCGT CCAGCGACGG TGCCTTTAAC CGAAACAGGA GCTTTGTTTA 300

AATTAGAAGT GGCTCTTCTA AAAGATAAAG TGACGTTGAC GTTGGACACG ACTGGACCAA 360

GTTTATTTAA ACGAGGCTAT CGCATTGAAA AGGGGGGCGC ACCTCTTAAG GAAAaCATGG 420

CAGCTGCTTT AGTGAwGtTG ACCAATTGGC GCAAAGaTCG TCCTTTTtAT GATCCAGTTT 480

GTGGTTCTGG AACAATTTGT ATTGAAGCAG CATTAATTGG GCCATAACAT TGCCCCTGGC 540

TTTAATCGTA GTTTCACTTG TGAAACATGG GATTGGGTTG ATCCAGCAAT TTTTGAAAAA 600

GTACGTAACG AAGCCGAGGC AAAGGCAGAC TAC 633

(2) INFORMATION FOR SEQ ID NO: 724:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 672 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 724:

ATAGTTTTAC TCAGAAAATA AAAAGAAATA AGGTATGAAA TGATTGTTGT TATAGTTGGA 60

TATAATTTTG AGAAACCGTC TGAAGATTTT AAAAATGCAC TGCCTATTAC TTCACTGACT 120

ATkGCTAwTA mCAAaTATAA ATATGGCATA GAAATtGCtC CTTTATTAaG TGATATTTAT 180

ATTTTAcmyT TTAGGTGGCT AATTCaATAA TAAATAGACT TGTATTTTTC TAGTGATTTG 240

CGAGGTATTG ATTTTTGTGA TTTCAAAAAT TTTAGGTAGG TTCTGAAACG AGTTGAAAAC 300

GAGTTTCTTC TTGTCTTGAT AATATAGGTA ATTATTCAAT AATATAGGTA ATTATTCACG 360

GCGAGTTATA TCTACGCTAA ACCGTTGATA TTATTGGTTT AAAACACAAA AAAAGTTGCT 420

TTTTTCAGCA TACTATTTTA TGATTAAGGT GTCGAATCAA AATCAAAAAT AGAGAGGAGA 480

AAAAGCAACT GTTTTTTGCT ATAAAAAACT CTTAAGTTGT TTGGTTCTGT TGCAAAGTTT 540

TAAATCTACT ATCAAATAAG GTAGAATAAT AGAAAAAGAT AGCAGGAGGA ATGACGATGA 600

ATCATTTTAA AGGAAAGCAA TTTCAGCAGG ATGTGATTAT TGTAGCCGTG GGCTACTATC 660

TcGTTATACC TT 672

(2) INFORMATION FOR SEQ ID NO: 725:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 599 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 725:

GGTAAACCAG GATGTCCTGA GTTCGCTTTT TGTACGGCTT CAATACTTAA TGTCCGGATT 60

GTATTAACGC CCAATTGATC TGTCTTGTCA AACAAAAAAA TCATCTCCTA ATTTAGGTTT 120

TTTTTACCTT TATTCATAGT GTAGCTTATT TCGACAATGA AAGCAATTAC TTTCTGTTGT 180

GAAGTCCTTT TTCACGCTGA ATTTTTTTTA GTTTTTCAGG AGTCACATCG GTCCCTTCTT 240

GATCCACGAC CTTCATTCCT TCGATATGGT GACGCATCCC TCCACGAAAT GCTTCTAAAT 300

ACTCTTTACG TAATTCTTGT TGTTCGACTT TTTCTTTCGC ACTAAGTTCT cCACTTTTGC 360

TTaTTCGCTA ATTCATTtAT ACGGGCAATT TTTtCTTtCG ATAACATCAA AATCCCTCCT 420

TACTCTCACt GTCTAwATtA TAAGAAcTGC CTtAAAAAAA CAaCtATATt CTGATATTtA 480

TAAAaCaCGA aCGCTtGTTt GATTcGTcGC ATTTATGGTA CACTAnTTTT nTTAnGGAAA 540

GAGGTGCGGG TTTGGCGAAC GACAGAACAA GCAATAGAAG ACTAAATATA TATACGACA 599

(2) INFORMATION FOR SEQ ID NO: 726:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 694 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 726:

TTTATATCCT CAGGCACGAT CTAGATGTAG CTCTTAGTAA ATCTTTTTTA AAGAAAAATT 60

GTCGCTTTTC TATTTCAAAT TTTGAGTATC TATTTTTATC ACACCATTCT GAAAAATTAA 120

GATATCCAAA AGCGCCGTAT TGTTCTCCTA AAATnAAAGC AACCCATGTT TTCAACGCAA 180

CCTCAATTTT CTCTATTGCA TGTAACAAAT GAATTCTTAA ATTTTTATCT TGATAGTATC 240

TTTTCAAAAC ATCGGAAAAA TATATTCCTT TGTAATCATA TACAGTATTT TCATTAACAA 300

TTTTCTTATT AGCTAAAGGC GCAGCAAACT GCTTAATTCG ATAATATCCT ATAACTTTTA 360

ATTTTTCTAC ATCTTTTGAG GTCATCTGCA TACCTCTTTC TTCAAAAAGT GCCATTTGTT 420

CATCAAACGA TAATGATAGT GGTTCCTTCA TATTTTTCCT CCAAAATAAA AAAACCCCCA 480

TGTCAGTACG TGACTGTCCT AAAAAGGGAT GGGGGGATTG TCTTCTTTTT TAAAGAATAA 540

TATACTTTTA AAATATTGTC AATAGCTTTT ACTTATTTGA ATGGACCATA CAGGACTCGA 600

ACCTGTGACC GAACGGTTAT GAGCCGTTTG CTCTAACCAG CTGAGCTAAT GGTACAAGTA 660

AGAACCACCT TATTGAGGAT AAAATAGTTC TTAC 694

(2) INFORMATION FOR SEQ ID NO: 727:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 928 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 727:

TATCCAAGGA GTGATAGGAC ATTAATCGAT CCATAATTTT ATCTGGTGCT GGCCCCCGGT 60

CATAGATATC CCCAATAACA TGTAGGTGAT CGATAATCAG ACGTTGGATT AGATAAGCTA 120

ATTCTGTTAC GAAACGTTTG GCTTCACCTA GTAAAATAAT TTTTTCAATA ATTTGTTGAT 180

AGTAACCGTT TTTGGTTGTT GTTTCATCAT ATTGATAAAT CAATTCTTCG AGAATATACC 240

CATAGGTTTC AGGAAGGGCT TTGCGAACTT TAGAACGTGT GTACTTCATC GAAGAAGAAC 300

GGACGATTTC AACCAGACGG CGAATCGTTA ATAGCCACCA TTCATnAGaG AGGGCGCTTT 360

CATTTTCTAA AAGTGTTAAT TtTTCTTCTG GaTAGTAGAT AATAAAACmA AGCTCGTCCa 420

TCTGATGGGC ATTTAATyCT kGTTTAAATA AAAATkGAAC TTTTTCTCGG aTtCGCCCAG 480

AACCaTTTCt TAAAATATGG TCAAAAGCTt CAAaTTCACC ATGTAAATCA CTGACAAAAT 540

GTTCTGTTCC TTTCGGTAAA TTCAAAATTG CTTCCAAATT AATAATTTCT TCAATTTGTT 600

GATCGCGGTT CAGTGTAGCC AAGTCATCTG CCTTCTTTCT TTTTCGTATA TACCAATTTT 660

AATTTAAAAG ATTAAAGTCT TGATAATAAA GGTTTTTTAT TCGTATGCCA ATTATTGTAG 720

GCTAATTGGC GAAAAAAAGC nACCGGAGAA AGTCAGAGGA AAATGGGGGA AATTATAnGG 780

AGCnACCGGG GTTTAAGCCA AGAATTTTAA TTAAACCAAA TGGAAGGGCC GGGAAACCAT 840

GTAATTGGGA CAGCCAAAAT GGATCCGGGA AATCCATTAG GAAATCGGnG GTTGGGACAC 900

CGTAAAGTTG GCGCCAAAAA TCCGGGAC 928

(2) INFORMATION FOR SEQ ID NO: 728:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 420 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 728:

GCCTGGAGCA GAGCATGCGA GAACTTnATA ATAAAAACCC AGAAAAATAT AAGGAAACAG 60

TAGCTGCTTT TCTAGATAAA TATATAAAAT AACAAATTTA TTCTTTACTA ATAGTAGCAA 120

CAACAAAAGA CTTGTTCTTA AGTTAACGAC TTAGTGAACA AGTCTTTTGT TGTCAGTTGA 180

TTTTTTTTAA TAAGTTGGAA TTTACCCATG CAATCTAGGG CATAATCTGA TAGGATAGTA 240

CTAGAAATGA GGGAAGGATA ATGGAAGAAC AATTAGTCGA TTATTTACAC TATTTACCAA 300

TAGAGCGGGG cTTATCaATA ACACACGGCG GArTTACGAA CGAGAtTtwA AAAAAtACGT 360

rGCCCTTTTT ACCAGGACCA AGGCnGCATT CTTGGGCCGA ATTGGACCGT TACCnGGnTT 420

(2) INFORMATION FOR SEQ ID NO: 729:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 729:

GATGGGAGGC AAAAAAATGA ACATTGTCCT GAACAATCCA TTGGAAACAG AAGCAATTGC 60

TAAAATCATT GGACAAAAGG CACAAGCTGG TGATGTAATT GTCCTGACTG GTGATTTAGG 120

GGCCGGTAAG ACGACAATGA CGAAAGGGAT TGCTTTAGGA TTAGGCATTT CCCAAATGAT 180

TAAAAGTCCC ACCTATACGA TTATACGCGA ATATCCACAA GGACGACTTC CGTTATACCA 240

TATGGATGTC TACCGTGTTG AAGAGGGTGC GGATGAATTA GGGTTAGATG aGTATTTTGA 300

AgGAGATGGG CTATCCGTTG TAGAATGGGG CAGCTTGGAT TGAAGAAGAA TTGCCAGAA 359

(2) INFORMATION FOR SEQ ID NO: 730:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 830 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 730:

TCCACTTATT TTTTCATTtT CCATTGACAG AATAGTTGTa ATTGTcCAGA wAATTGTATA 60

TAATAAAGAA AGTTATTTTC ATATTATTTT TTTTCAATTA AAGGAGAATC CTATGAGTGA 120

AACACAACAA ACAACCTTAA AAAAACAATT AAGTCGCGCC ACATTACGAT GTTAGCCTTA 180

GGGGGAGCGA TTGGCGCAGG TCTTTTTAAA GGAAGTGGCG AAGCAATTGG TATTGCTGGT 240

CCTTCTGTTT TAATCGCCTT TCTAATTGGC GGTGCTGTTC TCTTCATCGT CATGAGTGGT 300

TTAGGAAAAC TTGTTTTAGA TGGCGGArAC ACACATCATG GTTTGTCTGG CTTGGTTCGA 360

CCATTTTTAG GTGCACATTC TGCTGATTTT ATTGATTGGG TTTATTATTC CATGTGGACA 420

ATTAATATTA TCGCAGAAGC GGTTGCCGCA GCTTCCTTTT TACAACTTTG GTTTCCTAAT 480

ATTCCTGCTT GGTTTTTTGT TTTTATTTTA rCTATTTTAA CGACGCTCAT TwaTCTTTAC 540

TCGtTCGTTT ATTTGCAGAA AcArAAtATT GGTTGGCGTT TGCTAAAATT AGTGTCATCA 600

TTTTGTTAAT TATTTTTGGT GTTTATCTGG TGGGACAACA AATGCTAGGC AGTGGTGTCT 660

TCCCTACTTT ACAAAGTATC ACCGATCAGG TGGCTTCGCC CCACAGGCAT GAAAGGTATA 720

GTTAATTCTT ATTAGTAGTT ATTTATCCTA GGCGGTTCAG AATAATCGCG ATTACTGTCA 780

GTGAAGCAGA TGATCCTAAA AAAGCCATCC TAAAGCGATC GTGGTGTGAT 830

(2) INFORMATION FOR SEQ ID NO: 731:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 861 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 731:

TATTAAAAAA AATAGTTGnA AGAGTACTTT sAAATTTGTA AAsAsTAGAA ATGATTTTTA 60

AyATTAGTAA GsTTAAAATA GTCATTACTA TAGTTGAATA AACTAATCTT GAATAATACT 120

TTCTGTTATA TTTTATTGGA TTAATATCAA ATTTTCTACT TAGAAATCTA TAAGCTAtGA 180

TGGTTGTAAA CATCGTACAT ATAATTGTAG CAATCGATAA TCCATACCCT TTAAATAGAT 240

ATATACATGG AATTTGAATA ATAAGTTTTA ATAAAAGAGT TATTGAAGTG ATTTGCATAG 300

CTTTCTTATG AAAACTTAGT GCTTGTAAAA TAGATAATAC TATTGTAAAA ATCCCTAATA 360

GAGATGTGCT AATAATTGTT GCTGATAAAT ATCTACTTCC ATCTAAATTA AATCCAAAGA 420

ATAATGTATT AAGAGGATTG CATAATATTA TCATTCCTAT CGAAGTAGGT AGTAAAAATG 480

TAAAGGCTAA TCGTAGTATT TGTGAGGTAC CTTGTAATAA CTCAGGTTTT GATTTTACAG 540

TACTTATTAA AGGTAGAGCT GAAATAGTAA TTGATCCAAT AATACCTAGT AGTACTGGAG 600

CTAATTTATT AGGATTTGCA GAGGCTAATG TATACAATGT TTGAAGTTCg CTTAAACTAC 660

TTTGAGGAAA AAACAAAGAA TACAAAACCT TTAAAGAAAA TTGATCAATT AATTGTAAAA 720

TAGAAATAGC CGATCCTACA TATACAAACG GCAATGATTC CTTGATTATA TCTATAGCAA 780

GACCTTTCAT GTGAGATATA GATTTTaAAG GTAAGAAGTA AAAATCTTTA AGTTGAAACA 840

AATTTCTTTT TTTTCCGATA A 861

(2) INFORMATION FOR SEQ ID NO: 732:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 662 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 732:

TACGACCTAC TAACAAGCGG ACGTCAATCG TAATTTTTTC CAACGGTCGC TGTTGTAGCT 60

CGGCTTTAAC TTCTTGAGAA ACTtGCCACT CTAATGGCGA CATTTCTAAT AAATCCAACC 120

GTCTTTCCTC GGGAATGTCA AAGCAATAAG TTATTCGTTC ATCTACTTCC ACAGCTAATT 180

CCTCGGCAAA ACGTTGGACT ACTTTTgATT GGAATAACTT TGTTTCTTTT CGTCATTGGG 240

ATAAAAAGCC GCTCGTAATT CCTTTAAGTA ATTCTCTTCA GGAATAATTT TAATCACAGT 300

CCCGTCAGCT TTTAGTACTC GCCGAAATTC TTGATAATGT GAAGGGGAAA AAATATTTAA 360

AATCGTATCA AGCCCTTCAT TCGCAAATGG CAAATTCGTT AAATCCGCCA CACACCAAAA 420

CGCATCAATT GGCTGATTAC TTGCTAAATA GATTCCTTCC TTTGATAAAT CAAACCCGAT 480

TTTCAGTCCA GATAAGCCTG CTTGACTTAA CTCAGCAAGA AAACTCCCTT CACCACAACC 540

AACATCGACT ACTGTTTTAT TTTGAGGAAG ATAGTGCATA ATTTTGTTTA ArACAGGCGC 600

GTACATGCCA CTTTGAATCA TTTTTCCTCG AGAAGTAAAC ATTTTnTnAT nATATnCTGT 660

CT 662

(2) INFORMATION FOR SEQ ID NO: 733:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1067 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 733:

AGAATCGATA TCTGTAATTT TCATTCTTAG CAGTACCTCA GGGTCAACCT CATATCCTGT 60

AAAATCTTGC TTCATAGAAG CCACTAGACT AGCTCCTAGA TAAAGTAATT CGAAAGCATT 120

AAATTCATCT GATGTAGACG GTCTTTTATA ATTATATTGT TGAAAAAGAG AGCCAATATA 180

AACTGTTCTC GCTGAACCAT ACATCATTAA TTCCTTAATC ATCTTATTTA AAACAGCTGG 240

CTTCAAATTT TCAATTTTTT CAGTATCCGC TATCATATCC ATCCAACGAG AGAAAAGGTC 300

TTCTATTTTT TTACCTGAAA TTGAACGATA ATAGTTATCA ACTTGCAATT CCCTAGTTGT 360

TTGATGTTCC TTGTCTTGCA TTATATTACT ATTTTCATTT CTTATCTCTT CTAGTTTTTT 420

TTCATGTAGC CTTTGTGGCA ATTCAGATAA GCCATTTTCG ATAGCCTTTT TTATCCCTTT 480

ATAGCCAGGA TAAATAAAAA TTAAAATAAG TAACTCAAAC CAATTATTTT TTAACAATTC 540

AAACATTTTT TCCTCTCCTT ATAATTAATG GTAAATTTCC TGTAACATTA ATTATAAAAA 600

ATAAACGAAA AATAAGCTAG TTTGTATTAT CTCTTATTTT TAATATATTG ATTTCCATTC 660

CAATTTACCA ATGATTTATT GGATGTTTTA ATATTAATTA CATTTTTTAA ATTATCTGCA 720

ACTTCTCTAT CAATAGTTGG ATATAAATGT GGGTACGTTT CCTAATGTTA TTTGTATATC 780

AGAGTGTCCA AGCCTTTTTT GAACTGCTAA AGGATTAGCA TTCAACTCAT TAATCAATAA 840

ACTTGCATGG CTATGTCTTA ATCCTTTTGA TAGTGACTAT TAATAAAATT CATCAATTAA 900

AAGGAAGCGA TTTCAACATG TCAAAAGAAA CTTTTCTTAA GTATGGTTTT GTTGTCCTCT 960

TCATAGGAAT TTTTTTACTT CGTGCCCAAA AATTTATTTC TGAGCCCCTT TCGTACCTTT 1020

TTTCGGGTAT TCTTTTGCTT GTTGTTGCGT ATTTAATACT CGTGAAA 1067

(2) INFORMATION FOR SEQ ID NO: 734:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 446 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 734:

ACTTAAAGGA GGGAAAATAT GTTCGCAGAA AGATTGAAAT TATTAAGAAA ATCAAAGCCA 60

AATTTGACTC AACAAGATAT GGCAAATATk TTAGGCGTTG CCAAAACCAC ATACGCATCG 120

TATGAGCAGG GAAAACGAAC TCCTGACGCA GAATTACAAA ATAAAATTGC TGATTACTTk 180

GAkGTATCGT TAGACTACCT ACATGGAAGA GACGTTATTA ATAACTCAAA TCTAGATAAA 240

AATATGAATA GAAACAACGA GCTTTCTCCT GAACAATTAA CAGTAGCCGC ACATATAGAT 300

GATGATGTTT CTGAAGAAGA AATGAAGGAA ATACTATCTT TCATTGATTA TATAAAGAAA 360

CGCGATCACT ATAAGTAGGT GTAtTTGATG rATAAATTGG GAAAGCyTGA TGGcyAAwTa 420

TCCkGAACtA AACTACCnGA TTAGGA 446

(2) INFORMATION FOR SEQ ID NO: 735:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 624 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 735:

AATTCTTCAG CGTATACATC ATCAACTGTT TGTGTTGGGT CGATAACAAT GTATAAGTTT 60

TTAGTTTTTT CGCTGACTTT TGTAGATTGA TAGATGTTAT CCAAGCCCTC ATCTGTTTTA 120

TCTTTGTAAG GGAAATAAAC TAAATCTGGT GTGCCATTTT TGTAAAGAAT TTCCATACGT 180

TCGATATCTT CATATTTTAA AGCACGAGAG AACATCCCTA ATTCTAAGCC ACCTAAGTTA 240

ATATCGTGGG TTTGAACATG ATCTCCTTCT TGGAAAATTT CAATTTTGAA CCCTTCACAT 300

GGATGAATTT CAACAAATTC ACTACCATGG ATACGGCCAA AACTTGTTGT AATATGTTTG 360

ATCCATAAAT CACCAATArA ACGACGATCA ATCGTCCaAG TTTCACCATT TCTGAAATAr 420

AATTTAAGTC CTGTCAkTTC TCTAGCCATA GCACAArTCT CCyTACTTAA AAATAATTTC 480

TTACATTATT AGTTTACCAT ACTTTTGATA AACGGTCACA AACTTTCGCT GTAGATTCnT 540

AAATCTTTTG TnATACTTTT TAAAAAAGGA ATCCAATCnC AACTACCTGC ACGTTGGTGG 600

AGACCCTTTG GCAATTnGCC GTCC 624

(2) INFORMATION FOR SEQ ID NO: 736:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 800 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 736:

GTCTTGAAAA AAGAGGAGTT ATTAAGCTTT ATAAAGCTGT AGGTAACAGT GTTGAAAAGT 60

GGTCAGAAGT AGGAAATCTT GATTTTTGTA GTAGTGGTTC AGAATACATG ATAGAGTGCC 120

CTAATCTTTT ATTTGTAAAT GATACACCAA TTCTGATTTA TTGTCCTCAG GGTCTTGATA 180

AATCAGAACT TAGTTATGAT AATATTTATC CGAATACTTA TAAAATTTGT CAAAGTTTTG 240

ATGAAAGAAC TGCGAAACTT GTGGGTTCTT CTAAAATTAA AAATCTTGAT TATGGGTTTG 300

ATGTTTATGC AACACAAGGA TTTAATTCTC CTGATGGACG TACTTTAATT GTTAGTTGGA 360

TTGGGCTTCC TGATGTTGAA TATCCAACAG ATAAGTATGA TTATCAAGGC GCAATGAGTC 420

TTGTTAAGGA GCTTTCTATC AATAATGGAA AACTTTATCA ATACCCTGTA GATGCTATTA 480

AAACACTGCG AACTAGTTGT GAAGATTTTA CGTCAAAAAT AGAAACAAGT AATACTTATG 540

AACTGGAAGT GACTTTTCCG GCTAATCAGA AATCAGAATT ACTTCTTTTT GCTGATGATA 600

AAGGGAATGG ACTCAGCTTA ATAGTTGrTA CTAAAgAAGG AAAAgTTATT TtAGATCGGA 660

GtAAAGCAGG aATTCaATAT GCaTTGaTTT GGtAwGaCtC GTGatGCtTA ATAGATCnGG 720

AGAACCAATG CTAATATnTT GGAGAAATCC ATCATGAGAT TTATAATAAG GGAGAAGTAT 780

TACnGCGGnA TTCCGAGAGG 800

(2) INFORMATION FOR SEQ ID NO: 737:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 500 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 737:

TGTTGTGACA ACTTCTATAT CTTATCATTA TCTCAGTTGT TTGTCAACAA CTTTTTAAAA 60

CTTTTTTGTT ATTTTATAAC AAGTTTTTTT GATCAAACGA TTGATTTAAA AGCAATCTCT 120

TTAACGACTT TGTTATTCTA ACATGTTATT TATCTTTTCG TCAAGAAAAA TTTTTAACTT 180

TTTAAAACTT TTTTCTTAAC GAACTGTTTG TCAGAACAGA TATTAATATA CCAAGGATTT 240

TTCTAAAACG CAACTATTAT TTTGCAGAAA AATAACATTT TTTTCAATTT CGAACGATGA 300

ATATTTTCTG cAAAGaAAAC ATTCGTTTCT TatCTTATAA TCCCTTTTCT CTTACTTAAA 360

TGTTCGTTAA TCATTCGTAA AAAAAGCAAC GTACTTTATT GTACGTTGCT CTCTTCTTTT 420

ATTCCAGCGA TTGGGAATAA AAATTCCAAA GATCGTTCCT GGATGCAATA CGCTTTCTGC 480

ATCAATACGT CCTTATAGTT 500

(2) INFORMATION FOR SEQ ID NO: 738:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 562 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 738:

AGTAGAATCC GTCTTGATTT TATGCATCAC TCGTCGCAAA CATTTGTCTA TTTCTTCTTG 60

ATAATCTTCA CCAATTAGAA CGAGATTGTA TTCTTTTTCT ACTTTAGTTT TAGCACGAAT 120

AATTGTTCCT ACTTGCTCAT GAGCTTCTTG AATAGTATTA GAGAAGGCmG CAATAAATTT 180

TAAGCTATCT TTGTTTAGGA AAGTGTTCTC TTGAACTTCT GAAAACTGaT CGAGCAATAA 240

TTGTTGTTGC AATCGTTCTT TTTCAGTATC TATTGTATCT ATTGTATCTG TATTTAGATT 300

ATTAAATACT GTACTGaTTT TCATACCCTC TGaTTTATCA ACGTTTTGGc AGGGTATGAT 360

TTtCATACCC TCTTGATTAC CAAGGGATtC AGCGTTTGTT TTTCTTTGGG GGGTGGGTAT 420

GATTtTCATA CGCTCTGATT TATCAACGTT TTCTTGCATA GTATCATCGA ATGTTTGCAA 480

TTGATAAATA TCGTTTtCtG TTACTtCAaG TTCCGCTAGA TAAAGTCTAT TAGGATTGTT 540

CTTACCTAGT GTTTTATTAA AA 562

(2) INFORMATION FOR SEQ ID NO: 739:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 437 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 739:

GCATTATTTT TTCGTTTTAA AATAGTATAA ATAATCCTTG TAATCAAAAG AAGGCTACCT 60

AAGAAAAGAA ACGCGTTTTC TTTTGTTTCT CCTGTTTTTG GCAATTCCTT ATTACGAATC 120

GCCTGTTTTT CAATGGTTAT TGAAGTAGTT TCTGTGGTTG AATTCGTGGT AGTATCTGTT 180

GTCCCACTCG TTGGTGGTGT GGTACTTTCT ATAGGTTTTT CTGGCGTATA GTGATTAGTA 240

AATACTTGAC CTCCATCATA CTGAGTAGTT GCTACAAGAT TTCCTAAAAG CGCCTCGACG 300

GTTACTTGAA CTTGAAGGTt GTGCGTAtCG TAGGAAAcAC CTGTCTGTTG CGTGkTTAcT 360

kCCCGAATGG TGTATTGATA GTTTCCAGGA GTTCAAAGGT TAATGGTGAA AATTGAATTT 420

TCCCTAAAGT ATCGTTA 437

(2) INFORMATION FOR SEQ ID NO: 740:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 503 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 740:

TCATTTAAGA AAAATGAGCG TGGGACAAAA ATCACTTTGG ATTTTTGCTC CACCCTCAAA 60

AACTGATAAA CGGCGGGAAC AGAACCAACT CCTTCGGAAA TAAGCCGAAA TTCTCCAAAA 120

ATTAAAGAGC AATTTTCGGA AATTTCTTCT TATTTCTCGG AGCTAAACGG TTCTGTCCCG 180

ACCTCATTAT TTCTTATTTT GCTGATTCTT AATGGTATAA GTCGAAACGA CCTTTAGCCA 240

TCATTTCTTT GGTACCACCT GTTTCAAAGA GCGCTTTGTT CATGATAATT TTTTTCATGA 300

ATGAAGCTGG GTATCCTTTA ATGTCTGTTT TACCAACGAT TCCTAAAGCA CGTGTATTTC 360

CGACAGAACA AACAGAACCT TGTGATTTAA AGCTGAAAGG TTTAGTTGCT TCCCCTTTCA 420

ATTGTGCTTG AATATTTTTA GCAGCATGAG CGCCCATCTT CAAGGCGATT TGTGCTGTTG 480

TtGGGTAgGT CGAwTGGGTT CnG 503

(2) INFORMATION FOR SEQ ID NO: 741:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 673 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 741:

TATTGACAAT TAAAAATAGC GGTGGTATAC TTTTTTTAAG TTATCAAAAG TATACTAAGG 60

AGGAGTTGTC AATGGGAAAA GTATATGTTA TTGGAAATTT TAAAGGtGGG GTTGGTAAAA 120

CTAAAACAGT AACCATGCTT GCATATGAAA GTGCACTACA ATTAGGACGA AAAACGTTAG 180

TAATTGATCT AGACCCTCAA GGAAACGCTA CTAGAGTATT AGCTAAGACG GGAGATTTAG 240

CCGAAATCAC TTATACTGTT ACTGAAGCTT TTCAAGAAGG TAGTCTTGAA CCAGCAATCA 300

CCAATATCTC AGAAAACCTT GATCTCATTC CAGCTAACAC TGCTTTTAGA AATCTCACCA 360

AAATTTTAAT GACAAAATTT CCGTCTAACG AATTTGACCA AATAAATTAC TTAAATACAT 420

TATTAAAACC CCTTAAAGAG AAATATGACG CAATTTATAT TGACGkTCCA CCAACTATTt 480

CAGATTTtAr TGrTAATGCm ATGCyTGCmG CAGATTACTG tAwTAwTGkT CyTCaAACmC 540

mAGAGTATCT TTAGACGGTG CTCAAACATA TATCGCCTAT ATGCAATACT TAGCTGACAC 600

ATACGATAAT GGATTTACAn GGCCCTAGGG AAnAATTCCC TGGTATGTTG CGGnCCCGGG 660

GGTAnGGGAA GCC 673

(2) INFORMATION FOR SEQ ID NO: 742:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1476 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 742:

TTGGTAGGCT GGTAAAGTTn GnGGTTTAAA CTCmCCaTTA CGAtCtCTAG GTATtTCCCA 60

AkGTTAATTC CmCCATATTC GGTTTTGaTT GaTCGaAAGT AAGAACCGTT TCTCGAATTA 120

CCTGAATTAA AACCAGTGCG ATCATATTTT TCGtAATCTA AAAAAGCCGT TAATTCAGTC 180

CGTAGGAGTG TGTTTATCGC TTTTTCTAAG TGCGAACGGA ATAATTCATT TAAATCGCCT 240

TTAGTGACTA GAGTTTGCAC AATTTCTGTA GTAAAATCAT TCATAGGGAA GTCCTCTTTT 300

CTGTGAATTG GTTGTCGTTA ACTTTATTCT ACAGAAGCGA CTTCCTTTTT TGTATGGATT 360

TTTTCATTTA CACAAAATAT TTTACACTCT CCAAAAGAAA TACCATTTAA AATATTTTTT 420

TCAAAAACAG ATATAGTATA GTAAACATTA TTAACCCTAT AAATAAACTC TTCTTTTCCA 480

ACATTTCCTT TTTTATTATT AAAATCATTT TTTTCTTCAT AAGTTAAAAC TTCAGATAAT 540

TTTTGAAAAT ATACATTTAA TAGTAGAAAG TCATTATAAG AGGAAACCAA ACTTAGAATT 600

GGCTTCATAA CCATTGTTAC TATAGAAACA AAACCAATTA GTGATCCTAA TGACAAAGAA 660

TTGTTGATTA TTAATTTAAT ACCTATTATT AAAAATAGAG CTGGCATTAC ACTTTGTATG 720

ATTTCTGGTA ATATTCCAAA AATAGCTATA TATCTATTTT TATTTTTTGT AATTAATAGT 780

TGAGACGTAA ACATGTTTTT CCAATTTAAC AAAAAACTCT TTTCTGCATT AGCAGATTTA 840

ATGGTTTCAA TkcCTTCAAT TGCTTCTGTA ATAATTCTCT GGACGTTTCC TTGTTCCATT 900

ATTTCTTTAT CTACAAATCT TTTTATTGTA TGCGAATTTA TAATACTTAA AAAAGCTATT 960

AAAGAGATTA GAACAAGAGC TATTATTGTC AGTAAAATAG AATAGTTAAC CATTAAAAAC 1020

AAATATATCC CTAAAAAAAG ACTATCTATT AAAGTTGTTA TTACCTTTTG AGACAATATT 1080

TGAGTCAAGT CCAGACTCCT GTGTAAAATG CTATACAATG TTTTTACCAT TTCTACTTAT 1140

CAAAATTGAT GTATTTTCTT GAAGAATAAA TCCATTCATC ATGTAGGTCC ATAAGAACGG 1200

CTCCAATTAA GCGATTGGCT GATGTTTGAT TGGGGAAGAT GCGAATAATC TTTTCTCTTC 1260

TGCGTACTTC TTGAkTCAGT CGTTCAATTA GATTGGTACT CTTTAGTCGA TTGTGGGAAT 1320

TTCCTTGTAC GGTATATTGA AAGGCGTCTT CGAATCCATC ATCCAATGAT GCGCAAGCkT 1380

TTGAATATTT TGGTTGATCG ATATAATCAT GAATCAATCG ATTTTTAGCC TCACGCGCTA 1440

AGTTAATATC TGTGAACTTA AAAATTCCTT TAACAG 1476

(2) INFORMATION FOR SEQ ID NO: 743:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1081 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 743:

CCTGAACTGG AATTGTTGAA GTAATATCTC ATGCTGGTAA AGTAACGTTT ATTGAACAaA 60

AAATTAAAGa AAAATTTTAA TATAGTCTGA CTAGACATAC TAGAAGACAT CTGACCAATT 120

GTGTTTTTAC ACATTGGTTA GGTGTCTTTT TGTATTTTTA TTAGAAGGAG CGATTCGCCG 180

ATGTTAGACC AACACGGAAG AAAGATTTTG ATTCAGGAAT ATAAGTCTGA TTTAAAAGAC 240

GCAAGTCGTC AGCATCGACG AATTGCAAAG AAGAAGTATC AAATTGAAAA AAATGGCAGA 300

TTAGAAACAA TTGATGACCG CACTGCAGAA GATATAAAGG AACAGTCCAT TTATGCTGAA 360

ATTATTTCAT CTACCAAATA TGCGTTATAT TGGCTTGAAC ATGGAATTGA GCGACCTCTT 420

GATGAGGAGG CAGCAAAAAA AATACCTAAA TATCGCAGAG ATAAACATAT TACAAATATG 480

GATAAGATAT CATACGAAAT TTATTGTAAC CAGTATGAAT CTGCACGTAA TTACCCTATT 540

ACTGAAGAGA AACAAGAGAT GTTGATTCAA CTTAAAGAGT TGTTATCAAC GTTCaGCGAG 600

AGAGAACGTG ATTTGTTTGA TTATATCCAT AATCAGCAAC TTACTTATGC AGAAGCTGCT 660

GAAAAAATGG ATATTAAAGT AGGGACTGCT AAATCCATGT CACAACGAAT TAGGAATAAA 720

ATTGATGCTT ATTTCGAATA TGGACATCAG ATTTCATTAT TTTAAATTTC ATTTTTTTGT 780

AAACCATTCC CACCTATAGG TGAGAGGTAA AATTCTCCTA TTCTAAGTTG GTAGAGTAGC 840

TTAATAATAT TTACTTGTAA AArAATACAA GAGAGGAGGT GTTCCTCCTC CTCTAAATTT 900

CTACAAATTA CGAGTAAATT AGGTAGACGT GTAGCTCAAT AGGTAGAGCA ATTGATTTTT 960

AATCAATGGT GCGTGATTGC TTGTGCCAGG TTCGACTCCT GTCACGTCAA TAAGTGGGAA 1020

ACCACTTAAA TAAAAAATCG GTATATGTCA ATAAATGTTT CTACTACTGT GACACACGAT 1080

T 1081

(2) INFORMATION FOR SEQ ID NO: 744:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 486 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 744:

TTCGAGGAGA AAGTATGATT AACGCAGGGT TGTCCTGGAT AAAACACAGA TGTATCAATC 60

CCAGGATTTA ACGCTAAAAT TTGTTCTAAT GTCATTCCGT TACGTTCGGC AACTTGACGT 120

GCCCCTTCGC CACTTTGAAC GGTATCATAA ACAGGTTGTT GGTTCTGCGC TTGTTGCTGT 180

TGTTGATTTT GAGCTTCTTG TTGTTGTTGC TGCTCTTGTT GTTGCTGTTG CTGTTGTTGT 240

GCTTGTTGGT CCTGTTGTTG CTGTGCAGCA GCATCTGAAC TTGAAGGTGT TGTATTTTCA 300

GCTGGTTGGC TGCTTGCTGG TTCTGAGCTT TCCACTGATT TTGACGTACT TTCTTTTTTC 360

TTCTCTTTTG ATGATGATTG GACTAATGAA GAAGATGGCT GGCTATTTTG CTAGCACTTT 420

CGTTCGGTTC CTGGCCTGCA TCATCCATAA GAAAGTCCCA ATTGGAAnGC nATAATTAAT 480

AATAGC 486

(2) INFORMATION FOR SEQ ID NO: 745:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 527 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 745:

AGTGAATTTA CCnAATAAAT TAaCAGTTTT AAGAATATTT ATGATTCCGA TTTTTATTAT 60

TATTGTGTCG GTGCCAATGG ACTGGGGAAC GATTAGTTTT GGCGACACAA CGTTAGCCGT 120

TACCCAATTA GTGGGCGCGA TTATTTTTGC AGTTGCTAGT TTTACAGACT GGTTGGATGG 180

CAAAATTGCC CGTGCACAAG GCTTsGTGAC GAACTTCGGA AAATTTGCGG ACCCaTTAGC 240

GGACAAAATG CTAGTGATGA CCgCTTTwAT CGTCyTAGTk GGTCAAGGca AAGTACCtGC 300

tTGGATTGTA GCaATwATTG TTTGTCGTGa ATTGGCaGTA ACTGGkTTAC gGTTGTTAwT 360

AgyCGAmCAT GGCGAAkTAA TGGCAGCAGC TTGGCCAGGA AAAGTAAAAm CAGCyACCCm 420

AATGGTgCaA TATTtACyGk TCaTTAmCAA CATyCCATTT yCAGaCTcyA CCACTAGATC 480

AAATCATGCC ATATGCCTGC TTGAATTTTC ACCATTTACT CTGGGAG 527

(2) INFORMATION FOR SEQ ID NO: 746:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 772 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 746:

ACGTATCCAG ATAAAAACCG GTTGAATCAA ACGTCATnTA ATGGTATTnT AGACGAGTTG 60

TTATCCTTTT ATGAACAAGC AGGGGAAAAA GTGGATCAAT TTGGCCGAGT TTTTGAAGCG 120

GCCTTTGCTA CAACCGAAAA TGGCCATTTG TTATCAGGTC CCATGAAACG ATTATTGGGA 180

TTGTTGTTAC AAGTTTTATG GAGCCAGCAA GTCAATCACT TTGACTTTCA ATTTAAAAAT 240

TTTATCATTT GGTTTATTCG CTTAGGTTTT CCGGTGGCGT TTCCAAAAGA AATTTTAAGT 300

GCTGATTACG CAGAACAAGT TTTGCTTCAC ACAGAAACAC TAGGTCCTCC TATGGTGTTG 360

GAAAAAATCG GTCAACTAGG AGCAGCTTTA AAGCCAACGC CCGATCAACT GTTAACAACG 420

ATTGAAAGAT ACCAAGAAAT TTTAAAAGAG ATTTGAGGGG AAATTATGTC ATATAGTGaA 480

AAAATGGTTC AAGCATTACA AGCTGaAAAT TTaGCaGAAG CACAGTTAAT GTTTGaAGAA 540

GCaTTAAAGa AAGATGaTGa AAATACCTTA GCTGATTTAG GGGAAACGTT ATTATCTTTA 600

GGTTTTCTAG AAGAAGCCAA ACAAATTTTC CAACAGTTGT TGGAACAATT TCCCGATGCA 660

GATGGCTTAA ACATTCCATT AGCAGAAATC GCCATTGAAA ATAATGAAAT AGATGACGCA 720

TTCATCTATT TAGAAAAAAT TCCAGAAACT AGTGACAGTT ATGTCCAAAG TC 772

(2) INFORMATION FOR SEQ ID NO: 747:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1112 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 747:

GCCCTTTTAA AATTTGTAAC CATAATTCTG TTACATCTTT ATTCTCTTGG TTCGTTAAAG 60

TATACTTATT TTTATTAAAT ACCTCTACAT TAAaCCAGtC ATtAGTAATA AAaTTTyCrG 120

AaTaCATTAt CTCTtCACCT CATTACTTtC TATTAGAGAT AGTTAtGCAa TTaATTTtGk 180

GTAaCACTCT GGGTTATCTG TTAAGCATAA TATGTAATCt ACGCTTGTAG AATACAAAgT 240

GATAGCTTAA CTAAAATATC TACTGGTACC ATTGTTGCAC CACTCTCATA TCTGGAATAC 300

GCAGATTGAG AACAATGTAT ATAGTTACTG ACAAAAGATT GAGTATAATT ATTTTTAATT 360

CTCAAAAATA GCAACCTAGA TTCCATAATT TTTCCCTCTC CTGTATGCTA TCTTTGTTTG 420

TTAATAACAG TCAATATATT ATTCGGATTC TGCATAATTC TAAAATTTTT ATTATTTATC 480

AATCACGAGA AAATAAAATA TGTCTATACC TTTATAAAAG GTAAAAATTT GTATTTCTTT 540

ATTATTTGTT TTATAATATA AGTATCCATT GAGCAAAAAC GCTCAAAATA GGCTGATCGA 600

GTTAATCGAT CAGCCTATTT TctTTAGATT TATATTATTT ACTAAAACTA TACTGTCACG 660

AAAAACAACA TATATAATAT TATGTTGTTT TATTCAAACT ATAAATGATA aTATAyyTAA 720

ATATATAATA GAGACTGTCT CTTGGGGGAG ACAGTCTCTA AGGTTATAAC TTATTTTAGC 780

CTTTCATTTC TAACAAATAA TCATTCATAG AACTGAACGA TCAATTCATT CAGTCTGAAA 840

GAAAATAGTT AAATCTGAAA TAGCTATTTT TATTTTACAG TAtCTGCCTT GGCCTTTCAA 900

TTTCCTTGAT ATTACTTTTA TTTCAGGATT ATTTAACATA AACAAATACT TCTGAATCTT 960

TTACATAAAT GTCATGGTAA TAATTAGATT TCATTAATTC ACTTTCTTCT TTACTCAACG 1020

CGTTAATATC AAAGATCTCC ATGTTCACGT TTAAACCAGT GTACGTTCTA AATAAAAACT 1080

GATTTGGCCA ATAAAGGGCT TCATTTGGTG GA 1112

(2) INFORMATION FOR SEQ ID NO: 748:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 864 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 748:

GGGCTTCCAT TTCnGGnTAT GCnGATGGCG CATCAAAACT TGCACGTAGG TAGTCTCCCG 60

AAAGAGGaGT AAtGcTAAAA AAaTATGGtC GCmCTTtCGG AAAaGCTTCa CTTGGTACAT 120

TTkGATTGGC GTGATgTTcA TTATGTAGgC TTTAGCAATT TATTTTTTTC AACAGTTGTT 180

AGATCATTAT CAAAAAgTTT CCaACTAGGG TTGTTAGTAG CGTATGGAAC GACAATTATT 240

TTAATTCCTC TATTGTCTTA CTGTGAACAG AAACCGAAAG CTTaCTTaAC TAATGGCATC 300

TATTTCTATT TGAAGAAATT AAGTTTAATA AAAATGAGCA AGATTTCCTA TGAAGAATAT 360

CTAAAACTGG GAGCGGGGGC TTTACTACAA AAAGTAGAAG TTGGTGCAGC AGCTGGGAGA 420

AATATCCACT TGAATTTTTA TGGGCGCTTA TTTCGAGAGT TAATTCCAGA GACGTTATTT 480

AATTTATTTT TCATAGCATT GATTGATAAA AAATTATTAC CAGCAATTCT AATTGGTTAT 540

GTGATTGTAT TTATTTTAAC CAAAATACTT TTGAAAACGC TACAAAAAAT GAAAGAAAAA 600

ACACTTATTT CTGAAGAAGC AATGAATGCT ACTTTGATAC GAGGGATGAC CGAACTAGTA 660

ACTTTTAGAA TTAATCGAAA ATATAAAAnA GAGATTGAnA ATTACGCGTT AATGGCAGAG 720

GGAAAATAGT CAAAATATAA CAAAAATGGA CTATGGATCC ATGGAATTTT TCTTTGGGCT 780

TTTTTGGCCT TGTTAGTTGG CACTTATCCA AGTAAGTATT GTCCGTACTT AGGTTTTACC 840

TAACGnCGnG GCGCCCAGTT TAGG 864

(2) INFORMATION FOR SEQ ID NO: 749:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1376 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 749:

ACCAACCTTC AAGATAAGTT TTAAGTCTGT GCCCTTACAC GAGATTTTTT TACGGCAAAA 60

ATAATCCTTT GGTAGTTCAT CCAGCACAAG CACATTTTTA TATAACTGAT TAATTTGGTG 120

GAATTATAGT TATATCTATA TTGATTAnTA GCTGATCTTG CTAAGCATGG ATTTAATAAG 180

AATATTTTTG TTAAAAAATC ATATAnCCTT ACGTGTAGTG CTGCAGCACT CATTTCAAAT 240

TCTTCGCAAA TTTGACGGAA AGACATTTGC TTTTTTAAAC ATTCTTTCAA TGCTTCATTG 300

TTTATTAAAG CAAGGGATGC AAATATATTA GCTCTTATTT CTTTTGGTTC ATCTtCGCTA 360

GTGtATCCTT TTCCAGTAAT TAAATCTGAA AAACTTTGAG ACTTGCATTC CTTATTCATA 420

TCAAAAAAAT AATGTGcAGA TTCATGAAGT ATTGAGAAAT TTTGCcTTCC TTTAACCATT 480

GtTGGATTAT AGGaAATTCC GCATTTATCG TGAAATTTTA TTATGTTTCC TGaAAATATT 540

TTTcTkGAAA CATCACCGAA ACTATAGGGA ATTATGTCAA CATTTTCTAC GTCAATCATA 600

TAATCTTTGT AGTGTTTCCA TCTTAACTCT TTTAGTTCAA TTCCATAGTA ATTAGCAACT 660

TTTTCAATAA TAATATATGT CTGGTCATGA TATTCAAAAT AAGTATCGTA GTCTAAATAC 720

ATAGGGGTCA TCCAATCATT TATTTTTAAG TCTTGATTTC ATGAATTCTA AGTACTCTTT 780

TAATTCTTCT CTGAGCTGAT CTTTTTCGGA TTCGGTkAAa TCTtCTGTAT TTACACGGAA 840

GAAAGTTGTA ATTTCATCTy CTTCTAAGTT ATTAGAATTT GGATTATCAG TTCGTCCAAG 900

TAAATAAkCT ACAGATATAT TGAAGTAGTC CGCAACTTTT TTTAATTTAT CAGAACCAGG 960

AGTATTCTTA CGCCATCTAT TAATTACCCC GTTtCCAaAT CCCAATGTTT CTTCTAATTG 1020

TCTTATTGAC ATTTTCTTTT CGCTAGCAnG TAATTTTATT TTnTTCGTAT AAATCCACTT 1080

ATATCAACCT TTCTAGAAAT ACAAAGAAAA TATGTATACA AAAAATCTAC AAAATGGTTG 1140

ACAATGTAGA AnAAATGTAT ATAATAATTC TTGTAAGTTA ATTGGATAGA AAAAAAGCAA 1200

ACTAAAAACA TACCTTACAG CATTAAGTTT GGCGACCGAG TGCATAATAA AGGTTTTGTT 1260

ATAGGCTTAT TTAACTATGT TCTGATTGTA GAATATTTGT ATACATAAGT CAATAGTTTT 1320

ATAAATTCTA TCTAATTAAT TTACAAAAAA TACTAAAGAA AGGAGAAACT AATATG 1376

(2) INFORMATION FOR SEQ ID NO: 750:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 669 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 750:

TTTTGTAAAT GTTGTCATTC CATTTTATTT GCAGGATGCA CGCAAACTAA GTGCTAGTTA 60

TGCTGGTCTA TTAATGATGG TATTTCCGTT ATTAATGGTG GTTGGGGCTC CTCTGAGTGG 120

CTATTTGACG GATAAAATTG GCCCAGGTAT TTTAACATTT GGCGGATTGT TGCTCTTGTG 180

CTGTACGTCG TTAATGTATA TGTTTTTAGA TATGAATTCG CCTATCTGGT ATTATGTGAT 240

TGCAACGGCC ATTATGGGCT TGGGAAATGC ACTTTTCCAG TCTCCAAACA ATACAATGGT 300

TATGAGCAGT GTTGAAAAGC AAGATTTAGG TGTAGCAGGA AgTATGaATt CTTTTGCTAg 360

AAAnCTtAGG aAkGGyCATk GGGATTGCGT tGTCaACGAC CATTTTAtAt CGTGGCATGA 420

GCGAACCTAA TGGTGAACgA GTAACCaCGT AtCTGGctAA tCgCCCaGAt ATAtTTATTG 480

TGGGAATGCG TGAAcCTTTT TTGTCGCCTT TCTATTATGT GTGGCGGCCT TTATATTAAC 540

GATTTTACGT TTTCGGAAAA CAACCAAATA AGTAAAAAGA AACAGACGAG AAAGATTCAT 600

GATTTTTCTC GTCTGTTTCT TTACTAGCTA AGAAaATAAG CaATGCGATT GTATATTTCT 660

GATCCCnGn 669

(2) INFORMATION FOR SEQ ID NO: 751:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 500 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 751:

AAAATAATTA AAAACAGGAG TGATAAACGT GAACGTATTA GAATTCGGCG GTACTCGCTA 60

GCATCAGCAC CACAGTTACA GAAAGTCTTA CAAATCGTCA AGGAAGAGCC GCGCCGAACG 120

TTTGTCGTTG TATCCGCTCC TGGAAAGCGA ACCCCGCAAG ATATTAAAGT GACTGATTTA 180

TTGATTCAAT ACTATCAGCG CTACTTAAAT AATGAGGAAA TTGAATCGAC AATCTCTGCC 240

ATCATTCGTC GTTATGAAGA TCTTTTTGAT GAATTACACT TGGATAAAGC AGTTTTAGCT 300

GATATTGCTA TTTCGATTCG CCAATTAGCC ACTTTACCAA AGAAAACAAT GCTTTTTTAT 360

TTGACTATTT TTTAGCTAGC GGTGAGGACA ATAATGCCAA GTTGTTGCCA GTTTTTCAAG 420

CAAATGGCTT AGATGCTCGC TACATTAGTC CAAAGAATAG GnCTCCTAGT GAnCCTGAGC 480

CAGCATGCCG CATTTTACCA 500

(2) INFORMATION FOR SEQ ID NO: 752:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 346 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 752:

TGAGTTTTAC AACGAGTTGC TTAATGAAGA AATAACTAGT GCTCAAAAAA GAAAAATGAG 60

CAATGATATT AAATTCTTAG CGAGTATTTT AACTAATATT TCTTGCTACA GATACTCAAT 120

TACTTTAAAT GATGTTTTTT TTAGAGGTAA GAATAAACTT GGCTATTtGA TAGGGGAATT 180

TGGTGTtAAA ATAGTAAAAT CATATtCGAA CTTTATATTC TCATGTTTAT CAGGAGGCGA 240

GTCTATATAT AATGTGGTTT TAACATCCAA AAAAGGAAGT AAGTTAAAAG TGaTAGAtTT 300

TAATTCTGAT GaATATGAGA TGTCTTtATTG GTGTTAGACC nGATAC 346

(2) INFORMATION FOR SEQ ID NO: 753:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 685 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 753:

GTCGATTGGT TAGTACTTAC CaAAGGAATG AATCGATAAA AAAAATAGCA GTGATGGCGA 60

TAACCGCTTT GACTTCTGCG GTCGGTTTGA ATTTCTTTyT GATTCCGGCA AAAGTTTTTT 120

CGGCAGGGAT GAACGGAATT GCGCAAATCA TTGCTACATT ATTATATACA AATTTAGGAA 180

TCCATATTAA TACAGGGATT TTCATTTTAT TTTTAAACAT TCCCGTCTTC ATTCTAGGGT 240

TTGTAAAATT AGGGAAACAA TCAACGATTc TTAGTTTCAT CAATGTTATT GGTATTTCCG 300

TTGTCACTAT GTTTGTGCCA ATCGTAACAG TCACAACGAA TCCTTkGATG AACGCAATaT 360

GGGTGGTGTC TaGTCGGCGT TGGCGCAGGc TTcATgAAAA TGGGCTTtAA CACTGGCGGC 420

ATGGACATTA TTTCACTGAT TCTtCAAAAA CAACAGGGaA AACAGTTGGT AATTTtATGT 480

TCTTATTAAA TGGtAwTATT GTGCTGTTGG CTGGTTTTGT ATTTAACTGG GAAAGTGCGT 540

TGTATACTAT TATTTCTATn TACTGTTTGA GTCAGGTGGT CGACATGATT CATACCAGTC 600

ACCAAAAAGT CACAGCAATG ATTGTCACCA CGAGACCAGA AGCCGTTGCC TTAGAAGTTC 660

TCAACAAATC GCTCGTGGCA TGAnC 685

(2) INFORMATION FOR SEQ ID NO: 754:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 682 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 754:

GGTGAATGTT TTACCGTCGC TAAGATTCTC TTCCAGATTA TAAACAAGCT GTGGCAACTC 60

TGCGCCGGTC GTTTGTTCTA CATCAAAGAC AGGCGTTAAA AAATAGCGCG TCTCTTTGAC 120

AATTTCTCCA TTTTCATCCG TGACTTTTTT CCCATTCTTA TCGACTTTAT CTTTAAAATA 180

AGGGGCATAC ACATACAGTG CTTTGCTGCC TTTTTTTACT TGACGATCTA ACTTTTTCCA 240

CGCATTGTAA CCAGCTACAC GACGTATGTT TGGATTTTGT GCCAACAACA AACGAATATT 300

TTTTGACGAA TATTTATGAA ATTTTGAAGC AAAATCTAAA TAGTTTTTAA ACGTATCACT 360

TGTTAAATAC TCTTTGaTCC CGTCGTTCAA AkGCTGCGAT AATTTTTGaT AAtCTTTCTT 420

TTTAATTATC TCTTCCAAGG kGATTtCkGG TtGGCyTTTT tGCTTCATTt GTGTTTTTTC 480

AGACATTyCT TwATtCCTTt CTCTGGATCG TTTTTTTGCT TCTTGACGCT TTAATTTTGC 540

TTCTtTTyTC TTTGTTTCAT TCTTTGTTTT CTTTAAAATA TCGTTCCAAT CTGTkTTACT 600

TTTTCCTTCA ATTAGTTTAG GkTGGTCAAG AGTAATnGGG ACACCAGTGA ATTGGAAAAT 660

GGTGGGAATA AATTTCCTTC CC 682

(2) INFORMATION FOR SEQ ID NO: 755:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 345 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 755:

TTGAGGGCTT GTCCCTCTAC TCATCGGAGA TTAGAGAGGT CAAAAATTGA CGAGGTAGAG 60

AAAATATTTG AaAATATTTT TTAATAGGCA TGAGTAACGG CAACATCCGT CAGATTGGAT 120

GTTGCCGTTG cTCTTaCCTA CTAAGGAAAT ATATACGAAT CTGCTTAAAT GCTCTTTGAA 180

GCAGGCGAGT TACCGCCGAC GGGTCACATC CGAGGaTAAC TGCAATCTCC TTGTATTGaT 240

ACCCTTGCCA GAAGTGaAGT TCaATCACTC GcCGCTGCCG GTCtGTCAGA CGGCTTAgAG 300

CAGCCgATaA ATGCTCATTC TGAACAGAGT CCATAAAATC CTGCT 345

(2) INFORMATION FOR SEQ ID NO: 756:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 608 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 756:

ACGTTTTATA TATCTTTTTC ATTAGTCAAA AAGCGCTTTT TGACTAACGA TGCTACTCAC 60

TATAATACGT TTTAAATCGT TTAsCAAGTC TTTTTATTAG AGAATGTTCA TCTTAGATAA 120

GAAACCCACC TATTTTGAAA AAAGGAAAAT CAAAAATAAG TGGGGCAGCA GTCGAAACGA 180

CTGCTGCCCa CTtGcTaACT ATTAAGATAA TTCCCTGAAc AAAAGTAGCT CCTTTTATTC 240

CTGTTrAATT TAAATCACTT CTTCTTCCTC TTCGACTTCA GAGAAATGGA ACTTAATCAT 300

TTCAyTGACG AAAATTGGcA CAAAGGATAG GAAAATTGCT AACATCCAAT GATTTAAACT 360

AATATGCACT AATCCAAATA GTTCTTGTGT GAAtGGTAAT AAAGcAAkGG TtGTGGTAAT 420

TAGTAATGGC TAAAACGACC ATTTCAAAGA GCGACTTGTT GCTGATAATT TTACACGGAA 480

AACAGAGTTC GCACTACGAA CATTAAAGAC ATGTAAGATA GAGGAATAAG CTAAGATTAA 540

AAATGCCACT GTTTGTCCAA CTTCATAnCG TTTGAnAnAT AGGCTGAAAC ACCGGTACAA 600

GGCACCTA 608

(2) INFORMATION FOR SEQ ID NO: 757:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1672 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 757:

GATTCTCCCT CATAATCTAA CTCAGCTTGT CAACTTTAGC TCTAGnATTT TnCCTTTTAA 60

CATGGAACTn TCATTATTAT AGCCCCCTAT TGTAATTCAA ATGTAAAATT ATCTATTAAT 120

CTTGTTTTTC CTATAAATAC TGCTATTGCA ACTAAAACTG ATTTTTCAAT ATAATTTACT 180

CTTTGTAATG ATAAAGAATC AACAACTTCT ACATAATCAA TCTTTGCCAA ATTACAAGTA 240

TTTATATTTT TACTTATTAT CTCmATTATT TTTrAGcTAT CTCTCTCTCC aTTATTTAAG 300

GCTTCCTTrG CTAAAGTTAA rGATTTATTT AAAAtAGTAG CTGAAGATCT TTCCTCTAAA 360

CTTAAGTATG TGTTTCTAGA ACTTTTAGCT AGTCCATCCT CTTCCCTTAT AATTGGACAC 420

CCTACAACGT CTATATCAAT ATTAAGGTCC CTAACCATTC TTTtAATTAC AGCTAACTGT 480

TGTGCATCCT TTTCTCCAAA ATAAGCTCTA TCTGCTGGTA TTATGTTAAA TAAyTtACTT 540

ACAACAGTGC AAACCCCTCT AAAATGTATT GGCCTACTAG CTCCACATAA ACCTTCTGTT 600

AATCCAGTTA TATTAACAAA AGTACTTGCA TTGTCACAAT ACATCTCTTC AACTTCAGGA 660

TTAAATATTG CTGTTGCTCT ACCTCTTGTG CATACTTCTA TATCTCTTTC TAAGTCTCTT 720

GGATAACTAC CTAAATCTTC ATTTTTCCCA AACTGAGTTG GaTTTACAAA GATACTtACA 780

ACAACTTTAT CATTnTCTTT TGATGCTTTT CTAATTAAAC TTTCATGACC TTCATGTAAA 840

TATCCCATAG TAGGAACTAA TCCTACCTTT AAGCCTTGTT TTTTCCACTC TTTTACTTGA 900

CTTCTAACAT CATTTACTGT TTTAAATATA ACCATATTTC TTGACCCTTT TAATATAATT 960

TTTCTAAAAC ATCATCACTT ATTTTAAACC CATGTTCTGG ACCTGGAAAA ACTCCATCTT 1020

TTACTnCCTC TATATATTTT GTAAATGCTT TATTCATCTC TTCTGATAAA TTAGCATATT 1080

TCTTCACAAA TTTAGGGGTA AAATCACTAT ACATCCCCAA CATATCTTGA TATACAAGAA 1140

TTTGTCCATC ACACCCTGCA CCTGCTCCAA TACCAATkGT AGGTATTGAT ATCTCTTTAG 1200

AAATTAACTC TGCAAGCTTA GTTGGAACAC ATTCTAACAC TACAGCGAAT GCTCCTGCTT 1260

TTTCTACAGC TTTAGCAGCT CTTATTAATT CTTTAGCAGC TTCCTTGTCC TTTCCTTGTA 1320

CCTTAAATCC TCCAAATGCA TTAACAGATT GAGGTGTTAA ACCTATATGC GCCATAACAG 1380

GTATAGAAGC CTTTACTATA GCCTCTATTT TATCGCATAC TTCTATCCCT CCTTCTAATT 1440

TAACAACCTG AGCTCGTCCT TCTTTTATTA ATCTTCCTGC ATTAACTACT GAATCATATA 1500

CTGAAGTTTG ATAAGACATA AATGGCATAT CTGCAACAAC TAAAGTGTTC TTAGCTCCTC 1560

TTGTAACTGC TCTAGTATGA TGGAATCATG TCTTCCATTG GTACTGAAGT GTATCTTCCA 1620

TGTCCTAGAC ATACCATACC AAGAGGnATC TCCTACTAAG TATnCCATTT AT 1672

(2) INFORMATION FOR SEQ ID NO: 758:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 734 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 758:

TTCTCACACA GATAATACCA GTTGAAATTT TATTCGTCAA CTCAATAAAT GTAACTTTTA 60

CATTTTGATG AAGAAAAGCC TATTCTATCA ACAAAAATTT TTTATTAAAG GAGATTTTAT 120

ATGTTTTTAC AAAATAAGAA TGTCGTCGTA ATGGGTGTTG CCAACAAAAA AAGTATTGCT 180

TGGGGCTGTG CAAAAGCATT AAAAGATCAA GGGGCCAATG TTATTTATAC GTATCAAAAT 240

GAACGCATGA AAAAGCAAGT GGTTAAATTA GCAGATGAAA ATGATTTGTT GGTGGAATGC 300

GATGTTGCTT CTGATGCGTC CATTCAAGCA GCCTTTGAAA CAATTAAAAA TGAAGTGGGC 360

ACGATCGATG GACTAGTCCA TGCGATTGCT TTTGCCAAAA AAGAAGAATT GTCTGGAAAC 420

GTTAGTGrCC ATCACTCGTG ATGGTTTcCT ATTAGCACAA GATATTAGTA GCTACTCTCT 480

TTTAGCAGTC ACTCATTATG CGAAACCTTT GtTAAATCCT GGTTCAGGaA TTGTCACATT 540

GACGTATTTA GGtTCCGAAC GTGCGATTCC TAACTACAAT ATGATGGGGA TTGCTAAAGC 600

TTCTTTAGAA ACTGCCGTGA AaTATTtAGC yTTTGAATaG CtGCmGAaAA tcGTGTCAAC 660

GGAATCTCCG CAGGnCGGAT TAAAACATTG GCTGTGACGG TGTAAAAGAC TACGATCAAT 720

TGATTAGTAT TCnG 734

(2) INFORMATION FOR SEQ ID NO: 759:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1262 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 759:

ATTATTCAAC AGACGGAGAA GTAGATGAGT ATCTTACGCA ACTTATTTTG AAAGCTCGTC 60

AAGCAGGTAT TTTCTTTATT GTTGCAATGC AACGTCCAGA TGGAGAATTT TTAAAAACAG 120

CATTACGTGA CCAATTTATG TTTCGCATGT CTGTTGGACG TCTATCTGAA ACAGGTATTT 180

TGATGATTTT CGGTGATGAA AATAAAAATA AAAAGTTTAA ATATGTTGAA AAAATTGACG 240

GTCAGAAAGT CTATGGTCGT GGATATGTGG CTCAAGGAGG AGGAACTGCA CGTGAATTTT 300

ATAGTCCTCA AGTACCACAA GACTTTGACT TCATTGAAGA ATTTATCAAA ATTTCTAAAG 360

AGTTAGGTTA TGAAGATGTT CCAAAAGAAG TACAAGAAGA AGTATCTCAA AAAATATCTA 420

AACATATTGA TAAAGAAGCA CTGGCAGAAA TTGATGAAGA ATTTCAAGCA GAAAAAAATC 480

AGTTATCAGA GTTAAGTGAG AAATATTCTG CTTAAAAAGA AGTCACATAA ATTAGTTGGA 540

AAAGGAAGGT TTTTATGGAT TCTAATGAAA TAATTAAACG TGTCAGAGAA CGAGTATATA 600

GAGAAGTAAA GAAAAAATAT ACTAGAGACG ATTTGGATAC TCGTATTCAA GATGTATTAT 660

ATTATCGTTC AGAAACATAT ATGAAATTAG TTAGCTTTGC AAATGGTAAA AGAATTAAGA 720

AGTTAGCTGA TCCTAGAAAA TTTGAAAAGT TTATGGATAC AAAAGGAGTA AAGATTGTTG 780

CAGAAGTACT TGATGGATTG AATAATCAAC CAAAAATGCA AGCTATGGrA TACGAACAAA 840

AGGTACTTAC TAAAGTACGT CAATGGTATC AGAAAAAGAA TCACCCTGAG CTTGTAGATT 900

TGGAAGAAGA AGCATTTGAA CAACTTGTAG AAAAAAATAT CATCTATAAA AAAATGAAGA 960

AAAGGTTGTA TGAAGAACAG GATAATCAAG GTTTTGTTTA TTCGGATAAT TTTGATATGC 1020

AATTGATTCG TGATTCATGT GATATTGAAG AGGCTCTATA CTTAGATATC ACGTTGGGGG 1080

ATTACTAAGT GAAAGGAGTA AAAAGAAAGT GAAATTTAAA TCTCGAGTAC AAATGTTAAA 1140

CTATGTGATT GGGAAGAAAC CAGACATCAT GAAAACAGTG ATCGGAGAGC AGAAAAAGAG 1200

GGGGGTACAG ATGATAATAG GTTACTTGAT TATAAATTAT TAGTAATTTC nGnTTTAGGT 1260

GC 1262

(2) INFORMATION FOR SEQ ID NO: 760:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 776 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 760:

CACnGACTTA AAACTTATCT GGAAGTTTGT TGATAAnACT ACCACCAAAA ATGGAGGnAT 60

TATAGTATGT TTTTATTTGG AAATAAAGAA GAAAnAAAGG TAAAGAAAGC AGAAGAAAAG 120

CAAATAAAAG AATATTTTAA AAATAATCAT GATATTGTAA TAGGCGGTAT TTATTTTAAT 180

GATTCAGACA AAAAAATATT TATCCCAAAA AGTATTTCTG AATCAAGAAA ACAACAAGTT 240

ATCAACTATG ATGATTTGAT TAGTTACACA GATATTTTTG TTGGTGGCAA TATAAAAAAA 300

CATCACGGAA TAACTCGTGC AGTTGTTGGT GGCGTTTTAG CTGGTCCTGT TGGGGCTTTA 360

GTTGGTGCTG GAACAGGTGG AAAAGAATTT ACTTCCATTA AGCAATTAGG AGTTATGTTA 420

CATCTTCCTA ATAATCAGAC AGTAAAATAT ATGTTAATCA CAACTGAAAC AAAAACTGAT 480

TCTATGATTG GTAAAGGTTT AATGGATAAA TACAACGAAT TAATTGCAAA ACTGGaCCAA 540

ATACTAAAAA CTAATTCAAG TAATAAAGAT AACACTGTAT TATCTTCTGC AGACGAGATA 600

AGAAAATTTA AAgCCTTGTT AGATGATGGc ATTATTACmA AACaAGAATT TGAAATAAAA 660

AAACGTGAAw TGtTACmATA aAAATaACGC ACCCTCTCCG GCTAAGAAGT TAGtGCGTTA 720

AAAATaGAAC CaAAATAGGC TwATTTkGTT ACGCCTATTT TACCACAAAG AAAAGG 776

(2) INFORMATION FOR SEQ ID NO: 761:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 927 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 761:

ATTGCTTTTT CAATTTGGTC CTGAGGATTC CCAGACCCTC CGTGTAATAC TAGAGGTACA 60

TCAGAACCGA CAGCTTCAGC AATCGCTTGT AAGTGATCAA ATGCTAAACC TTTCCAGTTT 120

TCTGGATATT GACCATGGAT GTTACCGATA CCACATGCTA AGTAGTCAAT ACCTGTAGCA 180

ACCATTTGTT TACATTCTTC GATATCAGCT AATtCGCCAG TACCGATGAT TCCGTCTTCT 240

TCACCACCGA TTGAACCAAC TTCACATTCG ACAGAAATAC CTTTAGCGTG CGCTTTTTCA 300

ACTACATCTT TCGCTAATTT TAAGTTTTCT TCGAATGGTA AGTGAGAACC ATCAAACATG 360

ATTGAAGTAT AGCCAACTTC GATACATTCT AAAGCAGCTT CGTAGTCACC GTGGTCTAAG 420

TGAATTGCTA CTGGAACTGT GATATTCATT GAGTCTACTA AATTAGTAAT CATGTCTTTA 480

GCTACTTTGT AGCCGCCCAT GTATTTAGCA GCGCCCATTG ATGTTTGGAT TAGAACTGGT 540

GCTTTTTTAG CTTCTGCTGC TTCTAAAATC GCTTGTGTCC ATTCTAAGTT ATTTGTGTTA 600

TAGCCACCTA CTGCATACCC GCCTTTACGT GCTGCTTTTA AGAATTCTGC TCCTGATACT 660

ACTGGCATAA AAAATTCCTC CTAAGTTTTT AAAAACTTTA TTTGTTAGGT GTTCCCTAAC 720

CAACGGTTCT ATTTTAGCAG AATTCCTGAC ACTTTTCTAG TAAAATTCCC TTTTTCAGAA 780

AATTTTTGTC TATTCTTTCA CATAACCTTC ATTTTTATAA AAAAACCTCC AGTTTTCCCG 840

TTAAGACTTT GGGACCTGAn GGCCGACTTG CTATTTTAAT TnGTCTTGGT GGATAACGTT 900

CACCTTGTAG GATGAAGCnC AGGAAAT 927

(2) INFORMATION FOR SEQ ID NO: 762:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 528 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 762:

AATTACCAAA CCACTGAATT AGAAACAAGA CAAGCAAAAG AACGGGCCTT AGCAGCAGTT 60

TCmGCGATTG TTGAAGGAAA GCCAATCCAC GTTATTTTTT CrGGGAAACC TGGTGTCGGT 120

AAAAGTCaTT TGGCTATCAG TATTTTAGTT GAAGTCTTAG aACGCTCTGC ATATCaAAAG 180

TATTGTTTAT TTGTCAGCTA CTCTGAGTTA TTAGAAAAAC TAAAAATGTC CATGAATGAA 240

TCGGCCAAAA GCCAAGCAAA GGCTCAAGCG TATATTACTA GAATGAAAAA AGCAGACGTT 300

TTGGTTTTAG ATGATTTAGG TGCTGAATTA GGAATTArAA ATAAAGTTAG TACGGATTTT 360

AATAATGACA TCTTAAACCG AATTTTAGAA GCTAGACAGA ATAGAGCAAC TATTTnTACT 420

ACTAATTTTT CCGGAAACAA CTGGTGGAGG CCTATGGAAC ACGCATTAnn CCTCGTCTAA 480

TGAGCACGCC AATGGCTATG TnTCCATATG AAGACACACA GACAACGA 528

(2) INFORMATION FOR SEQ ID NO: 763:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 500 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 763:

GCTGAAGCTG ATCCAATATA TATCAAAACA CAGCTAGGAC ATAAAGAGAT AATACAGTCA 60

TATGAATACG CATCAGCTAC TGAAGCAAAT AGAAAAAAGA ACAAAGAAAA ATATGAAAAC 120

AAATACAAAG ATATTTTGTA AATACCCAAA AATTACCCAA AGTATATCTA AATTGTTATC 180

TTGATGTATA TATTAAAACA TATAGAATTG TTTTTAATTA AACAGCAATA TTCTTGAATA 240

ATTATCTATG AAGCTGGCAG GTGGCTCTTT ATTAAAAnGC TAAAACCTTT ATTTATTAAA 300

GGTTTTAGCT TTTTTCATTA CCGACATACC CAAAAAGTAC CCAATTTTnT TGTATTATTT 360

TrTATTTCTT GATAGACAAC ACGAACGAAA GTTCGATATA ATTGTATTAG GGAGGACGAT 420

ATCATGAAAA CTAATTATGT AGGAGTAGTT GAAAAGATTA AAATGTnAAG TATGTATCCA 480

AAAATTTTAG GTCGATTCTC 500

(2) INFORMATION FOR SEQ ID NO: 764:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 446 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 764:

ATTTATTTTT TTAGTACAAA AAGTCAGCAA TTTTTTGCGT ATGTTTGGGC CATTTTCCCA 60

GTCATCGCCT TTTTTTTGGT TAGTTTTTAT ACATTGGATT TTTTATTTTC CCCAAGCTAT 120

TTAATTATGG TGCCTGTATT TACTATTATG TTTAAAATTA ACTATAAAAA AGATTATTCG 180

TTTGTAGCGC TGCTCAAAAT GAGTATCAGG CAATTTGTGA TCAGTTTTTT AGCATTTATA 240

GCTAGTAGTT CCTTTAGCTG GTCTATTATT TTAAATAGTA GTGTGACATT TTTAATTGTT 300

TATTTATTTT ATTCATCTGC GACACCGATG CGGTATCATA GTTATTTGAT GATGTTTACG 360

TTTTGTCAGT TAATCCAGGT TAAAGGAAAT ACATTTTTGA TGCAATCTCA AGTTGTGGnC 420

TTTTTAATTG GTCTTTTTTT ATTGGC 446

(2) INFORMATION FOR SEQ ID NO: 765:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 594 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 765:

TATTTTTTCT GAACTGCTCT TGCTTTCTTC CGCTCCAGAA GTCTGTGTTT TTTTCATTTG 60

TTCTGGTGTG GGATAAATCT GTTCGCCTTC TTTTGAGAGA AGGAAACGGC TACGTGCTAA 120

TTTTGCCACA TAATCATCAT CTTTTAACAA AGCAACGTCT TTTTTTAGAT CCTTGACCTT 180

CTTATCAACT GCTACGGATT CAGCAATTGC ATCTGCTCGT TCTTGTTTGA AAGCACCTAA 240

GCGATGATAA TCTTTCATCA ATTGGATGCC AGAAAAGGCA AAAATAATGA AGGCCACGAG 300

AAAAATCGCT GCCAAGgTCT GCGTCTAAAA ATCAATTGTT TCTGTTGkTT TTGGAATTCG 360

ACATACTGTT CTTTCGTGTA GTCATTTTCC AATGCCGCAA CTTTCyTCGA GTTTTTTtCA 420

TTCTTTCCCA TCGTCTCGCT CCACTCGCTT tGGATTGTCC GTTTTATTAT ACCAACATCT 480

AAGAACCTnG TCTAATGTtA ATTTCAGAGA AGTTAAAGAA AATCAAGCTT TTTTCTAAAA 540

TAATCCTTGC CTTTGGAAAT AAACAGCGGT nGATTAnTAA ATTTTTTCCA AAAG 594

(2) INFORMATION FOR SEQ ID NO: 766:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 516 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 766:

TGATTATTTT CCCTTGGAAT ACCTTAGCAG GCACAACGGT GAGTCCATTT GTTATGGTTT 60

TTGAAAAAAT GAATATTCCT TTTGCTGCTG ACATCGTAAA TTTTGTGATT ATTTTAGCCT 120

TGTTTTCTTC TATTAATTCT GGAGTATACG CCTCTTCTCG CCTTTTATAT TTTCGCTTAA 180

AAGATAAAAA AGGACCCATG AGTAAACTAG GGGTTTTAAA CAAGCACCAA GTACCTCAAC 240

GTTCAGTTTT CTTCTGCGCC AGTGTTTTAT ATCTGGGCGT TATTCTTTCT TATTTTGTTG 300

GCGATGAATT ATTTGGTTAC TTAGCCGGTT CTCTCTCTTA CACTGTCTTG CTAATTTGGA 360

TTTaATTAGC GCTGCGGCGT TTGTnCTGTC TTTAAAACGA GGGTCTCTCT TTGAAAAAAG 420

TATTAATCTC CTTGCGTTAA TCATTTTAGG TCTCATCTTT ATTGGCATCT TATTtACAAA 480

tTCGTaGGCG TCACTCyTTT AACAGGCCTA CTGTAC 516

(2) INFORMATION FOR SEQ ID NO: 767:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 864 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 767:

TACGGCTTTG AGCGTGTCTT TACTAGTTTA TTGCAAATCA ATGATGATCG GGAAAAGGTA 60

TTAGCAGAGT TTAAAAAAGT CGTAGATTAC GCCAACCAAT TAGGTATGGA AGTGATGGTC 120

GATATTAATC CTGCATTGTT CGAACAaTTA GGCATCTCTT ATGATGATTT ATCATTTTTC 180

CATGATATGG GCGCTTACGG TATTCGTTTA GATCTTGGTT TTACTGGGCA AGAAGAGGCG 240

AATATGACAA GAAACCCTTA TGGtATTAAA ATTGaAATTA ATATGAGTTC mGGGaCAAGT 300

TATGtAGATA ATATTATGGc CTACTCACCa AACACaGaAA ATTTACTAGG CTCACATAAT 360

TTCTATCCAC ATCGCTATAC CGGCTTAGGC TATGAGCATT TTGTTTATTG TTCAGAAAAA 420

TTTAGAAAAT AtAATTTAAA TACGATGGCG TTTGTGAACT CACACGATGC TACATTTGGG 480

CCATGGCCGA CACAAGATGG GTTATGTTCT TTGGAAGACC ACCGCGATTT AGAAATTGCA 540

ACGCAgTTAA GCATTTAGTT TTAACAGGAT TAATCGATGA TATTTCTGTC GGAAATGCTT 600

ATGCCTCAGA AGCAGAATTA GCAGCAATGG CAGAAGCCTT TCATGCACCG TATCCTTCGA 660

TCAAAGTGGA TACAGAACCA GAAATCACAG AAGATGaAAG AATTGcATTG TTCGATAACT 720

TACACAGTTA TCGTGGaGAT CGTTCTGaTT ACGTTTTGCG TTCaACAaTG aCACGtGTTT 780

ACTATAAAGA TCGTCCTTTC CCAGCTCATA CAACCGAGAG ATATGGTTCG GGCGATGTCC 840

TTGnATTGAC AACGnnGGCT ATGG 864

(2) INFORMATION FOR SEQ ID NO: 768:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1207 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 768:

ATTTGCTTTT AGCAAATAGG AATCTTAAAC TTTAATTACA ACTAATTTAT TCTAATTTAA 60

GGAGAGAAAA ATGATTTCCA GCAAATACAC TATTAGAGAA CAAAATGAAG CTCAAATACT 120

TAATCAAATT ATAAAACATA GAGAAATATC TAGAGCCGAA TTAAGTAAGA TCACCTCTTT 180

AAATAAAGCA TCTGTTTCTT CAATTACAAA CAAACTAATA AACGATGGAT TCATCATTGA 240

AAAAAGGATC GGAGAAGCTA CAAATCGAGG AAGAAAACCT ATCATGCTGA CCTTTAACGG 300

TAGTTCTTCA GCCTTAGTAA TTGCAATTGA TCTAGGCTAT AACTATATAG ATGGCGCATT 360

AGCTTCTTTA GATGGAAAAG AAATCCATAG GGTTCAACTT ATTGATACCT ATGTTAACAA 420

AAAAAATGTC CAAGATTTAA tTCATAAAGT ATTCAATCAG CTTACTTCAG ATTGTCCTCA 480

TACCCATTAT GGCATAGTAG GTATGACAAT AGCTCTTCAA GGTCAAGTTC TAAATAATAA 540

AATTATCAGT ACTTCCTACA ATGATCTTGC TGAAATAAAC CTAGTTGAAA TGTTAAACCA 600

AGAATACAAc TTCCCAGTAT TCTTACAAAA TGAAGCTAAT TTATCAGCGT TGGGTGAATA 660

TACTTTTTCA TCAGATATAG AAmACTTAGC ATCCmTTmGT CTTCATAGTG GCATTGGTGT 720

CGGACTTGTG AAAAATGGTA AACTAGATGk TGGTAATAAm GGTTATGCTG GACAATTaGG 780

ACATACTATT TTGTTTCCCA ATGGAAGAAA ATGTACTTGT GGAAATCATG GTTGCTTAGA 840

AAAATATTGC TCTACACAAG TGATTTATCA CGAAATTTCG AAAGAGAAAA ATCTCAATAA 900

AATAAATTCA GATGTAGTTT CTATTCTATA TAATAACAAA GAATCAAAAG TTGTTAAAAT 960

GATTGAACAA TATGCCTACT ATTTATCTAT TGGTGTTAAT AATGCTATCA TGTTATATGC 1020

ACCTGAATTA GTAATTTTCA ATAGTCCATT AACAAAAAAA ATTCCAGATA TAATAAACAT 1080

AATCCAAAGA AATTTAAACA ATCAATTTAC AAAAGACGTA TTAGTTATTA ACAGCCCTAT 1140

CGAATGGAAT CCTGTAATTA GTGGAGCTAT CTCACTTTCA ATACAGAAAT TTTTAAATAT 1200

TGAACAA 1207

(2) INFORMATION FOR SEQ ID NO: 769:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 772 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 769:

nGATAATCAC TGCACCATCA TGTAGTGGTG TATTTGGAAT GAAAATATTA ATCAGTAATT 60

CACCAGTAAT ATCGGCAtCC AAATCAATAC CTGTCTCAAT GTATTCATCT AAACCGGTAT 120

TGCGTTCAAT GGTTATCAAA GCACCAATTT TCCGTTTTGA CATGTACTGA ATGGCTTTGT 180

CAAAGGCCTG AATAATACGC TCATCTTCTT GTTGCGCACC TTTATTCGGG CGGAAAAACG 240

AACTTCTCCC TAGATGTTCC AGCCCTCTTC GCACTTCTGG TTGAAAAATA ACAACCGCGG 300

CGATtACACC ATACATGATG ACTTGATTCA TtAACCAAGA TAATGTATGT AGACCAATTA 360

TCTCACTTAG AATGCGGATA ATGATAAAAA CAGCGACACC TTTTAATAGT TGGACTGCTT 420

TTGTTCCACG AACAAGCATG ATTAATTTGT AAACGAGGTA CCATACGACT AGGaTATCTA 480

TGATATkGAT AACAAAATCC TTAGATAAGA AATCCGTTGA GATGACTTGC TGCCAATAAT 540

TCmAATCTAA TAATTGATTC aATkGAAACG ACaTCATACC CCTCACTTCT CTCtCAAGAT 600

TyCCTAACAA GTATACCACA TTATTTTCTT TCATCACAGA ACGCAAGCGG AAACATCAGG 660

AAAATTTAGG AACTTTTTCA AATAACTTGA AATTTTTCAT CAAAAAATCG CATTTCTAAC 720

TAAAAAACGT GTGCTAAGCn AAAAGCTnCA GACACGTTTT TTTGGACTnC TA 772

(2) INFORMATION FOR SEQ ID NO: 770:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 706 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 770:

GACnAGCTTC TAAAACTCGG TCCACGTCTA ATTCGGAGTG CATATACTCA AAAACATTGA 60

ACGTATCACT ATTTGACAAC GTGTTTAAGT CTAAGACCTT AATCTTGTAC TTGCCTTTTT 120

TCAATAAAAG TCCAATCTCT GGCAATATCC CACCGTCTGG GTCTGTTGTG ATAAAGCTAG 180

CATGTCGTTG CATAAGATTG GTTTTGATAA AAGCTAACGT TTTGGCCGCT CCAGAGTCAC 240

CTAGAATGGC GATATTCTTA TTTTTTTGTA GCTTGATTGG TAAACGTTTA TTAAAAAGAG 300

AAATTTTTAC ATGCTTACTA ACAATAATAT TGTTTTCAGG AATTGGATCT TCATATTTTT 360

TCATTTCAGC AGGTGTAGCA AAACGAGCCG AACCGTATTC TTCTCCATGA CGGTAGATTC 420

CACGATCATT ATCTCGCACA TACATTAATA GCCCTAAAAG AAAACCAAGT ATTCCCACAC 480

TAAATCCTAA TAGACTTGGT GTTATAAAAA ACCGATAAAA AGTTTGTTCG ATATACGTTA 540

GAATATCTTT TGTAAAGAGA TAATTCGTTT TTGAAAAATA GTCTGTCCCT GGTAACAGCC 600

AAAAGAAATT TCCTATTCCA AAGCCAATCA CAAATAGAAA AATACCAAGG AGAGATACGG 660

GTTCAGTTCC TTTTTGGTAC GCTGCATCAC AGTTCCCTCC TTTTGG 706

(2) INFORMATION FOR SEQ ID NO: 771:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 635 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 771:

CCAAAGGGAG GATTTnATGC AAATTAGGCA GGTACATGTT AGGGCATCGA AAATTAAACT 60

GCAAAGAAnC ATTTACGATC GCTTTGGGGA CGATAGAATC AGCGGATAGT GCAATTGTGG 120

AAATTGAAAC AGAAGAGGGA TTAGTTGGAT ATGGAGAAGG AGGGCCAGGC ATTTTTATTA 180

CCGGTGAAAC GTTGGCTGGC ACGCTTGAAA CGATCGAATT ATTTGGCCAA GCAATCATTG 240

GCTTAAACCC CTTTAATATT GAGAAAATTC ATGAAGTGaT GGATAAAATA TCTGCTTTCG 300

CTCCAGCGGC AAAAGcAGCA ATCGATATTG CTTGTTATGA TTtGATGGGT CAGAAAGcTC 360

mACTGCCTCT TTATCAATTG TTAGGCGGCT ATGATAACCA AGTAATAACG GATATTACAC 420

TAGGTATTGA TGAACCAAAT GTGATGGCAC AAAAAGCCGT CGAAAAAGTC AAACTAGGTT 480

TTGATACATT AAAAATTAAA GTCGGAACAG GAATTGAAGC TGATATCGCA AGAGTAAAAG 540

CGATTCGTGA AGcAGTTGGC TTTGATATCA AGCTCCGCTT AGATGCGAAT CAGGCTTGGA 600

CGCCCAAAGA TGCAGTTAAA GCGATTCAAG CATTA 635

(2) INFORMATION FOR SEQ ID NO: 772:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1102 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 772:

TnATTAATAT CTATATTTGA AATGTTATCA TCATATAATT CAATTGTCGC TAAAAGTTCA 60

CCTACAGTTC CCTTACACCA ACTATTGTTT TTAGGTTCTT CTTCAAAGTA TGATTCTTTA 120

ATTTTATGAA GTAAAGAATT TGCCTTATCT ATCCTATTGA ATTTTGATAG AAGATGGACA 180

TAGCTATATA TACCGTGTCC GAATCCTCTG AAATTAAAAT AGGGTTCTTC GGATAATTTT 240

TCAAAAATTT CTAATGAAAA TTTTCTATAT TTTTCATCCT CTGTTATCTC GGATAACAAT 300

AATAAAACTT TTATTATACT ATTATGCCCG TGGATCCAAT CATTTTTTAA CTCGCCATTA 360

TTTATCGGTT TCTTTTCAAT CAACATGTCT GGCAATTTCT AnGGACTACA TTTAAAGAGT 420

TTATATCATT ATTTAAGCGG nAATCTACAA GTAAAGGATA TATTAAACTT CCTTTTCCAA 480

AAAAAGCTGA CAATATATCT TCAGATGGTA TTGTATAAAT TGAATTTTTT ATACATTCAA 540

TTACATAGTC AtATTTATGA TycTttGTAA TGTATTTAAG AGCCACATAA AATATGAATA 600

TACCAGGCAA TCCATCGwAC ATATTATTAT TTAGGATTCC CACATTCCAA TCTTGATCCA 660

ACTTTATGTC AATCCAATTT ACAGTATTGG TTTTTTTATT AAAAATAGCT CTCTTGAAAA 720

TTTTTTTCTC TATCTTCTGA CATGCTTGAA TAATTTTACC GTTTAGCTCT AAGCCAGTAT 780

AAATATATTT ATTTGAATTT TGATTTTTCA AATCATTAAT ATATTTATAT GGATTATAAA 840

TATTcAAAGC AATTTCTAGC CAAACGGTTT GAATAGAAAT ATCTTCrTCA CArrGaTCAT 900

TTATTTTATT TAGGCATCTA TTCAGAGCGC TTnCTTGATA AAAATCTTCT ACTAAGCACC 960

CGTCACTAGC TATTAATGAT GTTTTTGAGA TATTGTTATA AAAAATGGGT ATGTCTCCAT 1020

CTATTAAGTC TGAAAATTCG TAATGAACTA CCTTTTTATT CTTATATGGA TAGGCCCACA 1080

TATTATGAAG TACTTTTTCT CT 1102

(2) INFORMATION FOR SEQ ID NO: 773:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1081 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 773:

TATGATTGTT TAGGTTGTAG TTGTTGCAAT TCGGCTTCTG TTAAGGGCCG ATATTCCCCT 60

AGTTCTAACG TTTCATCCAA CCATAGTTCG CCCATTCGTA ATCGTTTTAA ATACGTCACA 120

GGTTTGCCTA CAGTGGCTAA CATCCGTTTG ACTTGATGAA ATTTCCCCTC TTGAATCACT 180

AAACGGACTA CGGAAGTCCC TTCTTCCTCA TTGATCGTCT CAATGACTAA ACGACTTGGT 240

TTAACCATTT CTCCATCTTT CAAGGCAAAT CCTTGCGCAA ACAGTTGACA ATCTTCCGCT 300

GTCACAAGTC CTTCAACTTC AGCTAAATAw TCyTTATCCA CATGCTTTTT GGGkGaTAAT 360

AATTGATGCG CTAACTGGCC ATCATTtGTT GgTAGAAGCA AGCCyTCGGk AyCTTwATCC 420

AAGCGACCAA CTGGGrACAA GTCGGCACGG kAATCCTGAT CAGTtAACAA GTCAATTACC 480

GTTTCATCTC GCTGATCTTC CGTCGCAGAA ATGACACCTT GTGGTTTATG CAACATGTAA 540

TAAAAGAATT TCTGATACGT AAGAATTTCT CCCATCAAGG CAATTTGATC TGTTTCyTCA 600

TTtACTTGTG kTTTGGCTTC TTTtACGACG TCACCATTCA CCGkAACGaC TTTTTTCTTT 660

AACAATTGTT TGATTTCTTT ACGACTACCG ATACCAACAT CGGCTAAAAA TTTATCTAAA 720

CGCATCGTCA TTCCTCTTTC CTTTAAAAAA ATAGCTTGGG ACATAAATCG AATCGATCGA 780

TGTCTCArGC CTTGTCTCTT AAGAGAATCT GCTGAAAGGA GTTCAATCAG CTATTATTTA 840

ATATTGAATT TTGCACGTAA TTTTGCGGCA TTTTTCCCTA ATAAACGATC AGCTAAGCGA 900

ATCTTCAAGA CCGAATAGCC ATAAAAAGCG CCTCCTGCAG CAGCCACAAT TACAACGATA 960

ATGAAGGCTk GGAAAAnGTC TAGTTGGATC TAAAAATAGA TACAAGAACT GACGGAAAAT 1020

CAAAGnGATA ATCATCATGA TGACTGTnAT ACAGAACATT AACAATGTTC GCCGTAAGGT 1080

T 1081

(2) INFORMATION FOR SEQ ID NO: 774:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 471 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 774:

TTACTAAAAA ATTGTTTAAG ATTGTCCATA GTACTCCCTC TGTTTACGTT TCTTGGCTAA 60

GTCTACCATA TTCTTGCAGG TACGACCTAG GTCTAAAGCA TGAAAAATTC TTGAAAAACA 120

ATTTCAATGG TCTTTTCGCT GTAGCAAAGG AAATATGGCC TTCACTAGAG GTAAATAATT 180

GCTTACGCTT AGAAACGAGT CTTTATTTAG CCTCTATTCT ATCATGCTTT TCTTCGTTTT 240

TTATGTAAAT TTGTTTAAGA ATCTCTCAAG GAACTGCCTT TCTAACAAAT TTCAGCTTTT 300

TTTCCGTTCT GACTTGACCT TTTGAGTAGC TTTGTTACAC TAAATAGCGA ATCAGTGAAA 360

GGTGGGATTT GGTGGAATAT TCAATGAGAT TACCTGAAAA TTTTTCAACA ATGACTGTAA 420

AAGAACTATT AGAAAATGAn GGCTAGTCCT CGAAAAGTGC GCCATTTTTT n 471

(2) INFORMATION FOR SEQ ID NO: 775:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 480 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 775:

ATGGAAAGAA AGTAATTAAA GGAGCTTCAA AATAGTTTAA TATATAACTT TTAGGAGGGA 60

ACCCAAATGG AAATTATTTA CCCACCTTTA GTTGAAGAAG GATTAAAGTA TTATCTTGAA 120

ACGACCCAAC AATCGCTAGA TAAAAGTACT TTTTATCGTT CGATGGTAGA GAGAGGAATT 180

ATTACTGAAA CGGGTTTACC GACCCAACAG GCAATTGAAA ATGGATTAGT CAAAGACTAT 240

TACGAGGACC AAGGGTTATC ATTTGATGAT TTTTTAAGAA TCTATCCAAT TTTTGAAGAG 300

TATGACGAGG AGTTATTTCA ATGTATCGAT GGTaTTGGGA GATACCTGTg ATATGAAgAA 360

AATTTAGCTc ACAATgGAAT CTGGCGAGTa cTTTgAgtGC ACAcAGtcAA GCGTtCTgAg 420

AcGtAgaTgA gaGGAAtTtG CGcAAAATAT TAGAGGCGAC GGGACGCGAC CCTGCAAAAA 480

(2) INFORMATION FOR SEQ ID NO: 776:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 485 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 776:

GTGAGAACCA ATCTTAATTG GTGATCCAAC AGCATATTGC AAATGCTGCT ATGGCTCGTG 60

GAATCAACAC ATCAAACTTT ATTATTATCA ACCCAAATGA TTATGAAAAA TGGGATGAAA 120

TGGTCGAAGC ATTTGTTGAA CGCCGTAATG GAAAAGCTAC TAAAGAAGAT GCTGAAAAAA 180

TCTTAAAAGA TGTCAACTAT TTCGGTACTA TGTTAACTTA CATGGGTATT GCTGATGGTA 240

TGGTTAGTGG TGCTATCCAC TCAACAGGTG ATACAGTTCG CCCTGCGTTA CAAATCATTA 300

rAACAAAACC AGGCGTGAcC GTGACAAGTG GTGCTTTtCTT AATGATTCGT GGTCGTGACC 360

AAGAAAGATA TkTATTCTCA GACTGCGCGA TTAACGTGAA CCCAACTGCA CAAGAATTAG 420

CTGAAATTGC TGnTGATAGT GCCAAAACAG CTGAATnATT TGACATTGAA CCAAAAGTTT 480

CATTG 485

(2) INFORMATION FOR SEQ ID NO: 777:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1388 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 777:

GTAAATTACT TCATnATTTG GnAACTACGG ATAATTTACC TAnAnATGAC AATTGGCCAA 60

TAACTATTAA AAAAATTACT TTTTTAAATC AAATATCAAA TGACAGCATA AGTAATTTGT 120

ATAAATTGTG GGTATTACCA CTATACTATA AGGCTATATC GGCTGATAAA CCTAATAATA 180

TGAAATACTT TAGTGTATTT AAAAAGTGTA TTCCATTTAT TTCAGATACT CTTTTTACTT 240

ATAAGTTATT ACTAAGATAT CAAGAAACAC AATACGAAGT TTTAAGAGAT TTAGAAAGTC 300

TATACAGTTC AATTTTTCGT AATTTGTTGA AATTTAAAGC TAATGAAAGT CCTGCTATTA 360

TAGTAGAAGA ATTGAAACGT GAAAGATCGG GCATTCTTAA TAATTCCCAG GAAATAATAC 420

AAAAAGCTAT ACAAGAATTA AGCTATTCAA AAAAAACAAA TTCCCCAATG GATACGGATC 480

TAGTCAGGTA TCTAATTGAA AGACTAAATG AGAATAATTT TAATACTCAA AGTGTTGTAG 540

GTGGAATTGA cACATAATTG AgATGAAAAg GamCGAAtGC AGTAGTAACA TAGGAATTTA 600

GTaTTTTTAG AAGGAAtTAT AATGATATGG CTAATAAGAT TAAACAAGAT TTACTTAAAA 660

AACATAATGA TGATGAAAAC AAGGTTGAAA GACTACTTTT AGAAAAGAAG TTTACTGAAA 720

TATACACTCG TAGTGAGTAT CCTCAAGTGA AGGAGCTATG CCGTACTTAC CCAGATGGGG 780

ACTTTACCGA AGTAGAAGTC AAAGAAAGAG CTAAATATAT GGCAAAGTAT TTTATCGATG 840

AGTATTTAGA AATCAATTAA CAGTTAATGT AAGTAAAATA AAACAATAGT TATTGAATCG 900

ATTACATTTT TTTGAGAATA CCAAATGGAT CTATTGATAT GGGGAGTCAA CCTTAAATTA 960

TTACCAATGT AAGAAAACGA TTCATACAAC TCGAAACAGA GTAAGTCAAA AAGGATATTA 1020

CAAAGAGTTA CACTATTTAT GAAAGAGAAA AAATTTAAAA CGTGCCAACA ATGCTTTTAT 1080

AAATTAAGCA ACGTTGGTTT TTCTGTTACT TAAAATCTAA TATCTTTAGT ATAATACCTT 1140

AATGTATTCT TAGAGTACGC CATAAGTGGT TAAAATACGC ATTCAATCTA GTATTCATGG 1200

TGAACTTTCT AAAAATAATA GTATTAATTA CAATTATTTA CTTGAGCGTT AAAAACAAAC 1260

TTAAAGAAAA ATGACTGTAG ATTTCTGAGC TACTTTAAAT TCAAAATAAT TCGGraAACT 1320

GAAATTTTGT TGkTGCaAAA CAGGACCGTT TGGTCCGGAT AATGGTACTC GGAAAGGCCG 1380

GTTTTTGC 1388

(2) INFORMATION FOR SEQ ID NO: 778:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 735 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 778:

ACTACnGAAG AGTACnATnG AGCACCTTGA TCCCGCACAA AATCGGTGCC AGATCCTTAT 60

ATGATTGATA CCACACCCAT GGCTGCCACC ATTAAAGCGG GCGAGACCAT TTCCATGACT 120

TCGAAAAATA TGCGACAAAA AGGTCAAATT CTTTTAGAGA AGACTGGGGT AGAAACAGGT 180

ACTGATCTTT GGAATGACAA TTATTCTCTA GCTGGAAATA CATTTGCCAT TCGTAAAGAC 240

AGCCCAGCTG GTGAAATTGT CCAAGAAATA ACAACGGATG AAAAAGGTCG TGCGGAAACA 300

CCAAAAGAGC TTGCTAATGC TTTGGAACTG GGAACCTATT ACGTGACAGA AACTAAATCT 360

AGTAATGGTT TCGTGAATAC CTTCAAACCA ACAAAAGTCG AGTTAAAATA TGCCAATCAA 420

ACCGTGGcTC TTGTTACCAG TAACGTAAAA GGGCAAAACC AAGAAATTAC TGGGGAAACC 480

ACTTTGACAA AAGAAGACAA AGATACCGGT AATGAGAGTC AAGGGGAAGC TGAGTTTAAA 540

GGAGCTGAAT ATACTCTCTT ACTGCAAAGA TGGTCAGCTG TAAATnGAGT GAGCTTTAAA 600

ACAGATTAGT GAGGGACGAA GCTTCTGTGA ACAGTGCTTG GCTTAGTGAA GACCAGTGCG 660

TAACnCTGCA TTACGGATTC TGCAGACCAG CCTGAGTTCT TGTGACGGAT CTGAnnTAAA 720

GTGTATACGA AAATG 735

(2) INFORMATION FOR SEQ ID NO: 779:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 255 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 779:

GCTGTACGGC AGCCAGACGT TGCTGTCCGG CACGGTGATA CTGTTCAAGC AGCTCGATAC 60

CGATATACCG GCGTCCGGAC TGAAGGGCGG CCACACAGGT TGAGCCGCTG CCTGCAAACG 120

GGTCCAGCAC AATCGCGTTC GGGTGTGTGA AGCTCTCAAT CAGTGGTTGC AGGCTGGTAA 180

CAGGCTTTTC CGTCGGGTGA TGACGATTGC CTGAATATTT CCAGCCCAGC ACGTCGGGCA 240

ATGGCTTCTG AGGCA 255

(2) INFORMATION FOR SEQ ID NO: 780:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 677 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 780:

AGATGCTGAT TTCTTTGCAT TTGAAGCGTG TGAGTATCGC CGTCATTTCT TGGCTTATTC 60

ACCAGATTAT GCGATTATGA CGAATATCGA TTTTGATCAT CCAGATTACT ACAAGAGCAT 120

TGAGGACGTT TTTTCAGCGT TCCAAACAAT GGCTCATCAA GTCAAAAAAG GAATTTTTGC 180

TTATGGTGAT GATAAGTATC TTCGCCAGTT AGAATCAGAA GTGCCAGTTT ATTATTATGG 240

CGTCAcGAAG AGGgATGATA TCCAAGCCCG AAATATTCAA CGAACAACGG AAGGCTCATC 300

TTTTGATGTT TATCACAAGG ATGATTTTGT AGGTCATTTT GTCTTACCAG CATTtGGCCA 360

TCACAaTATC mTGAATGCGC TAGgTGTGAT tGCtGTGGcT tATTTTGrAA AACTtGrTAt 420

GCmAAAAGTC GCaGAaGAAa TGCTAAGTTT TAAAGGTGTA AAACGTCGTT TTAGCGAGAA 480

AAAAGTCAGT GACATGATTA TTGTTGATGA TTATGCGCAC CATCCAGCTG AAATTAAAGC 540

AACGATTGAT GGCGCTCGCC AAAAATATCC TGACAAAGAA ATTATTGCTG TCTTCCAGCC 600

ACATACATTT ACACGAACAA TTGCCTTAAT GGATGAATTT GCTGAAGCAC TGGATTTGGC 660

AGATGAAGTA TTCTTAT 677

(2) INFORMATION FOR SEQ ID NO: 781:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 373 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 781:

TTTAGGGATG AGACAATTAC CGATGGGCGT TTCTTATGCG GTTTGGACAG GAATTGGTAC 60

AGCAGGAGGC GCTATTTTAG GAATGCTCTT AGGTGAATCA AAAGACCTAC GACGGATATT 120 tTTTATTTTT TTAATCATTG TTTCTGTGAT TGGTTTAAAA TTAATTGGTT AGAAGGCGAG 180

AGAACATTGT TTGGATTATT GAAAGACAGG TGAGTAAAAT TTAAAAAGAG ACAGAACCAA 240

GTGCGGTGTT CTGTCTCAAA ATAGCTTTTA TTTTACAAGG TTACTATAAC GtGTCGCAAA 300

TTCATTGTTG TAGGACTTGC CGTTGACATC TkGACCAACG TCCCAATTTG CAGACCAAGT 360

CATTAAGCCA CGG 373

(2) INFORMATION FOR SEQ ID NO: 782:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 706 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 782:

ATGCATGAAA TTAAGTAAAC GACAAGAGCA AATTATCGCT GTTGTAAAAG AGCATCAGCC 60

AGTTAGTGGT GAACGCATTT CGGACATACT AGCCGTTTCA CGTGCGAmAT TGCGTTCTGA 120

TTTGTCTTTT TTAACACTTT CAGGAATTTT AAAAGCAAGT CCGAAAGTTG GTTATACTTA 180

CGAAAGTGAT AATATGGAGG CTTTTTTCTT TTTTGATGTT TTTCAAACAA AAGTTCAGGA 240

GATCATGAGT CCACCaTTGA TGGTTGCCCA AGACACCTCA ATTCGAGATG CGATTACCAA 300

CTTGTTTATG TATGATGTCG GTTCTCTTTA TGTCAtGGAT GAAGCmAAAG AaTTATTAGG 360

GGTTTTGTCA CGAAAAGATT TATTACGGGC CTCTTTAAAT ACAAATATTG ATGGCACACC 420

TGTTGCTGTT TGTATGACAA GAGTGCCACA TGTGAAAACC TGcACACCAG AATTTACTAt 480

TCTTGAAGCT GCTGACACTT TGCAAAAATA CGAAGTGGAT TCTTTGCCAG TCGTTGAAAA 540

AGAAAATCCT AAAAAAGTTA TTGGAAAAAT TACGAAAACG AAAATTTTAA CCTATATTAC 600

GCAACAAGCG AAAGAAGCCG CACAAAACAG ATAAGGGAGC TTAATTATGA AAAAAGAGAT 660

CATTGTTTAT ACTATTTCCG ATTCACTTGG AGAAACATCA CAAAAA 706

(2) INFORMATION FOR SEQ ID NO: 783:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 499 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 783:

TGAACAAAAT CGACTTCTTC ATAAGTCACA TCCTTTATCT ATTAATACAG TTAAAAAATC 60

TGTATTACTT ATAGCATAAA TGAAAGCGAT TAAAAAGTAA AGTTCTCTTG AAAATTCCCA 120

TGTTCAACTT TTAGGCTAGG TTCTTTTTAT ATGCTAAAAT AGAATAGAAA CGGAGGTTTT 180

TaATGGAATG GaTTGAAATT AaGCACGCAA CaCaAAATAA TTTGrAGrAT ATCTCTGTCA 240

ATATCCCTAA AAAGCAACTA ACTGTTGTTA CTGGACTTTC GGGTTCAGGA AAGTCCTCCT 300

TAGTATTTGA CACATTAGCC GCTGAATCAC GTCGGGAACT AAATGGATAC CATTTAGTTC 360

GnTTTGGTTC CAAAATTACT TTACCCCAAA TATGGGTCGT CCCAGAAGTT GGAAAAAAAT 420

CGAGAATCCT TCCTGGTTGG CCAATTGGTC CATTGGACCC nGAAAAAAAG TAnGCAGGGn 480

AATTCnCCGT TCCGACAGT 499

(2) INFORMATION FOR SEQ ID NO: 784:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 715 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 784:

TATTCATTTC TCCCCTACTC TCTCTGGTTT TTATGGACGT GTATAgaTTA CTgTTTTTTT 60

ATAGGaAAAG AAATGGGCAT TCTTGTATAT TTTGATGTAC AATATAGTAT AGAAAGCGGA 120

TACAAGTAGG AGTGATTTTT TATGAACTAC ACAAATAGTC GAACGTTTAA TCCTGAAATT 180

TTATATGCCT TTGATCCTTG GAACGAAGAG ACCCACCCTT ATAATTGTCA TCATCATGAA 240

TTTCTGGAAA TATCTATTCT TTTAGAAGGT GAATCTGAAT ATATTGTTCA AGGGCAACAG 300

TACCATGCCA CTGCTGGTAC AGTGTTTCTT TTTAATCCGC GAACGGAACA TGGTGAGCAA 360

CAAAAAGCAG GAACTTACTC ACATCAACTA CACATTGGCA TTTCAAATCT CTATTTAGAA 420

GGCTTAGCGC GAAACGTATT TCCAAATAAA TCAGCTTTAT TAGATTTAAG TCATTTACAT 480

GGTGCTTTTT TAGAAAAAGC TTGGCAAATT GTTCATGAAT TAAACCATCA AGAAGTTGAA 540

TCAGCGTTAC AAATTAAAGC TTTAGTTATT GAACTATTGG TCTATATTTT AAGAAGCTTA 600

GCAGTTGATC AAGAAAATAA AATTGAAACT CGGCTATCGA AAACGGAAAA AAGAAAACGA 660

AATTTAGTCA ATCnCAACCA AAAATnGTTG CGATGGTCCA TGGCGAAACG GCGAT 715

(2) INFORMATION FOR SEQ ID NO: 785:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 694 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 785:

ATTAAAATTG GCGCATTTTT CAAAAAAGCC CGTGCTATGG ATAAGCGTTG TCGCTCTCCT 60

CCAGAAAGTG AACTTCCGTT TTCTCCGACC AATGTATGAT AACCGTCAGG TAGTTTTTCA 120

ATAAAGTCTT CGCAGTTAGC TAAACGAGCT GCkTTTTTCA CTTCCTCATC TGACGCTTGA 180

CTATTTCCAA TACGTATATt TTCCAAAATA CTTGTATTAA ACAaAGTTAC CTCTTGCAAA 240

ACAATAGCGA TTTTGCTAAA CAnAGAAGCT GGAGATACAC GTTTAATATC ATACCCATCA 300

ATTAAAATAC AGCCTTCATC ATAATCATAC AGTCTAGACA CTAATTTCAA GATACTAGTT 360

TTCCCGGAAC CACTAGCACC TACTAACGCA tCACTTCTCC TTGTTTAGCC GTAAAAGAGA 420

TATGATCTAA AATTGGGGTA TTGTTATCAT ATGAGAAAGA GACATCTCGC AACTCAACGT 480

CAAAAGATTT CAGAGGACTA TCTGATCCCT CTTGGAnACT GGTTTCCTTC ATAGCTCTAA 540

TTCGCTGTAT TTTAGGnGCT AAATAAAATA TTTCAAGCAC TGCTTCTTTC ATAGAATCGA 600

AGGAATCCTT TAnTTTTATT GCGGCTAGTA AGTAACCAAC GACGTAGAnG TATAGTCACT 660

TCACCCGTTA AAAGTAAGTG AACACCTACT AGAA 694

(2) INFORMATION FOR SEQ ID NO: 786:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1800 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 786:

AAGGAACTGT nAACAGAGCA ATTATTTTGG CTAGCTGAAT CTTTTTGGCC ACAAGTGTTT 60

CCAGAAGAGG AAGACTATTT TTTTGGCGTG AGCGGTGGCC TCACATTTTA TCAAAAACAA 120

AGATTGGCCT CGGCACAACT ATCTATAATT ACGCTTGAAA AAGAAAAGAC AGAGGTTCCA 180

ACAATGCCTC GTTATTTTGA TTACTTGGGC AAAGAGCAAT CCTTGCCAGA AGCCTTCAGC 240

GCTTTATATG ACTTGCTAGG GGACGTAAAT CCTGAATACT ATCTTTCGCA AGCGAGAAGA 300

GTCATTAAAC AAGCTCTGCG AGGCCGCAAA GTCAGTACTG TGCCTAATAT TTTTCAAGAA 360

AGTTTGCACT TGACCCAAGT GATTACGATT GACCAGGACC ACTTAAAACT CCTTTTACCA 420

GTAGCAACGG AACAAGCCGA GCCTTTAGAA GCACAGTCGG ATATTCTAGC ATTTTATTAT 480

GAAAAAATTG CTAACCGATC GGCCATCGAA CGCTTAGTTT TTATGCAACA ATTGATTGAG 540

CAACTAGGTA CGAACTCTCT AAGTTATCTT AGAATAAATT AAAGTTACGA TAATTTTATT 600

ATGTAAACTA AAAATCAATA ACAAGAAATG GCGCCGCTTC TTAGTGAGAA ATGGCGTCAT 660

TTCTTTTGCG AACTTTTTTC GTCCAAAAAC CACCGGAAAT GAACTTAGGC TCATAGGAAA 720

TGATAAATGC TCTCGGCTCT AATTGATTGA TTAATTTGTA CAATGTACGT TCGTTTTTTC 780

TTGGGGAAAG GATTTCTAAA ACCATTCGTT CTCCTTCGCG TCCGTAGGCG ACACTTTGCG 840

TAACACCATA GCCGTGTTaC ACGAAGtGTT tCTGGTAAAT GAAATTGTTC AGTAGTAGAA 900

GGGAGGATAA CCGATACCAT AATATAACCC AAGGCCAAAT AGTCTTCGAT TTTAATTCCT 960

ACACTAATGC CAACAGCGTA ACCAAGCGCA TAAACAAGTA AATTTAAAGG ATTGTCTAAC 1020

CGATTCAGAA CCATACTCAG CCCTAAGACA TAAATTGTTA TTTCAGCCAT ACTAACCAGC 1080

GGTGCAATAA CCCGATAACC TTTCATCGTT AACATAAAAC GAATAGTATT TAAGGTAATA 1140

TACGCAAAAT TAATGATAAA AATCATAGCT AACATTTTCA AATCAACGAC CATAAAATAA 1200

GTACCTCTTT TCAAAATCTT TACATCCAGC CCTCATTATA GTACAGaAAA ACTGAGAATC 1260

AAAGGCGAAA CACTTATAGA AATGAATTCT TTTTTGAATA CTCGCTCATA TTATTCGCCT 1320

AATTTAACTG TTATTTGATA TAATCTTAGT ATGAAGTAAG AAGGAGGTAA GCGATGAAAA 1380

TTTTTGTTTG GATTTTGTTA TTTTTATTAG TAATTAACGT TATTGCTGCC TTAATTACTG 1440

TTTTTAGAAA ACCTCGTAGT ATTTCCAGCG TTTTAGCATG GATGATGACG TTGATTTTCT 1500

TGCCTGGTAT CGGCTTTATT ATTTACTTAT TCTGTGGGCG AGGTATCGAT GGACAAGAAG 1560

TTTTTAAATT GTCTGATGAT GAAAAACAAA TGGTTCAACG AATTAAAGAA AAAGTGGACG 1620

TTGATAACAA AAAAGCAGGG CGCGATAAAC GTTATGATTT ACTTTATGAT GCGAAAGTTT 1680

TAAATCGTTA TTTCCGGAAT ATGGACGCAT CACCGTTGGC CAAAAGAAAT AGTTTGCAAT 1740

TATTCACAGA TGGCCAAGAA AAATTCCAAG CATTATTTGA AGATATTCGT GCAGCGAAAG 1800

(2) INFORMATION FOR SEQ ID NO: 787:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 976 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 787:

ATTTGAATGA TTTGTAAAAT TTTTGGTACA AATAGAATTA GTCCTACTAG CACTGCGAAA 60

CCGGCTGTTA ATAAAACAGC TCCTGCGGCC ATGTCTTTAA CTTTTTTTCC GATTGGATGA 120

AAGTGAAAAT CCGTAAACAT ATCTACGACA TTTTCAAAAA TTGTGTTGAT AATTTCTGCC 180

AACCAGACAA GAAAAACGGC CAATAACAAC CACAGCCATT CAGCAAGTGT TAAGGAAAAC 240

AAGAAACCAA ATACAATGGC CGTGACACCA AACGCTACAT GTTTGCGCAT ATTACGTTCT 300

TCTCGAAAAA CTGTTTTGAT TCCTTGCCAG GCAAATTCTA ATGAATTAAT AAAATGTTTA 360

TTTTTTCCTG TTTGCTCATC TTTTAAGGCC ATACGCATCT AAAaTCTCTT cTGTAAcCAA 420

cATTTyTTTy CacTTCAGGT TCCATATGAT CGTACCCaTT CAAATGTAAA AgCCATGTAC 480

TGCCAAAAAG CCCATTTCTC GTTCCAATGT GTGCCCATAT TCAACTGCTT GTTCTGCTGC 540

TCGCTCTGTT GAAATAATTA AATCACCTAG ATtACGTGGT AGCTCTGCCA ACTCTTCATC 600

ATCAAAAATA ATTGGCAATT CGTCTTCCCC TTCATCCTCT AAAGCAAAAC TGATGACATC 660

TGTAGGCATA TCTTTGCCCC GATAGTCACG ATTAATCATT TGGATACCTG CGTTATCTGT 720

AAAAGTCACC GACATCTCCG TATCTTCGGc AATTTCTAGA AAACCCGCTG CAAATTGCAG 780

CAAGTTTTCG ATTTCTTTGA TTTCTTCTGT TGGTACTTTC TCGGTTTCAT CAATGAAAGT 840

AATATCCATT AGTCATCTCT CTTTTCTTGT TCATCTTTCA TTTCTTCTGC TTCAGCTTTC 900

ATTTTTGGAT ACTTAATCCG TGAATGGAAG GTACTGCATA ATCCATTAAT TAATGATTTT 960

TCAATGATGC GAATCT 976

(2) INFORMATION FOR SEQ ID NO: 788:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 562 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 788:

TTAGGAATGA CCAATTTTGA AAGGGGTTCG ATTGATGAAT TCGATGATAT TTCTAGCATT 60

GATCAATACT ATCGAGCAAT AAAAGAAGGA TTTACGCCCA AAGAAGCCTT GTCACTTGTC 120

AATTTGCGTA GTCGAGACAA TGCCCGCACG CCTTTTCCAT GGAATGATTC TATGTATGGC 180

GGTTTTTCAA GCGTCAAGCC TTGGTTAGGA ATGGTGGATA ATTATAAAGA AATCAATGCC 240

GAAGCAGAAA TAAAAAATAG TCAAAGTATT TTTCATTTCT ATAAGCGGAT GATTGCATTT 300

CGTCaAAAGA GTCCCTATAC AGACATCTTA CTTTaTGGrA CATTTGAGGG tCTGTCaAAT 360

CTTCCmGaTA ATGtTATTGC tTATAAACGC mAATTaAATG aAAAAACGAT TTACGCCtTT 420

TTTAATTTkG GkGAGGcGGk tCCaATCCCy CmTTAtTTAG kTAATGGCAC AGTTATTTTT 480

GATACACAGT CTGAAGAACG AGAGCAATTA GTTCAGCAGG GAACATTATT AAAAAGCTAT 540

CAAGGACTAT TGATTACTAT AA 562

(2) INFORMATION FOR SEQ ID NO: 789:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 981 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 789:

CAGTATTAAA AGTACTTnTC AATTACACAC TTAGTGTnCC AGTTTTTTTG GGTCTTTTCA 60

ATAAAAATGA TTCTTTTTTC AtCGTTTTCT TtCAAaTCyA AACAAAGCAT TCGTTTGTTA 120

GAAACAATAG CTACTTATTT CCGAATAATT AGaTTTAATT AAACCAAAAG CAAAGGAATT 180

CACGAACGTC TTTcAATTAA CTTCTTGTTT AAATGGGAAT TTCAGGTCCT CTCACAAAGC 240

ATTTTTTACT TATTTTTCAA AAAAAACAAC GAAAGAATAG ACACAGCCAT TCTCTCGTTG 300

CTTCATATCC TTTTATTTGT AATCTTCTTC GTAAAAACGT CCTAAAATTT TAACATTATC 360

TGAGATGCTG ACAAAGGCAT TCGGATCACT TTCTTTCATG GCTGCTTCTA AACTTGGCAA 420

CTCAAAGCGT GTCACAATAG TTAGCAGAAC AGTTTGGCGG TCGTGGCGAT ACGnCCCTTC 480

CGCTTCATGA ATAATTGTAA TACCGCGGCG CATTTTCTTT TGAAtCTCAT CCACGACTCG 540

ATCTGGGkTT TTAGTGaCAA TCATAACTTG CATTTTCTTT TGCTTAGTAT AAACGGCATC 600

TGTGACCTTC CCACTTACAA AAATTGATAA CGCACTGTAG AACATGTACT GCCAGCCAAA 660

TAAATAACCT GCAACGAAGA CAATTAGTGC GTTAAAATAA ATTGAAATCG AGCCAACGGA 720

ACGCCCCGTT TTTTTCCGAA TGGTAATGCT TAAAATATCT AAGCCCCCAG AAGAAACCCC 780

GTTTTTTAAG ACAAAACCAA TTCCTGATCC AGAAATCGCC CCwCCwAAAA TAGCACAGAT 840

AATTGGATCT TGGGTCATCa CCGTTTGCGG AACAATATGC ATAAATAATG AAGTTAAAGT 900

CACGGTAATG AAAGTAAAAA TTGTAAACTT CTTACCTATT TTTTTCCACG CAATAAAGAA 960

TAACGGCACG TTTAAAAGAT A 981

(2) INFORMATION FOR SEQ ID NO: 790:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 432 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 790:

TTTTTCGAGA CGATACAGTT TATAGTTTCG TCCTGTGTTA CTAATGAGAA TCGAACTAGC 60

ATTTTTGGGT ACATACTTAA CATTCTAATC TTATCAACTA CTCCTACATA ATTAGTTTTC 120

ATACTATCGC CCTCCTAGTA ATATAATTAT ACCGAATGTA TGTTCGTGTT GTCTATAGAA 180

AAATTTTGTA AATCTATGAA ATGATAAAAA CAAATGGTGA TACATTTGGT GATACATCTG 240

AAATATCTAT GAAAAAAATG TAAATAGTGA AAGACTTCAA TAACTAATAA ATCATTGATA 300

TATAAGTAAT CGGAAAGAAT GAAATCTATG GAAATGTTTa TCTAACAAAC GCAACATGTt 360

CGCtTATTAG GTAAACCAGG TTTgAAGAAT TAGCmAAAGA TTTAAATGCA CGTCTATAAT 420

TgTAAtAAnA GA 432

(2) INFORMATION FOR SEQ ID NO: 791:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 533 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 791:

CTGAAAAAAG CAACACTAGA AGTTTTGAAT TTACCTGCGT TGACAGGTAT AACTGATGAT 60

AATTTACGTC GTCTACTCAA CAATCTAATT ATTGAACTGT ACAAGTACCA GGCGGAAGAA 120

GAACGTCGAT ACATTCGTGA AACGCAGCAA CAGGGAATTA CACTGGCTAA AGCAAAAGGA 180

AAATATAAAG GTGGAAAGCC AAAGTACcGA GAAAATGATC CACGTCTTCA GTTAGCATTT 240

AAATTGTTTT TAGACGGATG CACAGATAAA GrAGTAGAAC AACAGACAGG GATTAATAGA 300

AGrACTTTTA GrAGATATAG AATGAAATAT GGAATAACAA ATTCTCTTAG ACATAAGAAT 360

GATTAAAACT ACTAATCCAA TGTAAAACAG ATAATATATT AaAGAGGGCT ATAGAGTAAA 420

ATGCAAAAGA AGTTTGAAGC AAAGCAATGC TCAGTAGACG nGTTACTAAA GGCGAnATAA 480

GCGATTATAT AAGGTTTCGA ATTTTCAAAG AACATATGCA TGGAAATATA AAC 533

(2) INFORMATION FOR SEQ ID NO: 792:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 287 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 792:

CGAGTTTTCG CTGGCGCGGG CCACCCTGCT GCTGGATCGC GACGCCAGAT CGGAGGTCGA 60

ACAGGGCGTG CGCGATGCCT GGGACCTGGT CGGCGGtTCG CGtGCGCAAC TGGCAGAACT 120

TCCTGGCCTG GCAGTTCCAC TACGAGCACT ACATCCTGGC GCGCCTGTTC GTGAACCGCC 180

GCACCGCGCG CGTGCTGGAC TTCGGCCATC GCGAATTCGG CGAGAACCGC ATCCTGCCCG 240

ACGAGGAGGC GCATCGCATC CGCATCACGC AGTGnTGGCT GCGCAAC 287

(2) INFORMATION FOR SEQ ID NO: 793:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 379 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 793:

AATAAAATGG CCATAATAAC AAACGCAATG ACTAAATATA AATTACTATC TTTCAAATAA 60

TTTTTCATAT AATCCTCTTT TCTTGGTTTC AAGTTGCCCC GTGCCTGCCA CGGAAAATTC 120

TTTTTTATTA TAGTCGGTTT CGAGGGAAAA AGAAACAACT TCTCTTAGGT CTTATGATAT 180

ACTGAAAAGA AAAAAATGAA AGAAGGATTG GATATGACCT ATTCTATTAC ATGGGATTTA 240

GACAGTATTT TCCCCAAATGG GAGCGAGTCC AAAGAACTTA GCCAACGAAT GACTCAATTA 300

ACCGAACAGA CAAAAGAGTA CCATCAATTA ATCACAGAAT GGACACCAGA AAGTGCCCCA 360

nGCAAnGAAC AATTTGCTG 379

(2) INFORMATION FOR SEQ ID NO: 794:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 906 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 794:

CCAAAATTAT CCGACTACGC CATAGATCAT TTAAGAAGGA TAATTTTTGG AAGGGAAAGA 60

GAATCATATA TGAAGTGAAG GAGGTCAGGA CAGAACCGTT TATTTTCGAA GGAATTGGTT 120

CTGTTCTCGC CATTTATCAG GTTTTGAaCC TGGAACAAAA ATCCAAAGTG ATTTTTGTCC 180

CAGGCTCTTT CGCTCTTTTA TTGAGTAAGA AGTCAGTTTA GAATTTATAT TAAGTTTCGC 240

TGTACATTTT AAAATTATCA TCAACAGAAA ATACAACAGG CGAATTTTTA GATAAAAAAA 300

TTAGTAAAAC AAGTATTAAA TAAATAGTAT TTTTGAGGAT AAAGTAAGAG AAATGTTGTT 360

TTTTTAATAG TTGTTAACGG AAAACAAAAT TATTTTATAG CCATAAAATA TATTTTTATT 420

ATAATTATAT TATAAmaymm tGaGCATATG ATTTGCAGAT AAGTTTTTCA GGGCCTATTC 480

TGAATAGGAA CTTAGCTATA AAAATAAAAG AGGAGAGTGG TTTTGATGAG AAAAAATCAG 540

AAGATTGTGC CAATTGTGAC AGCTTCAGAC GAAAATTATG CTCCTTATTT GAATGTAATG 600

ATGACAACTG TTTTAGAAAA TTGTCATGCG GAGAGACCTG TTCATTTTTA TGTGATTGAT 660

GATGGGCTGT CGCTATCTAG CAAAAAAGCA CTCCGAGaAA CGGTTAGTAG TAACTCACAA 720

AATGCAACTG TTGAATTTCT AACAGCAGAC AAAGAAGTCT ATCAAAATTT TTTAGTGAGT 780

GATCATATTA CAACGACCGC ATACTTACGC ATTTCTTTAC CGTCACTTTT ACAAAAATAT 840

AGCTATAAAA AGGTTCTTTA TTTAGATGCA GATACGCTGG TTTTAGATGA TATTGTCCAG 900

TTGTAT 906

(2) INFORMATION FOR SEQ ID NO: 795:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 232 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 795:

GGGCAAGAAG cATCCAAGTC TTGATTCGCC GAGGAGCAAG TCTTGAAGAT GCACAAGATG 60

CAGTGTCGCA AACGTATACA ACGATTTTTT CAATATTACC TGAAATTACA TCAGAAAATC 120

TACGCCCGTG GTTTTTCCGA GTAACATTTA ATTACTACAT TACGATGTAT CGAAAAAAGA 180

AACGAGAACG CACTTTTGTG GCAAACAGCC ATCCTGTGCA ATCAATAGAA AA 232

(2) INFORMATION FOR SEQ ID NO: 796:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 600 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 796:

TTTATTACAT ATCAAGATTC TATCACCTTT ACAAAGAATG TTGCATCTTG GTATACAGCT 60

GTATCTTCGG TATTACCATA AACTTTAGCA GCTGTATGAA CATCAGTGAT TAATGTATTT 120

TCGTATAGTt CTTGkGCTGC ACTTACTCCA CCGCTATTAT GCCCATCGAA tAGaCCTTtA 180

ATTTCTACTT CTACTGTTtC ATCTGCACCT TtCTCATTAT CAGCATCAAA TAAATTAGAT 240

TTaTTTTTAT CTTATCTCCA ACTTTAAGGT TATTTTTTTC AGCTAAATCT TTATGCATTA 300

AGATTTTATT TTTATCTTCA TTTTCTAAAT GCTTTCCTTC TACTAATTTA TATGCTTCTG 360

ATACAAATTT TGTTTCTTTT GTTGAGTCGT TAACCCCAGT TAACATAACT GTTCTTTTAA 420

AATTCTTTGC TCTTTCAGGk GATTGaTTcG CAAGAGTCTC TTGAGTTTCA ATAATATCAT 480

AATCAACTAA ATCTGCAACA CTGTTTATTC TTTTTACATA GGAGTCTATA CTATCTGTTT 540

GAGATATCTT TTTAATATCT TCACCTTTTA CATTTCCCCC ACCTCTAGGT GTTCCTGGAT 600

(2) INFORMATION FOR SEQ ID NO: 797:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 636 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 797:

TTTTGGAGGA CAAGATGATT CGTATACTCG TACTATTACT CGCCGTTGTC ACAGGCGTAC 60

cAGTTATTAT TTGATGAAGA AAAGTGCAGC TTTTTTACCA CTCTTAAAAA AAGAAACTGC 120

AACAGAAAGT CAACAGTTTA TCGAACGTTT TGGGCGTTAT TACTTGATTA TCGCCATTCT 180

AGGTGTTCTA GCGGCTATTT TTAATCGCCC GTTATTGTCA ATTGGTTTTA TCTTTTTTGT 240

TTTATTACTT TCAACGTTGT TTAGTTTAAC TTTTGCAAAA AAAATGTCAT AAAAGACGAT 300

GAATTTAAAG AGGTCGGGAC AGAAGTGTTT AACTCCGAGA AATAAGAAGG AATTTCCGAA 360

AATTGTTCTT TAATTTTTGG AGAATTTCGG CTTATTTCCG AAGaGTTGCT TCTGTTTCCG 420

CCGTTTATCA ATTTTTAAcG GTGGGACAAA AATCCAAAGT GATTTTTGTC CCAGGtTCTT 480

TTTTTACmAA AAAAATGCCA ATATTCAGaA AAAAAAGAAG CGGnTTCAGT TcCkTTTTtC 540

nTTGnTTGnT TTACnAttAT aTngAAmCAt TAAtGaTTtT GTTaCGCTGA AAGGAGCGGT 600

TATTATGTTA CAACAATATC AnAAAATTAT GATAGC 636

(2) INFORMATION FOR SEQ ID NO: 798:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 658 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 798:

TGGGTAATTT TTAGGTTTCA CTTCGAAACG ATTATTGACA TACTTTCGTA AATACGCACC 60

TGTTTGTAGG GAAAGTGTTT GTTCCTCTTT AAAAGATAGA AGATTTTTGA CTGTAAAATT 120

AACTAGCATG CTGTTCACTG AGAGATCCCC TCATAATTTC CCCAAAGCGT AACCATGTGT 180

GAATAAATTT TGAGCTAGTA GGGTTGCAGC CACGAGTAAG TCTTCCCTTG TTATTGTGTA 240

GCCAGAATGC CGCAAAACTT CCATGCCTAA GCGAACTGTT GAGAGTACGT TTCGATTTCT 300

GACTGTGTTA GCCTGGAAGT GCTTGTCCCA ACCTTGTTTC TGAGCATGAA CGCCCGCAAG 360

CCAACATGTT AGTTGAAGCA TCAGGGCGAT TAGCAGCATG ATATCAAAAC GCTCTGAGCT 420

GCTCGTTCGG CTATGGCGTA GGCCTAgTCC GTAGGCAGGA CTTTTCAAGT CTCGGAAGGT 480

TTCTTCAATC TGCATTCGCT TCGAATAGAT ATTAACmAGT TGTTTGGGTG TTCGAATTTC 540

AACAGgTAAG TTAGTTGCTA GAAyCCATGG cTCCTTtGCC GACGCTGAGT AGAnTTTAGG 600

TGACGGGTGG TGACmAtGnA kTCCgTGTCG AGCGCTGAtT TTtCGGsCTT TAGAGCGA 658

(2) INFORMATION FOR SEQ ID NO: 799:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 826 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 799:

TAAAATTAAT ATTTACTAAA TTAAGATTCT TCAATAATTT TTTTGCTTCA TCTACAGTAG 60

AGCATTGGCC CAATACCCAA GGAATAAACT CAAATGGAGA AACATTTTCT TTTCCTTCTT 120

CAATTTTTTT ATAATCTGCA TAGCCTGAAA AGTTTAATCC AGCCATTCCT AATCCTTTTT 180

CATTTATTGC ATCATAATAA AGCGGATAAT CAGCAATCCC AGCAGCAATT CCAAyTATTG 240

CAAAATGATG ATCTAAATTT CCAACTTCTC GAAATGAAAA CTTATAATTT CTCGGCGTAA 300

TAGTAACCAC CTCATTATAA GAmATTTCAT AATCAAAATT CCTTCCAAAG TAATGATCTT 360

TTGATACATA AGTAATTGCT GTACACATAA TCAATTCCTC CTATATACTA ATTCAGTCAC 420

TAAAATAGTA CACCCAATAC TCTTTTTTAT CAAATTTAAC CTATCCACAA AATGGCTAAA 480

CAATATATGA ACATTCAATT TGGTTCTGTT GCAAAGTTTT AAATCTACTA TCAAATAAGG 540

TAGAATAATA GAAAAAGATA GCAGGAGGAA TGACGGTGAA TCATTTTAAA GGAAAGCAAT 600

TTCAGTAGGA TGTGATTATT GTAGCCGTGG GCTACTATCT TCGTTATAAC CTTAGCTATC 660

GTGAAGTTCA AGAAATCTTA TATGATCGTG GGCATTAACG TTcTCATACG ACGATTTATC 720

GTTGGGtGCa AGAATATGGc CAAcTACTCt AtCCAAtTTG GGAAAAAGaA AAATAAACma 780

TCCtTTTAwT CccTGGGAAA tGGGATGaAc gkACcTCCaA ATTAnA 826

(2) INFORMATION FOR SEQ ID NO: 800:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 413 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 800:

AAACCAGATT GGTCTGCTCT GACTCAATCT GGACACAAAA TCATTATTTA TTTAATTAGT 60

TTTGTTACTT TATTGATTTA TTGGAACAAT CATCATCATT TATTTCAAAG TGTCCGTAAA 120

ATTGATGGAC GTGTCCTTTG GATGAATAAT TTTCTAATTT TGACGTTGAC GTTTTTCCCG 180

TTTGTTACTG GTTGGGTGGG TGAACATCCG TTTTCTTGGC CGCCACAAGC TTTGTATGGC 240

TTAGTAGTAT TmGGmGmTGA TGTTGCTTAT TATCTAGTGG TGCGGGCATT GATTCAGGCA 300

AATGGGCCCA AATCCGAAGT ACAACGkTTG TTkGGCCACT AkCCAAAATT AACGCTTtTCG 360

ATTGTGTTGA ATATTAnTGC GnTGATTGTG GGTAAAATGG TAGCACCAAT CGC 413

(2) INFORMATION FOR SEQ ID NO: 801:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 414 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 801:

TCTTCATTCA CATCAAATGC GCAACTTACA CGAATACTGT TGCTTTGATG GAATTTGTAC 60

TTTAATAAAG CACTTCCTAA ATTACCAACC CCGATTAATG CGACATTGGT TAACTCATCT 120

TCATTCAATG TTTTTGCAAA AAAGTTCATT AAGTTTTCAA CATCATAGCC GTACCCACGT 180

TTACCTAATT CACCAAAATA TGAAAAGTCA CGACGAATAG TCGCACTGTC TACTTGTACC 240

GCTTCACTTA ATTCAGTTGA TGAGACTTTA TTTTTTCCTG TATCATGTAA CATTCTTAAG 300

TAGCGATAAT ATAAAGGAAG GCGTCGCGCC GTTGCTTTTG GAATAACTTG ATCTTTCACA 360

ATTGGACCTC CAATACTTtC ACTTAAATCk TATCCAtTAA CATAATAGCA GTTT 414

(2) INFORMATION FOR SEQ ID NO: 802:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 749 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 802:

GCCAATTTTT ATAGAAAAAA AACAAAGAAG AATAGAATAC TGGCTAATTG CTTTTnTAGG 60

AATTTTATTT TTAACAACAT TTATTAATCA TACGTATGCT TTTTCAGAGA ACATTAAGTT 120

GATCATTTGG CAATGTATCT TTTTCTTTTC CGTTTACGAA GTAGGGCGTT CGAATGATAA 180

ATCTGTTTTT AAAGCTTTTG AGTATGTACT TTTAATTGTT TGGACCGCAT TAGTCATCGT 240

TGGGTTATAC TTATTTTTTG CGCGAATCAG CTTTTCAAAA CCAGTAGAGT CTTTGTATTA 300

CGGAATGAGA ATCGGTTTTT TTGAAAATCG ATTATATGGT GTTTTgTGGA GCCTAACTAT 360

GCATGTACCA TCTCATTAGk TtGTATTTTG GTTGGCGTGC GGAATTTtAT TAATACCAAT 420

AGCAAATGGT tAAAAGTGGT TtGTGCAATG GTGATTTTtC yACAATTtAG TTATGTAGCA 480

TTGTCTGGAT CAAGATCAGG AGTTATTCAG TTGATTTCAA TGGTAATTTT CGGGATGTTC 540

TTTACAATTT GGTTCTTTCa AAAAAATACA TCCAAAGCAG TTATAAAAAA AGTTGTGCAA 600

GCAGTATTAG TTTCTTTTTT CTGTGGTGCT ATTGTTTTCG GGGGCCTTAC AGTTAATAGA 660

AAAAGGGTAT ATTGCTACTG CAAATTCAGT GAATATGGAA GTACCAAAAG TTTTAGGATA 720

ATAAAGGAAA ATGGAATTCn AGGGAAATA 749

(2) INFORMATION FOR SEQ ID NO: 803:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 624 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 803:

AAGGCTTCGT AGGTTGTATC AATCAGCTTG GCACCAGCAT CCACAATCAT GGTGGTACCG 60

GCTACTTTGA TTTCTTTCAT TTCTGTTAGC ATAATCACGA CATCGCGAAT ACCGCCATCA 120

CGAACAATCA AGTTGCTAGC ATTGCCTAAT ACTAACCATG AAAGTCCTTG TTCTCGGCAA 180

TACGCCACAA TCTCTTCAAC TTCTTTTTTT GTTTTGGGTA AGGCCAACAC ATCAGCTGGG 240

CCGCCTGTTT TGGTAAAGGT TACGTTTTTT AATGGTTCAT CTACTAATAA AGTAATTTCG 300

TTTAATGTTT CTAACATAGC TTTAGTGTTC ACGAAAGTTA GTCCTTTCTT TTCAAAATAT 360

CTGTCTTATT CTAGCATGAT AGAAAGAAAG TTGCTAGTCC TCTTTAAAGA ACATGCCGTT 420

GTTGTTGGCG TACATTTCTT CACATAACGC ATGACTTTCA TAGACTTTTT CTTGCTTTAA 480

CTCTTGCTCT TGCGGTTTTT GAATGGCCTG AATGAAAGCA GTAACCATTT GGTAAAAACC 540

TCTTTTTTCC AATGTTGGCG TCCAATCTCC CATCTCCTTT ACCTGGTATC CTCTTCCGGT 600

TGGAGGGCAA ATGCGTAAAT TTTC 624

(2) INFORMATION FOR SEQ ID NO: 804:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 570 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 804:

TTTTTAGTCA TTTACTCCCT CCTTTTCGAG CAGTACTTAC CTTACTGTAG CATAGTTTTT 60

GATTGGCGCC TATCAGCAAT TTATCTCCAG ATAATAGAAA AGCCCCCCAG CGAAAGTGrA 120

TGGCTTTTCT ATTGTTAATT CTTTCTGATT TgTAGATAAC TTACACCTAA AACGGCAAAT 180

GTACTTGCCA GTATTCCTGC TATTAGCATA AACGGTGTTT TTGTTGTGCC TGTTTTGGGA 240

AGTTCTCGTT GCCCGTTTGC TTCATTTAAT TTTTGAATTG GTTCACTAAG CAGTTGGGGA 300

GCCAACTGTG GGCTTCCTTC ATTTTTCAAT GTATTTAATG GAGCGTTTAG TATTTGTGGT 360

TGATTTTTTG ATGTAACATT TTTATTTTTC TCCACAGTTG TTTCTGTGGA TGGCTGTTTT 420

TTCTCAGTTG TAGTAATTTC TTTTGTTGGT GAGACATGAA CTTCTGGCTG GCTGCTTTCT 480

GCCGTTcCGT TATACCATTC TCTTCAGTgT kACGCTTGGt TnTCAGGtAt GATTGGTGCA 540

ATCGTGCCTA GTGTTCACTT CCATCTGGAT 570

(2) INFORMATION FOR SEQ ID NO: 805:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 463 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 805:

TGCATTTTGC CTGGGACAAT CCAGACGAGG ACTTGACCGG GTATTTCCGG CGCTTTCTGG 60

AACTGACCGC CATTAAGAGn GACCGAAAGC GGCGCGTCTA CGTCCTGACC AATTACGGCA 120

GCACCCATGA GCAGGACTTG TACCGGGTGG AAACCCTGCG CGGCATGrGr TATGACCCCT 180

ATGTGATGGT CTACGACCGC CCGGCCGCAC CGAAAATCAC CCGCCAGCTT CAGCGCTGGG 240

TGAATAATAA GCGGATTTTT TACACGGTCA AAAACTTTGC CGATTACATT CCAAACAGAA 300

CAGGAGGGAT AACGTGAAAA TCGACCGATT AAAAACCGAC CTGGCAAAAG CCAGGGAAAA 360

AGCCGCCGAG TGGCAGGCCC GCGTGCGGGA CTTGGAGAAG CAGATCACCG AGCAGGAAAA 420

CCTTGAAATC TTGCAGGTGG TACGCAGCGT AnGGCTTTCG CCG 463

(2) INFORMATION FOR SEQ ID NO: 806:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 611 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 806:

AAGnATTnCC GGACGCCGAG AAATTAATTT TAGAAGATGC GGGTGTAATA CCGTTACTTC 60

AAATTGGAAA TGCTAAATTA AGAAATCAAA AAATTTCTGA AATGAAAGTA CACTCTATTG 120

GTGCTAAGTA TGATTATAAA ACCATGGAGA TAAAATAAAA AACCTCTCTT AAAAGAGGTT 180

TTTTTATTTG TCAGATTTTA AATAGCCAAC ACCTTCAAAA GTATCAGCAA TTATATTTAA 240

AGAAGAGGGA ACTTCAAAAT TTTCTTTTCT AGCTATTTCT GTTACTTCTT TAACTAACTC 300

TTTATAAGTA TCATTTTCAC CTAATTGAGA GAATAAATTT ACGATGTTTA CAGCTTTCAA 360

ATATTCCTCC ATATTTCGTT TTTTCAAGAA ATTTCTGAGA TGTTTTAAAT AAATCATttC 420

TAAATTAATA gTaCCATCTA AAATAAAAGA AGGGACTTTT CGAACAATAT CAAATTGTTC 480

AAGGAAGTAA TCAGCTTTTT CAAAATCCTT TTTTTCAAGA AAGACAGTAA CCGCATTCTT 540

GATACATAAC TGAACTGCAT GGTCATGGAT ATCTCCATAA GATTtGGTTA ATtGGGAAAg 600

AGTTTTTCCT G 611

(2) INFORMATION FOR SEQ ID NO: 807:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 552 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 807:

AAATGAAGAT AAATTTACTC ATTGGACCCC TAATGTTTAT AGTTATGGTG CATATGCTGA 60

TGATAATAGA ACTATTGTTA AAGGCCACAA CGAAAGAAAC TTGCAACAAA TCAATACTTT 120

TGTAATCGAT TTTGATCGCT CACTAGGTGA AATATTAGAT TCTCAAATGA TTCTCGATGC 180

TGCAATTGAT CTGGAACTGA TGCCAACGCT AATTTTAGAA ACTCCAGGGG GATTTCAAGC 240

GTATTTCATT CTTGAAAATG CCTGGTACAT TTCTTCAAAA AATAATTATC AATCAATCGA 300

AGTGGCCAAG AGAGTATCTG AAAATTTAAG AAAAGCATTT GCTGAAGTAT TGCCTTCGGT 360

CGATTTGGGG TGCAACaTTT TGGTATTGCA CGTATTCCTA GAACAGACAA TGTAGTCTAT 420

TACTACCCAG CACTGACACA TGATATGCAA CAACTAaTCC AGTGGTCCAT GAAATTTGAG 480

GCCAAAAAGA ACACCTnAAA AAACCGGAAT CTTAATGTTG GTTGCnGGTA AAGATGGAAT 540

CnGGATTTAA AG 552

(2) INFORMATION FOR SEQ ID NO: 808:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1217 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 808:

GGCAGTAGAT ATTTTAACAT CTGCAAAAGT CAACGAAACT TCTAATATAG TAGAAGCAAA 60

AATTGCGGAC GAACAAGTAA CTGGCAAATT AACCGCAACA AAAACCGTTA ATAACGCCAA 120

ACCAAAACTA GGGGAAGAAA TCGAATACAC CATCAGCTTC CGCAACACTA TCGAAAACGG 180

CATCCTAAAC AAAGTAGTTA TCACAGACCA ACTGCCAAAA GGACTGACAT ATGTCAAAGA 240

TAGCCTAACA AGTGTTGGTG ATGAACCAAA GCCAACCAGC TTGAAAGAAA CCAACGGCAC 300

AATCACAGCG GAATACCCAA GTATCACCGA TATGAAAGAA CGGAGTATTC GCTTCAAAGT 360

GATCGTTAAC GAAGAAGCCA AAGCGGGCGA AACAATTCTC AACAAAGCCA AAGTAGATGA 420

CACAGTTAAT CCACCAGAAG AACCAGAGGT GCCTGTGGTA CCAGAAACGA ACGCAGGCAA 480

ACTAGCCGCA ACAAAAACCG TTAATAACGC yAAACCAAAA CTAGGCGAAA CTATCGAGTA 540

CACAATTAGC TTCCGCAAyA CCATCGAAAA CGGCGTCCTA AACAAAGTAG TTATCACAGA 600

CCAACTACCA AAAGGACTGA CGTATGTCAA AGATAGCCTA ACAAGTGTCG GTGATGAACC 660

AAAACCAACC AGCTTGAAAG AAGCCAACGG CACAATCACA GCGGAATACC CAAGTATCAC 720

CGATACGAAA GAACGGAGTA TTCGTTTCAA AGTGATCGTT AACGATGAAG CGAAAGCGGg 780

CGAAACAATT CTCAACAAAG CCAAAGTAGG TGATGGCATC AATCCACCAG AAGAACCAGA 840

AGTGCCCATT ACGCCTGAAG AACCAGCGAA AAATaAAAAG GAAACAAACA AAGTTGTTAC 900

AGACCAGAAC AAACCTACAA AAAATAGTAA AAATGAAATT GCCATAAACA AGAAAGAAAC 960

GTCTAAATCA TCCTATTTAC CAAAAACAGG AGAAAAAGTT CAAAAAATAT TTGCGTATCT 1020

```
GGGTGTAGGA CTGATCTTAA TTGTATTAAT TCTCTATGTT ATAAAGAGAA ACAAAGAAAA    1080

AGAGGAGTAA ACAAAATCGA TTGTTTTAGA ACGACTAATT AGCTGAAAAT TTAnGnGTTG    1140

ACGACATAAG TCAAAATGAC CTATGTCGTC AACTCTTTTT GGATTTATAA CAAGGAAATC    1200

CAACATATTG TGTGTGC                                                   1217
```

(2) INFORMATION FOR SEQ ID NO: 809:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2405 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 809:

```
GCGGCGAGCG AAcGGaAGAG CCCAAnCCAA CAAGCTTGCT TGTTGGGGTT GTAGGACTCC      60

AATATGGTAG TCTGTTAGTA TAGTTGAAGG ATTTGGAAAA TTCCGCTAAA GAGGGTGAAA     120

GCCCCGTAGA CGAAATGCTG ACAACACCTA GGAGGATCCT GAGTACGGCG GAACACGAGA     180

AATTCCGTCG GAATCCGCGG GGACCATCCC GCAAGGCTAA AtACTcCCTA GTGACCGATA     240

gTGaACCAGT ACCGTGAGGG AAAgGTGAAA AgCACCCCGG AAGGGGAGtG AAATAGATCC     300

TGAAACCGTG TGCCTACAAC AAGTCAAAGC TCGTTAATGA GTGATGGCGT GCCTTTTGTA     360

GAATGAACCG GCGAGTTACG ATTGCATGCG AGGTTAAGTC GAAGAGACGG AGCCGCAGCG     420

AAAGCGAGTC TGAATAGGGC GAATGAGTAT GTAGTCGTAG ACCCGAAACC ATGTGATCTA     480

CCCATGTCCA GGTTGAAGGT GCGGTAAAAC GCACTGGAGG ACCGAACCCA CGTACGTTGA     540

AAAGTGCGGG GATGAGGTGT GGGTAGCGGA GAAATTCCAA ACGAACTTGG AGATAGCTGG     600

TTCTCTCCGA AATAGCTTTA GGGCTAGCCT CGGAATTGAG AATGATGGAG GTAGAGCACT     660

GTTTGGACTA GGGGCCCATC TCGGGTTACC GAATTCAGAT AAACTCCGAA TGCCATTCAT     720

TTATATCCGG GAGTCAGACT GCGAGTGATA AGATCCGTAG TCGAAAGGGA AACAGCCCAG     780

ACCACCAGCT AAGGTCCCAA AATATATGTT AAGTGGAAAA GGATGTGGGG TTGCACAGAC     840

AACTAGGATG TTGGCTTAGA AGCAGCCACC ATTTAAAGAG TGCGTAATAG CTCACTAGTC     900

GAGTGACCCT GCGCCGAAAA TGTACCGGGG CTAAACATAT TACCGAAGCT GTGGACTACA     960

CCATTAGGTG TAGTGGTAGG AGAGCGTTCT AAGGGCGTTG AAGGTCGATC GTGAGGACGG    1020

CTGGAGCGCT TAGAAGTGAG AATGCCGGTA TGAGTAGCGA AAGACAGGTG AGAATCCTGT    1080

CCACCGTATG ACTAAGGTTT CCTGGGGAAG GCTCGTCCGC CCAGGGTTAG TCGGGACCTA    1140

AGCCGAGGCC GATAGGCGTA GGCGATGGAC AACAGGTTGA TATTCCTGTA CCAGTTGTTT    1200

TTGTTTGAGC AATGGAGGGA CGCAGTAGGC TAAGGAATGC ATGCGATTGG AAGTGCATGT    1260

CCAAGCAATG AGTCTTGAGT AGAGTTAAAT GCTTTACTCT TTAAGGACAA GTTGTGACGG    1320

GGAGCGAAAT AATAGTAGCG AAGTTCCTGA TGTCACACTG CCAAGAAAAG CTTCTAGTGA    1380

GAAAACAACT GCCCGTACCG TAAACCGACA CAGGTAGTCG AGGAGAGTAT CCTAAGGTGA    1440
```

GCGAGCGAAC TCTCGTTAAG GAACTCGGCA AAATGACCCC GTAACTTCGG GAGAAGGGGT 1500

GCTGACTTCG GTCAGCCGCA GTGAATAGGC CCAAGCGACT GTTTATCAAA AACACAGGTC 1560

TCTGCAAAAT CGTAAGATGA AGTATAGGGG CTGACGCCTG CCCGGTGCTG GAAGGTTAAG 1620

AGGATGGGTT AGCTTCGGCG AAgcTCAGAA TTGAAGCCCC AGTAAACGGC GGCCGTAACT 1680

ATAACGGTCC TAAGsTAGCG AAATTCCTTG TCGGGTAAGT TCCGACCCGC ACGAAAGGCG 1740

TAACGATTTG GGCACTGTCT CAACGAGAGA CTCGGTGAAA TTTTAGTACC TGTGAAGATG 1800

CAGGTTACCC GCGACAGGAC GGAAAGACCC CATGGAGCTT TACTGTAGTT TGATATTGAG 1860

TGTTTGTACC ACATGTACAG GATAGGTAGG AGCCGATGAG ACCGGAACGC TAGTTTCGGA 1920

GGAGGCGCTG GTGGGATACT ACCCTTGTGT TATGAACCCT CTAACCCGCA CCACTAATCG 1980

TGGTGGGAGA CAGTGTCAGA TGGGCAGTTT GACTGGGGCG GTCGCCTCCT AAAAGGTAAC 2040

GGAGGCGCCC AAAGGTTCCC TCAGAATGGT TGGAAATCAT TCGAAGAGTG TAAAGGCAGA 2100

AGGGAGCTTG ACTGCGAGAC CTACAAGTCG AGCAGGGACG AAAgTCGGGG CTTAGTGATC 2160

CGGTGGTTCC GCATGGAAGG GCCATCGCTC AACGGATAAA AGCTACCCTG GGGATAACAG 2220

GCTTATCTCC CCCAAGAGTC CACATCGACG GGGAGGTTTG GCACCTCGAT GTCGGCTCGT 2280

CGCATCCTGG GGCTGTAGTC GGTCCCAAGG GTTGGGCTGT TCGCCCATTA AAGCGGCACG 2340

mgAGCTGGGT TCAGAACGTC GTGAGACAGT TCGGTCCCTA TCCGTCGCGG GCGTTGGAAA 2400

TTTGA 2405

(2) INFORMATION FOR SEQ ID NO: 810:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 537 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 810:

AGAATATGCT AAAGTATTAA GAAAGAATAG TCTGGAGTTG ATAAGAATGA ACTTTAAAGA 60

AAAAATTACA GAAATGGTTG AAAATGTTTA CGATTTAGCA CGCGATGGTA AGTATGATGT 120

CAAAGTTGAC TTAATTCCTC GTAAAAAAGG CAACATCAAA AGTGGTGTTA AAGTCACTTC 180

TTCAAGAAAT ACAGCCCACA ATATCAAATG GCGTTCAAAA AACTAAkTCG TCTTTAACAG 240

AAACACCGGA GCAGACAAAA TTATTGTCTA CTCCGGTGTT TTTATTTTTC TAATTCGATA 300

CCTAAGACAG TTAAATGCCC ATTACCAACT AATTTGATAA CTAATTGGGC TAAGTGGTTG 360

GGTGTCATAT CTGTTTCGCT nGTGGTCCAC CAATGAAGGG TACCGATGAA GATAGAGGGC 420

ATGTATTCAA TnACAAAATC nTATCGGnAC ATCAATAATC ATTTTCGGTA ATACGCAACT 480

GCGGAAAAAA ATCGTCGCCA TATTTTCCTT CGAnAATTTC AGCTAAGCGT TTTCGAA 537

(2) INFORMATION FOR SEQ ID NO: 811:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 799 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 811:

GCTTTTACTT TTAAACTGAG CGGTACTATA ATTTCTGCTG AATTTGTTAA ACTTTCTGCA 60

CTAGTAATTT GTGGGGATTG TAGTGAAACT AAAGCTGCTG TAGATAGTAG CAAACCGCCA 120

AGAATTATCT TTGATTTCAT AATATCATTC TCCTTAAATT CTATTTATTA AGCTCAAAAA 180

AATAGTTTTT ATCTTAACAT TCTAGCTACT TTAATTCTTA ATACTACAAA AGGAAAACAT 240

TGTTATATTG AATAAAAACA TGTCCCTATC TCCTTTCTAA TTAGttTCCA CAAAACGTAA 300

TGATTACGTT TTGTGGAAAC TAATTTTACA AATTTTTTTA AGAATGTCAA TAAAAATGAA 360

GGTTCATTTT ATTTAAGTGt CAAAAATAGA TTAATATAAA ATTTAACTCG ATTATTTTTC 420

AGATAGCTAA AAAGCAATTT CATCTTTATT AGCTAAAAAG TAGTCAAGTC CAGACTCCTG 480

TGTAAAATGC TATACAATGT TTTTACCATT TCTACTTATC AAAATTGATG TATTTTCTTG 540

AAGAATAAAT CCATTCATCA TGTAGGTCCA TAAGAACGGC TCCAATTAAG CGATTGGCTG 600

ATGTTTGATT GGGGAAGATG CGAATAATCT TTTCTCTTCT GCGTACTTCT TGATTCAGTC 660

GTTCAATTAG ATTGGTACTC TTTAGTCGAT TGTGGGAATT TCCTTGTACG GTATATTGAA 720

AGGCGTCTTC GAATCCATCA TCCAATGATG CGCAGCTTTT GAATATTTGG GTGATCGATA 780

TAATCATGAA TCAATCGAT 799

(2) INFORMATION FOR SEQ ID NO: 812:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 480 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 812:

TTTGCAAGTT AGTTTACTTT TAGGCTTCCT TTTAATCTGG ATTCATTCAT TTATGGAAGA 60

TATTTTATTA GCGCCTCATT GGATGCCGAT AGTTTATAGT TTCCTTGGGC TAGCCTTCTA 120

CTTCCGGCCA GAGAAAAAAC GAGGAAGACA TGAAAGACCA ACGACACCTA AAAGAAGAAA 180

ACGAGTGAAA CAAACAAGTC CTGTTAGCAA TGAAGAGCGC CCGATAGCAC CTGTTGACGA 240

AGAGGGCTGG GaCCAACCTG AAGAATTAAG TCGTGtCCAA AGACATCGCm GATAATTTTc 300

ATCCTATAGA ACTTTCTTGT CCGCTGACAT TATGGkrATA TAgGGGTTGC ATGTTATTAC 360

AAATAATGTA ATTACTGGAA GTGAAGAGAG GAAGTTTTTA ATGACATCAA AAAGTGAATG 420

GAGTAATGCA AGACGGATTG TAATTATTTT AGCAGACGTT TTGCTCTATA ATTTATCAAT 480

(2) INFORMATION FOR SEQ ID NO: 813:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 698 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 813:

TTAAGCGATA AATCAGCTGA AGTCCTTTTA GATTTAGATG CTAAGAGCTT CCCAATGAAC 60

GGCAAAACAG TTCGCGTTGC CCAAGTAAAT ACCGTCGATT TAGCAGAAGT TTTAGATCGT 120

CAAGCTGAAT TAGAAGCAGC TATGGCCGCA GAAAATGCAG CCAATAATTA TGACTTGTTT 180

GTCTTAATTA TTACGAATAT CCTAGACAGT GATTCTGAAT TATTAGCGAT TGGTGCGGAA 240

CAAGCCAAAA TTGAAGCAGC GTTCAATGTA ACATTAGTTA ACAATCGTGC GTTCTTACCC 300

GGCGTTGTTT CTCGTAAAAA ACAAGTAGTA CCACAATTAA CAGAAGTCTT TAACTAAAAC 360

CGTArGgscC TAAGcAACGC GATTAAATTC CTTGTTGCTT AGGCGCTTCA ATAGTAAAGG 420

GGCGTTAACA TGAATGATTT AGCAATTTAC TATCATTCTC CGCAACAAAC GAAACAATAC 480

AATCATTTGC TCAACCAATC ACTTTTTTTC CCTCTTGTTA CTTTTATGTA TGAACATCGG 540

GAAGAAAGAA TCATTTTAAG ACAATTAAAA GCGGCCTTTG CAACAGAAGC CAAGTTAGAA 600

CAATTTTTAT CAGAAAGGAT TGATTGTCAG TTAATCATnG GGAAAATCGT CAATATCGTT 660

TGAATTTTCC TATTATACGG CAGCTGAAGT GCGAGTTG 698

(2) INFORMATION FOR SEQ ID NO: 814:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 756 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 814:

CTTAATCGTT TTTCTTTTAT AGGATATCGC AnTAATCGTA AnACATAATT CATnnGAAAG 60

GATTAACCCA TGACCTTAAG AAAATCAAGT TTAATTATTT TAGTCCCATA CATTACGCTT 120

TCTTCGGGCC GTTACTTTTT TCGGGTATTT CCCCCAATGC AGCGATTAAT GTGACAACTG 180

CTCTCTACAT TATTGGCGCC CTTGTAATGA TTGGCATCTT TTTAAAAACA ACAGAACCTT 240

CTCCTTTAGA GGAAACAGCA ACCTTGAAAT CACCTATTTT TATTTTTTTA CTAGGTGTTA 300

GTGGGATTTT CATTGCCATG CTGATTCAAG GGGTTACTTT TGCCATTGAA GTAGCCATCA 360

CAGGAgAACA AGCTACTTCT CAAAATACAC AAGCTATTGT AGCGGTGATT TTGGCAAATC 420

CTTTATTTAT TTTGGCAACT ACTATCGGTG GTCCGATTAT GGAArAAwTC GTCTTTCgTT 480

ATGCATTTAT CCATTTaATC CAsCCTTTtA CAAATTTCyG GawTGcGGsT ACTGtTAnTT 540

CGGCGATTTT TTCACTCGCA CATGCCGATG GTCACTTTTT TGTTTATTTC TTTATGGGAT 600

TTTTCTTTGC TCTTTTGTAC AAGCAAACTG GAAAAATTTG GACtCAATTA TTGCCCATTG 660

CGGCATGAAT ACGATTGTGA TTATTGTTCA ACTACTTTTG CATAATGGTA CAATTCAATA 720

AGAAAAGGGC CTnTCGTCTA AACGACTAGG CCCTTG 756

(2) INFORMATION FOR SEQ ID NO: 815:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1602 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 815:

GATACGCTGC AAGCGAcrGT TTCaAGTGCC yGcGTTTTCT GGTTGTTCCA AGTCGTTAAA 60

CCTTGGACTT CTGCTAACGT TAACGAAGGC TTTGCTTGAG GTCGACTAAA AATTAAGTCC 120

ATGACATATA AAAGTTCTTT TCGTCGGGCA ACGATTTGTA TCTCATCCGT TGGCAGTTTT 180

AGCATTTGCC ATTGCTCAAA AATATCATTT GCGGCTTGGG CAACTTCTGA CGCATTCGGT 240

CCCTCTAGGC GAAATGTCTC AACTGGCCAA TTCACCATTC CTGCGTTGAT TAAAGGATAG 300

TTAGCTAACT CAACATATTC TGAAATGCTC GCTTCGGCTA ATGGAAATTG TTGCAAACCC 360

GTTTGATAAT AGACATGCTC ATTTTGTTCA AAGTCAGCAT CAATCCCAGC AAAATAATGA 420

GGAAAGACCT CAACAATTTT CAACAATTGC TGAAAAGTCT TAATTGAACG TTTCGCTGAC 480

TGATGTTTTT CAAACGCAAT CACACCTTGT TGCTGATATT GTGCCTGTTT TACAAAGGAA 540

AACCCCCAGC TTTCGCCATC TAAGTTCATT CGAATCACTC GATGGTTCGT AGTCGCAGGA 600

AATTGCTGAC TTCCCTGGTA GCCTTCTGTT GCCATGCACC ATTCACAGCG GGGATAGCTT 660

TGTGCAGAGA GCGTTGCTTT TGACGCCTCA CTGTGGACCT TATTTGCTAA AAAATCTCCA 720

TAGACTGTTG AAAAAACAAC AGGTTCTTCC TGTTCAATGA CCGTACCGTT TTTTTGACAA 780

AGTTGGTAAA AATACTCGGT AGCCTCTTGT GGCTCTTTGG CATAATGTTG TGCAAAAAAA 840

GCATTAACCA CAGAAGGTGG CGGTGTTAAC AAATCCATCA ATTGCACCAT GAACTGTTCT 900

TTTTGTTGCT CTGTTTTTAC AAGTTGGTTG GCACTCGCAA CTTGACACAA TTGTTCCGCC 960

AAATCAGCGG CGGGTGTTGC CACAGGCCGT ATGTCTACTT CGCCTAATTC TTGTTCACCA 1020

ATCATTGATA ATAAACGGTT CTGAAGATAT AAGCGATCCA ATTCCATCCA ACCACCTGCT 1080

TGAATAGCAA GTGTGGTAAA ATCGGCAATC ATTTGACTCG TGGTCATCGA TTCATTCCTC 1140

CTATTACTTT TCTTGATATC CATGTGGATG AGCTTGATGC CATTGCCAAG CGGTGGCGAT 1200

AATATCCTTG ACCTCAGTTA CTTCTGGTTG CCAGCCTAAA ACTCGTTTGG CTTTTTCACT 1260

AGAAGCAATT AACGTACTAG GATCCCCTGC TCGGCGTGGG GCGATTGTTG CAGGAATTTC 1320

TTGGCCTGTC ACTTCGCGAG CAGCATCTAA CATTTCTTTA ACAGAATAGC CGTTGTTGCT 1380

ACCCAGGTTA AAGACGTCAC TTTCGCCACC ATTTTTCAAG TATTCTAAAG CCAAAATATG 1440

TGCTGCAATC AAGTCTTCAA TGTAAACATA ATCTCGAATG CACGTGCCAT CTGGTGTATC 1500

ATAATCATCC CCAAAAATAC TTAGCTCTGC TCGTTGGCCT AATGCCACTT GTAAAATAAT 1560

TGGCACAATA TGCGTnTCTG GCGTGTGATC CTCACCAATT GA 1602

(2) INFORMATION FOR SEQ ID NO: 816:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1008 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 816:

ATTTCCTTTG GCGTGGCTTC TCTCGTCATT TCAAAAATAG ATGCATTCTT TTTTATCTGC 60

ACAGACATTA CAACACGCAT TGAAACAGTT GAGTGCTTCG GATGTGGATA TTCATGTGTT 120

AACGATACCT ACCAATAGCT TTTCATGTTC TCTCTCCTAA TTGTTTTTAT GTATATTTGT 180

TTATTGTAAG ACAAACACTT TAATCCCTGT TATACTCATT GAAAGACAGC AACTCCTTTT 240

TGCTTCATGT AACACTTCCA GTTATTTCAA AACATAATCT GCTGTCTGGC CACTAGATAT 300

TTTATCTGGT GGTTTTTCAT GCGAAAACAA TCCAAATATC CGACAAAACT TGACAGCTAt 360

GTkACACTTG ttTTAGGkAG CACTCTkTCA TAATAGCkAA AGTTCATAAA CTACAAGTGA 420

CACGAGATTT TCACTAACGC TACCTAGCCA CTAGATCCCA TTTCTAGTGG CTTTTTTTGC 480

ACAAAAAATA AGTTTTGATG TGAATAAACG CTCTTTTATT CTTTATTCTT ATTTGATATA 540

ATTTTTTAGG TAGCAACTCC TTTTTGTAAA TAGCAACCAA CAAAAATTTT GCACGAATGC 600

TACCTAGCCA CTAGATCCCA TAGTCTAGTG GTTTtTTATG TACGArAAAA rGmCCACTTA 660

CTGAGTGAnC TTwATAyCGm CGTGsCAGGA GTkGAtCtGC AgCCATkGAT TAAAAAwTAm 720

CkGTyCTTAT TTATTAAkGA mCCTTTyCTT TTTTGGTTCA GwTTCAAAAA CTATTTCTAA 780

mCTCAAAwTA wTAGTCTGTA AAGACmACgC CATAAmCAGC CCTGAGAAAA ATGTCCATAT 840

ATAAAATATT AGGATTGAAG CAAAGGTCTC TTTATACTGA AATATTTGGA GCATCATTGA 900

TAGTATCAAA GTAGTAAATG CACTTAATAA AGATATAAAG AGTTGATATT TTAATTTTTT 960

TCTATAGCAG ATCCTCGTAT ATTCTTCATA AAACTAGTAT CTTTTATA 1008

(2) INFORMATION FOR SEQ ID NO: 817:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 645 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 817:

GTCATAGCTG TTTCCTGTGT GAAATTGTTA TCCGCTCACA ATTCCACACA ACATACGAGC 60

CGGAAGCATA AAGTGTAAAG CCTGGGGTGC CTAATGAGTG AGCTAACTCA CATTAATTGC 120

GTTGCGCTCA CTGCCCGCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT 180

CGGCCAACGC GCGGGGAGAG GCGGTTTGCG TATTGGGCGC TCTTCCGCTT CCTCGCTCAC 240

TGACTCGCnG CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT CAAAGGCGGT 300

AATACGGTTA TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA 360

GCAAAAGGCC AGGAACCGTA AAAAgGCCGC GTTGCTGGCG TTTTTCCATA GGCTCCGCCC 420

CCCTGACGAG CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT 480

ATAAAGATAC CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG cGCTCTCCTG TTCCGACCCT 540

GCCGCTTACC GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC TTTCTCATAG 600

CTCACGCTGT AGGnnTCTCA GTTCGGTGTA GGTCGTTnGC TCCAA 645

(2) INFORMATION FOR SEQ ID NO: 818:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 687 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 818:

ATGCAAGCCA GTGGATTAAA ATATACTAAA AAACGAGAAT TACTGATGAG CTATTTGATT 60

AAGCGAAATC nGTATGTTTC TGCACGTGAA GTGTATGAAT TTATGAATGA AACCTTTAAA 120

GGAGTCAGTT ATGATACGGT TTATCGTAAC TTGCATGACT TTGAGCGCTT AGAATTACTG 180

GAAAAAACAG AATTAAATGG CGAGCAAAAG TTTCGCTTCC GTTGTTGCCA AGAGGTGGAA 240

CATCATCATC ACTTTATTTG TACGGTTTGT GGCAAAACGG AAGAAATTCA CATGTGTCCG 300

ATGAATTTTT TTGAAGAACA GTTAAAAGGT TGTTCAATTG AAGGACATCG CTTTGAAATC 360

CTTGGTCGTT GTGCCGATTG TTGCGAAAAA TAGCGAATAA GACGAAAAAT CAAACAAAAA 420

TTTCGTATTT TATCAGTAGA GTGGTTGACG GTTATTTATG CTTTACTATA ATATAACAAG 480

GACAGTATTT TAATTTGTGG GTTTTTCCCA CAAATAAGAA AAACGGTTAC TTTTTAAGCT 540

CGGAGGGGAG GGAATAACAT GTCCAAAAAC AGTCGTTCGT AAAAACGAAT CnCCTGGATG 600

AnGGCTCTTC GTCGCTTCAA ACGTTCCGTT TCCAAAGCnG GGTACTTTTA CCAAGAATCC 660

ACGTAAACGG GAATTTCCAC GAAAAAC 687

(2) INFORMATION FOR SEQ ID NO: 819:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 370 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 819:

TnGTAnGCGT TCTTTTTTTA AGAATACCAA AACATTATTT AAAATAAAGC AATTTTTATT 60

TTTAGTAGTA AGATAGTAAA GTAGTAAGAT AGTAAAGTAG TAAGATAGTA AAGTAGTAAA 120

ATAGTAGAAT AGTAAAGTAG TAGAATAGTA AAATAGTAAG ATAGTAAAAA ATGTTTATTT 180

TTTAATAAAT ACGTCTAAAA ATCCTTTTGT AGAAAGGTTA ATATCATTCT TTTAAAACAA 240

TAAGATAGTA AAATAGTAAG ATAGTAAAGT AGTAAGATAG TAAAATAGTA AAGTAGTAGA 300

ATAGTAAAGT AGTAGGATAG TAAGATAGTA AAATAGTAAA TTTTATAAAT ATnATTTGTA 360

GTAAGAnAAT 370

(2) INFORMATION FOR SEQ ID NO: 820:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 334 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 820:

AAAGTAAAAA TTGGCTTAAA TATGTTGTTA ATGATTATCG TTTTAGGCGT AATCTTTTAT 60

GTCATGGATA ATTCATTAAG TGATATTTTT GCTCAACTGA TGGAAACGAG TTGGCTGGTG 120

CTGATTGCTG TTATTTTTTT TGGAGTAGTT TATCAATTTG CAGAAGGGCG CTCAATCAAG 180

GAAATAGCTC GTTATTTTAA CAAAGACTTT ACGACAGTGG ATGGTtTTTT TACTTCTTGT 240

TATGTAGCAT TTaTCGGATT AATTCATTGG CACGGGGAAC CTTGCCTTTC AGAAATTTAA 300

TTTTTAATAG GAAAAAGGGG AATTCCCGGT TTCC 334

(2) INFORMATION FOR SEQ ID NO: 821:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 493 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 821:

CTTCAGACTC ACACCCATTC TACACTGGTC GTCAAAAATT nTTGACAAGC CGACGGACGT 60

GTGGACCGTT TCAACAAAAA ATACGGTCTC AAAGACGAAA ACGCAAATCC AGACGCATAG 120

TCTTTCAAAA CATTGTTATG ACAAGGTTTA TAGAGGTTTT CCACTGTGGT GGGAAACCTC 180

TTTTTTTGTT GGATAGGCAG TTTTAAATTC AGAGAAATTA TTTAAATATT ATTTTATTTA 240

ACTGAAATAT AGCTAGTCCG TATACCGTAt AATTTAAATT ATTTGGATAG TATACATATk 300

GTAAAGTGCA ATTGTCTTTT TAGAATCTTT AAGGTATAAT TTGTTTTAAT GAGTTTATAT 360

GTAAGCTTTt tCtcTTTAAA GmCTTtGAyC AtAyCTTTTT tATAagaGAA aCTATgTTtG 420

TGAAaaGGaT ATAAAAGTGn TAAAACAATT GTACTTAACA GTATTAAGAT GGGAGACCGT 480

ACAAGGCATA AAA 493

(2) INFORMATION FOR SEQ ID NO: 822:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 493 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 822:

TTTCAATTCC CTCTTCATAT ATCGAAATAA GATAACTATA ACCTTCCAAT AAATAATTAA 60

TTTGTTCATC AATAGTTTTA TTCCAAGAAC CTTTCTCTTC TTTTAATTTA GAATTTATAA 120

AAAACTCTTC TAGAGTTGAA TACGGAGATC CGTCAAGCGT AATCAGTCGT TCTGAAATTA 180

CATCCAACTG ATCATTAATT TGCTCCATAA GATCATCCAT TTTTGGATGT AGCGTTAGAA 240

ATCTTCCTCC ACGCATATAC CAATGAACTT GATGAATTCT TACTGAAAAT ACGCTTAAAT 300

CGGCTACCAA TTGATTTAAT ATTTTTkTTG TTTCACAAAA TTTCATAGTA TCACTCCTAA 360

TTTAGTGTTA TTTAAAAAGA AATCATTCCG ATTTACGTAC AATTTAATAT AATAAATATA 420

TCTTGTCAAA AATAAAATGA ACGTTTATAT TAAAAAAGTG TTTATAAAAA TCTTTACAAC 480

TTATTCATCC TTT 493

(2) INFORMATION FOR SEQ ID NO: 823:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 272 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 823:

TAGAACGGAa CTAACAGAAT CATGCAGATT gCAACnTTTG GTCTGATTAA TGAGATGGAA 60

TCAGTGAAAA AATAATGAAt TTTGTTATCT CAAAGCtTCG aTTcTTCAAG TATTATTGAG 120

ATATaTTTAA AAGCTCGTAG ATTTATCTAT AAAcGTTTCT TCaCAGTACc AtCAAAATCG 180

GGAAACCGGG TGGCGACGGT GAATAATGTG CCAATCGGAG ATAGGCATCG GATAGTATCA 240

AATTCACTAC GCATGATGCG TAGCAATTAC AT 272

(2) INFORMATION FOR SEQ ID NO: 824:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 252 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 824:

CGCGGCCTAC GAGAACACGA GGTTTTCnTG AACTATCnTC ATGTTGACTC AATACATAAC 60

CACCGTAACG CCCTAATTTA AAAGCTAATT CAGGTGTTAA CTCCTTATTT GCAATTCCTC 120

TTACACCATC TGTTCCAAAA TATTTACCCA TTTTGATTCT CCTTAGTCAC TTGATTCTCT 180

GTATTTTCAG TTGAACCAGC AGTTGTACTT TCTACTGTTG ATCCCTCACT TGTTGTTTCA 240

GAAGTATTTG AT 252

(2) INFORMATION FOR SEQ ID NO: 825:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 476 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 825:

GACATCTAAG GTGCCTTCAG CTTTTGAATA AAAGGCATTA TTCGGnGATC GTAACCCCAT 60

TTTGTTACTA TCAGTTTTAT AAATGGTATA AGGTCCTTTG CCTTTTAAGG ATAAGTGAAG 120

TTCTTCGCCT GCTTGTTTAT TTTTTAGTGT ATAACTTTCT TTTAGCGGTG TGACTGATTT 180

TTCCGCTTGC TTGTAATAAT AGACGGCGCC AGATCCACCA ATTGCCCAGC AAAAAATTCT 240

TGTTTGGCTT TTGGTAAAAA CAATGTGATA GATAAGAGGA TTTGTTCTTT AAAGATTGAC 300

CAATCTGCGA ATATTTCTTC ATTGGTTAAA ATCATTTCTA TTTTTTCTGC TATTTTTTTC 360

CAGAGGATCA TTCGCCAGTT ATTTTCACCC ACTTCcATAt TCTTTAAACc ATGCTTAATA 420

aCtGrAGrAT AaTGGGGAAC ATTTGGckGA CAtCgCsGCA AATnAAAAAT CCATTA 476

(2) INFORMATION FOR SEQ ID NO: 826:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 413 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 826:

TTAATGACAT GTGACAATTG AATAATTTGT GGGTCGGTTA ATTCTTTGGA ATCACGAACA 60

CCTAATGATT TTAGTTTGCT AATCATGCTT TTGTCAACAT AGGnCAGCAC AGACAGTTAC 120

TGGACCAAAA TAGCTACCAk TACCGrCTTC ATCGGaACCT ATCACAGACC AGTTGCCAAA 180

ATCAGCAGGC AGTGTAGTAG TCTGGGGAGA AACTTTTTTC TTTGGTGTGG TACTCGTTCC 240

TTCCCAACGA GCGGCCTCTT TTTCAGCTTG ArGTCCCTGA AACATGACTT TACCTGATTG 300

ATAAGCGGTG ATAGTAGTGG TGCCAACCTT GGCCACAAAG ACAGTATAAG GGACCGTTTt 360

ATTTAAACGa TgGGTTGtAA ACAGTGGTCA TCcTGCTAAT GnTTnTACTA ACT 413

(2) INFORMATION FOR SEQ ID NO: 827:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 207 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 827:

TTAAAAGATA CAATGTAATG sAGACAGGCA CATTAAAaCC GACAATTTTT GTTGAGAGGG 60

TTGTTAAAAT TTGTGCCAAG CCTGTTAAGC CACTGGAATA AATATTTCCT GGTTGATAGA 120

AAAAGTTCAT GGCAtTGATG CAAGAATAGC ATAGACAATC GACACAGAAA ATTTTGTTGT 180

ATAGTCATGA ACTGGCAGTC CGTCATA 207

(2) INFORMATION FOR SEQ ID NO: 828:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 325 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 828:

AACGTTTGGC TAAATCACCT AGAATTGAAT CAGGCACATC CATCCATTGT GTAAAGTACG 60

TAGAGAGGAC GCCATCATCC AATTTnAnGG GGGCTTATAA nGTGAAGTCC CCTTTAAAGA 120

ATGGTACAAG CAGAGAGGCT TGTAAATCAT AGTCAAATTC AGGGTTTTCA AATAATTCTT 180

TCGCCCGATG CAATAAATGA TCAAGAATCA CTTCCATCCC ACGAGAAACA GGGTGAAAAT 240

AGACTTGGAC ATACATTTGG TAACGACTAA CGATATAGTC TTCCACCGCA TGCATCCCAT 300

TCATAGCAAA GGCAATGCCT CCTTT 325

(2) INFORMATION FOR SEQ ID NO: 829:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 957 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 829:

TTGACGGTTn TTTAAGTCTT TACTTCGTAn TTACAACTTC TTCCATTAAC TTATnATTTC 60

CTCGCTATTC CTATCGTTGT CTAACAATAA GTTGATTCGA TCATATTCTT TGTCTTTATA 120

CTTCACTGTT TTCTTACGGA TTTTATTAAC ATCACTTGTT CCTAGTTCTG AGCAAGTAAG 180

GTTCCnTACA CATTGGTGTA AAAAACACGT GAAGCCAATT AAATTTCCTG CACTTTTTGC 240

TCGTTGATCT GCAGATAATT TTTGCTCTAC TTCATTTCCG TTCTTATCTT TCCCTTTCTC 300

TGTCATTGGT ACACTAACAT CACCTAAGAC CGGATTAATA TTTACAAATG CGGTCTCAAC 360

TTCTTTATCT ACACTAAaCA TCATATCAAA CAAAGAATTG GACGTGTTGG CATCTTTTAG 420

GACGGAAAGA CCCGTAAAAA TAAGAATCGT CATTAAAAAA ATACCAAAAA ACCGTTTGTA 480

TCGTTGTTGT CCAATAAATT TACACCTAAA ATAACAAAGA CAAAGGCAAT nCCAATCGTT 540

CCCGCAATGC CCAACATGTT AGTTGCTATG CTAGACGTTA GGTTCATAAT CGGTTGCTTA 600

ATAGCAGTTA TAATATCCAT ACTAAATAGT GTTTTTACCA TAACTGCATT AAACTCTCCG 660

AGTCCTTTTA CCATTGACCA GACAAAAGTC TTTATTGTCG CAGAAGCATT TGTGATGGCT 720

TCTTCAACnC CTGTCCATGA ATCTTGTTTT TTTTCTTTGG TCATtAACTC GAAAGCGTTG 780

TCTTTGTAaG TATCGTAAAT CAAATCTACA TTAGTGGTTG GCTCTGTAAC AATAACTTTA 840

TTTTCATCAG GTTTAGGTTC GGCATTAACC GCTGTAaGTT TGgCGTAAAC ACcAAAAAAA 900

ACCnGGCCTA CTAAAAAATn GTAAGGACTG GCTTTTACCC CGTGGATTTC TTAATTC 957

(2) INFORMATION FOR SEQ ID NO: 830:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 261 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 830:

GGATTTTnAC TCCAAAAAAA GTTTAGCATC CGCTTACATT GTTGCCAAGA AAAATGGAGA 60

ATTTTGGAGT TTTTTAATTy TTATAGTGTA TAATAAACAA AAATTACATT CTAGAGGAGA 120

TTTTATGAAT TACATTTATT TAATTAATGA ACATTATCCC TTATTAACAA AATCGGAACG 180

AAAAGTTGCT GATTTTATTC TTAACTCTGG AGAATCTATT ATTTACAGTA CTATGAATGA 240

nATTAAAAnA AAAGCGAATG T 261

(2) INFORMATION FOR SEQ ID NO: 831:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 451 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 831:

ATTGCTATTA AACCAAATGG GAACACTATA TATTCTCCAG TAGATGGTAT TGTCCAAGTC 60

GTATTTGAAA CTGGGCACGC TTATGACCTC AAATCAAATA CTGGTGCTGA AATATTAATT 120

CATGTTGGTA TTGACACAGT ATCATTGAAT GGTAAAGGAT TCACTAAAAA AGTTGGTGCT 180

AAACAAAAAG TAAAAAAGGG AGAGGTTCTT GGAACATTTG ACAGCACAGT AATTACAAAT 240

TCAGGACTTG ATGATACAAC AATGGTTATT GTAACAAACT CTAAAGATTA TTCAGAAGTT 300

ATTCCTATAA CTAAAAATAT AGTTACTGAA gGTGCAGCCT TATTGAsrAT TAAATAGCTT 360

ATTAAAAGTT AATAGACTCT CtATACmAAG AGAGTCTATA TTTTTATTAT AGTTAATTAT 420

AACTATTCTA AGCTAATTAA TTTGACATAA A 451

(2) INFORMATION FOR SEQ ID NO: 832:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 612 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 832:

CCCTTTATGC CCTnTCCGnC TGTAATGTTG TCCGGTTGTG GAACCnGTAT GGAAAAgrTA 60

TTGACTGCCG tGGTGGTAAA TGGGAACTAG AAAATAAAAG TCCAACTACT ACTTACACTT 120

TTTTTGATGA TGAAACTTTT TCGAGGTATA ATTCAAAAAT TAGTGATAGT GGAACGTACT 180

CTTACGATGA AAATAATAAA AAACTCACTT TGGATATAAA AAATAAAGAA CAATTAATAA 240

TGGAAAATGT TGAATATAAA GACGGTAAAT TAAAAGGTGA AATTGGAGGC GTAGAAGGAA 300

CTCTGATAAA AAAATAAAAT AAAGAGGTGA TCTTTTGAAT AAGAAATCAC TGGTAAAAGC 360

AGTCTTACTA TTTTTAGTAG GCTGTTTTTT TGTGTTTACG CAAACTACAG CGGTTAATGC 420

CGAACCTAAA CCTGATGAAA ATAAAGTTAT TGTTACAGAG CCAACCACTA ATGTAGATTT 480

GATTTACGAT ACTTACAAAG ACAaCgCTTT CGAGTTAATG ACCAAAGArA AAAAACAAGA 540

TTCATGGACA GGGTTGAAGA AGCCATCACA AATGCTTCTG CGACAATAAA GACTTTTGTC 600

TGGTCAATGG TA 612

(2) INFORMATION FOR SEQ ID NO: 833:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 995 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 833:

CACTTACGGC AAAACTTTGA ATCGACCGTG TCAATTTTTT CACGAATATC TTCCAATGAA 60

TCTTTGAGGC CTAACCACCG GATAATACCT CTTTCCTGCC AAATCTCCGC TTCACGTTTA 120

TTACACATCG ATTGTGaACG TcGATAAATC GGcTTcTcTA ACGATGaAAG ACTTCCTAAT 180

AATCtCAAGC AATCGTTCTG AGATAGTGTT TCGCTTAATA CAATTGCATC AAAATAATTA 240

AAAAAAGTTA ACTCATGTTC TATAGTTTCT TCACAGCAGT TCGTACAAAA GACTTCATAA 300

TTTAAATGTT GTAATTGCTG CTGAAATGTT TGCTCAACTA AAATGTTCTT TGTTAATACC 360

AAAATCCTkC kCATTTcTTC TTTCCTCCTA GAACTCTTTC TGCTACCAAT AACTAAACGC 420

ACCAATCAAA ATACGTTACC AAATCGCGTG TTTTCTCCAT CGATTGGTTA ATAAACACCT 480

TTTTATCTTc ATCTCTCTTC GCTCCTTTAT TCTTTATTTT GGTGGTTATT TATATnATAT 540

nAGAaATGgA TTTTgTTTGk GTGTAAAATT TTTTTAtCCc CGTAAAAAAA AATGaATTyC 600

CTTTAATATC AACTAATAAA GCkGATTTTT aACAAAAAAg TGCCGAAACT TCCTACTTTT 660

TGCTAAAACA CGCCAAAAnT CTGTAATTTT ATTATTTTTT TCTCGAGAAA ACATCAAAAT 720

GTTGATGTTT AACACTAACA TTTTGATGTT TTCGACAAAA ATCTTTACAA TTAATCGGTA 780

ATTTCTCTA CTTTTAACGA TAACTACTTT TAAAAAAGAG TTATTTTAAA AGTGACTTAA 840

ATATTTATCA TATAAAAAGT AAGGGACTAA AATCTTTACG ATTCTAGTCC CTTACTTTTT 900

ATAATCCACA ACGTTTTTAG TTAGAAGAAT TCTTTTTTGC TTTCCTAGCA GAACGATCTT 960

TTTGACGACG TTCAATTTTA CGTTTTTTCT TATTG 995

(2) INFORMATION FOR SEQ ID NO: 834:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 753 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 834:

```
CGTTGGGGTA GCGAAATTGG TTCTTGGTTA GTTCAAGGGA GTAGACACCT TACGGATATG      60

GGGTTATGGA CGCATATTCT TTCTGGATTC ATTGTATCAG TAACAGGGAT TATCATTGTT     120

TACTCTTTTC ATAAGAAGCT ATCTCTTTTA GCTTTAACAG CAGCCATGTT GGTTGGCTTT     180

AATCCTTGGT TTTTACAAAA TTTAAGTTTC AGGTTCGACA GTCCTTACAT GTCTTTGAGC     240

CTTTTATTTT CTGTTCTTCC ATTTCTATTT TGGAATAAAA ATAAAGTCTT ATTTTTAGTT     300

GTTTCTATTT TaAGCaTTTT TTTGATGTGt AATACTTATC AGGCGTCTTC GGGGATCTAT     360

GTCGTGATGG TGTTAGCATT ATCACTGAAG CAGCTGTTAG ACAATCAATC GTTTATTGCT     420

GTGCTTAAAA AAGCGATTGT AGCTATGATT AGCTATATAA GCGCCATGTT TCTTnATCTT     480

ATTGAAACAA AATTTAATCC TGAAATAGCA ACGCGAGGCG GTATGnCAAC AATCGCGAGT     540

GTTAAGGATA TACCCAAAAC AATCCTGGTA AATAGTCAAA TGTATCTGTC AAAAaTAaCT     600

GAACAGAGTA CAAAACTATG GaTACTTTTG TTTTTTATTC TAGTAATTTT CTTTGTTTTA     660

AGTACGGTTC TTAATGCGAA AGTATCGCCT GTTAAAAGCT TTTTATACGC AGTATTATTT     720

CTATTTTTAG GCTCTATTTT GAGTTAnGGC GTC                                  753
```

(2) INFORMATION FOR SEQ ID NO: 835:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1076 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 835:

```
TGAGCGTGGA GCAAAAATCC AAAGTGATTT TTGTCCCACG CTCTTGTTTA TTACGGAAAA      60

ACCATCAATT TTTTATTAAA TTCTGATATT TATTATTGCT TTTTCATTAT AGTTGGCGTA     120

ACTTTAGAGT ATAAATTTGT GAAGCGAGGG AGTCCGTGTG AGAAATTTTA AAAAATCCGA     180

TAAATTAAAT AACGTAAGTT ACGATGTCCG TGGTCCTGTT TTAGAAGAAG CAGAACGCAT     240

GCAAGAAGAG GGCATACGTA TTTTGAAATT AAATACAGGG AATCCAGCGC CATTTGGTTT     300

TGATGCCCCm AATgAAAwTG TGCGCgACAT GATTGTAAaT GTtCGCGATy CAGAaGGTTA     360

TtCCGATyCA AAAGgTATTT TttCTGCAaG AAAaGCGATT GAmCAaTATT GCCAATTAAA     420

AAAaTTtCCy AATGkAACGA TtAATGAwAT TtACACkGGT AATGGCGTGA GTGAACTAAT     480

TACCATGTGT ATGCAAGGCT TGTTAAATAA CGGCGATGAA GTACTTGTGC CAATGCCAGA     540

CTATCCTTTA TGGACTGCAT CTGTTTCTTT GGCGGGCGGT ACACCCGTTC ATTATATTTG     600
```

TGATGAGCmA GCAGAATGGT ATCCAGACAT TGrTGaTATA AAaTCmAAAA TAACTyCGAa 660

TACAAAAGCA ATCGTTATTA TCAATCCCAA TAATCCGACT GGTGCTTTAT ATCCGAAAGA 720

GTTATTACTA GAAATTGTCG AAGTGGCACG TCAAAATGAT TTAATTATTT ACTCCGATGA 780

AAtTTATGAT CGCTTGGTCA TGGACGGTTT GGTksaTGTA CCGATTGCTA CATTAGCTCC 840

TGaTTTAwTT GTCGtAACAT TGAATGGTTT GTCAAAATCA CATCGAGTGG CAGGCTTCCG 900

ATGCGGTTGG ATGGTGkTAA GTGGAGATAA ATCACGAGTA AAAGGGTATA TCGAAGGGCT 960

AAATATGCyT GCTTCCATGC GGTTATGCTC AAATGTGTTG TCACAmCAAA TCATyCAAmC 1020

kGCATTAGGT GGCTATCAAA GTGTCGATGG TTTACTTTTG CCAGGTGGCA GAATTT 1076

(2) INFORMATION FOR SEQ ID NO: 836:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 501 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 836:

AGACAGCTTA AAGGTGAGCA AATGGnGGTT GTTACGCATG TTAnTGATGG CcAGCATCAA 60

TTTAAAATCA ATGATTTTAT CGATTTTACA GGCAACAATC ATACATCAAA AGATAGCATG 120

ATTGGGATGC TAGTAGGCAT ATTAGGTGGT CCTCTGGGCA TTCTTTTTGG CTGGTTTGCT 180

GGAAGTATGT ATGGTGCAAG CAAAGACGCC AAAGAAATTC AAGAAGCACA AACGGTTTTT 240

GAACATGTGA TTCAAAAGAT TGATGAAGGA CAAACGGGAT TGTTATTAAT TGCAGAAGAA 300

GAAGACAACC GTCCGCTCAA CCAATTGGTT ATGTTTGACT TAGGTGGCGA AATCACGCGG 360

CTTGATTTAG AGGAAGTCCA CCAAGAATTT ACGATGCGAA CGAATTGCCC AATGGAAGCG 420

GAACCATCCG TGGGCAGCCC AAAAAAAGGn CCACCCCCnA GGAAGCCACC TTCCGGAAAG 480

nAGGATTAAG GGnGCCAAGA C 501

(2) INFORMATION FOR SEQ ID NO: 837:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 856 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 837:

AGGTTCATTG GTTGTTTCAG GGCTTACTAC TGGTTGTTCT TGTGCTnGAT TTTTCACTCC 60

TACTGTTGGT TGTTGGACGT TTGTGGTATC ATTTGGGGTA GCTGAACTCT GTTacTCmAT 120

TTTTTTcrGA CGCTACTTCT TCTGCTTTGG TCGGTTGAGA AGTAGTGTCT TTTTGTACTG 180

CAGATTCTTC AGATACTGCC GTATTAGGCG TTGCACTACC TAACTGCGGA TCGGGGTTAT 240

CGGGGTGCAC TGTCGTTGAT CCTGGTTGCG TATCTAATTC CGCTGcTTGt ACATTATCAG 300

TAGCTAATCC TACAACTCCT AACACGCCTA GAAAAAGAAT AGGAGCTACG ATCCAATGCT 360

TCTTTGcCTT ATACATTTTA AAACGTTTTT TTACTTCAGT CTGTTGATTC ATGTATCATT 420

CTCCTCGATT TTTTATTTTT TTGTTTGTAA AATAACTCAC GCAACACAAG GCGCTCGTTG 480

CTTTTTCATT GTCATTTCCC TCCTTACTAA AAAAAGAGAA CCGAAAGTAT CTTCGATTCT 540

CTTAGTTATT TTATATTCTT ATAAAACACG ATATATATCT TTGTTTTGCT ATATCAAACA 600

ATTATTAACT ATTTTTTTTA CGTTTTTTGT TCTTGAAATT GATCGTACCT GCTGCCATCA 660

ATAATCCAAT AATAGTCAAC CAAATACTTT GTTGTTCTCC CGTTTGTGGT AAGGATTCTG 720

GTGCTTTTGC TACTTGTTTC GTCTTACGAG GTTCTGAATG TGTTAATGCT TGTGCTTTTG 780

GTTGTTCTTT TTGTTTAGAA AAAGATAACT GTTGGAGACC AGGTGTTTTG TCTGCTGAAG 840

TAGCACTTGC TACTAA 856

(2) INFORMATION FOR SEQ ID NO: 838:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 261 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 838:

TGTGCCTCTT GGGATTTCGC TGTGCCATTT TTTACTTTCA AACCCAGGTG TGCCACCGTG 60

AAGTGTTGTT TGAGTAGGAG TATTAATCAT CGTTTGATAC TTTTTACCCG CTAGTTcAAA 120

GGAGCCATCG GCAATTCGGC CAGCTGgTTC GACCAATtGT TGcACCAAAA TATaAGTCTT 180

TAGACAAATA TTCATCTTGC TGAGGGGAAA GCCTAGCACG AGTGATCGGG ATCATTTnTG 240

TAAGGGAACG TCGGCGCAAC A 261

(2) INFORMATION FOR SEQ ID NO: 839:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 477 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 839:

ATATAGCTAA AAATGATATA CCTGATGCAG ATATTATTAT GTACAATGAG ATGGATATGG 60

CATATCTTGA TGAAAAAATC AAGAATACTG GATTGCCTGT ATCATATATA GATTTACAAA 120

CAAAAAATGC TGATAGTATA AAGAAAAAAT TAGAAAAATT TGCTGAAAAT AAAATACAAT 180

AATAACTACT ATAAAGTTAC TGATGAAAAG AGCCAACAGG GTATACCCTG TTGGTCTTTT 240

TGTATGTTCA TGATATACTT GAATTTCTAA GAGGkCaACA GAGTTGAGCA mCAGAATTTT 300

AAAATAAAGT AATTTTTTrG AGGrGAtTTT CTATGGCAAA AATTGGATAT GCnCGAGTAA 360

GAGTGGGATC AAAATTTAGA TCGACAGTTA GAATTATTAT CnCAGTGTGA TAAAATTTTT 420

ACTGTAAGCA AGTGGGAGGA ACTAAnCGnA TGGTTCCAGA AATGTGAATT TATCGGG 477

(2) INFORMATION FOR SEQ ID NO: 840:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 354 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 840:

GCAACAAGTT CGCTATATTC AACAATTATT GGAATATACG GCAGAAACAT TTGCAGAGAA 60

TAATGATGGA CAAGAAATTA TGAATCGAAT TAAGACCGTT TATGAAACCT ATCGCACAAA 120

ACCATTGCCG CAAACAAGAr GTGAATTTGA AAATCGGGCG CGACTCTTTC AGTTTTTACA 180

ATTGTTTCAG TCATTCATTC AAGTCAAAGC GGACTTTGCC AAATTACAAC TAGATAAATA 240

AACAGAAAAA GAAACCTTTT TAAACAGAGG TTTCTTTTTT GTnCTCTTAA AAATTCAGGC 300

AAGTTCTTTA GTATTCATTG AGAATGTGTT AAACTAATGA GAAAGAAAnn TAAC 354

(2) INFORMATION FOR SEQ ID NO: 841:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 458 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 841:

AAGCCAGAAA TATGTTGTAA AAGTTGAGCA CTCGCTGTAT TCACATCAAC ACCAACTTGG 60

TTTACCGCAG TTTCTACAAC AAAATCCAAT TGCTCGGCAA GACGTTTTTG CGATACGTCA 120

TGCTGATATT GACCGACCCC GACTGCTTTG GGATCAATTT TTACTAATTC TGCCAACGGA 180

TCTTGCAAAC GACGTGCAAT GCTGACCGCA CTTCGTTCTT CGACTTGTAA ATCAGGAAAC 240

TCTTTACGAG CCACTTCACT CGCAGAGTAA ACGGAAGCCC CTGCTTCATT AaCAATCACG 300

TAAAAGACGT CACGCTTCAC CGCTTTTAAT TGCTCTGCCA CAAACAATTC AGATTCACGA 360

CTTGCCgTTC CGTTTCCAAT AGCTACCATG TCCACtTGAT AGTCTTCArT GATCTTTTTA 420

AAGGCTGGAC CAGnGGCTTC ACGTTTCGCA GCAGTCGC 458

(2) INFORMATION FOR SEQ ID NO: 842:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1482 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 842:

ATTTGkCGTA CTTTTTCTTC ATAATTTGTC GTACTTTTAG TCGCAACTGC TTGAAGACTA 60

GCCAATGTCG CTTTATGAAC AGATAGTtCT TTTTCTGCTG TTGCTTGTGT TGTTTTGGCT 120

GTATTTAAGA CTTGCTGACT TGCACTTTTT GCTTGTTGAG CGTCTGATTC TGCTTGTTGT 180

GCAGTAGCTA ATTTTTCTTG ATTGGTTGCT ACTTCTTTTT GAAGGGTTTC TTTACTCTTA 240

AAGACTGGAG CAGTGCCTTC TTCATAGGTA TTTGCTTCAA CTAATTTTGA TTGATTATAG 300

CTAATGATAT GGATTTTTGC CGAAATATCA TTTAAATCTC CTGAAATAGA AACTGCCATG 360

TTTGTTGTGT CCCCTTTTAG CAATGAATCT ATATGAGAAT ACGCTGAACT TTCATCGTCA 420

AAGAGCATAT CAGCAATTGT TTTTCGAGCA GCTTTTTCAA ACTCTAGTTG AGAAATAGTG 480

CCATTAATCG TTTCAAAATA GCCCATACTT AGATTTTCAT AAAAGTTTTG CCCTGGATAC 540

TCTTTAAAAC CATTTTCTTT TGCTGCACGA TTAATTGCAT TTACGTCATG ATCAAACTCT 600

TTGGGATTAT CATATTTTGC TACATCCCAA GCAAATCTCA TAGCTTGGTT AGAAATCTTC 660

AACTTTTGTA ACCCAAGCTT TTCGCGAACA TCATTGATAA GTCCTACAAC GTAATCGCTC 720

ATTTGTTGCT TTTGTTTTyC GGTAGGATTT TTGACATTGA TCATTTCCGT TrCTTTTGCC 780

rCATCTGCTT GATTTTCAGG AAAAACTTTA TTTAATGCTA ATGCTTCTTT TTCCATTGCT 840

TGCTTTTCTT GTTCAGwTAA TTTTTTGTAA TAATCAGGTG TGAAGTTAGC AGGyAAGTTA 900

AkTCCTTTAT GyCCyTTTAr CTCTTCTTTT GAkyTATCTA AAGTATTCTG CGCTTGCTTT 960

GTTGCTGCTT GCTTGTCTTT yAAAGCATTt TCTTTyGmAG TTAAATCTGC CTGTTTCGCA 1020

GCTGTATCTG CTTTTGCGTC TGCTACTTTC TCTTGACTAG TTGTCACAAC rGTTTGTTGG 1080

TCTGCrACTG CTTGATCGTC CTTCGTyTTT TGATCTTTAG CTTTkGCTTC TTCTTTCTGT 1140

TTGTTAGCCA GTTCAGCTTC TTTTGCTGCT TGTTCTTTTT CTGCCTTTAC TTTTTCCGCT 1200

TCCGCATTTy TCGCTAATTC TTCAAGCTTG GCTTGTTCTT GTTTTGCAGC TGTTACTGCT 1260

TGTTGATCTT TTTCATTTTG CACTTTAGCA GCATTCGTTT CTTTTGCTTT TTCATCGACA 1320

ACAGCTTGTT GTTGGTTAAC GTCTGTTTGA GCTTGATCCA CTACTTTTTG TTGTTCATCA 1380

ACAGCCTGTG TATCCGTAGC CACTTGCTCT TTAGCTTTTT CAATGGCTGA AGGTGTTGCT 1440

TCATCCACGA CTTTTTTAGC TTCGTCTACA ACTGCTTGTT GG 1482

(2) INFORMATION FOR SEQ ID NO: 843:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 315 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 843:

GGAATCATGG CTAAAAATGA CCCTAAATAA GGGATTAAAT TTAAAAAGCC TGCTGCAATA 60

CCTAAAGTGA CTGCATAATC TAAGCCAATG ACAGAAAAAC CAATCATAAA CATAATTGCC 120

ACTGCGAAAG CGACAGTTAa TTGTCCTCTA ATATAAGAAG AGACTTGATC ATTGACTTCT 180

GCAgAACTTT TAgGTTGGTT TACGCATCTT CACTGGTAAA AACTTCATTA AATtGGCGCT 240

AgTTTTTGCC ATCTTTCAGT AAATAAAATA GAATAAcGGC AtCGTAATGA TGGCGACAAA 300

GATCGTAGCC ACCGC 315

(2) INFORMATION FOR SEQ ID NO: 844:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 232 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 844:

TTTTATTCTA TTAACCAATC CGGCGCATTG GAATATCATT GTTATCTGGA TGAACCAATA 60

AATATTGAAT AACATCAATA TTGCkTGCTT GGAATGAGGc TGCACATGCT TGCAAATATA 120

AgTCCCACAT TCGATAGAAg CGCTCGCCTT TTTCGTCAAC AATTTCTGCn TCTATATTAT 180

GGAAGTTTTT TGTCCAATGT TCCAACGTCA ATTGAnAATC TCTGCGCAAA CT 232

(2) INFORMATION FOR SEQ ID NO: 845:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 897 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 845:

TCGCTGCATT TACATCTGGA TATGGGCCCA AnATATAACG CTTTGTCTTT TAAAACCTTT 60

CGAGTAATCA TTAAACGAGG ATATTTTTCG TTTGTGATTT TGATAAAAGG ATAGGTTTTA 120

TCGTCTTTTA ACATGATGTn ATATTTGGGA TTATTTTTAT GGATTAAATT AATTTCTAAT 180

AAAAGTGCTT CAATATTCGA TTCTGTCACA ATATATTCAA AATCATCGAT TTCACTGACT 240

AATCGTTCTG TCTTAGTATC ATGGCTACCA CGAAAATAAG ATCGGACACG GTTTTTCAAA 300

ATTTTGGCTT TCCCAACATA GATAATCGTG CCGTTTTTAT CTTTcATTAA ATaACAACCA 360

GGTTGATCAG GAAGCAGTGC TAATTTATTT TkTATTCGTT CGTTCATACG ATACTTCCTT 420

TCTAACAAAA TGCTTCGtCT ATTATACACC TCTTTTTGTA ATTAAACAAA AAAAGCCGCT 480

ACTTTATGTA CGAAAGAAAC AGCCTGAAAA GACAAtCGCT AATGATTTCc TCTTCAGACT 540

GAATCAAACT ATTCGTAATG ACTAATCACG TTTATAAGTG TTTGGCAATC ATTGCTTGCA 600

ATTGGTCTTT TGTATGAACA CCGACTGCTT TTTCAACAAC TTCGCCATCT TTTTTCAATA 660

GCAATGTTGG GATGCTCATG ATTCCAAAAC TTGCTGGTGT TGCAGGATTT TCATCCACAT 720

CCATTTTGAC GATTTTCACT TCGTCTTCAT CATATTCTTC AGATAATTGT TCTAAGATGG 780

GTGCTTGCAT ACGGCAAGGG CCACACCAAG TTGCCCAAAA GTCGATTAAG ACTAATCCTT 840

CATCTGTTTC TGTTGCAAAA TCTTTATCGG TAATTGCTTG TGTCATTTCA CGGGATC 897

(2) INFORMATION FOR SEQ ID NO: 846:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 740 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 846:

```
TCGAACTGCC GACACACGGA GCTTCAATCC GTTGCTCTAC CAACTGAGCT ACCACGCCTA      60

AAACAAAAGC TTGTTCAAAT AATTTTCAAG TAACAACATT AAATGGCGGT TCCGACGGGA     120

TTTGAACCCG CGATCTCCTG CGTGACAGGC AGGCATGTTA ACCCCTACAC CACGGAACCA     180

ACCGAGTAAT GGAGGTTAAC GGGATCGAAC CGCTGACCCC CTGCTTGTAA GGCAGGTGCT     240

CTCCCAGCTG AGCTAAACCT CCAAAAAGGA ATGTTACTAT TAAAATAAAT GACCCGTACG     300

GGACTCGAAC CCGTGTTACC GCCGTGAAAG GGCGGTGTCT TAACCGCTTG ACCAACGGGC     360

CAGTTTGTAA ATATGAAACG GAGAGTAAGG GATTCGAACC CTTGAGACAG TGTTTACCGC     420

CTACATGATT TCCAATCATG CTCCTTCGGC CTCTCGGACA ACTCTCCATA GCTGAAAGAA     480

GCACATACTC ATGATATTTC ATATGTTAAA CTCCGGCAGT AGGACTCGAA CCTACGACAT     540

CATGATTAAC AGTCATGCGC TACTACCAAC TGAGCTATGC CGGAATAAwy GCGTGGCGAC     600

GTCCTACTCT CACAAAGGGA AACCCTTCAC TACAATCGGC GCTAAGAAGC TTAACTTCtG     660

TGTTCGnCAT GGGAACAGGT GTATCCTTCT CGTATCGCCA CCACACTGGG TGTTGTnTCT     720

TATTGAGTTG AATCTTCATT                                                 740
```

(2) INFORMATION FOR SEQ ID NO: 847:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 615 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 847:

```
CATTAATACA GGTATTAATC TCTATAATGA TGGTCAACTA GACGATGCGC CTGTAGCTGG      60

GGAATATAGT AAGCAATTAG AAAATnACAA AGATTTCATA CGAGAATTAT CTGCTACAAC     120

TATGTTTTTA GAGGTAAATC AACGAAATAA AAAAAGTATA ACTTCAAATA AACATGCTCG     180

ACAAGCGATT AATTTTGCGA TTGATCGTGA GGCTATATCA AACAAGATAT TAACAAATGG     240

AAGTATACCA GCTAAAGGTG TAGTACCGAG TAAACTAGTA TATAATCCAA AAACAGGAAA     300

AGATTTTACT AATTCAAGTT TAGTGTTTTT GGATAAGAGT AAAGCTAAAG ACAGCTGGGA     360

AAAAGCTAAA AAAGAATTAA AAGGTACTGA TTTATCTATC GATATTATGG TCAATGAAGA     420

AGATTTATCA AAAAAACTAG GTGAATATCT TCAAAATGAA TTGCAAGATA CTTTGGATGG     480

ACTTAAAGTT TCAGTTACTG CAGTACCTGC CACGTTACAA ACTGAACGAT TAAATTCTGG     540

GAACTTTATG ATTGCATTGA GTGGTTGGCA AGCAGCTTTG CCGATCCAGT TAGTTTTTTA     600
```

GCAAATTTTG AGAGT 615

(2) INFORMATION FOR SEQ ID NO: 848:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 926 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 848:

CGTTGTAACG CCTTTTGCAA CCATGCTTGC TTTACGCAAT TGACGGCGTT GTTTTTTAGT 60

TTTACCGTGG AAACGGTGAG ATGTAAAGGC ACGGCCACGT TTTAAGCCGC CTCCACCAGT 120

ACGTTTTACA CGTTTTGCTA ATCCGCGGTG TGTTTTTTGT TTTGGCATGA CTAATTTTCC 180

TCCTCAAAAT CTTTGCTTCA AGTGAAAGAC TCTAACAGTC ACAACTGTCG AATATCCTTC 240

TTGTAGAAAT TACCAAAAGT CCGAATGCTC GGTCTCTTCG TTTACTTGTC GTTTTTCGGT 300

GCCAGCGTTA GAAACATGCT GCGTCCGTCC ATTTTCGCTT TTTGTTCCAC TGTGGCGATG 360

TCCGCAGTTT CTTCAGCTAA GCGATCAAGA ACTTTCTGAC CAATCTCTTT ATGGGTAATG 420

GCACGGCCTT TGAAACGGAT AGAAGCTTTC ACTTTGTCTC CTTTTTCTAA GAACTTACGC 480

GCATTACGAA GTTTTGTATT GAAGTCATTT ACGTCAATTG TTGGACTTAA ACGAACTTCT 540

TTGACGTTGA TCACTTTTTG TTTCTTGCGA GCTTCACGCT CTTTCTTTTG TTGTTCGAAA 600

CGGAATTTTC CATAATCCAT AATTCGCGCA ACTGGTGGTT TCGCTCCTGG CGCTACAAGT 660

ACTAAATCCA AGTTTGCAGA TTCAGCGATT TGTAATGCTT CTGCTTTGGT TTTTACACCT 720

AATTGTTCAC CATCTTGTCC GATCAAACGT rACTCGCGTG cACGAATGCC GTCGTTAACC 780

ATCATATCCT TtGCTATGGT CATTCACCTC CAAAATTTTn AAGAGAAAAT CCTAACGGAA 840

GTCTGCTTTG TTTGCTAAAT GCAATAAAAA AACAGGTTCC TATCCAGAAC CCGCCAACAG 900

TCTGCCTTAC TTAAAAAAAT AAGGAA 926

(2) INFORMATION FOR SEQ ID NO: 849:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 206 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 849:

ACTCAAACTA TTTCGAACGA TAATCGTATA CCAGCTGAAT TTGTCCGTCA TGTTGATGGA 60

GATACCACCG TGTTAAGAAT AGACGGAAAA GAACAAAAGG TTCGCTTTTT ATTAGTAGAT 120

ACACCCGAGA CTGTAAAGCC GAAAACAAAA GTACAGCCAT TTGGTTTGGA AGCCAGCAAC 180

CGTACAAAAG AGCTTTTGTC TACTGC 206

(2) INFORMATION FOR SEQ ID NO: 850:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 310 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 850:

ATCAAATTCT TTTTCTTACA ATAATTAAAA ACCATATTTA GAACCTGAAC AAGAGATTGT 60

AAACTATTTT TAGCATAACC AACATTAACG CCATCAATTC CCTTTTCTTT TAAAATGAAA 120

AGGAATTTTT GTAGTTCTTC TGATGATAAA CTACTTAATT TTTTACCACC AATCCATCGA 180

TTAAGTATTT TCACCTGTnC TTCTCGTTTA TCATACGTTG CCACCTTTTG TTCGTTTAGT 240

GATCGATAAT ATTCAAACCA CTTTTGACTA ACAACACTAA ATTTATCATT ACTAGCAAGC 300

CTnTnTTGCT 310

(2) INFORMATION FOR SEQ ID NO: 851:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 257 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 851:

CCTTTGAACA TTACTGAAGC CAACACGTAA TCGACTTCTT CCCACGGAAT TTGGATAAAC 60

TTACGTGGAT CTTTGGCATG GTAAAATTCA AACCCGCGGT CACCTATCAT AATTTGTCCA 120

TAATCTGTTA ATCCTGTAAA TGCCGTCGCA TCTATAACTA AATCAACTTT TGTGTTTAAT 180

GATTGTACCA TGTTATCTkC TCCATTTCAT TCAACTAATT CCTCTTTATT ATACTTTTCT 240

TTCCTCATCT AAACAAA 257

(2) INFORMATION FOR SEQ ID NO: 852:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 731 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 852:

CACACCTTAC TCAGCAAATC TCGTCACAAA TAGTACTACA AAAACCTTCG TGTTATTATT 60

AAATAGCCAT GAAACTTTTG AGGCTTGGGT TACGAGCTTT ATTGATAAAG CTCAAGAAAA 120

TGGCTATAAC TGCTTGATTT TAAATAGTCA CGGTAGTTTA GAACAAGAAC TGAAAAATAT 180

TACCTCGCCA ACTTTAAAAC AAGCCAGTGG CATTATTTGG GACCCAATCA ATGAAAACAG 240

TAAACAATAC GCGAAATTTA TAGAAGAATT AACGATTCCT TACTATTTCG TCAATCACTC 300

AACTAACGAA TTACCTTCTC TAGCGGGGCT TTATGAAGAA GCTAGTTATG TTTTAACCCA 360

AGAATTATTA CAGAATAACC ATAAAAAAAT TAGTTATTTC GTCACTAACC CmAAATwTAA 420

AGrrGCmATT ATTGCmGGcT ATAAAAAAGC CTTGTTTGAT GTCAATTTAC CCTTTAGTAA 480

AGAAAATATT TTAACTACTA TCCCTGCTGG ATTTCAGTAA TCAGCTTATC ACAGATGGGA 540

TTACGGGTAT CCTTACTGGA TGACTATACC CAGGCTGTTT TTTTGGAACA GCTTTTAAAG 600

CAGAGCGGTT TGCGAACCCC CGATAATTAC TCTTTACTCG CAATTAAGCG GAAAAATGGA 660

TCGGTGCCAT TTTACCGGAT CCnAnTTCCT nCGnCCTAAT TAGATACCGA GACGGTTGGG 720

TAGCCAGTTA G 731

(2) INFORMATION FOR SEQ ID NO: 853:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 142 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 853:

TCACTCGTTT ACATACGAAA GACAGCTGCG ATTTTGTAGA mGCATCTGGT TCTTTCTGCC 60

AAAGCAAGCG ACAAATTTGA TCACGATCCA ACCCTTCGTG GGCATGCTGA TACAAGATTT 120

TTAAAAAGCG TTGCTCGTTT TT 142

(2) INFORMATION FOR SEQ ID NO: 854:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 301 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 854:

AACATACACA TTTCCTTCTT TAACCAATGG ATTTAATTGA TCTAACACAC CAGCTAATTC 60

GCCAGAGTTC TTCTCTGCGT AAAACTTGGC CCCAAATAAA CCACCACCAA TAACAATTGC 120

TGTCACCACT AAAAATTTAA TAAATTTCAT CACATAAATC CTCCCGAGTA ATTATTCTCT 180

GTTACTTCTT CCAAAGTATA GCAAGTTTTA AACAATCCTT CCTCGGGCGC AACGACCAGT 240

TTTTCCTCAG ACTTATGTCG TATnTnTGTT ATTTTGAATA AATAACAAAA CACTACTCnT 300

C 301

(2) INFORMATION FOR SEQ ID NO: 855:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 330 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 855:

TGCAGCAGCT GCAGTTGCCG ACTGAAAATC CCCGTTGTGA TCCCAAAATA AAGCTTCTAA 60

CATTTTTGCA CCTCGATAGT TTTATTTTAA TTATCTACTA ACTATCGATA TCTTTCAACT 120

TAAGATTATC CTCAATCAAT ATTTTTAAAT CAGAAAATTA ATTCAATTTG CATTCTAATC 180

CATCAAATTC TkGTTAATAT TATTTTGAAT GTCTGCTTCA TCAAAGAAAC CATGCCCACA 240

ATAAACCAAG CTTACAATCA TCAATTTCCC TTGGCGTGGC TCCCTCGGnC ATTTCCAAAA 300

ATAGGAnGCA TCCTTTTTAA TCTGGCACAG 330

(2) INFORMATION FOR SEQ ID NO: 856:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 582 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 856:

AAGGTGTTAG AAAAAATGCT ACCTAACTCT TCTGATGTTA CTTTTTACAA ATTTTTAAAT 60

AAACAGCATA GTTTGTATGA ACAACAATTT GAATTTTTAA AAGAAGTAAT TCAATTTAGA 120

AAAAATAATG AATATGACCA ATATAGCCAA ATAGGGTGTA ATTTATTACA CTATGATTTA 180

TTAATTTCAG GATTACTTTT TGATGTTAAG TCTTATGATT ATTCATATCA ATACTTTTTA 240

AAATTAGATT TAGCTAACTA TTTTTCTGAA TTAGAGATTC CTGATGAATA CAAAATGTTA 300

ATAAATAAGT ATCGGGAGAA TGAATCATGA TTAAGAGACC TCATGkAATG CAAGAkAGTA 360

TACmAGATgT GGTGkTgCAT GCATAGrAAT GrTATGkAAA TTCyaTAATA TTAATATAGr 420

TAGaaGATAT ATCCAAGAAG AAACTGGCTA TGGAATGATA GGAATATCAT TAAAAGCTAT 480

GGAAAAATTT TTTTCAAAAG TAGATGCAAA CCCAGAAATA GTAAATATAT CTAAAATTAA 540

TAGGTTAAAC AAAGAAAATA GGGAAATGAT TAATAACTCA CT 582

(2) INFORMATION FOR SEQ ID NO: 857:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 484 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 857:

TGCAnAAATT AAAAGAATTG CAATAGCTAA ATAGTTAGTT tGATTAAAAT CTATTACAAA 60

TTTTTTTAAT AGTTCTAAAA CTGTAGATAG TGCAAGTAGA GCTATAAATA TTAATATTAT 120

ATTTAAGGAA GTTTGTATAT TTTTTTTTAA AGTTTGACTA TTGTTATTTA CACTAAAAAA 180

TGTTATTAAT GACTCTTTTT TACACATTTC AATAACTTTG ATTCTAAACA GTAATATTTT 240

TTTTATATTT GTTTTTAAAT CGACTTTTCG TTTTTTTTCA TCTAGAATGA CTTTTCCATC 300

TCTGGTTAGT TCAACAACAC CCAGAAACAG TTCTTCTTTA TTACTATTCC ATAATAATAT 360

CGGCCCCTGA CCATGATTTT TATCAATATC AAAATCAGAw ATAATAACTT TATCTGCAAC 420

TTTTCTTAAA TCATTTTCTA GTGTTTCTGA AGTTAAGTTC TTATTTAATT TATGGATCAA 480

TGTG 484

(2) INFORMATION FOR SEQ ID NO: 858:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 587 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 858:

GCGTAACTCA TGTGTTCACC CCCTGTCCGT TGCTCTCTGA AACAAATGGT ATAAGGTAAA 60

GGGCAGGTGA ACCTTGGTCT TTACGAAAAA AGTTTTGAnA ATTTTTTTGT GCCTGATTGC 120

GCATTTTGCT ACCTCCGAAA ATTTGAAAAT TAGTGAAATT CCAAACCTTT TTGGAGGTTT 180

AATGTCTGAT TAGCATATCT GTATACGTTT GATAATGCGC TTAAATGATT TTGAAGCGGC 240

GCGnTCTGTG CCACACTTTT TTTGACGGGA TACCCGCCGT GTATATAAAA ATAAAACCAT 300

CCATGTACTG TCCCTTTCTA CGGACAGACA CAAATGGCAA AGGAAGTAAG CCCTGATGGC 360

GGTATTCTTC TTTGCGCTGG CGTTCCATTT CACGCTCATA GTCGGCAAAG ACCTGATAGA 420

TTcTCTGGAT ACCTCGACCG TGATGGGCTT ACCGCTACCA TCAAAAATCG TAACGTAACA 480

CATGGcTCCC TCCTGTGCAG CGTCATCCAg GCAGAGAAGA ATGCCAkGGG GGAAATGTCG 540

TCATCAGGAT CAAGGGGAGC GnCCCATGCT GCACAACCGA GGGGCAA 587

(2) INFORMATION FOR SEQ ID NO: 859:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3549 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 859:

GCCnCCnTTC AAAGGAGTnG CACCATCAGG AAAATTAGCT ACTGCACAAC AAGCAGAATC 60

AGCGCTCAAC AGCAAAAAGT GCAAGTCAGC AAGTCTTAAA TACAGCCAAA ACAACACAAG 120

CAACAGCAGA AAAAGAACTA TCTGTTCATA AAGCGACATT GGCTAGTCTT CAAGCAGTTG 180

CGACTAAAAG TACGACAAAT TATGAAGAAA AAGTACGACA AACTGCAACG GCAGAAAAAA 240

GCCTTCAACA AACAAAAGAT CAATTGGCAA CAATCAATGA GTTAATTCAG AATCGAGCTG 300

CTGTTTTAGA AAAAGCGAaA ACAAACGTTT CGGAAGCACA AGCAATTGAA CAAACGTCTG 360

CTAAAGTGCT GAAAGAAAAA CAAGCAGCCC AAAAAGCAGA AGAAAACACA TTGAATAGCT 420

TGAAGGAAGT ATTGAATTTA GCAAAAGAAA ATTTAAATCA AAAACAAGTT GCATTTAAAA 480

CAAGTACACG TTCATTGTCT CGTTTAGAAA ATGCTCAACC AACATACGAA AAAGCATTAA 540

ACGAGTTAAA CAAAGCAGAA GCAGCAGCGG TCCAAGCACA AGAAGCCTAT GAAAATTCTC 600

TGAAATCATT GGAGGAACTT AAAGAACAAC AAGCCGTTGC TACACTTGCT TATACACAAG 660

CACAAGAAGA CCTTTCTAAC GCGAAGTTAG AGCTACAGCA GTACCAGGGC GTATTAAGAG 720

AATTAGAAGC aCAACAAGCC GaACAGCAGC GACAAGAAGC GTTGCAAGAA CAAGTAGCAA 780

AAGAACAACA ACGCCTTGAA CGAGAAGCAA AACAAAGCCA AACGTTAGTA GCAAGTGCTA 840

CTTCAGCAGA CAAAACACCT GGTCTCCAAC AGTTATCTTT TTCTAAACAA AAAGAACAGC 900

CAAAAGCACA AGCATTAACA CATTCAGAAT CTCGTAAGAC GAAACAAGTA GCAAAAGCCC 960

CAGATTCTTT ACCACATACA GGAGAAAAAA ATAATAAATG GTTAGCTATA GCTGGTCTGA 1020

TATTTGCTTT GTTAGGAGCT GCGGGTATTA TAAGTTTTAT TAGTAGAAAC GAGAAGAAAG 1080

TAAAAAATAT CTTTAAAATT AAATAATAAA GTCAAAGGAG CGACATATTT CGCTCCTTTT 1140

TATTGTAATC ATTGGAATAA ATTGAAAAAT AAAAAAATAG CCAACCAAAA TGGTTGGCTG 1200

ATTGTATTAA AGGCAGAAGA ATCTACCAAA TTTAGGAGCC AACCAGTTTC CGCTGGTCAA 1260

CAAAAAATAA ATGGATTGTC TCGGATGGAG ACACTTAAAT AGTACCTGAC AATAAAATTT 1320

TAGGCAACAA AAAAGTCAGC CGTTTTCCGC CGACTGACTA AACGAATAAG ATTTGTAAAG 1380

AAGTGATTTG TACTTTTTAC ACAATTATGA TAGCACGTTA AATGTTTTTA ACGCAAAAAA 1440

AAGCTAAGGG AATCGAAAAa CTTTCGGTTC TCTTTTTGTG AGGAGGAATG GAAATAATGA 1500

TGAAAAAGCA ACGAGCTCCT TGTGTTGCGT GAACTATTTT TACAAACAAA AAATAAAAAA 1560

TCGAGGAGAA TGATACATGA ATCAACAGAC TGAAGTAAAA AAACGTTTTA AAATGTATAA 1620

GGCAAAGAAA CACTGGATTG TAGCCCCTAT TCTTTTTTTA GGTGTATTAG GAGCTGTAGG 1680

ATTAGCTACA GATAATGTAC AAGCCGCGGA ATTAGATACG CAACCAGGAA CAACGACGGT 1740

GCAACCCGAT AATCCCAACC CGCAGTCACG AAATGAAACA CTTAAAACGG CAGTATCTGA 1800

AGAAGCAGCA TTACAAAAAG ACACTACTTC TCAACCAACC ACAGCAGAAG AAGTAGTGCC 1860

GAAAGGAATT GCTGCTGAAC AAAGTTCAGC TACCTCAAAT GATACCACAA ACGTCCAACA 1920

ACCAACAGCA GAAGCAGAAA AATCAGCACA AGAACAACCA GTAGTCAGCC CTGAAACAAC 1980

CATTGAACCT CTAGGGCAGC CAACAGAAGT TGCACCAGCA GAAAATGACG CTAATAAATC 2040

AACGTCAATT CCTAAAGAAT TTGAAACACC AGACGTTGAC AAAGCAGTTG ATGAAGCGAA 2100

AAAAGATCCA AACATTACCG TCGTGGAAAA ACCAACTGAA GACTTAGGAA ATGTTTCTTC 2160

TAAAGATTTA GCTGCAAAAG AAAAAGAAGT AGACCAACTA CAAAAAGAAC AAGCCCAAAA 2220

GATTGCCCAA CAAGCAGCTG AATTAAAAGC CAAAAATGAA AAAATTGCCA AAGAAAATGC 2280

AGAAATTGCG GCAAAAAATA AAGCGGAAAA AGAGCGCTAC GAAAAAGAAG TGGCGGAATA 2340

CAACAAACAC AAAAATGACA AAGGCTATGT GAATGAAGCA ATCAGCAAAG ACTTAGTTTT 2400

TGATTCAAGC ATTGTGACTA AGGATACTAA AATAGACAAG ATTACAGGTG GCAAGTTTAT 2460

TAAAGCTTCC GATTTTAACA AAGTAAACCA AGGACAATCC AAAGATATTT TTACAAAATT 2520

AAGTAAGGAT ATGAATGGAA AAGCAACAGG CAACTTCCAA AGCTCAAAAG TGGCTGCTGT 2580

TGAATTTGGC CCTAAAGGAG GATACGCGGT TCTTTTAGAA AAAAACAAAC CGGTCAATGT 2640

CACGTATACA GGATTAAATG CTAGTTATTT AGACCGAAAA ATCACAAAAG CAGAGTTCAT 2700

TTATGAACTT CAATCTGCAC CAAGTCAAAG CGGCACCTTA AATGCAGTAT TTTCGAATGA 2760

TCCTATTATT ACAGCATTTG TTGGTACTAA AAATGCCAAT GGGAAGGACG TAAAAGTTCG 2820

CTTAACCATT AAATTATATG ATGCTAATGG CAAAGAAGTT TTACCAGATA AAGACCATGC 2880

CTTTGCCTAT GCGTTATCGT CATTAAACTC TAGTCTAGGA ACAAATTATA GTGTAGAACA 2940

TGCGGAATTT GTTTCAGACT TTGGCTCAAA AAATGAGTTC AAGTATATTA ATGGTTCCTA 3000

CGTAAAGAAA CAAGCGGATG GGAAATTCTA CTCAACCGAA GACCTTGATT ATGGGACAGG 3060

TCCTTCAGGA CTAAAAAATA GTGATTGGGA CGCTGTAGGT CATAAAAATG CGTACTATGG 3120

TTCTGGTGTA GGTCTAGCTA GAGAGGGTGG ACGTATTTCT TTCTCTTTTG GTATGACAAC 3180

CAAAGGAAAA GTCAACCTAT CTGGTGCGCA ATGGTTTGCC TTTAGTACCA ATTTAAATGC 3240

GAAATCAATT AGACCATACC AAAAGAAAGG GAATCCaAAA GAACCAGAAA AAGCAACAAT 3300

TGAATTCAAT CGATACAAAG CCAATGTCGT TCCTGTTCTT GTGCCgAATA AAGAAGTCAC 3360

TGATGGTCAG AAAAATATCA ATGATTTAAA TGTGAAACGA GGCGATtCTT TACAATACAT 3420

TGTGACAGGG gATACGACAG AACTTGCCAA AGTAGATCCG AAAACAGTGA CAAAACAAGG 3480

GATTCGGGAT ACCTTTGATG CAGAAAAAGT GACGATTGAT TTATCCAAAG TGAAGGTTTA 3540

TCAAGCAGA 3549

(2) INFORMATION FOR SEQ ID NO: 860:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 308 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 860:

AATTCTAAAA TTTATTTTAA AATAAGAATT TGCATTGGCA TTTACTTTGT CAGCGGGTAT 60

ACTTATACTT GATGTATTTT GCATTTATTT CAGAAAGGAA TGGTGACATG CTTTTTGCTA 120

ACCTATGGAA AGCAATTATT CTTGGTATTA TCGAAGGAAT TACCGAGTGG CTTCCAATTA 180

GTAGTACCGG ACATTTAATT TTAGTGGATG AGTTTATTAA ACTCGATCTA AGTAAAGATT 240

TTATGGAAAT GTTCAACGTC GTTATCCAAT TGGGAGCGAT TATGGCGGTT GTnATtTTAT 300

ACTTCCAC 308

(2) INFORMATION FOR SEQ ID NO: 861:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 175 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 861:

CACTATCACA ACAAATCAAC AGTTTGGAAT GCTCAGGAAA AGATTTAGCA GATATAAACC 60

CAACTAACAC TACAGGTTGA AGAGTCCAAC GCTTGCAAAA CAAGAGATGA AATTACTGCA 120

TGCAGCATAA AGCTGCAGCT TTCAGGACCA ATAGCAGGGG ATCTCCCCAT CCACC 175

(2) INFORMATION FOR SEQ ID NO: 862:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 198 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 862:

TTCCTTTAGT GAAGAGAGTC TTTGGTTGGT GGAAAAAGAC GTAAAGAAAC AGCGAACCAC 60

ATCTGGGAGT TGACTTTTTG AACCCTTAGT AGGAAAGTCC GGGACCACTC GTTAACAGGC 120

GAaGTTATGA ACTTCATAAG GtCGTGTTGC GAAACACGAA TAAACTAAGG TGGAACCACG 180

ATAATTGATT ATTCGTCC 198

(2) INFORMATION FOR SEQ ID NO: 863:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 372 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 863:

TGATTATTTT AATAGATTGG TGTAGTTTAA TATCAAAATC AATATTATAG GTGATAATAC 60

ACATAGAGAA ATAAArATAA TCAAAAAATA GAACCATTTT TTAGAAACTA GCATACCTAA 120

TCCTATCGTT CCTAGTGTGT ACAGTACTAT CAAAATAAGA TAGATTTkkG TTCCATTTCC 180

TGAAAAATCC GTCTTGCTCA CACTAGTTAT CATTCGAAGA ATAGCATATA TAATCCCCAA 240

AGAGCACAAC ATCCAAATAA AGCTACTAAT ATATTTAAGG AGATTCATAA CCAGACCTAC 300

TTTCTATTAn ATTACTnAGA GTAGCCTTTC CAAGTTTGTG TGTGCGTCGC TCTAACATCA 360

AAGTAATCAG GA 372

(2) INFORMATION FOR SEQ ID NO: 864:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 249 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 864:

CATTGGTTTA AAAATGACAA ACGGTCTTGA AATTTTATTA CACATGGGTA TTGATAnGGT 60

AGAGTTAAAC GGGGCACCGT TCACTATTAA AGTGAAAGAA GGnGACCAAG TGACTGCAGA 120

TACGGTGGTC GCAATTGCTG ATTTAGAAGC AATTAAAGCA GCTGGTAAAG GAACTGAAAT 180

GGTTGTCATT ATTACCAACA TGGATAAAGT AACACAATTT TCGTTAGAGA AAACTGGCGT 240

GGTTACAGC 249

(2) INFORMATION FOR SEQ ID NO: 865:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 683 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 865:

ATTTTCCGAG GAAATATTTA TAAAATGCAT ATTTACCGCC tAAAGCAAAC GGTGGCGATA 60

TGAGTTTACC TGTTACCTAT TTGTTAAAAA ATGCCACAAA ATTTATTTAT GTCAACACTT 120

CATTTTTCAA GGTATTTTTT ACTATTAAAG AACTGCCTTA CATTGCTGTT TTAGATGAAA 180

ATAATTATTT TTACGGCATk TTAACACACA GCACCTTATT AAATATTTTA GCGCAATCTT 240

GGAATGTAAA ACAAGGGAGC TACGTTTTAA CAATTGCTTC AGTTGGTCAA CAAGGCGACT 300

TAGCAGCCAT TTCAAAAATC ATTGCCAAAT ATAGCAGCAT CGCCAGCTGT ATTACATTGG 360

ATGTCGATAG TGAAGAATTT GTTCGTCGCA CACTTATTAC TTTACCAGCC GGCACAACTG 420

CCGAAACGTG TACAGCCATT GTAGAACATT TAGAGCGGAA AAACTTTAAA GTTGTAGAAT 480

TGGAAAATTT AGAAgACGAG TAAAATTTAC GACCTTTAGC TTAGGATTTA mAGCTAGGGA 540

GGTAAGCCAA TTTGACTTAC CTCCcTAGCT TATTTTTTTA GACTTCTTCA CCAATAATTC 600

GAACTTCCGT TTGCAATTCA ACATCAAATT TTTCTTTAAT CACTTCTTGG ATATGTGCAA 660

TTAATTCCAC ATAATCTGTT GCG 683

(2) INFORMATION FOR SEQ ID NO: 866:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 716 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 866:

AATGTAATAA TGCCCATTTG AACTTGTGCA TTTGGTTGGT TTTCTTTATC AAAAAAAATC 60

GTCATTTCTT CTGATTCTGC TGATTTAATA AGCAATAGCT TCCCCTCCTT TTTTAAATCT 120

TGTATCTTCC ATTATTCCTT TAAATAGAAT CAAAGTCAAT CTTTTTATAG AAAATTCTGA 180

ATATTCTACA CAAAAAAAGA GAGCATGAAT TCTCTCTCTT TTGnCTCTTT AAACTTTTAC 240

AATCCAGCcT TCTGGCGCTT CTTTGTCACC AAACTGTATG CCAACCAGTT CATCATATAG 300

TTGcTTAGtC ACTGGtCCAA CTTCCGTTTC ACTATAAAAT ACATGGaAAT CTGTGCCGTT 360

TTGaATCCCA CCAATAGGAG AAATAATCGC CGCTGTGCCA CAAGCACCCG CTTCGGAGAA 420

GTCCGCTAAT CGATCAATGT AAACGTCTCC TTCAAGCGCC GTCATCCCTA AACGTTCTTT 480

TGCTAGCGTC AATAATGAAT ATTTAGTAAT ACTTGGCAAA ATAGACGnTG ATTTCGGCGT 540

AATAAAAGTA CCATCTTTGG TTATGCCAAA GAAATTGGCT GCACCAACTT CTTCAATTTT 600

CGTATGCGTA TACGGATCTA AATAAATGCA ATCACTAAAT TCTTTCTCAT GTGCTTCTGC 660

ACCTGGTAAA AGACTAGCTG CATAATTGCC CCCCACTTTA GCTGCCCCAG TTCCTC 716

(2) INFORMATION FOR SEQ ID NO: 867:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 582 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 867:

TACGTTTCAT CCATTTTCCA TGAATAAAAG GATTTTTTAT TTTTCTTTTT CCAAATTTGA 60

TAGAGTAGTT TGCCATATTC TTGCACCCAA CGATAAATCG TCGTATGAGA AACGTTAATG 120

CCaCGATCAT ATAAGATTTC yTGaACyTCa CGATAGCTAA GGTTATAACG rAGaTaGtAG 180

CCCACGGCTA CaATAATCAC ATCCTGCTGA AATTGCTTTC CTTtAAAATG ATTCATCGTC 240

ATTCCTCCkG CTATCTTTTT CTATTATTCT ACCTTATTTG ATAGTAGATT TAAAACTTTG 300

CAACAGAACC AATGGTATTA TAACAAGAAT TACTAATGTT AACTTTGCAC TAAGGTTAAT 360

TAGGACTATC CCATATAAAA GAACTTCAAA TATTGATAAC ACAACTTTGG GTATTGTAGA 420

ACTAACTAAA GTATACATAC TACTTGCATC ATTTGATAAT ATACTAGCGA TATTACCAGr 480

AAGAGAATGC TCAAAGAAAG AAGTTCGCAT ATTTAGTATT TGTTTTATAA CTTTAGCTCT 540

TACTTTATTT ACAATAATTA TTCCTATTTT TTGTAGATAA AA 582

(2) INFORMATION FOR SEQ ID NO: 868:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 422 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 868:

TGATGTCACT TTCTGTAACG GTACATCGTA AAAAGGATCT TCACGACGTC CTAAAAAGTC 60

TTTTCCTTGT TGCGGGAACA TCGCATATAA TAGGACATCC TCTTCAGATT TTGCATATTG 120

TTGGATTTCC TTTTCTAAAC TATGCAACTC TGGTTTCAAT AAATCCGcgG TCGACAAtGA 180

TTACCTCTTC ATCCCCAtGA TTTTTGCTTA ATCTCTcgCG ATAGTGcGGT GATTGCCATA 240

TcTCCGCGCA CATAAyCTTG ATTCATtCGk gATtAAgGAT ATCTTCCCCA GAAATCACAT 300

CCTTAAGGCT GTGGTCCAAC CATTGTGAAA GCGGCGGACT AGGGGGAACC CAAACTGCCT 360

ACTTAGGCTC AGCAAAACTC TCAATTGCTG GAACTGCCCG GAATGCTAAA AGGTGAnACA 420

AC 422

(2) INFORMATION FOR SEQ ID NO: 869:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1069 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 869:

TACTCAGGGA AATCGAATTT tCTTTCTCTT CCTGCAGGTA CTAAGATGTT TCAGTTCTCT 60

GCGTCTACCT CTAATCAGCT ATGTATTCAC TGAAAAGTAA TATCCTATAA AAGATATTGG 120

GTTCCCCCAT TCGGAAATCT CTGGATCATA GCTTACTTAC AGCTCCCCAA AGCATATCGG 180

TGTTAGTCCC GTCCTTCATC GGCTCCTAGT GCCAAGGCAT CCACCGTGCG CCCTTATTCA 240

CTTAACCTTA TCAACCTTAC GGTTGGGTCT TrATCACTTC TTCTAGCGAT AGAAGGTTAA 300

TCAATAAATA AGCAATTGAA CTTATtAAAA AACTCATTCA ACGCGGTGTT CTCGGTTTGT 360

TTTGATTCTT TTTACTTCAA TATCCAGTTT TCAATGAACG AATCTGATAG CAAAAGCTAT 420

CAATGGAGCc TAGCGGGATC GAACCGCTGA CCTCCTGCGT GCAAAGCAGG CGCTCTCCCA 480

GCTGAGCTAA GGCCCCGAAT TTATTTGATT GAGAGTAGAC CTCTCAAAAC TGAACAAAGT 540

AAAGACGAAA TGTGTATTTC CGTAATATTC CTTAGAAAGG AGGTGATCCA GCCGCACCTT 600

CCGATACGGC TACCTTGTTA CGACTTCACC CCAATCATCT ATCCCACCTT AGGCGGCTGG 660

CTCCAAAAAG GTTACCTCAC CGACTTCGGG TGTTACAAAC TCTCGTGGTG TGACGGGCGG 720

TGTGTACAAG GCCCGGGAAC GTATTCACCG CGGCGTGCTG ATCCGCGATT ACTAGCGATT 780

CCGGCTTCAT GCAGGCGAGT TGCAGCCTGC AATCCGAACT GAGAGAAGCT TTAAGAGATT 840

TGCATGnCCT CGCGGTCTAG CGACTCGTTG TACTTCCCAT TGTAGCACGT GTGTAGCCCA 900

GGTCATAAGG GGCATGATGA TTTGACGTCA TCCCCACCTT CCTCCGGTTT GTCACCGGCA 960

GTCTCGCTAG AGTGCCCAAC TAAATGATGG CAACTAACAA TAAGGGTTGC GCTCGTTGCG 1020

GGACTTAACC CAACATCTCA CGACACGAGC TGACGACAAC CATGCACCA 1069

(2) INFORMATION FOR SEQ ID NO: 870:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 144 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 870:

CCTGATGAAA TTGTCGACTT TTATGAAGAC TTACAAGAGA TTGATTTATC AGGAAAAATC 60

TTCGGCGTTT GTGGTTCAGG AGACACGTTT TATGATGATT TTTGTAAATC GGTGGACGAC 120

TTCGAAGCAG TCTTTACTCA AATT 144

(2) INFORMATION FOR SEQ ID NO: 871:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 871 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 871:

TCnGAAAGTT ATCCACAGAT GTGGATAATG TGGACAATCA GTGGCGAATA TGTGTTTGTA 60

TTGAAAAAGC TGTTAAATAA AGTAATTTTT TAGTTTGTTC TGTGGATAAC TCTGTGTATA 120

TGTGTATAAT TATGTGGAAA CGTTGTGGAA AGTAGCCAAA CTTCGTGCTA AAATAAGAAA 180

GAAGAGAAAA CGAGAAAGTA GTGGAATAGG TGAAAATTAA ACTAGTAACA GTCGGAAAAC 240

TGAAAGAAAA ATACCTGATT CAAGGCATTA ATGAATATCT GAAACGTTTA AATAGCTATG 300

CCAAGATGGA AATCATTGAA GTACCAGATG AAAAAGCCCC AGAAAAACTC AGTGACGCTG 360

AAATGCTTCA AGTCAAAGAA AAAGAAGGCC AACGTATTTT GGGGAAAATT AATGACAACG 420

AATACGTCTT TGTTTTAGCC ATTAATGGCA AACAATTATC TAGTGAAGAA TTTTCAAAAG 480

AnATTGAACA ATTAGGGATT AGTGGCAAAA GTAATCTAAC CTTTGTCATT GGTGGCTCCT 540

TAGGTCTCAG TGATTCAGTC CTTCAGCGAA GTAACCAACA AATTTCTTTT GGCcGCTTGA 600

CTTATCCACA TCAGTTAATG CGTTTGGTAT TAGTGGAACA AATTTATCGC GGTTTTCGGA 660

TTATGAAAGG GGAACCGTAT CATAAATAGG AAATAGGCTG GTTGCTCAGT AAGTTCTATG 720

AATCAAGnCA TTGCTTGAAA AATATATGAA GTAAAATATC TTTAGTCTGT TTGTCTGATA 780

GAAAATAAAA AAATTCAATC AAACTGGATT GAGAGAAACG TtACCGTTCA AACAATAAAT 840

AAAAaTGAgA GGAATGGCAA AAAmCAtGGC A 871

(2) INFORMATION FOR SEQ ID NO: 872:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 345 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 872:

AATATTTTTA TTAGCGTGTT CAAATCCTGA TTTCTTGTAT CAATTATTTT CTGTTCACTT 60

TCAATAAGAA AAAaGTCTTT ACTTCCATGG CTAATAATAa GTCTCTATTT TtCAtAATAT 120

gCAAAATtAT ATCTTAATag AAAaTATTTT TATTTTAAAG ATAAAAAATT CTTCATCCTG 180

CAATACTTTA TATTCTATAT TGTAATTATT CAGAATATTA CTTACAATAT ATAAACCTAA 240

TCCATTACTA TTTCCCTTAT TTAAATCAAA TTTAACATCA AATATTTTAT CCATATTCGA 300

AATTTTATTG TTACCGTATG AATTCTCTAT ATATAACCAA TCATT 345

(2) INFORMATION FOR SEQ ID NO: 873:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 443 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 873:

nCCCCCATTA TATGGAATAA CACCTGGAnC TATATTTCCA AAGGATCCCT AAAATTAnGG 60

TTTATGTCCA TGGTTTCAAA TGTATCTAAA TACTTTTCCG ACAAnCCTTC aCkGAGTAAA 120

TCaAATTGTA AAAGTAACTT TTGGCACTGC ACCTCCaGTG CCTTTTCGTC CAAATGATAA 180

TCTATTAATT GAGAATATAA TTTTTTAGTT TCTTTrGTTC TGTTGCAAAG TTTTAAATCT 240

ACTATCAAAT AAGGTAGAAT AATAGAAAAA GATAGCAGGA GGAATGACGA TGAATCATTT 300

TAAAGGAAAG CAATTTCAGC AGGATGTGAT TATTGTAGCC GTGGGCTACT ATCTTCGTTA 360

TAACCTTAGC TATCGTGAAG TTCAAGAAAT CTTATATGAT CGTGGCATTA ACGTTTCTCA 420

TACGACGATT TATCGTTGGG TGC 443

(2) INFORMATION FOR SEQ ID NO: 874:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 325 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 874:

ATTTGTTTAT GAGCATCAAG ACATTCTCTT CTAGAAATGA AAAGTAGGTT ATTTATCCTT 60

TCGTAACATG TTAGGCTTAT TCAACTATGC TTCCAATCGA AATTTAATTA ACTTAACAAC 120

TTAAAATGTT CTCTTTATTG ATTTTCATAA TTTaAATTGT TCAAATTGAA TTAGATAAAT 180

AGTAATACTT CCCATATTAA AATTTATTTT TTTTCAGCCC TATAATACTT ACAATTACCG 240

TGAATAATCC TAAATACATG GGAAGACCAT TATACTTTTC TCCAGTTTTC GGTAACTCCA 300

TAGTATTTTT ATTTCGATTA CTnCn 325

(2) INFORMATION FOR SEQ ID NO: 875:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 428 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 875:

GATCAATCCA ATTATTTGTA TTGnATATTC ACATCGGTCT TGCAACATTG CGTAAGCTAT 60

TTnACTATCA TTAGACATAG TTTTATAAAA TTCTGATGTA AAAAGTACTT TAGGTAATTG 120

GTAATACAAA TTTCTATAGG TTTCATTAGC TGAAATAAAT TGAAAATCGT TCATAGCTTT 180

ACCTCTCTGT TTTTGTTAGT TTATAGTTTA ACCTACTTTT ATTTCCCTTG CAAATACTGA 240

TATAACAGTA TTTGCCTGCT nTCTTAGTAT ATCACGATAT TGCGATATAA CGCAATTGTA 300

ATATTGGAAT ATCGTGATAT TACAATATTT AAAAGATTAT TATTGATACA ACAGCGTTTG 360

TAGAAAAAAA ACAACAACAA TATTATGATA TTGTTGTTGT CTAATATTGT AATATTGTGA 420

TTCCACAA 428

(2) INFORMATION FOR SEQ ID NO: 876:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 155 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 876:

GGAAAGnAGT AGTCTATGAT AAAATAGTAA AAAGAGCAGA AATGAGGACA AcgAATGAAA 60

ATTGGAGTCA TTGGTTTAGG AAATATTGCA CAAAAAGCGT ATTTGCCAAC TTATAGTGAA 120

TGCCGCCACT TGGCAGAGTT TGTATTAGCA ACTAG 155

(2) INFORMATION FOR SEQ ID NO: 877:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 277 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 877:

GCCGAATAAA AGAAGTTTCT GATGATATTG AAGTTGTGGC TATCAATGAT TTGACTAGCC 60

CAACAATGTT GGCTCACTTA TTACAATTTG ACTCTACGCA TGGCACCTAT CCTGGCACCG 120

TGACAGCCAC TGAAAATGGC ATTGTGGTTG AnGGCGAAGA nACCCGTGTC TACGCAGAAC 180

CAGATGCAAG TAAAATCCCT TGGGTTAAAG AAAACGGAGT AGACATTGTC CTTGArTGTA 240

CAGGTTTTTA TACCTCTgAA GnAGAAGCAC AAGCTCA 277

(2) INFORMATION FOR SEQ ID NO: 878:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 364 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 878:

AAAAGTATAC TAGTATACTG GTATACTTTC CTTTAGCAAA CCTTTTAACA AGCAAGATTT 60

CTAAGGTATA CTGGTATACC TTTGAATAAA AGTTTAAGTA TACCAGTATA CTGnCACGAT 120

AAAAGCAACA GTATACCAGT ATACTTACGA TCATATTTTG GAGTATACTG GTATACTATT 180

GCTAAaGAAA ATAAAAAATA ATGATTTACA AATwTAATTA TGTACGGTAT TTTTTAAGAG 240

TATTGAATTT ATAAATATTA AGATAAAAAA ATAGAGACGG TAACTTACTT GATCGCCAAA 300

TCAAATAAGT TACAGAATCA AAAATGTACA CCCATTTTCA ATTCTTTTGC GTCCCTGGAT 360

GATA 364

(2) INFORMATION FOR SEQ ID NO: 879:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 211 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 879:

CATGAGTGTT GCCTGGCAAT ATTTGTGGGG ATTTTTTGAT TATTGTTAGC ATTTTTCTAG 60

ACAAGAACAA CTGAAAAAAG AGTTTCTTTT TTCCATTATG TTTTGTTTCT GAGTTTAACC 120

AAATCACTCA TTATTTCAAG GAGGAATTAG GTTATGTCAG TaCAATTAGA AGTCAAAGAC 180

AGAGCCTATC CGTCCAAGGT TCAACTAAGG A 211

(2) INFORMATION FOR SEQ ID NO: 880:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 418 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 880:

ATCTTTTATT TAGAATGAAT AGGgATCTGA AaTaTAAGaT TTTTTGATGG AGAAAaCCAG 60

GATGATTTAA TAGTTATyCA TATACATATG ACTCAACACG nAATAGATGG TGATTAACAG 120

AGAGTTATCC AAATTATTAT ATATAGAAAT AATTCAAAAT TATTTTAAAG AAATTCTAAC 180

TTTTCACTCG ATAAAATATA GTGTAAAAAC GCAACTTGAT TCAGAGGATG AAAATGAAGT 240

GTTCAGATAT GAGTTGAAAA ATACTTATTT CAATTGATTG AGTAAAGCTG TTTTATCATT 300

ATTAGTAATA TGAAGATTTT TATATCAGAC TTAGAGAAAA GCTTTACTAC ACTCTTGTTA 360

ACTATAATGC TTTTAGCGAG AAATAATACC TAAAGTGTAA GCTAAAAATT TATCATAT 418

(2) INFORMATION FOR SEQ ID NO: 881:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 408 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 881:

ACATCTTTTT TnCTGGnACA TTAATTGnGC TACTTTTTTT AATTGTGTCG AGTTAAATCT 60

CCAAGGAACG AACAGTAATT TTAGCGCATG AATTAACCCC TAGTnAGCTA CGTCGATTTT 120

GGCAAAACTG ATTTATGAGT ATCTAAAAAA CGAACAAATC ATCGGCCTGG AAAAATTCTA 180

TACCTTTAAC TGTGCCGATT ATGCCAATAA TCCCGAATTG CTATCATCTA TTCTCTTCGG 240

CCATACCAAA GGCGCATTCA CTGGGGCTGA ATCAGAAwAA CrAGGACTCT TGGCCCAAGC 300

TaATAATTCC GTACTCTTTT TAGATGAAGT ACATCGATTA TCCAACGAAA ACCAAGAAAA 360

ACTCTTTCAG TTTATGGATA CAGGGACCTT CCGCCCTATT GGTGAAGA 408

(2) INFORMATION FOR SEQ ID NO: 882:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 465 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 882:

CTTCAAAAAT CTCGCGATTG TGATCGCGAA TAATATACGT CATTGTAGCT TCTTCTGGCG 60

TTCCAGCAAA CGCTGCTAAA TGATAAAAGC CTTCTTCACC TGCTGTTTTT TCTGGAACTT 120

CGTCAGCAGG TAAGGCATTA TGGAAATCAA TTCCTAATTG TAACGCATTA ATCATTGTAT 180

TTTTGGCTGT TCCTGGATGA ACGTTTTTCC CTTGAATTGT AATTTCTGCT TGCGCCGCAT 240

TAAACGTTTC AAATTGTAAT TCGCCTACTG GCCCACCATC CATTGTATAA GCAAAATCCA 300

CGTTAAATTG AGCGACATCA AATTTGTCGG GCACCTACAC CAATTTCTTC GTCAGGACCG 360

AAAAGCCACA CGAATGGGGG CCATGTTTAA TTTCnGGGGn GATTAAATTA AATATTCCCn 420

GGGCGGTCAA TAATTCCGCA ATAnCCGGAT TTATCAACCA GCTCC 465

(2) INFORMATION FOR SEQ ID NO: 883:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 341 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 883:

AAAGTTTTGA TGATAAAAAT TACTATTGCT AACTTATTAG AAGGGGTCCC ATTATTACTC 60

TTATTAACAT TATACAAATA ATCTTTCCAA TCACTCCTGC AAACGGGCTT CGTCTCTAAC 120

TGAAGGTCTA TTACATTAGC ATTGTGCGCA CATGCATTCC TAACAAAATT TATACATTTT 180

AGCCATGATA ATAATTCAAT ATCCGTACAT CTAAACATTT TTGAAAGTTT TCGTTTATTT 240

CTAGTAGACA TAATTTCTAC TAATCTAACA AGTTCACCAA AAGTAAGAAC TTCCACAGCT 300

AACCAAATAC TTGGaAAATT ATCTTTTTCT TTAnTCCTnn T 341

(2) INFORMATION FOR SEQ ID NO: 884:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 197 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 884:

AAAGATGGAT TCTTCTTTAC TCACACCTAA AAGAnGTTTT TCTGAATCCT TTAGTAAGAA 60

AAACCGGTGA CTCTATCGTT AACAGAGAGT AGTTGTCAAC GCAACTACAT AAGGCTTATC 120

TTGTGAAAGA TAAGTAAATT GAGGTGGCAC CACGAATGCG ATTCGTnCTC TTGGCTCACA 180

GCCAAGAGGC TTTTTTG 197

(2) INFORMATION FOR SEQ ID NO: 885:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1548 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 885:

CGGAATTGGA CTAnCCnATT GAAAACCATT TCTTTTTnTC CTAAAAAAAT CGAAGATGGG 60

TAATAAAAAG AATGGGTAAT CTGATAAAAA AACGAACTTT ATCTAATCGG ATAAGAAAGA 120

ATAATCGGTA TTGAAAAAAA GTTTAAAAAA AGTGAAAATA ATGATTGACC AATGCTTGAA 180

ATCTTGGTAT ATTATTTACA TGGTCGCAAG GCGAGTCACA CAGCTGAGCG ATTCAGAAGT 240

CAATAATTAA ATAACAATTC TTGATTTCAA ATCTCTCGAA AAAAGTTTGA AAAAACAACA 300

AAAAAGTTGT tGACaAATAm CAAACGAGAT GATAAGATwT AGAAGTTGTC ACAACAGACg 360

ACaACAACCG AGCAAAACAA CTTAAAAAAA CTTTTTTAAA AAAGTTCTTG ACATTCrAAy 420

GAAAGTTTGT TAwGaTATAA AAGTTGCTGC GAGGTAACGT AGACCTTTGA AAACTGAACA 480

AAGAATAGAC GAACCAAATG TGTAGGGCGT CTTGATTCAA TTCAAGACAA CAAACATTTT 540

TAACAAGCAA GCAATATGCT AGCAAACAAA TTGAGCTTAA CAATCGwAAG ATTGTTCTAA 600

CTTTTATTAT GAGAGTTTGA TCCTGGCTCA GGACGAACGC TGGCGGCGTG CCTAATACAT 660

GCAAGTCGAA CGCTTCTTTC CTCCCGAGTG CTTGCACTCA ATTGGAAAGA GGAGTGGCGG 720

ACGGGTGAGT AACACGTGGG TAACCTACCC ATCAGAGGGG GATAACACTT GGAAACAGGT 780

GCTAATACCG CATAACAGTT TATGCCGCAT GGCATAAGAG TGAAAGGCGC TTTCGGGTGT 840

CGyTGATGGA TGGACCCGCG GTGCATTAGC TAGTTGGTGA GGTAACGGCT CACCAAGGCC 900

ACGATGCATA GCCGACCTGA GAGGGTGATC GGCCACACTG GGACTGAGAC ACGGCCCAGA 960

CTCCTACGGG AGGCAGCAGT AGGGAATCTT CGGCAATGGA CGAAAGTCTG ACCGAGCAAC 1020

GCCGCGTGAG TGAAGAAGGT TTTCGGATCG TAAAACTCTG TTGTTAGAGA AGAACAAGGA 1080

CGTTAGTAAC TGAACGTCCC CTGACGGTAT CTAACCAGAA AGCCACGGCT AACTACGTGC 1140

CAGCAGCCGC GGTAATACGT AGGTGGCAAG CGTTGTCCGG ATTTATTGGG CGTAAAGCGA 1200

GCGCAGGCGG TTTCTTAAGT CTGATGTGAA AGCCCCCGGC TCAACCGGGG AGGGTCATTG 1260

GAAACTGGGA GACTTGAGTG CAGAAGAGGA GAGTGGAATT CCATGTGTAG CGGTGAAATG 1320

CGTAGATATA TGGAGGAACA CCAGTGGCGA AaGCGGCTCT CTGGTCTGTA ACTGACGCTG 1380

AAGCTCGAAA gcsTgGGGAG CnAACaGGAT TAGATACCCT GGTAGTCCAC GCCGTAAACG 1440

ATGAnTGCtA ArTGTTGGAA GGTTTCCGCC CTTCAGTGCT GCCAGCnAAA CGCCnTTAAG 1500

CACTCCGCCT GGGGGAGTAC GACCGCCAAG GTTGAAACCC CAAnGAAT 1548

(2) INFORMATION FOR SEQ ID NO: 886:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 486 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 886:

AAAAAACAAC CATTGCCTAT GTAGAACAAC GTAGCGCCTT GGATCTTAGT TTCCCAATTA 60

GCGTTTTTGA AACGGTCTTG CTAGGAACCT ATCCAAACTT GGGGACTATT AAAACGCCCA 120

GGAAAGAAAG AaAAGCAAGC AGCCATGGCT GCATTAAAAA TGGTGCAATT AGAAGACTAT 180

GCGCAACGCC AGATTGGCGA ACTTTCTGGT GGCCAATTAC AACGTGTGTT TATCGCCCGT 240

GTTTTGGCCC AAGGTGCTGA GGTGATTTTT TTAGATGAAC CTTTCGTCGG CATTGATATG 300

TCTAGTGAAA AAGTGATTAT GGATATTCTT AAATCATTAA AAAATCAAGG TAAAATGATT 360

ATCATGTTCA CCATGATTGC ACAAAGTGTC CCACTATTTT GATGAATTAA TCGTTTGAAA 420

AAnCGGCTAA TTGCTGCnGG GCCTGGTGnA CCAAACATTA CTGCAGAACG CTCCCAAGAA 480

CATACC 486

(2) INFORMATION FOR SEQ ID NO: 887:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 887:

CAGACGCGCA CAACCATTAA AACAGGTGAC GATTCTCTTT TATGTGACTG TGCCAGAAGC 60

CTTTCGnGAA CGGGCGACAC GGTGTTTGAA GCTTGTCTCC CCAAACTGTC CAGTGATTCA 120

ATCATTGGCT GAAAGAATTA CTGTCGAAGA AACAGTTGTA TTTAAAAAGT AAACCATTAG 180

AGACATTATC GTTACGAGCG CTTTTCCCTT TTTGGATGCC TCGTTTCTGG GAATGTCTTT 240

TnTTAGAGGA GCGTAGGGAA ATAGCTGTAG A 271

(2) INFORMATION FOR SEQ ID NO: 888:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 316 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 888:

GCAGGCAGAG GAGGCACCTG TAGGGGGTGG TGGGCTGGTG GCCCAGGAGA AGTCAGGAAG 60

GGAGCCCAGC TGGtGaCAAG AGAGCCCAGA GGTGCCTGGG GCTGAGTGTG AgAGCCCGGA 120

AGATTTCArC CATGCCTCAC AGCTCCGACA GCAGTGACTC CAGCTTCAGC CGCTCTCCTC 180

CCCCTGGCAA ACAGGACTCA TCTGATnGAT GTGAGAAAAT TCAnAGGAAG GAGAAAAATT 240

CGTATTGCCG CCCAGAAGAG CCGACAGAGG CAGACCACnG AAGGCCGAAA CCCTGCAACT 300

TGGAGAGCGA AGACCT 316

(2) INFORMATION FOR SEQ ID NO: 889:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 173 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 889:

AGTCAAGACC ATAGCCTTCT AAATTGATTT CAGCAATATC AGCTAaGGGC GTGCGCTGCA 60

TCGATGTCTT CwTGTGTGCA AGTTGGTGAT nTAAATAATA ATGTATCAGA AATAATCGCT 120

GAGACCATCA TACCAGCAAT TTGAGCTGGA ATTGTCACGT TATTTTCTTT AAA 173

(2) INFORMATION FOR SEQ ID NO: 890:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 160 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 890:

AAGAAGGCGG CAAAAAAGCG TTTGAATTAA TTAAAGAAGG TATGAACAAT GGTGCGAAAG 60

TTTTAGGCTT AGCAACAGGT AGTACACCAG nAACATTATA TAAAGArATG ACTGCAAtGA 120

tGTCGACTTT ACAGAAATGA CTTCTGTAAA CTTAGATGAA 160

(2) INFORMATION FOR SEQ ID NO: 891:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 403 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 891:

AGGTGTTATT TTAACATTAG TAATAATTGC AGCATTACTT ATTTTATCTA GTAGCTTAAA 60

GTTGCAAGTT TTATTGGTTA AGTTTGCCAA TCGTTGGTTT AAACGACCAA AACTAAGAGA 120

TTGGGTGGAT AATTGTAATT TACAAATCTA TTCCTTGCGG GCAGCTGTCC AAACAATTAC 180

ACAAGAyCGC TCAGCGTTGC TTCGAATCTT TTCATGGAAT GTATTTAAAT TATTATTTTG 240

GTACATTATT CCTTACATTG TTTTAGTAGA AAATCACCCG AATATTGATT TATTATTAGT 300

GATGTCCTTT ACCAGTTTTG CTGGTTATTC TTTCCAGGAn TTATTCCCGA CACCTGCTGG 360

CATTGGTCCA TTTGAAATTG TTTACTTATT GGTTGTTTAA ACC 403

(2) INFORMATION FOR SEQ ID NO: 892:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 499 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 892:

nAAATACGAA TTCGAGCCCG GTACCCAAnA AACAATTCAA ATTAATTGAT TGTCGCAGTG 60

AAGGAACTAC CAATTATTAC TCACTTTCTC ACGATACCAC CAAAATTGAG CAACTCCAAC 120

AACTATTGAC CCACATTACG TCGATTATGC CAGGGAGGAA GCAAGCATGA AAAAAATTCT 180

CGTGGTCCTC ACAAACGTCT CCCGTTATCA CGGCACCGrA raACCAACGG GsCTCTGGcT 240

AGGTGAAGCC ACGGAGTcGT GGAAGAAGTA ACCAAAGCAn GATTTTCAGT TGACTATGTT 300

AGTCCCCAAG GAGGCTATGT CCCTTTGGAT CCCCGGAGTA TGAAATATGT AGATTCATCC 360

ATTATGGCAG TTTATGAAAG TGCTGATTTT CAAGAACGAG CGCTAGCACA TTCCCTCTCT 420

CTTGAAGAAA TTGAGCCAGA TGAATACGCA GCAATTTACT TTACAGGAAG CCATGGCGTT 480

ATGTGGGGAC TTTCCTGAT 499

(2) INFORMATION FOR SEQ ID NO: 893:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 303 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 893:

CTGATGCGtA TTTTCTCCTT ACGCATCTGT GCGgTATTTC ACACCGCATA TGGTGCACTC 60

TCAGTACAAT CTGCTCTGAT GCCGCATAGT TAAGCCAGCC CCGACACCCG CCAACACCCG 120

CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC TTACAGACAA GCTGTGACCG 180

TCTCCGGGAG CTGCATGTGT CAGAGGTTTT CACCGTCATC ACCGAAACGC GCGAGACGAA 240

AGGGCCTCGT GATACGCCTA TTTTTATAkG TTAATgTCAT GATAATAATG GTTTCTTAGA 300

CGT 303

(2) INFORMATION FOR SEQ ID NO: 894:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 894:

ATCAATTACG CCTTATCAAG AAAAAGGACG TCCAAAGCAA CCAGAAAAAG CAACGrTTGA 60

ATTCAATCGA 70

(2) INFORMATION FOR SEQ ID NO: 895:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 238 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 895:

GGACCATTTT AAGGAAATTA CCCGGGGGGG AAAATTTTAA AAAGCCCTTT TGGGGGCCGG 60

CCAAAACCCA AATTTTTAAG GAAAGGTTTT TGGGGGTTAA ATnTTTTAAA TTTGGGGGGG 120

GTTGGcaATT TTTTTtCCCC CAAAACCCAA TTTTTTTGGG GGTTCCCTTG GCCCCCGGAA 180

CCAAAAATTG GGGGCCTTAA CCGGGAAAAA ACCCCCAAGG AATTTTTTAA ATGGGTTT 238

(2) INFORMATION FOR SEQ ID NO: 896:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 456 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 896:

TTTCAAATAG TATAGTTTGA ATGCTCAAGT ATTTCAGCAA CTTTTATAAC ACAGGAGGTA 60

ATAAATATGG CTGTACCAGC GAGAAGAACA TCGAAAACAA AAAAAAGATT GCGTCGAACA 120

CATGAAAAAT TAAAATCACC TGAAATTTCG TTTGATGAGA ATCTTGGTGA TTATCGTAAA 180

AGTCATCATG TATCTTTAAA AGGCTATTAT GGTGGAAAAA AAGTGATGGA TAAAAAATAG 240

GAGGTAATAG TATTGAAGAA AATGAATATT ATTTATCCAC CACTTGTGGA ACAGAGCTTT 300

CAATTTTATC AAGATTATGA ACAAGAAAGA TATGATAAGT CTGAATTATA CAGAATAATG 360

GTTATGAAGA ATATTATTAA TGAAAATGGC ACTCCAACAG AAGAGGCTCT TAAAAAAGGA 420

TTAGTTAAAG CTTTTATGAA GATATGATTT ATCTTT 456

(2) INFORMATION FOR SEQ ID NO: 897:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 308 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 897:

TCAAAAAGCG GGTTGGCTTT GAGCGGTCCA GTTGTGAGAC GATTAAACAA ACCTGAATGG 60

ATTCAATTGT ATTTAGATGA ATCAAGATTA GCATTGTTTG TAATTCCCTG TAAAGCGACG 120

GACGAAGGCG CTAGAAGTTG TGTAAATCCG AAGTCAAAAA AGAAAGCAGG GTATCGTAAA 180

AGCTGGTCAG GTAGCATTTT AGAGAAGGTA GCAAAAGCCA GCAAGATGGA TATTGAAAAT 240

CATCGTTATC ATGTTGAACC AGnAAGCGTT GAAGGCTATC CAACTGACTC TTGGTTTTGA 300

TTTAACGA 308

(2) INFORMATION FOR SEQ ID NO: 898:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 62 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 898:

CAAGTGCTTG GTACTTTAGC CACTGGTTCT GTTACCGGCA CGTTGCTTGG TCCTCTTTTG 60

GG 62

(2) INFORMATION FOR SEQ ID NO: 899:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 484 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 899:

TTCATTAATC ATATTCGTnG CTGCATTATT TTTATTCGTT ACATAGCTGT CAATTTGGGA 60

AAnGGnGTTA TTTTTTTGTG AATCAATTGA TAAGTTAATT GTATAAGTAA CATAAACGCA 120

CAAAGCAGCC ACCGTTAACA AAATAATACT TGTTAACAGT AGACTGTATT TTTGTACTAA 180

TCGATTTAAT AAATATTTTT TTTTAAACAG TTTTTTGAAG ATTCGCTTCA TTTTGTTCCA 240

ACTTTCCATC AATAAACTGG CGATTTTTGC GCGTGACATA GGCACACCAC ATCATAATTA 300

AGGGCAAACG AAGCAAATAG TAGCAGTnCT TTCATTTTCC AGTTAAAGCA ACGATACTAA 360

TTACACCGAA TAAGACTTTA AAAAATACAC CGAGGTTTAA AAAGATACTG ATAAATGGAA 420

GGCTCATTAA ATCAATAGGT TGAGATAAAA TAGCTCCCTC CTATAGTAGC CTATAAATGT 480

AAAA 484

(2) INFORMATION FOR SEQ ID NO: 900:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 181 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 900:

AAAGTCTCTT TTCAATCATT TAGTATACAC CTGAAAAGAA GAATAAACTA TCATTAATTA 60

AAAGAAAAAG TGAAAsAACT AGCGGATTTT TTGTTCAGCT GATTATCTTA AGAATTAGAA 120

AATGAGGTCG GGACAGAAGT GTTTAACTCC TAGAACCAAG TAGGTACTGT TTAACATCTA 180

n 181

(2) INFORMATION FOR SEQ ID NO: 901:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 845 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 901:

TTCCGTGAAA TACGGCGTTT GGACAGGCAA AGAATTAACA AATCAAGGGA GCTTTCCAAC 60

ACCTGGCTCT TGGGAAGAAA TGAAAGCGCA TTTATATTCG GTTTATGCAG ACAAAAGAAA 120

TGAGTCGATT AGTGGTGTTG CTTTTAGTTC ACCTGGAGTA GTTGACGAAA AAAGTCAACA 180

GATTTTAGGG ATTTCCGCCA TTCCATATAT TCATCATTTT AACATTTATG AAGAATTAGA 240

AGCGTTGTTT GGCTTACCAG TGACTATCGA AAATGATGCG AATTGTGCTG GTTTAGCTGA 300

GATATATGAA GGCGCGGCGA AAGGGAAAAA AGAAGTGCTT TTTGTCGTTA TTGGGACAGG 360

AATTGGCGGA CGATTTTTCG CAATGGTGAA TTGTACAAAG GCGCTCATTT ATATGGAGGC 420

GAGTTTGGAC TGAACTTTTT AAGTAATGGC CAAACaTTCA GTGAAATTGG CACAGCTGTC 480

AAAATGGCTC AgCGCTATTG TGAACGAATT GGAGTCGAAA AACAGGCGGT GACAGGAGAA 540

GAAGTTTTTG AACTGGCACA ACGAGGGGAT GAGATTGCTC GCGAAGAAGT CAATAATTTT 600

TATGATTATT TAACACAAGG ATTATTCGGC TTACAGTTTT CTTATGATCC TGAAATGATT 660

GTACTCGGAG GCGGTGTTTC TGCTAAAGAA GGCTTATTAG CAGAGATTAA TCGCCGGATG 720

CTAACGCATT TACAAACCTT TGAATTAAAA GATTTTGTTC CAGAAATCGT GACCTGCCAT 780

TATCAAAATG ATGCCAACTT GAnTGGCGCA CGCCCAACTT TCAAGCCAAA ACGATTGGGA 840

CTATA 845

(2) INFORMATION FOR SEQ ID NO: 902:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 330 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 902:

ATATCAAATG AGAAAAAGTT AAATTGATCA TCCTCTTCCG TCAACAATTC TTTCATCAAT 60

GTTGTTTTAA ATAATTCnGT TAAATAAAGT TCTGTTCCTA TTAATAAATA CACTGGAGCG 120

ATTTTTTGCT CTCGCACTTC TTTCAAAGCT TTTTGTAACT CCATTGGCTC GCTCCTTTTC 180

TCGTTTCTCT TGTTTATCCT ATCATGTTAC TTCTGGAAAT ACTAGTCTTG TTCkTTAACA 240

AATTTGGCTT TAGAAAGCCC TTGCCaAGGt AACCATTGAT AAGAAATCAT TCCTGGAAGA 300

TCGGTCCGAn AATAATnTAT GGTGTGCTCC 330

(2) INFORMATION FOR SEQ ID NO: 903:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 276 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 903:

ATAAGGGGCG TTTTTCTAGT GCTTAGTCGA AAAAAATAGT TGATTCATAA GTTCATTGTC 60

TTATGAATCA ACTATTTGAT AACTTTCTCT CTTGCTTATT GCGCATTCAG GATCCACATG 120

AATCCGACAA AAACAAAGAA CAAGCCATAC ATAATTGGCG AAACTTCnTT CCCACGTTTT 180

GCCGCAATCA TTGTAATTGG ATAGAAGATA AAGCCTAAGG CAATCCCATC TGAAATGCTG 240

TATGTCAATG GCATTCCTAA CAAGATTAAG AAGGAA 276

(2) INFORMATION FOR SEQ ID NO: 904:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 124 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 904:

CGCTTAAAAC TTCTCCTTCA ATAGGTGCAT AAATTTCACC ATTTTCTGGA ACTACTGCAT 60

AGCCATCkCC CATCATTTTT TGTGAGAAGG CAGGATCATT CACTTCACTG GTTGGGATAA 120

TGGG 124

(2) INFORMATION FOR SEQ ID NO: 905:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 143 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 905:

GGTTCCtCTA TGCCAACAAA TAACCCGGTT TGTGTCACAT AGCTAACCaT TCGATAGTTG 60

CCAGAAGCAA GCnCAATTCG GCTCATTGTA TCTTCAACAC GGTACATTTC TGCATCGCTT 120

TCTATCATAA TTTTACCAGC TAA 143

(2) INFORMATION FOR SEQ ID NO: 906:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 482 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 906:

TATCAAGAAA AGTCTAACAA AACTAGATAC AGAGAGATAG AACTAGGGGA AAAAAGTCCT 60

ACCATTGAAG AATTAATTTC TTTTACAGAT ACTTTAGGTA TAAGAtTATC TGAATTTCTG 120

TATCGGGGTG ACATTACACC TCTAGCAGCA TCTTACTATG GTAAAAGAAA AATTGAGGTT 180

GCTTCTTTAA TTAATAACTT CCCTGAATTA GAGAAACAAT TTCCTGAAAT ATACAAAGAT 240

CGCTTTAAGA ACCTACAGTG TTACACTTTA TTTACCTTAT GCTTAATAAT CTCTAAATCT 300

TTAAATAATT CCTTGTATAC TTTTAAAAAC AAAGATATTA GAGAACTTAA AAATTTTTAC 360

AAAAATAGAG AAGTAATATT AGGTGCTGGA CTATGCTATA CTAAGTCATC TTTACAGGGC 420

AGTTCCCAGA ATATTCGAnG GAACTGGATT TTCTAACGGG AAAACnCTTC CCATTAACCA 480

nC 482

(2) INFORMATION FOR SEQ ID NO: 907:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 147 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 907:

TGGAGAACGT CTGTATCTCA CTGAGCAAAA ATAAACGGAA TAACCCTGAA CGGATAGAAT 60

TAAAAAAGTA TTCTCCAAAA TTGCGCAGAC GGGCAATTTT CAAAGAAGTG AAATAGGCAA 120

GTGGtAAAAA GGrGACaAAA TGGGGCC 147

(2) INFORMATION FOR SEQ ID NO: 908:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 239 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 908:

AAGGATGCCC TACGAGCTTT TGCnAAAGCT GGCGGACAAA ATGTAACTGA ACAACAAGAA 60

GAAGGCTTCT TTGATAAAAT GAAAGATGCG TTTGGCGGTA AGAAGAAAAA ATAAACCAAA 120

CGAGTCAGAG ACAAAAGCAG TTAGCTCCAA GTGTTCATCA CTCTCGAAGG CTAAATTGCA 180

ACGAGTAATT GTAGGTGCTG CTTTTGTTCA ACTGATTATT TTAAGATTTA GnGAGTGGG 239

(2) INFORMATION FOR SEQ ID NO: 909:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 70 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 909:

AAATAGAAGG ATGATTAGAG TTAGCTTGTT TATCACTAGC TTTACTATCC GAACTTTGTT 60

TAGAATTTGT 70

(2) INFORMATION FOR SEQ ID NO: 910:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 78 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 910:

CGCATTGCCA GCAGTAGAAA CGCTAGGAAA CTGCGaATGT TATCTGTTCT GATAAGACAG 60

GAACACTGAC ACAAAATA 78

(2) INFORMATION FOR SEQ ID NO: 911:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 887 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 911:

TCATGAAAGG GATTCGTTCC ACGGAACnTT CCGAaGaGGA atGCcTmCTT AGTCGATTTA 60

TCATTAmAAA TTAAAGAAGC GATTAAAAAT GGGTATTATC CGCAACTGTT AAAAAACAAA 120

TCTTTAGGAA TGATTTTCCA ACAATCTTCA ACTCGAACAC GAGTATCTTT TGAAACTGCA 180

ATGGAACAAT TAGGCGGACA TGGCGAATAT TTAGCACCTG GTCAAATTCA ATTAGGGGGA 240

CATGAAACGA TTGAAGATAC CAGCCGTGTA TTGTCACGTT TAGTAGATAT TTTGATGGCT 300

CGTGTGGAAC GTCACCATAG TATTGTTGAT TTAGCTAATT GTGCTACCAT CCCAGTCATT 360

AACGGTATGT CTGATTACAA TCATCCAACC CAAGAATTAG GTGATTTGTG TACCATGGTT 420

GAGCATTTAC CAGAAGGCAA AAAATTAGAA GATTGTAAAG TTGTCTTTGT CGGAGATGCG 480

ACACAAGTTT GTTTTTCTTT AGGATTAATC ACAACAAAAA TGGGCATGAA TTTTGTTCAC 540

TTCGGTCCAG AAGGCTTCCA ATTAAATGAA GAGCATCAAG CGAAATTAGC AAAAAATTGT 600

GAAGTTTCAG GCGGTAGCTT TTTGGTCACA GATGATGCAA GTTCTGTTGA AGGCGCTGAT 660

TTCTTGTACA CGGATGTTTG GTATGGCTTG TACGAAGCGG AATTATCTGA AGAAGAACGT 720

ATGAAGGTTT TCTATCCAAA ATATCAAGTG AATCAAGAAA TGATGGATCG GGCcAGGksC 780

CAATTGtAAA TyCATsCaCT GTTTaCCAGC AACmCGTGGG GAAGAAGTAm CTGATGaAGk 840

AATTGACGGT AAAAATTCTA TTTGTTTCGA TGAAGCAGAA AACCGTT 887

(2) INFORMATION FOR SEQ ID NO: 912:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 742 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 912:

GAACAGCATC CGTTACGATT TAAGACCCTG GnCGGTAACT ACTCGTCACT TTTTCTAACA 60

TTGACGAAAC GATAGACCAG GAnACACTAC CATGGCTAGT GATTAGCGAC AAAAATTCTT 120

ACGAACGATT TTTCTTTACa GCTCAaGAAA GCACGCmAGA ACAaGCTTTT GAGAAAATGT 180

TAAAaGAACT AGAAACCGAC CaTGACAGTA CACTAGAAGT TGTTCATAAC TGGTTATGCC 240

GACAAACAGA CGAAGCGCTT TTTCTAGGTA TTCTCAAAGA AGATAGAACG TTAAAAGGTG 300

CTATGACTTA CTGCATAGGA AAAGCCCAAG AACAATCCAA AGACCAATCA AGCGCCATGG 360

TGGCGGACGA AGTGGTTTTC AGTTGGATAA AAGAGTATTT CTTATTAGAG AAATTACCAG 420

AAACCAAAGC AGTTGGTAAA GTAACGACCA AAAAATCAGA AACAAAACCT GAAAGAAAAG 480

AGGTGGACGA ACAAATTGAT TTGTTCGAGT GCGTATGAAT AAAAAACACC AAAACCATTA 540

ACCCCACCCA AAGCCTTTTT TAACTGGTGT ACTGCACAAA TACCAACGTA TGAATGGCAA 600

AACAAAAAGG AGAcAATTCT TGCGAGTTCT CGCAAGAATT GCCCcaCCAT AAAAAAACGA 660

TTAACGAAAT ACTCACGATT ATCGTTTCCA ACGAAATTTT ACTCCTTTGT GATAAAGTnC 720

GACTTTTGCA AGCnCAGTTn TC 742

(2) INFORMATION FOR SEQ ID NO: 913:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 461 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 913:

TGAACTTTGT TTCGCTATCA AGTGAAGAAA GATATCGAGG AATGTTAGAT CCATTGCTAT 60

TTCTGCCAAG AGAAGAAGCT ATTCAAACCG CACGAAACGT GTTAGAAAAT TTTGGTGAAG 120

TTACGACAGA CAGTCACACT GCAGTGAkaA GAAGACGCTr ATTTTAGATT CTGTCAATGC 180

GGTGATGAAA GGTAAAGGaA AAAAaCACCT AACGAAArTA ATCGAGGTTA TTCGCGAAAA 240

AGACCCTAAA TTAGCCAATT TAATCTCTGG CCACAATGTT GGGTTAGGTA AAATTCTTTT 300

AGGGGAACGA TTACAGCGAA CCTATTCGCT TTGGAAAATC CAAATCCATG TCTTAnGTAC 360

CCAGGGGCTT TAAATCCCTA CACAGCGGAA ATTAGACAGC GGACGATTAA ACATGGAACC 420

AATCGCCGGC TGTCCATTAT GGnAGTTATT ATGGAAATGT C 461

(2) INFORMATION FOR SEQ ID NO: 914:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 521 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 914:

CTTACAAATC TATCGTTTAT CATCGCGGAA AACGGCTGAC TTTTTTGTTG CCTAAAATTT 60

TATTGTCAGG TACTATTTAA GTGTCTCCAT CCGAGACAAT CCATTTATTT TTTGTTGACC 120

AGCGGAAACT GGTTGGCTCC TAAATTTGGT AGATTCTTCT GCCTTTAATA CAATCAGCCA 180

ACCATTTTGG TTGGCTATTT TTTTATTTTT CAATTTATTC CAATGATTAC AATAAAAAGG 240

AGCGAAATAT GTCGCTCCTT TGACTTaTTA TTaATTTTAA AGATATTTTa CTTTCTTCTC 300

GTTcTACaAT AAAcTaTAAT ACCgCAGCTC CTAACAAAGC AAtATCAGAC CAGCaTAGCa 360

ACCATTATaT TTTcTCCGTA GTGGTAAgAA CyGGGcTTgC aCTGTTCGcT ACGAGATCgA 420

AGGTAAGCTG gCTTGGcGTC TTGTTAGAAA gAAcGTGGgA CAGGTTGcGC GAGAGCCTGC 480

ACACGTGCTG TTGCCCGTCA GGTGTGCTTG CACGnCTnAC C 521

(2) INFORMATION FOR SEQ ID NO: 915:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 576 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 915:

TCCTTGTCAC ACAGCGTTTA TGGnTTATGG TTAAGGnGGT TAAAAAGGAC TACTGTCTTA 60

ymAATTTGCG TTCCGGCCAA mCCAAAaaGA aGCmCmGrAA AAAGTTTTAG AATTTTTTGG 120

TATGGATACG AmCCCAGCAA ATATTCaGTT AATTAATGAA CTAAAATCAG GAACTTGTTT 180

ATTCCAAGAC CATCGAGGGC GTAgcCAnCC TATTGCAATT GATGTTTTAT TCGATAGTTG 240

GTTAATGGCC ATTTCGTCTA CTAATAAAGA GGATGAAGCT ACTCAAaTGG CGTTGGCAaT 300

GGAACAAGGA TCATAAAAAA GGAGAAGTGA GCATGAAAAA AGTACTACCT TTTATTGCCT 360

TAGTCGGCTT GTTATTGTTG TCaGGTTGTG GAmCAGATAT GAAAAAGATA TTGACTGCCG 420

ATGGTGGTAA ATGGGAmCTA GAAAATAAAA GTCCAACTAC TACtTACACy TTTTTTGaTG 480

ATGAAmCTTT TcCGAGGTAT AATTCCAAAA ATTAGTGATA GTnGGACCGT ACCCCTTACG 540

TGGAAAATAA TAAAAAACTC nCTTTGGnTA TAAAnA 576

(2) INFORMATION FOR SEQ ID NO: 916:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 750 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 916:

CGTCTGCAAC TTTGGTTTCG CTTTTCTAAA ACAACAGCTC GATTCTGAAT AACTCATTGA 60

TTGTTGCCAA TTGATCTTTT GTTTGTTGAA GGTTTGTTTC TGCCGCCGCA GTTTGTCGTA 120

CTTTTTCTTC ATAGTTTGTC GTACTCTTAG TCGCAACTGC TTGAAGATTA GCCAATGTCG 180

CTTTATGAAC AGATAGTTCT TTTTCTGCTG TTGCTTGTGT TGTTTTGGCT GTATTTAAGG 240

CTTGCTGACT TGCACTTCTT GCTTGTTGTG CGTCTGATTC TGCTTGTTGT GCGGTAGCTA 300

ATTTTTCTTG ATTGGTCGCT ACTTCTTTTT GAAGGGTTTC TTTACTCTTA AAGACTGGTG 360

CAGTGCCTTC TTCATAGGTA TTTGCTTCAA CTAATTTTGT TTGATTATAG CTAATGATAT 420

GGATTTTTGC TGAAATATCA TTTAAATCTC CTGAAATAGA AACTGCCATG TTTGTTTCAT 480

TTGCATCTAA TAAAGAATCT AAATGGGAAT AGCCAGAATG TCCATCGTTC ATCAACATGC 540

TTACAAGTGC ATTtCGAGCA GCTTTTTCAA AGTCATACAT AGAAACTTTA CCtTCtgTGT 600

CGTAAATCTT CCCaTACTTA GGGtTTCCAT AAAAgTTTnG TCCaGGGaAT TCCTTtAAAA 660

ACaATTTcCT TTgcCGGCCA CCGAATTGGA TCCGCCAATT TACGGGCCCA nGGATTCCCA 720

AAATTCCCTT TGGGGGGATT AAnCCCAAAA 750

(2) INFORMATION FOR SEQ ID NO: 917:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 828 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 917:

ATTCTGGnTT GGnATGTTAC CATTGCTTGG TTGCGCTGAT TGTGAAATGC TTAAAAATGA 60

CAAGnAAAAT ACCGTCTACT TGACCAAGCA CGGTTATTTT TAAATnATAA ATTGGTCACC 120

GTCTTAAACG GCTCTACATC AGAGAGATAT GTTGGAAGCA TATCTCTCTT ATCTTnATCA 180

TTATAGCACA ACAACCCTAA AAAGACCATC AATAAAAAAT GCCTGAGCCT CCTATTTCAA 240

GGCTCGGGCA TTTTTATTTT AGGGACCAGT CACCAAATTC CAAGTGACCG CAGCATTGTA 300

CTGTTGTCCT TTGGTTCCTT GGTTGGCTGG TACTTCTAAT TTAATGTTCT CAAACGTAAA 360

ATCCAGTTGG TAGACGTCAC TACCTGTGAA CTGTTGGTTT GCGACAACAC TGGTTGCCAC 420

ATTATTCGCC GTTAAGCTCA AAGAACTAGT CTTACCAACT GCTTTGATTT GTTCGGTTGC 480

TTCGTTATAA TTTGTAAAAC TTGAAACATT CGCTGGACCT AACAATAAAC GAGTCGCTGT 540

TGGCAGACTA TCTGTGGTTG CTTTCGGTTG AGATAGTTGC GCTGTCATCT GCCAATTCGG 600

TTGATTTTTC TTAATTCGCA AGTAAGGGAT TGCGTGTACC ATTTTCGTAG TAATCGGCCG 660

CTTTTTTCAA GCCTGATTGT TGCGTTTTAC TAGCAATATT CATTTTCCAA AATCAAAGGT 720

TGGGCACGCT GATnAAGCCT TCnGTTGGGT GTAATGGTCC nTGGGGTTGT AACCGTCAGA 780

CGTAACGCAT CGGGGCGGCT CATGGTCGGC AAATTCCCGT AACACAAG 828

(2) INFORMATION FOR SEQ ID NO: 918:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 497 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 918:

AGTTATCTTA ACGACTAATT TTGCAGATAA AATTGGTTAT TATACCATTG AAAATACAGC 60

GCAAGAAGGT AGTCTAGCAA TAATCAAAGT AGACAAAGAT TCAAGAAAAC GTCTAGCAGG 120

TGCAGAGTTC AAATGGCGTG ATACGGTGAC AGGAAAAGAA GGAACAGTTG TAGCTGACAA 180

AAACGGACAA GCCCTGATAA CGAATTTGCC AGTCAATCGT GTATACGAAA TTACAGAAAT 240

TAAAGCACCA AACGGCTATA TCTTAAATAA CAAGACTTAC CGAGTTATCT TAACGACTAA 300

TTTTGCTGaT AAAATTGGTT ATTATACCAT TGaAAATACA GCGCAAAArG GTAGTCTAGC 360

AATAATCcAA GTAGACCAAG tTCagAAAaC GTCTAGCaGG TGCAGAGTTC CAaTGGCGTG 420

ATACGGTGAC aGGaAAGagG AACAGTTGTA GCTGACAAAA CGGACAGnCC TGTTACGATT 480

GCCAGTCATC GGTATCG 497

(2) INFORMATION FOR SEQ ID NO: 919:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 796 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 919:

TGATCTGCCA CAGTTTGCGC TTGCTTCAAT GCTTCACTGT AGACTTTCCA ACTGCTTGCG 60

GTGTAGTCTT TTTCTTGGTG TTTCTCGTTT TCTTTGATGA TTTTCAACAA GTTGGTTTTA 120

TCTACTTCTT TAGTTGGCAC TTTTACCAAT TGCTTCACTG CCGAACGTAA CTCTGTTTCT 180

GCTTGGTCTA CTTCTGCTTG CGTTGCTGTT GTTTGATCTG CCACAGTTTG TGCTTGTTTC 240

AATGCTTCGC TGTAGACTTT CCAGCTGCTT GCGGTGTAGT CTTTTTCTTG GTGTTTCTCG 300

TTTTCTTTGA TGATTTTCAA CAAGTTGGTT TTATCTACTT CTTTAGTTGG CACTTTTACC 360

AATTGCTTCA CCGCCGAACG TAGTTTTGCT TCTGCTTGGT CTACTTCTGC TTGTGTCGCT 420

GTTGTTTGAT CTGCCACAGT TTGTGCTTGT TTCAATGcTT CGCTGTAGAC TTTCCAGCTG 480

CTTGCGGTGT AGTCTTTTTC TTGGTGTTTC TCGTTTTCTT TGATGATTTT CAACAAGTTG 540

GTTTTATCTA CTTCTTTAGT TGGCACTTTT ACCAATTGCT TCACTGCCGA ACGTAACTCT 600

GTTTCTGCTT GGTCTACTTC TGCTTGCGTT GCTGTTGTTT GATCTGCCAC AGTTTGTGCy 660

TGTTTCAATG cTTCGCTGTA GACTTTCCAG CkGCTTGCGG TGTAGTCcTT TtCcTGGTGT 720

CyTTtTGTTC CTGAAATTAA TTTTCTAACG CTCTTCGGTA ACTTCAGGAG TGGTTTAACn 780

ATGAATGAAC GCGAGC 796

(2) INFORMATION FOR SEQ ID NO: 920:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 523 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 920:

AACTTCAGAA AATATTAACA GGnAATAGCC GAATnAAGTA ACGTCTACCA AGAAGAAGGA 60

CTCTACCGTT CACAnGCGTG TTCAAATCCT ATTGTTCCGA GGTCATTACT TTACATTCGG 120

CTTTTCTTTA TAATTCTGGT TCACGATTGG CATCTTTGGA ATAAACGTAC ATTAGCGGTT 180

CTTCACGACC AACCGCATAG GCTGCTTGTG GCACATAGGC ACTCATGAAA ATATAGTCCC 240

CTTTTTCCAC TGGATACCAT TCGTTGTCCA AATTGTACAT CCCTTGGCCA CTAATTAAAT 300

ACGCGCCATG CTCTTGGACA TGCGTTTCAA TGTAGGCATG ACTAGCACCT GGTTCAAAAG 360

ATAAGATATG CATATTCATA TCAAAGTCAA ACTCTTTCGG CAATAAACTC CAAAGTAACA 420

CGTCTGTCAT cCTTCATATT CTTCGGGTTG TTGGkCaTGA ATTGACCCAA CCACTTTaTA 480

AGGtTGATGG cCTGCCAACG GGCTGAnAAG CGGnTTTTAA TAT 523

(2) INFORMATION FOR SEQ ID NO: 921:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 427 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 921:

TTTCATCAAA CGACGTTCTT CATGTTTAGT CATTTTTTGT TTTCGTTCTT TTTTATCGGG 60

AATTTTTTCA GTAATCGAAC AGAATTTACA TTTGAAAGAA CGACCATTTT TTCCTTCTAT 120

GATGACCATT TTTTTGGCTT CTGGCCGCCC GACTAACGGT ATGCGGCCCT TAGCAGATGA 180

ATTGGAGATG GCTCATTATA ATGGACACTT ACTTTCCTAT AACGGCGCTT GTGTCACCCA 240

TCACGGcAGT CAGCAACAAC TATTTAATCA GACAATTTCA AAGTCATTGa GTCAACAAAT 300

TTTAGAGCAT TTGAAACAAT TTGaCGTAAT yCCAAkGATT AATGATGrAA CATTaTGTAT 360

GTCAATGATG TTTTTCATAA CACCTTACAT TTAGAGACTG GTGATTTTAA TATTATTGAA 420

TATGAAn 427

(2) INFORMATION FOR SEQ ID NO: 922:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 650 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 922:

TTGATTTTTA nGGATnGGTT CCGGTGGCAT AAGGACAACT AGGTCTTTCA CTGGCAATCG 60

TCAATGGCCT ATTGGnAAAG ATGCCCCTAT CTGTATTATC TGAAAAATCA nACAATCATA 120

ACCAATTAGA AAGGTGGTTA TTATGTATAA CTTATTAACC AAACAGGAAA TTCAGTTACT 180

TTCCCTAATT GAATACTTGT ATGACAGTAA AGAAAAAGTC CCCATGCAAG TACTCCGACG 240

AAAATATGAA TTTTCCCATT aCAATATCAA CAATTTATTG AATCAATTAA CTTTGTTAAT 300

TTCTCGAGTC AATACACATG AAAATGTACA TATTCGTATT ATTAATAATC AACAGTCCAT 360

AGAATTAGTC GCCGATGAAA ATATCCCGAT TGAGTTAATG AAAGAAGCTG TTGTCCGCGG 420

GTCACTAACC TATATGTTAG CCCTGGATTT ACTTTTAAAA CGCTACACTT CAGCCAAAGA 480

TTTTTGTGAA GAGCACTTTA TTAACTTTTC nATTTTTAAA CAAGTCAGTG GCCGGTTAAA 540

TAATCCATCC AGCTAGGATT AATTGGGTAA TTAAACCTAA AAAGGnGGGG AAAAATTTGn 600

GGGAACGAGG AAGGATTTCG CAGGTTnCTT TAATCCCTAT CCTCCATTCC 650

(2) INFORMATION FOR SEQ ID NO: 923:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 494 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 923:

GGGCCCCCAG TTCCCCGTAC TTTAACGGAA TCATCAGTAG GTCGCCTCTG TCGTGTTGGA 60

CTnCTnGCTT TTCGCCACTr rTTTTAAATT GACTCCCAGA GAAATACACG aaCACCAGGA 120

AaTTTCATTT AATTGTCCTC TGaTAATTTT AGGTAAATAA TAGACAAcTT GTAAAAAACT 180

AACTTAGGTA AAATGATAAT GTGGACTTCG CCGTCATCTG GCGTAcATTG GCATCGAAAT 240

TGGTAAAGCC ACCAGCAGAA TTCGTCATGG TCATCGTTAA TAATTGGGCT TTGCCATGCC 300

AACGTCTTTC ATCGCCATCA ATTTTTAGTT TGTATTTTTT CTTTTTCATT AATAAGCGGA 360

AAAATTGCTT TGTAAAGGCT AAAGGACCAT ATkTTTGTTT CTCCTGTTGG GTCACATTTG 420

CTGCTGTATC TGCTAATAGT CCAATCGTTA AGGTCGAAAT CATCACATCT TGATTTACCA 480

TCCCAAAGTC AACT 494

(2) INFORMATION FOR SEQ ID NO: 924:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 484 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 924:

TTTGGCTGTA TnTAAAGGCT TGGCTGACTT GCACTTCTTG CTTGTTGTGC GTCTGATTCT 60

GCTTGTTGTG CGGTAcTAAT TTTTCTkGAT TGGTCGCTAC TTCTTTTTGA AGGGTTTCTT 120

TACTCTTAAA GACTGGTGCA GTGCCTTCTT CATAGGTATT TGCTTCAACT AATTTTGTTT 180

GATTATAGCT AATGATATGG ATTTTTGCTG AAATATCATT TAAATCTCCT GAAATAGAAA 240

CTGCCATGTT TGTTTCATTT GCATCTAATA AAGAATCTAA ATGGGAATAG CCAGAATGTC 300

CATCGTTCAT CAACATGcTT ACAAGTGcAT TTCGAGCAGC TTTTTCAAAG TCATACATAG 360

AAACTTTACC TTCTTGTGTC GTAAATCTTC CCATACTTAG GTTTTCATAA AAGTTTTGTC 420

CAGGGAATTC TTTAAAACCA TTTTCTTTTG CTGCACGATT GATCGCATTT ACGTCATGAT 480

CAAA 484

(2) INFORMATION FOR SEQ ID NO: 925:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 633 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 925:

GCAACnTATT TnACACCTTG GATTAnCAGG TTGGCCCATG TGTTCTCTAC CGTCAGATAA 60

TCAAATATGG AACCCCnCGG CAATCGCTCT TCTTGGCTAC ATCCGAATCA TTTTAAAAAT 120

ACCTGCACAA CTAAGTTATA TATTTTGTCC GCTTTAAATG GTTTTGAAAA CAAACCGATA 180

TCAAAAACAG TTACTCGCAA ACTGTCTGAT TTATCTTACA AAGATTTATA TTATCAAGAG 240

ATGAAACTTT TAAAAATTAT TCGTTAATAT ACTTATAGTG GTCTATCAAA AAAAGAATAT 300

TTATTCCGTT ACATAAACTT TGAAAAAATG AAAAATAAGA CCTTCAACAT ACTCTATCCT 360

GTTGAAGGCC tTCTCTCTTA TTTAACCCTT TAATTTTTAC GACCtCTAAA TAATCATACC 420

CAAAATGTGC ATACGCAATG GCGTGCAATT CGTCCTCGTG GGTGTCGTTT AATGAAnCCT 480

TTTTGAATAA GAAnGGGTCA TACATACCTC CACCGTTCGG TTCTCACGAT ATTACAGCAA 540

GGACTAACCA CGGACGCGCG ATAATCATAT GTTCACATTT GTCGCATATC AGCTGAGACA 600

CTGCGAnGTA CnTTGCGCAC GCGCGTTGCA CGA 633

(2) INFORMATION FOR SEQ ID NO: 926:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 566 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 926:

TAGTTTTAGC AGCAACGTCG GTCAGCATTT CAGTAGAAGT ATTAAGAGAG TTGAATGTTT 60

TAAGTAGTAA AGAAGGTGCA CGnnTTCTTG GGGCTTCCGT CGTTGACGAT ATCGTGGTCG 120

TGATTATTTT AAGTGTCGCT GTCGGCATGA TTGGTGCAAG TACAGGCGGC AATACAGAAG 180

TCAGCTTTAT TGTTAAGTTA ATTGAACAAG GCTTATTCTT TATTGGGATw TTCTTTTTGG 240

TCCGCTTTAT TGCTCCCTAT kTACTACGTT TAAGTCAAAA AATGAACATT GGGTCGTCTG 300

TTATTATCAT GTCATTGATT ATCTGTTTGG GCATGGCTTA TTTGGCTGAT TTAGTTGGTT 360

TAAGTTCTGT CGTAGGTGCT TTTTTTGCAG GGATTGCTGT GGGACAAACG GATGCAAAAA 420

CAGAAATTGa TTTCAATATT GAAGCTATTG GCTATGCCGT GtTTAATCCC gyCTyCTTgc 480

CAGCATgGCy TAtCAGTGAC ATTwACACGT AnGAAAGGAT TAnCTTTTAT TTGGGAATGA 540

CGATATGGCG AATTTAACCG AAAnTA 566

(2) INFORMATION FOR SEQ ID NO: 927:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 490 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 927:

TTTTTGTTGG CGTGATTGGG TATGCCAGTT ATCTTTTTAC CCAAAAGTTA TTCCAAGTTC 60

CTTACCTCGA TTCATTCGCA GCCGCTTTTG TCATTGGTTT GTTGGCCTAT TTGGCTGTGC 120

ATTTTCACTT AGCGGTCAAC ATCGATAATA TCATTATTGG TGCGGTAATG CCGTTAGTTC 180

CAGGTGTGGC GATTACGAAC TCTTTTCGCG ACATTTTAGC GGGCCACCTC ATCAGCGGGA 240

CGGCGCGAGG GACAGAAGCT ATTTTTATTG CTGGGTCTGT TGGCCTGGGa ATTGCCTTGA 300

tTTTTAAATT ATTTaTGTAA GGAGTGTTCA AATGGAcTAA TTTTyCATCT cyCTTTAGTT 360

TTTTAGCGAC TGTGACGTTg GaTTATTACG AcATcCTCGT AAAcAcTGGg GCTGgGCATA 420

CTGGCATGAC AGGCTGaTGA tTATAwGcTA ACGcaACtTg wgCaGCaACt TGCAACTTTA 480

GGACGCCGGA 490

(2) INFORMATION FOR SEQ ID NO: 928:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 460 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 928:

ACGCTTTTTC TGGTGTGACG GCcAGTTGCG GAAACCGTGA CGATTGAAAG TAGTCCGACC 60

GCCGAAAGTA GTGCCAAGGA AGAGACGCAA GCAAGTAGCG TGAAGGAAGA AACAACGAAA 120

GCCAGTACGG AAAATAGTCA AGTAACAACT GACACGAGTC AGGAAGAAGC AACGAAAGAA 180

GCGGAGAAAG AAGAACCGCA AGCAGAAGTG GAACAAGCAG AAACACCAAT CATTCCTAAA 240

CCAAAAAAAA TCAATATGAG GCACTTATTC ATTTTCTGCA GAACTTATCA GTTTGGATTG 300

TGATGATCAG GTCATTATAA TCCAGTATTT ACCATTACGT TACTATGCCA AGGTCATGGA 360

GGCAGATGGT ATAATCGAAG GGnCTAGTTG GTCAAGGATG CTTCACCGTG GAACTTAAGG 420

ATGTATATGC CACACTCTGn GTATGCnCCA GnCGCATTGA 460

(2) INFORMATION FOR SEQ ID NO: 929:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 394 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 929:

CCTATGTAGG CTGCGATATT CAAAGCTCCG CAGCTTGCAT GGACAAATCA CTGGCCTACA 60

TTCTTACnAA AAAATGCGGG CATCGCCGTC CCCGAATTTC GGTTTTCACA ACGCGGCCTT 120

CCATCATGAT ATGAACAACA TCTGGCGTGA TATAGTTCAA TAAGCGTTGA TAAtGGGkAA 180

TAAtTAACGC ACCAAAGTTC TCGCCGCGCA TTTCATTAAC CCCTTTAGAA ACTACTTTTA 240

ACGCATCGAT ATCTAAGCCT GAATCAATTT CATCTAAAAT GGCAAACGTT GGCTCTAACA 300

TCAATAATTG TAAAATTTCA TTTCGTTTTT TCTCGCCGCC AGAAAAACCT TCGTTTAAGT 360

AACGCTCAGC CATTTCTTCA GGCATATTTA ACGG 394

(2) INFORMATION FOR SEQ ID NO: 930:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1317 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 930:

CTCCAGCATT GTTGGTCATT TTTTCTAATT CTnCATAGGA GTGTTCTGCA TCACTAAATT 60

TAACCATTTC TACTTTATGA AACTGGTGTA AACGAATTAA TCCTCGGGTA TCACGACCTG 120

CACTACCCGC TTCCGAGCGA AAAGAAGGAC TTAAAGCAGT AAAATAAATC GGTAACTCTT 180

GCGCATCTAA AATTTCATTA TTATAGTAGT TAGTTAACGG GACTTCCGCT GTGGGGATCA 240

GCGTCAAATC ACTTTCTGCA ACTTGGAAAA CATCTTCTTT GAACTTAGGA AACTGACCGG 300

TGCCAAACAT TGCTTTGCTA TTGACCAAAT AAGGAGGAAT CACTTCGGTA TAACCATGTT 360

CATTAACATG TTGATCTAAC ATAAAATTAT AGACCGCTCG TTCTAAACGA GCCCCTAAGC 420

CTTTGTAATA AAGAAAACGG CTACCAGATA CTTTCGCGCC TCTTTCAAAA TCTAAAATTC 480

CTAAGGCTTC ACCGATTTCC CAATGTGGTT TTGGTTCAAA TGAAAAAGTT TTCGGCGTAC 540

TCCACCGCCG AACTTCTACG TTGGCGGCTT CGTCTGCGCC AACAGGTACG CTCTCATGAG 600

GTAAGTTAGG CAAACGTTCT GCAATGGCTG TGCATTGTTC TTGCAAATGA ACAACTTCTT 660

GATCTAAAAC TTTAATTTGC TCGCTAACTT CTTGCATATT TAGCAATTGT TGTGAAGCAT 720

CTTTTTGTGC ACGCTTTAAT TGTGCAATTT CTCCTGACAC CTGATTACGG TGTTTTTTCA 780

GTTCTTCAAC TTTAACAAGT AACGCTCTCC GTCCTTCATC TAAAGCCAAA AAGCGTAATA 840

AAACTTCCTC GTCCACACCA CGTGTTTTTA ACTTTTTCTG TACCACTTCT ACATTTTGGC 900

GAATTTTTTT CATATCCAGC ATCTTATTTC CCTCTTCATG ATTAAATTTT AAACAATGGT 960

GAATTTTACT TTTTCAAAAA TTTCATTCTT CATTCGTTGG CCAGTCGCTG TGCCAGCTAA 1020

GCCACCTAAA CCCGTTTCTC TCAATTCTCG TGGTAATTGA CGGCCGACTT TATTCATGGC 1080

TTCCACCACT TCATCTGCTG GTATTTTATT TACAATTCCA GCTAAAGCCA TATCTGCAGC 1140

GATTAACGCA TTGCCTGCAC CTACTGTATT TCGTTTGACA CAAGGAATTT CGACTAACCC 1200

AGCCACTGGG TCACAAACTA AACCTAATAA ATTCCCTAAA GCCATAGCAA AAGCTTCAGA 1260

AGATTGGCGT GGTGTTCCGC CAGCCGCTTC AACAGCTGCA GCGGCGGCCA TCGCTGA 1317

(2) INFORMATION FOR SEQ ID NO: 931:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 724 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 931:

CCCTTGTCAC ACAAGCGTTT ATGGTTTATG ATAAAGTAGT AAAAAAGAAC TACTGTCTTA 60

CAAATTTGCG TTCCGGCCAA ACCAAAAAGA AGCACAGAAA AAAGTTTTAG AATTTTTTGG 120

TATGGATACG AACCCAGCAA ATATTCAGTT AATTAATGAA CTAAAATCAG GAACTTGTTT 180

ATTCCAAGAC CATCGAGGGC GTAGCCAACC TATTGCAATT GATGTTTTAT TCGATAGTTG 240

GTTAATGGCC ATTTCGTCTA CTAATAAAGA GGATGAAGCT ACTCAAATGG CGTTGGCAAT 300

GGAACAAGGA TCATAAAAAA GGAGAAGTGA GCATGAAAAA AGTACTACCT TTTATTGCCT 360

TAGTCGGCTT GTTATTGTTG TCAGGTTGTG GAACAGATAT GAAAAAGATA TTGACTGCCG 420

ATGGTGGTAA ATGGGAACTA GAAAATAAAA GTCCAACTAC TACTTACACT TTTTTTGATG 480

ATGAAACTTT TTCGAGGTAT AATTCAAAAA TTAGTGATAG TGGAACGTAC TCTTACGATG 540

AAAATAATAA AAAACTCACT TTGGATATAA AAAATAAAGA ACAATTAATA ATGGAAAATG 600

TTGAATATAA AGACGGTAAA TTAAAAGGTG AAATTGGAGG CGAGAAGGAC TCTGATAAAA 660

AATnGAATAA GAGGTGTCTT TGAATAGAAA TCCTGGTAAG CGTCTnCTAT TTTTGnGGCG 720

TTTT 724

(2) INFORMATION FOR SEQ ID NO: 932:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 342 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 932:

AAGGTACCTC CAGTAGTTGT ATTGCCACTT GCCAATAATT GTTCACTGCT ATTGTAAAAT 60

TGAACCGTGA CATTTGGATC AGCAGTTCCA GTGACTGTAT AGCCATTGGC TTTGTTTCCT 120

TCAACCTTCG TAATAGTTGG CGCCGCTAAA ACAACATCCG CTGGAATAAC GGCTGTAGCT 180

GGTTGACTTT CATTACCTGC GCCATCTTTG CTAATAATCG TAaTCGTwTC TCCTGGTGTC 240

ACTACGcCAG CtGGTAGAGT CACCGTATAT TGGCCGGTTT CGTTAGCAGT TGTAGCAGCA 300

ATGATGGTTC CGTCTGCGTC ACGGACATCA ATAGTGGTTT Tn 342

(2) INFORMATION FOR SEQ ID NO: 933:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 341 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 933:

AAATTGCCAG ACCTACTTTT TATGGAGACG ACTTTGGAAA TTATTAAAAC GACGAAATAA 60

CTTCATTCAG GATTTCATGC CAAAAGGGAT TACGAAAGCC TATGGTATTT CCTGTTAGCC 120

AAAGATTTAG GAATCCGTGG CAGAAGAAAT CATGACCCTA GGGACGAAGA AAACGATTTA 180

CCGATGATCG AATACGCGGC TAGTGTCGCC AGGCAAATGC GATTCCATaG TTAAAGAAGC 240

CGCAGATGTC GTCACAGACA CGAATGACCA GnGGCGTTGC CAAAnCGGTC GAAAAATATA 300

TTTTAACGCC GTTGGAAAGG AGGCCATTAA GAnGGGGATT T 341

(2) INFORMATION FOR SEQ ID NO: 934:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 519 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 934:

GACTTAATGC TATTGAAGCC CGATGGATGG TGnGTAAGGT CnCGAAGCTn GnGGAATCAT 60

AAAAGTTAAT AAGCCTGGTT TTCCTAATGG GGGCGTCACT TAGGATAGTT TTCTTCCATT 120

CTCCAATCCT GCCGGATAAG AAGGTATTCG CTTCCTCCTT TGGAGTTAGC TTGATAAACA 180

GCCCTTAAAG TCATCCACAG ATTTCTTTTC GGTCTTTGAC ACGTACTTTA TGAGCGATAT 240

TACGAGATAC ATGGATACAA CAATGCTGAT ATTTTGCTTT AGGATAAATT TGATGGATAG 300

TATCTTTCAT GCCTTTTAAG CCGTCCGTAA TAAAAAGCAA GACTTCTTGa ACTCCTCTGG 360

AGTTAATATC CTGTAGCAGC TCATTCCAAA CGTATGTgAT TCAGTTGGAG CAATCGCATA 420

ACTCAGTACT yCTTaGTGCC GTCCTCCCCG AAAACCATnG GCCAAAATAA ACCGCTCCTT 480

GGGATAACGG TTTGACGTTT nAGTGGAATG nAAGAAGCG 519

(2) INFORMATION FOR SEQ ID NO: 935:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 390 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 935:

CATATGGGGC TATACGTCAC GATTGGCATT GGAGATAACC CATTGTTAGC CAAGCTGGCC 60

CTAGATAACG AAGCAAAAAA TACACCAGGT TTTGTTGCAG AaTGGCGATA CGAAAACGTA 120

GAAGAAAAAG TTTGGACGAT CTCCCCTATC ACTGAATTTT GCGGGATTGG cmAGCGTATG 180

GCArCTCGTT TAAAGATGTC aGGAATTGAr TCgaTTnATG ATTTAGCACA TGCAGATCCG 240

TACATTTnAA AACAACGTTT TGGTGTCATG GGTTTACAGC TGTACGCTCA TtCTTGGGGG 300

ATTGATCGAa GTTTTTaGGA GGAAAAAAGA CAAGTTGCTA AAGAAAAGTC nTTTGGAAAT 360

AGTCCAAGTA TTACCnTAGA GATTACCGCA 390

(2) INFORMATION FOR SEQ ID NO: 936:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 443 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 936:

CTATTAAAAA ACGTCAAGAA GAAGCAAGGA AACATGGTAA AGCAGAAACT GGCTTAAAAC 60

AACGGCCATT AAATAAAGAA CAAGAAAAAC GTATGCnACT ACGCCGTGAA TTGAAGCCAA 120

GACCTCAACA TAAAGCGCCA AATGCTATCG AAAAAGCACA AGAAGGATTA AGACAAGCGC 180

ACGAAAAAGG ACGTATGCAA GAACAGAAAT CTGAAGAACA ACTCAAACGG CAGCAACATG 240

AATCAGCTCG CCAAAAAGAA TATGAAAAAG ATCGATTGAA AAAACAGGAA TCACTTAAAA 300

AAATTAATGA TCAAAGCGAC GAATCAACCA tTTAACAGCT TraGAAtCaA ATCGGCGTAC 360

TGGACAGCGA GTTAAAAACG CGAACAGGnC TTCAACACTT TTGAACGCCA GGGGGAAAGn 420

TCTTCACCAC AATTAnCCAA AGT 443

(2) INFORMATION FOR SEQ ID NO: 937:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 228 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 937:

AGGAATCACT TAAAAAAATT AATGATCAAA GCGACGAATC AACCAATTTA ACAGCTTTAA 60

GAAATCAAAA TCGGTCTGGA CAGCGATAAA GGGGGCGAAA CAGGTCATCA AACACTTATG 120

AAACGACAAG GAGAAAAAGT ACTTCAACCA ACAAATAAAC CAAAAGATAT GAAAAArCAA 180

GTGGTTGnTC GTAGTCCAAG ACGAACAGGA CAACCAACGA CTCGTCAA 228

(2) INFORMATION FOR SEQ ID NO: 938:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 330 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 938:

AAGATGTTTA GTTGTAGAAA ATAAACAGGT TTGGGAAGAA AGTGGGACAG GAGTATTTAG 60

CTTCAAGAAT GAAGAAGCTA CTCCTGTCCC GCCGTTTATT TTAGGATTGG GaCGGATGTT 120

CTAAGTTCTT CCTAAAtCGT TTCTGAAAAA AtGAAgGAGT GAGGGAGAAG ATGACAGrAC 180

CAAtGTGgAa TtGAGAaGTt AGGATtaGAA aGCGCATCCt GAaGGgCcGG TtATTTtAAa 240

CcAAACCAAT AAAAGTGAAC AATTGATTCC GACCAAACGA GGCGAACGGG CCGTnGCATA 300

CAGCGATTTA ATTnCTATTA AATnCAGAAA 330

(2) INFORMATION FOR SEQ ID NO: 939:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 314 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 939:

TACGGTCAAG ACCGAGTATG GGGAACTTCA TCTCCAGATT CCGCGCGACC GCAATGGCGA 60

GTTCAAGCAA CAGACTGTTC cTGCTTaTAG ACGGACGAAT GACACGTTAG AGGAGAnCCG 120

TCATTCACCT CTTTCCGAAA AAGTATTAnC CATGTCGGAA TCGCAGACTT GATTGAGAAA 180

TGTTGGGGCA TCATACACGC CCAAACCATG TCCATATACC AAATCATTAC AGAGGGTACG 240

CGTTAAnGGG GGACTCATGC CGTATGCGTA TTATATGCGC ACGATATCGT AACGGAACGC 300

GCAAGAGCAT CTAC 314

(2) INFORMATION FOR SEQ ID NO: 940:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 564 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 940:

CATCATATCA AACAAAGAAT TGGACGTGTT GGCATCTTTT AGGACGGAAA GACCCGTAAA 60

AATAAGAATC GTCATTAAAA AAATACCAAA AAACCGTTTG TATCGTTGTT GTCCAATAAA 120

TTTTACACCT AAAATAACAA AGACAAAGGC AATACCAATC GTTCCCGCAA TGCCCAACAT 180

GTTAGTTGCT ATGCTAGACG TTAGGTTCAT AATCGGTTGC TTAATAGCAG TTATAATATC 240

CATACTAAAT AGTGTTTTTA CCATAACTGC ATTAAACTCT CCGAGTCCTT TTACCATTGG 300

ACCAGACAAA AGTCTTTATT GTCGCAGAAG CATTTGTGGA TGGCTTCTTC AACACCTGTC 360

CATGAATCnT GGTTTTTTTT CTTTGGtCAT TAACTCGAAA GCGTTGTCyT TGTAAGTAtC 420

GTAAATCmAA TCTACATTAG kGGTTGGCTC yGTAACAAtA ACTTwATTTC ayCAGGGTTA 480

GGGTCGGGCA TTAACCGCTG GAGTTTGCGG TAAACACCAA AAAACAGCCT ACTAAAAATA 540

GGTAAGACTG GCTTTACCAG TGGT 564

(2) INFORMATION FOR SEQ ID NO: 941:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1150 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 941:

AATAATTGAA GACGAAAAGA GGACGAACTG ATTGTCCTTC TTTTTTTAGT TATGGAGGGG 60

AnCCATTGGA AGCAGTAGTA GTAGAAAGAG AGGCAAAGGG GATGAAAGAA ATTGCCATCC 120

AAGAAAAAGA TTTGACCTTA CAGTGGAGAG GAAACACAGG TAAGTTAGTT AAAGTTCGAT 180

TAAAGAATAC ACGTGCAATG GAAATGTGGT ACAACAAACA AATTACCGAA GAAAACATTC 240

AAGAGATCAC CACGTTGAAT ATCATTAAAA ATGGAAAATC TTTGGCATTA GAAGTATATC 300

CAGAAAAAAG TATCTATGTG AAACCAAATT TAGGCAGAAT CAATGTGCCT GTCTTTTTTA 360

TCAAAACACC TATTAACAGA GGAGTATTTG AAGAGATTTT CGGCGAAACG TTAAAAGCAT 420

AAGTAAAGGA GAAAGAAAAG CCAATGTATG TATCTTTTGT TCTAAGTTGT TTGCTTssaT 480

TTTCTGCCTA CAGTCTATTA AATAGGCTAA ATTCGTTGGA ATTTGTGGAT GTTTGGTTAG 540

ACAAAGrAAC ACAAAAAATC ACACTAAAAC GCTGTTTTTA TGATACGTCT TTCAAGAAmC 600

AAaCACTAAA aGAGTtAGca CGAGTaTATT TCCAATTAAa wGAAATaATC aaCGTGCAAA 660

TAAACAAGCG GTCTTTAAAT ACGAATGACA TACGTAATGT ACGAGAACTA GAGGAAAAAC 720

AACAAGAAAT AAAACGATTC ATGTTAGACG TTTTAGAAGA TGCTTATTGG AAAGAATTAG 780

CAAATATGCC AGAAGACCAA CGACACTTAG ACGATTGGGA TTTCTTTTAA AATGGGTTAT 840

TGAAAGGAAG CAATAATAAA AAAAATGATG CCTAACACAA AAATATAGGA GGGAAAAAAT 900

GATTATTTTA TCAGGACTTT TTATTTTAGg GATAGGAATA GTTGGAGGTT ATCAGCTAGC 960

CACGTTACCT AAATTGATTG AAATGAAACA GCACAAAGCA ATACAGAACC ATTTCAATGT 1020

TAAAGGAAAT GAGTACACCT ATTATCAAGA AGACAGTGAA AATTACATTT TATCCTTGGA 1080

AGATACAGAA TATCGAATTA AATTTTCCAg AACACCCCGT TAAAAGTGGT ATTCCTGAAT 1140

TTTAGAnCCn 1150

(2) INFORMATION FOR SEQ ID NO: 942:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 496 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 942:

TTTAATTCCA CGGAGACnTT GATTATGGAC GGTTCTTTCA GGATTAAAAA ATAGTGATTG 60

GGACgCTGTA GkTCATAAAA ATGCGTACTA TGkTTCTGGT GTAGGTCTAG CTAGAGAGGt 120

GGACGTATTT CTTTyTCTTT TGGTATGACA ACCAAAGGAA AAGTCAACCT ATCTGGTGCG 180

CAATGGTTTG CCTTTAGTAC CAATTTAAAT GCGAAATCAa TTAGACCATA CCAAAaGAAa 240

GGGAATCCAA AaGAACCAGA AAAAGCAACA ATTGAATTCA ATCGATACAA AGCCAATGTC 300

GTTCCTGTTC TTGTGCCGAA TAAAGAAGTC ACTGATGGTC AGAAAAATAT CAATGATTTA 360

AATGTGAAAC GAGGCGATTC TTTACAATAC ATTGTGACAG GGGATACGAC AGAACTTGCC 420

AAAGTAGATC CGAAAACAGT GACAAAACAA GGGATTCGGG ATACCTTTGA TGCAGAAAAA 480

GTGACGATTG ATTTAn 496

(2) INFORMATION FOR SEQ ID NO: 943:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 466 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 943:

AAAATGTAAA TCCCATTCTC CAGGCACCAC GTGATGTATG GGGTnCGGGA TGACTTGCCG 60

GAATGTTAAA ACTAACGGAT CCCCCAATTT CGTCGCACGA TCAATGGCTC TGCACAAAAT 120

TTTAGCAAAG CCGTACACCG CATCATGACG GTAACTCATT GGCGTAGTAC CTGCATGATT 180

GGCTTGCCCT TTTAAGTTAA CCGTGTAGCG ACGTTGCCCT ACAATACTAT GAACAACACC 240

GACTTGGAGC TTTTCGTTCT CTAAAATATT GCCTTGTTCA ATATGAATTT CGACAAATGC 300

TTCAATGTCC TCACGTCTAG TTTGCTCTTG CCgAAAATCA AAGCCTTGGC GATGCATCTc 360

ATCAACAAAC TTTTTACCTT CGTTGTCAGA GATATCGACT ACTTCTTCAC GTTTGGCTTC 420

ACCAACGACA TTTTTGCTTC CCCAGAAAAC CGTTGGAAAA CGACTC 466

(2) INFORMATION FOR SEQ ID NO: 944:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 444 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 944:

TTAAAAACTT GGGCCTCCCG ATGGCACGCn GCCCCAAATG CTCAAAAATC CCCnGATTTT 60

TAGAGGCCCn TCCnCCGTAG CCACGGTGTA TTATCCCGGA CTAGCGAGTC CATCCTGGTC 120

CACGAATTGC GCAAAAACAG ATGTCGGGAT TTTAGTGGCA TGATTTCTTT TGAATTAGCC 180

AAGGATGAGC AAGTCGTTCC TTTCGTAGAA GCGTTGGAAA TTTTTACTTT AGCGGAAAGT 240

TTAGGTGGCG TAGAAAGTTT AATTGAAGTA CCTGCCATTA TGACTCACGC ATsTAkCCCC 300

AAAGAAAATC GTGAAGCCGT GGGAATTAAA GATGGACTCA TTCGTTTATC TGTAGGAATT 360

GAAGCAGTCG AAGATTTATT GGCTGATTTA GAAAAGGGAT TACAAGCATA ArCAAAAGGs 420

CCTCGTCTTT AGACGAAGGC GTTG 444

(2) INFORMATION FOR SEQ ID NO: 945:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 568 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 945:

GATTTCnGCA CAGTCCAAAG TTATTGGTnA TTATTCCACC GGCnGATTGG AGGGAAAGAT 60

TTTTTAGTTG AAAAAGTATT GGTnCCTAaT GAAGTCAGkT TGaCgTwTmC CCACAATGAT 120

CGcATGATCt TTGGCGGTGT AACcACCGAC GACTGAGTcG TTAGAGATTA TTTTGAAmnA 180

AGAATTAGGT GTCGATTATT TcTTGGAACG GcGTGAATTA GGAGTCATTA ATATTGGCGG 240

TCCGGGTTTT ATTGAAATCG ACGGTCGAAA AGAAGCGATG AAGAAGCAAG ATGGCTATTA 300

TGTCGGGAAA GAAACTCGAC AAGTGATTTT CTCTTCAGAA GATGCAGCTG ATCCTGCGAA 360

ATTTTACATT AGTTCTGCTC CGGCGCATCA CAAGTACCCA AATGTAAAAA TCAGTATTGA 420

TGAGATTAAG CCAATGGAAA CTGGGGAAGC ATTGACATTA AATGAACGTA AAATTTATCA 480

ATATATTCAT CCCAATATTT GTGAAAGCTG TCAGCTGCAA ATGGGCTATA CGATTTTAGA 540

GCCGGGCAGT TCTTGGAATA CCATGCCG 568

(2) INFORMATION FOR SEQ ID NO: 946:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 486 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 946:

CTTTCCACTC CGCTTTTCGC AATCCTCAAG CTGGCTGtCT gTTCGCAcTG TAcgCTGtGA 60

tCTTATCATT gcACTTAgCA GCATctTTCt TTGcTTTTTC tCGACAACgC yTGTTGTTGG 120

TTAACGTCTG CTgAGCTTGA TCCACTACTT TTTGTTGGTC ATCAACAGCT TGTGTATCCG 180

TAGCCACTTG GTTTTTGGCT TTTTCAAkGG CTGAAGGTGT TGCTTCATCC ACAACTTTTT 240

TTGCTTCGTC TACGACCGCT TGTTGGTCAG TCACTGcTTG TTGACTTTGA TCCAATGCAC 300

CTTTGTTTTG ATCGACCACT GCTTGTTGGT CTTTAACAGA TTGATCAATG GTGTCTTTTT 360

CTTTTTTCGC AGTGTCAGCA ACTTGTTGTT TTTGATCGAC AATTGCTTGT TTCTCTGTTA 420

CTTGTTGCTG TTTTTCAGTA ATGGCTTGCT CCGTTGTTTC TGTAGCTTTC ACTGCTGGTT 480

GTTCTG 486

(2) INFORMATION FOR SEQ ID NO: 947:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 148 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 947:

ACATAACGTT CCACGATCCC ACAAAAATGG CCCAATGCAC ACCCCTCAAT GATTCCTTTC 60

AAGACGCTAC TAAAAAAGCT TATCACGATA TCGCtTCATA GAAAtCGCTt GTTgCATGAA 120

CTGCTGAACG ATATCCCGTT AATAAATA 148

(2) INFORMATION FOR SEQ ID NO: 948:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 587 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 948:

CTTGGTAAnn CTTCTTTGTG AAGCCAAGCC TTTCAATCGG GAAATATTTA AATGGAACTG 60

CCCCCCATGT CTTTCAACAA TTTGATGGCG GTTTCGATAG GCACTTnCCG CCGTGGCGCA 120

CCAGAGCTTA ACGGACCTGT CGCAATGTTG ACATAGCCAA CTTTGCCTGT GGGTGAAACC 180

AAGCCATTAA TCACkGTTTC GTCTTGACGA AGTAAGGcCC CGCGAAGCGC CGACACCTGT 240

GAAGACTTGG TTCACGTGTT GCGGTTGTAG AACTTCTGAT AAGCGAGCCA CCATTTGACT 300

TTGATTAGGA TCGCCAGCCC CTAAGCCGAC AGATAAAGCG TTGTTGGTTG CTGCTTGGTA 360

TTTTTTCATG TCGTCAATAG CTGCTTCATC GGTTTCATAG TTTTTTGAAA GGACGCCTAA 420

CACAACATGC CCTTCTGCTG CTTCATAACA GGCTTGGGCA TTTTCTACTG AATTGGCTAA 480

GACGTTTAAA CAAATGCGTT CTTCTAAATA ATTAGGTTTT AATGACATGA CTTATTTCTC 540

CTTTTTGTCC ATAATTTCTT GTCAATnTTG CACGATTTTG TTCATTn 587

(2) INFORMATION FOR SEQ ID NO: 949:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 468 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 949:

ATTnCGTTGG GGTTGCCACC AATTATCCCC CCCnCAATTT AAGTGAAGTG ATTGATsCAm 60

CAAGTTTATT GATGGACaAT CCtGATGTAA CGACGAATGA ATTGATGGAA GTGTTACCTG 120

GACCAGATTT TCCAACAGGC GnTTTAGTGA TGGGGAAATC AGGGATTCGC CGAGCATATG 180

AAACAGGGAA AGGTTCGATT ACTGTTCGTG CAAAAGTTGA ATTGACTGAA ATGCCGAATG 240

GAAAAGAACG TATTTTAGTA ACTGAATTGC CTTATATGGT GAATAAAGCC AAATTAATCG 300

AACGAATTTC TGAATTACAC AGAGATAAAC GAATTGAAGG AATTACTGAT TTGCGGGATG 360

AATCTTCTCG TGAAGGCATG CGGATTGTCA TCGATGTTCG TCGAGATGTG AGTGCCTCTG 420

TTGTGTTAAA CAACTTGTAC AAAATGACTG CCTTACAAAC ATCTTTTG 468

(2) INFORMATION FOR SEQ ID NO: 950:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 950:

CGGGGAAATG AAGTACAGGC AGCAGAACAA GCGCAACCAA AAACACCTGA AAACAGTTCT 60

ACAGA 65

(2) INFORMATION FOR SEQ ID NO: 951:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 730 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 951:

ATCAAGCAGA TGCAAGTCTA AACGAGAAAG ACTTAAAAGC TGTTGCTGCA GCGATTAATT 60

CAGGAAAAGC TATAGACGTG ACTGCTTCTT ATGTTCTTAA TTTAGATCAA AACACCGTCA 120

CAGCaATGAT GAAAACCAAC GCAGACGGTT CCGTTGTTTT AGCAATGGGG TATAAATATT 180

TACTTGTCTT GCCGTTTGTA GTGAAAAATG TAGAAGGCGA TTTTGAAAAT ACAGCTGTTC 240

AGCTGACAAA TGATGGCGAA ACGGTAACAA ATACAGTGAT TAACCATGTG CCGGTTAGTA 300

ATCCTTCCAA AGATGTAAAA GCAGATAAAA ACGGTACAGT TGGCAGTGTT TCTCTACATG 360

ATAAAGATAT TCCGTTACAA ACAAAAATTT ATTATGAAGT GAAATCTTCC GAACGTCCAG 420

CTAACTATGG CGGAATTACC GAAGAATGGG GCATGAATGA TGTCTTGGAC ACGACCCATG 480

ATCGTTTCAC AGGTAAATGG CACGCTATTA CAAACTATGA CCTTAAAGTA GGGgACAAAa 540

CGTaAAAGCA GGAaCAGATA TTTCTGCCTA CATTCTTTTA GAAAACAAAG ACAATAAAGA 600

CTTGACGTTT ACAATGAATC AAGCATTATT AGCAGCGTTA AATGAAGGAA GCAATAAAGT 660

AGGCAAACAA GCTTGGTCTG TATATCTGGG AAGTCGAACG GGATCAAAAC AGGTGACGTA 720

GAAAATACGC 730

(2) INFORMATION FOR SEQ ID NO: 952:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 758 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 952:

TTTCAATAGG GACTGTTGGC TGTTGTGTAC CATTTCCAGC GTGATTTAAT TCCCCTGTTG 60

GCGTTACAAC AACGTCGATT GGCTTGCTTG GTTCTGCTGG TTGTACTGGT TTTTCAGGTT 120

CTGTTGGTTT AGTTGGTGTC ACAGGTTTTT CGGGTTCTGC TGGTTTGCTT GGTGTAACTG 180

GTTTTTCAGG TTCTACTGGT TTACTTGGCT CCGTTGGTTG TTCTGGTTCA CTTGGCTCTG 240

TAGGAGTTGT CGGTTCTGTT GGTGTTGTCG GTTCCGTTGG AACTACTGGT TTTTCAGGTT 300

CTGTTGGTGT TGTCGGCTCT GTCGGTTCTG TTGGCTCCGT TGGTGTAATA GGTGGCTCTA 360

CAGGGTCTAC AGGTGGCTCA ATTGGCGGTG TTGGATCTGT TGGCTTTTCT GGTGTACTTG 420

GTTCTGTTGG TACTTCAATG GGCGGTGTCG TAGTATCATC AGGGATAATT TCTTGTGCAA 480

AAGCAGAAAT TTCACTTCCC CCAAATAGTA GTGTGGCACT AAATAAACTA GCAAAAAGAT 540

TTTTTTTCAT CATTATTTCC TCCATTTTAA TTTTAATTTG TAAGTTCAAA ATCTTCTTCT 600

GTCGGTCGTT CCATTACGAC CAGTTGTTCT TTTGTTcTTC AGCTGTAAAC GTAATTTTCT 660

TTGTCGATgC TCGTCTTTtT CATTGATTAA TACAACAATG AACGcTaTCG TCCATkGTTg 720

kCTTCCTCTT TACTCCCAGG GAAACATACT TCTCGCTC 758

(2) INFORMATION FOR SEQ ID NO: 953:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 499 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 953:

AATGCCACGG TTGAnATCCG AAATGCAGGA GGnGCCGTGA TaGGTACAGG TACTGCTGAT 60

GGGACAGGGG CATTTACAGT TACCATTCCC GCAGGTGAAG CAGGTGCGAA TGAAACGTTA 120

ACCGCCGTAG CGAAAAACGC CAGCGGTACA GAAAGTACGC CAACAACGTT CCAAACGCCA 180

GCGGATCCTA ATACGCCCGT GGCGACGCCA ATTGTTGAGA CTGTAACAGG TAGTACAACA 240

AAAGGCTATG AGGTCAAAGG GACTGCTGAA GTTGGCACCA CCATTGAGGT TCGCGATGCA 300

GCTGGCACGG TCCTTGGTAC TGCAACAACT GGAACTGACG GAAAATATAC AGTGACTTTA 360

GATTCAGGAA CAGCAACAGC AAATCAAACG CTGAGCGTTG TAGCGAAAAA CGCTAGTGGC 420

ACGGAAAGTC AACCAGCAAC GGCGACAAAC ACCAGCTGAT GTCACTGCAC CAACAGTTGA 480

TTACATCACA GGnAACTTT 499

(2) INFORMATION FOR SEQ ID NO: 954:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1753 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 954:

CTGTTAGGCA TTCATTTACT ATTTTGAGGT GTCTGGCTGA ACAnTTGTTG TTCCTTTTAT 60

TTCGTGGTCC TCAGCGTAAT GGCACTTCTG GTCTTCTGGT GGATGnGCCA TCACCTACTT 120

TGGCTTTGTG AGAATGTTCG CnCCTTCGCT TCAnCGTTAA CGATCACTTT GAAACGAATA 180

CTCCGTTCTT TCGTATCGGT GATACTTGGG TATTCCGCTG TGATTGTCCG TTGGCTTCTT 240

TCAAGCTGGT TGGTTTTGGT TCATCACCGA CACTTGTTAG GCTATCTTTG ACATACGTCA 300

GTCCTTTTGG TAGTTGGTCT GTGATAACTA CTTTGTTTAG GACACCATTT TCGATGGTGT 360

TGCGGAAGCT AATTGTGTAC TCGATAGTTT CGCCTAGTTT TGGTTTGGCG TTATTAACGG 420

TTTTTGTTGC GGCTAGTTTG CCTTCTTTCG CTTCTGGTAC CACCGGCACC TCTGGTTCTT 480

CGGGCGGATT AACTGTGTCA TCTACTTTGG CTTTGTTGAG AATTGTTTCG CCCGCTTTCG 540

CTTCTTCGTT AACGATCACT TTGAAACGAA TACTCCGTTC TTTCATATCG GTGATACTTG 600

GGTATTCTGC TGTGATTGTG CCATTGGCTT CTGTCAAGCT GGTTGGTTTT GGTTCATCAC 660

CGACACTTGT TAGGCTATCT TTGACATACG TCAGTCCTTT TGGTAGTTGG TCTGTGATAA 720

CTACTTTGTT TAGGACGCCG TTTTCGATGG TATTGCGGAA ACTGATTGTG TACTCGATCG 780

CTTCCCCTAG TTTTGGTTTA GCGTTATTAA CGGTTTTTGT TGCTGTTAAT TTGCCTTCTT 840

TCGCTTCTGG TACCACCGGC ACCTCTGGTT CTTCGGGCGG ATTAACTGTG TCATCTACTT 900

TGGCTTTGTT GAGAATTGTT TCGCCTGCCT TGGCTTCTTC GTTAACGATC ACTTTGAAAC 960

GAATACTCCG CTCTTTTGTA TCGGTGATAC TTGGGTATTC CGCTGTGATT GTGCCGTTGA 1020

TTTCTTTCAA GCTGGTTGGT TTTGGTTCAT CACCGACACT TGTTAGGCTA TCTTTGACAT 1080

ACGTCAGCCC TTTTGGTAGT TGGTCTGTGA TAACTACTTT GTTTAGGACG CCGTTTTCGA 1140

TGGTGTTGCG GAAcTGATTG TGTACTCGAT TTCTTCCCCT AATTTTGGTT TGGCGTTATT 1200

AACGGTTTTt GTTGCGGTTA ATTTGCCTAC GTTCGTTTCT GGTACCACAG GCACCTCTGG 1260

TTCTTCGGGC GGATTAACTG TGTCATCTAC TTTGGCTTTG TTGAGAATTG TTTCGCCTGC 1320

CTTAGCTTCT TCGTTAACGA TCACTTTGAA ACGAATACTC CGCTCTTTTG TATCAGTGAT 1380

ACTTGGGTAT TCCGCTGTGA TTGTGCCGTT GATTTCTTTC AAGCTGGTTG GTTTTGGTTC 1440

ATCACCGACA CTTGTTAGGC TATCTTTGAC ATACGTCAGy CCTTTTGGTA GTTGGTCTGT 1500

GATAACTACT TTGTTTAGGA CGCCGTTTTC GATGGTGTTG CGGAACTGGA TTGTGTACTC 1560

GATTTCTTCC CCTAATTTTG GTTTGGCGTT ATTAACGGTT TTTGTTGCGG TTAATTTGCC 1620

TACGTTCGTT TCTGGTACCA CAGGCAnCTC TGGGTTCCTC GGGCGGAATA ACTGGGGCCA 1680

CCAACTTTGG GCTTGGTTGA nAAATGGTTC GCCTGCCCTA nCCTCCTCCG GTAAAGATCA 1740

ACTTGnAACG AAA 1753

(2) INFORMATION FOR SEQ ID NO: 955:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 500 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 955:

GAGGATCCCC GTGACTATGC TACTGCAATT CTTGTTGTAC CAACCGGGAG CGTCAATAAT 60

TTTGTGTAAA TAACTCGTCC TTCTGCAAAA TAATGGGTTA CTCAGTAAAC ATTGAAGCTA 120

ATGTATCCGT CACTTGTTGG AAGCCTTTAT GGCTTCTATT TAAGAATTTT TGATTGTATG 180

TATCAAAGAT GGATACTAGG AAACGCTCTA GTGATTCTTC ATTTTGAAAC TGCTCTTTTC 240

TGCGGCTGTA TTTCTTGATT TGTTTATTGA AAGATTCGAT TAAATTAGTT GAGTAAATGC 300

TTCGGCGGAT ACCAGGCGGA AACTCATAAA AAGTCAATAA ATCTTGATTT TTTATCAGTG 360

ACTGCGTCAC TTTAGGATAA GTTTTCTGCC ACTTCTCAAT CATACTCCCT AAAAAGGTAT 420

TTGCCTCTTC CTTTGAGCTA GCTTGATAAA CAGCCTTAAA ATCATCACAG ATTTCTTTTC 480

GATCTTTGAC ACGTACTTTA 500

(2) INFORMATION FOR SEQ ID NO: 956:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 656 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 956:

ACGACACGTG TAGGCACGCT AGAATTAAAA GTACCCAGAA CACGTGATGG CCATTTTTCA 60

CCCACAGTGT TTGAACGTTA TCAACGAAAC GAAAAAGCCC TCATGGCTTC AATGTTGGAA 120

ATGTATGTAT CAGGCGTTTC AACTCGTAAA GTATCAAAAA TTGTGGAAGA ACTTTGTGGT 180

aAATCCGTCT CTAAGTCCTT CGTTTCTAGC TTAACAGAAC AGCTAGAACC TATGGTTAAC 240

GAGTGGCAGA ATCGTTTATT ATCAGAAAAA AATTATCCTT ACTTAATGAC CGATGTACTC 300

TATATAAAAG TACGAGAAGA AAATCGAGTA CTCtCAAAAA GCTGTCATAT AGCGATTGGA 360

ATAACCAAAG ATGGCGACCG TGAAATTATC GGCTTCATGA TTCAAAGTGG CGAAAGgAAG 420

AGACCTGGAC AACATtTTTT GnATACCTAA AAGAcGCGGT TTACAAGGTA CGGgAACTCG 480

TTATTyCTGA TGCGCACAAA GGGATTAGTC TCTGGCCATT AGGAAAATCC TTCACCACGT 540

AGGnGCCAAA AGATGCCAAG TTCACTTCCT AAGAnATATC TTTACCACCA TTCCTAAAAn 600

ATTCAAATCT TTCAGAGAGC GGTTAAGGAA TTTTAGTTCC AGTATTACTA GCGCGG 656

(2) INFORMATION FOR SEQ ID NO: 957:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 656 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 957:

CCCTTATTGG GGAAaGgCGG CAAtAAAAAA GTAATTGTAA AATACTGTTA GGGAATCATT 60

AGATAAAGTC CGTATAATTG TGTAAAAGTA AAAAGGCCAT ATAACAGTCC TTTTACGGTA 120

CAATGTTTTT AACGACAAAA ACATACCCAG GAGGACTTTT ACATGACCCA AGTACATTTT 180

ACACTGAAAA GCGAAGAGAT TCAAAGCATT ATTGAATATT CTGTAAAGGA TGACGTTTCT 240

AAAAATATTT TAACAACGGT ATTTAATCAA CTAATGGAAA ATCAACGAAC AGAATATATT 300

CAAGCAAAAG AATATGAACG AACAGAAAAC CGACAAAGTC AACGAAATGG CTATTATGAG 360

CGCAGyTTTA CGACACGTGT AGGCACGCTA GAATTAAAAG TACCCAGAAC ACGTGATGGC 420

CATTTTTCAC CCACAGTGTT tGAACGTTAT CAACGAAaCG AAAAaGCCCT CATGGCTTCA 480

ATGTTGGrAA TGTATGTATC AGGCGTTcAA CTCGTAAAGT ATCCAAAATT GTGGGAGACT 540

TGTGGTAATC CGTCTCTAGn CCTCGTTCTA GCTAACAGAC AGCTAGACTT GGGnACGGGG 600

CGATCGTTAT TCAGAAAAAT TCCTCTTATG nCGTGACCTT TAAGTCGGAG AAACGG 656

(2) INFORMATION FOR SEQ ID NO: 958:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 477 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 958:

CTGTCGTTTC TTTTTGGTAG TACTCACCTT CTTTATTTTT GTGTGTATTT TTAAGCTTAT 60

CGAACAACGC TTCATACTCG TTATTTATTT GCGCCATTTC TTTATCAGTA CCGCCCATGT 120

CTGGGTGGTA CTTTAAAGCT AGCTTTTTAT ATACGCGTTT CAATTCTTCT AATGTCGTTA 180

CATCTTTGAT AAATTTCATG CTAATCGCCC TTTCTGATTG CTGCTGAACA GCTAAGCTGC 240

CCAGCGTTCC AATTTTTCAA TTGATTTTTT ATCATTTTCA AGGTCTGCTT TTGTGTACCA 300

TGCTGGTATC TCTTTATCAG AAAATAGACT GTAGATAAAT TCTAAGATTT CCATGCkTtC 360

TCTAGCTTTA ACAGCTGATT CACATTGATA ACGTAATATC tTACAyCTgg CTAAGTCTAT 420

CGACATAgAA TAATGAAACG TTgGtTCCTc yACCATTtGt TAACcGtTTT TCCnTnG 477

(2) INFORMATION FOR SEQ ID NO: 959:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 628 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 959:

GTAAAGCCCC nGCCCACCTT AGAACCACTT AGTGGTTTAA TTCCATAAAC TATTTTTAGG 60

AGGCTTACAG AAATGAAAAA AGCAAAATTA TTCGGTTTTA GTTTGATTGC ATTAGGTTTA 120

TCCAnTTCAC tTGCAGCATG TGGTGGTGGC AAAGGCAAAA CCcTGAAgCG GCGGTgGcAA 180

AGGGGATGCA GCGCATAGTG CTGTAATCAT TACAGATACA GGCGGCGTGG ATGACAAGTC 240

GTTCAACCAA TCTTCTTGGG AAGGATTGCA AGCTTGGGGT AAAGAACATG ATTTACCAGA 300

AGGTTCAAAA GGGTATGCAT ATATTCAATC GAATGATGCA GCTGACTATA CAACCAATAT 360

TGACCAAGCG GTATCAAGTA AATTCAACAC AATCTTTGGT ATTGGcTACT TGCTAAAAGA 420

TGCAATTTCT TCTGCAGCAG ATGCCAACCC TGATACAAAC TTTGTTTTAA TCGATGATCA 480

AATCGATGGG CAAAAAGAAT GTCGTTTCTG CAACCATTTA GGAGATAATG AAGCCAGCTT 540

ACTTAGCCGG TGTTGGCTGC TGCAAATGGA AACCAAAACG AACCAAAGTC CGGTTTTGGT 600

TnGGTGGTTG AAGAAGGGGT CCGTAATT 628

(2) INFORMATION FOR SEQ ID NO: 960:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 566 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 960:

CTTTACCTCn TGnGAAATAG TACCTGCACT GCCACAATnC CAGGCGTTTC CATTCGTAAT 60

CGATGATGGA GCCAGGAATT GCTATCCCAT AGCCATTGGC GCCCTTGGTC ACATCGTTAT 120

CGCCAATAAA TGTTTTTCCA GTCGCTTTTT CAATCCCATG AAACGTTGTT GGCGTAAGTT 180

TTGTAGCATT ATTGACCGTT GGATTAATCG TCAAGACCTG ATAGGCTTTG ACATTTCCAA 240

ACCAAGTCCC TGTGCTATCC ACCAAAGCAG TTGTATCTGT CGTTGGTTTA ATCCATAAAA 300

GTTCATCTTT TGTTGGATCG ACAGCTTCAA AATCCGCAAT GTTTACTTTA CtGTACCAGG 360

AATTGTTTCT AATGTTGCAT GATCCTGaAC CGTTAGTTTG CTAGCTGAGA TTAAATTGGc 420

TGTGTTGGTG aTCCCAATTC CAGCGTTTTT AGATAGATGA ATkGATCCGA ATTCATGGnA 480

AGTCGTACnG GGGTCGCGGC AGTAGCAGAA CGCCATTTTG GATGTnCCAC CATAGCCATA 540

AATTAAnGCC ATGGGGGCAT TAATCC 566

(2) INFORMATION FOR SEQ ID NO: 961:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 534 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 961:

GTGTTCGATT TTTATGATGA TACCAATGTG AAACTAGTCC GGCCCTTTGG CGATACGATG 60

GGGCTACATG AAGAATTCCT ACCAGCTAGT AAACGGTATA TGAATGACTA TATCCAGTAT 120

GTAACATCTG ACTTTTTAGC AGGACTTGGC TTTGGCGCCA CTCAAATGTT AGGAGAGTTG 180

GAAgGGATTT ACTTTGGCTA CAATGCGGAT ACGGGGCGTA ATGTTTACCT GAAACCCGCA 240

TTAGCCTCAC AAGGAGTAAA GGGCTCAGTA ACCAATGCAT TAGCTGCGGC TTTTCTTGGT 300

TCTCTTGGTG GCGGAAAATC CTTTAGTAAC AATCTCTTGG TCTACTATGC TGTTCTTTTT 360

GGTGGCCAAG CTGTGATTGT CGATCCAAAA GGAGAACGAg TGGTTGGAAA gAAAACTTGC 420

CAGAAATTGC nGATGAAATC AATAtCCyCA cTTGACyAGt AAaCCrGAAA TCAAGGCnTA 480

CyTGATCCCT AATGGGATTA TGAAGAAAAG AAGATCAGAA GTTAGCATTA TTCC 534

(2) INFORMATION FOR SEQ ID NO: 962:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 462 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 962:

TAAAAATTCT AATTTGATTT TATCCAGCCC AATTTCTTTT TTCCCTTTTT CCCAGAATTC 60

TTTTGCTTTC GCTGTATCGT AGGTTAAGTA GTTATTTTTC TTAGCAGCTA GTTTTGTGAA 120

ATCTGTTCCA TCTGGCGCAA TCGTTTGTCC CGGCGGAATT ACTTGATCTG TAGCCGTTGA 180

TTGATCTTGC AAGACTTCkT TTACAAAACT TTCTTTATCA AAAGCTTGTG CCAACGCTTT 240

ACGAATATTG TTGTTCGCTA AAGCCGGATT TTCTTTTCCA TCACGAACGC TGTTCATTTT 300

TAAAAAgTAT GTGACGAATT GAGGAATTGA AAGAAAGGCT GGATTAwCmA CaTAACCAGG 360

AATAAATTCA CCAcTaGCTT AwTCACTAcA TCCAtTCATc GAGyCAAmCA AtTCACCGcT 420

GTCGTGGGGC CCTAATACTG AACTTTCGCT CnGTAATTAC TT 462

(2) INFORMATION FOR SEQ ID NO: 963:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 661 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 963:

GCATTATTAG CAGCGTTAAA TGAAGGAAGC AATAAAGTAG GCAAACAAGC TTGGTCTGTA 60

TATCTGGAAG TCGAACGGAT CAAAACAGGT GACGTAGAAA ATACGCAAAC AGAAAACTAC 120

AACAAAGAAC TTGTTCGTTC TAATACGGTG GTGACGCATA CACCTGATGA TCCAAAACCA 180

ACCAAAGCCG TTCACAACAA GAAAGGGGAA GATATTAACC ATGGAAAAGT TGCTCGTGGT 240

GATGTTCTTT CTTATGAAAT GACGTGGGAs TTAAAAGGGT ACGATAAAGA CTTTGCTTTT 300

GACACAGTCG ATCTTGCGAC AGGAGTTTCT TTCTTCGATG ATTACGACGA AACGAAAGTG 360

ACcCAATCAA AGACTTACTT CGTGTCAAAG ATTCTAAAGG GGCAGACATT ACGAACCAGT 420

TCACGATCTC TTGGGACGAT GCCmAAGGCA CGGTGACmAT CTCTGCCaAA GACCCmCAAg 480

CCTTTATTCT AGCGTwTGGT GGGCAAGAAT TGCGTGTAAC GTCCCTACAA AAGTCAAAGC 540

CAATGTTTCT GGTGATGTTT ATAATTCAGC GGAACCAAAT ACATTTGGTC AACGAATTAA 600

AACAAATACC GTTGTCAACC ATATTCCAAA GTGGACCCTA AAAAAGACGT GGTTnTAAGT 660

G 661

(2) INFORMATION FOR SEQ ID NO: 964:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 585 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 964:

AAGAACCGGA TATTTTTCTT TGATTTTATT CCAAAAGACA TTGTACATTT TCACATTTTG 60

GACATAGTAA AGTTCGGCAA ATAAATGTGT AGCAAACAAA TAATAAAAAA TTTGCGGACG 120

AATTTTCTCC AATTGCGGCG TCATGTGTAT TTCTTCTAAT AAGAAATGTT GGACTAAATG 180

AGTCAACATG TAAACAGGTG CTTCAGTCTC ATCAAAATGC GCTAAAATTG CTTGTGCAGT 240

GGCTTCATCT TGATAAAATT GGTCATTGGC TAACAAGAAA AAGTACACAA AAGTTAAATC 300

GTTTTTGTTA AGGAAAAGGG TTGTAGCTGT CATCTCkkCT ATAAAGGCTg GTAATGTTTA 360

TTACCTTCTA AATAAGTtTC TTGCTCGGCT GAAGGTCGAT ATTCGGCGCC TTGCGCACTA 420

CGGACTTTTG AGATAGCGAC CATATAwTTA ACTCGGCGTG kCACCGTTTC CGCAACTTCT 480

ACGTGAAAAA AGCCATCATT TTGGCGGTTG TGCGACTTCC GTTCTGTAAT AGnCAAAAGC 540

CAGTACACGG ATTAnTCATA AAATAGTGTA CCAAAGGGCA ATCGC 585

(2) INFORMATION FOR SEQ ID NO: 965:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 721 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 965:

nCGGGAAGGn AACTGTAAAG CCCCTTCCCA TCAGCAGTAC CTGACCTATC ACGGCACCTC 60

CGCATTCGnA TCTCAACCGT GGCATGGCAT CGGCAGTTCC TTAAACTCGT AACCTGCTGT 120

GGAATTACCT GTCACTCCTG TGATTGTTGG TGCGGTTACG GTTGCTTCAT CCGCTGGTGT 180

nTGGAACGTT GTTGGCGTAC TTTCTGTGCC GCTGGCGTTT TTCGCTACGG CGGTTAACGT 240

TTCATTGGCA CCTGCTTCAC tGCGGGAACG GTAACTGTAA ACGCTCCTGT CCCATCAGCG 300

GTACCTGTGC CTATTACGGT GCCTCCTGCA TTTCGGATTT CAACCGTGGC ATTGGCATCG 360

GCAGTTCCTT TAAtTCGTAA CCTGCCGTTG AATTACCTGT CACTCCTGTG ATTGTTGGTG 420

CGGTTACGGT TGCTTCATCC GCTGGTGTTT GGAACGTTGT TGGCGTACTT TCTGTACCGC 480

TGGCGTTTTT CGCTACGGCG GTTAACGTTT CATTGGCGCC TGCTTCACCT GCGGGAACGG 540

TAACTGTAAA CGCTCCTGTC CCATCAGCGG TACCTGTGCC TATTACGGTG CCTCCTGCAT 600

TTCGGATTTC AACTGTGGGC ATTGGCATCA GCAGTTCCTT TAACCTCGTA ACCTGCCGTT 660

GAATTACCTG TCACTCCTGT GATTGTTGGT GCGGTTACGG TTGCTTCATC CGCGGGCGTT 720

T 721

(2) INFORMATION FOR SEQ ID NO: 966:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 513 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 966:

AGAAAAAGTG TGAGCCTTGT GCTCACGCTT TTTTGTGCAA TTGAAGTTGG TGAATCTCAG 60

CGAAAAGAGA AAGGTACCTT AAATAGTGTG TTTCTGGTCA ATAGCACAAT GAAATATCTA 120

AATACATTAT GAACTGCTTT TCGTTTTATG AAGAGCAGTT TGTGCTATTT GAGAAATTGC 180

TCAATTTAAT CATTAAGTAG TTTCATCAAT GGTGACTATG CTACTGCAAT TCTTGTTGTA 240

CCAACCGGGA GCGTCAATAA TTTTGTGTAA ATAACTCGTC CTTCTGCAAA ATAATGGGTT 300

ACTCAGTArA CATTGAAGCT AATGTATCCG TCACTTGTTG GAAGCCTTTA TGGCTTCTAT 360

TTAAGAATTT TgATgTATGT ATCAAAGAtG GATACTAGGA AACGCTCTAg TGATtCTTCA 420

TTTGAAACTG CTCTTTcTGC GGCgTATTct GATTGTTaTg AAgATCGATA AATAGTgAGT 480

AAAgCTCsGC GaTACCAGcs GAAACTCATA Ann 513

(2) INFORMATION FOR SEQ ID NO: 967:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 534 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 967:

CAGTTATCCC CCnGAGACCC AGAGGTGCnT GTGGTnCCAG AACGAACGCA GGCAAACTAG 60

CCGCAACAAA AACCGTTAAT AACGCTAAAC CAAAACTAGG CGAAACTATC GAGTACACAA 120

TTAGCTTCCG CAATACCATC GAAAACGGCG TCCTAAACAA AGTAGTTATC ACAGACCAAC 180

TACCAAAAGG ACTGACGTAT GTCAAAGATA GCCTAACAAG TGTCGGTGAT GAACCAAAAC 240

CAACCAGCTT GAAAGAAATC AACGGCACAA TCACAGCGGA ATACCCAAGT ATCACTGATA 300

CAAAAGAGCG GAGTATTCGT TTCAAAGTGA TCGTTAACGA AGAAGCTAAG GCAGGCGAAA 360

CAATTCTCAA CAAAGCCAAA GTAGATGACA CAGTTAATCC GCCCGAAGAA CCAGAGGTGC 420

CTGTGGTACC AGAAACGAAC GTGGcAAATT wACCGCAACA AAAcCGTTAA TAACGCCAAc 480

CAAATTgGGG AgAATCGgTA CACmATCAGt TCGCACACCA TCGAAACGGG TCTT 534

(2) INFORMATION FOR SEQ ID NO: 968:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 791 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 968:

TAATAAAATC CGTAAGAAAA CAGTAAAGTA TAAAGATAAA GAATATGATC GAATCAACTT 60

ATTGTTAGAC AACGATATGA ATAGCGAAGA AAATAATAAG TTAATGGAAG AAGTTGTAAA 120

TTACGAAAGT AAAGACTTAA AAAACCGTTC AATTATGTAT TATAACAACT GGACCAATAC 180

TTTTTATGGC TTGTTTTATC TGGTAGTTAA CTTCATTCAA ACAGTGGTGT ATTTCCTGTT 240

AAGTTTCTnA CGATTAATTA TTGCGGTCAT TCAATTGTTC CTATTGCCGT TGTTACCGTT 300

ATTGTTGTTC GCAGGACTCT TTTTAACTGA AACGAACGTC TTTGCGAATT ACTTTAAAAC 360

CTTTGGCATG ACTATTTTTA TGAAGGGAAT GGTTGGTTTT GCGACTATTT TCTTTGCAAG 420

TTTCTTATCA CTAGGGTTTC AATTAAGCAA TCAGACAGAA AACGTGTGGC AGAAGATATT 480

AACGATACTT ATTTATCTTC TTACACCGCT TGGCTTGTAT GTCTTCCGGA AATTCTTTGC 540

TAATCTTGTC ACGGGACGTG TTTCACTTTC TGATGGTGTC GGGTTTATCG GAAaTCCTTT 600

TGGCACAGAA GCTAATATGC GTAGAGCGGC AAAAGAACAA AAGCaAGAGa ATAAGGAGCG 660

CAnGnAACAG GCACAAGAAG AACGTAAAAG CAGCTATTAA AAAACGTCCA GAAGnAGCCA 720

AGGGAACCAT GGTAAAGCAG GAAACTGGCn TAAAACCAAn GGCCATTAAA TAAAGGACCA 780

AGGAAAACGT G 791

(2) INFORMATION FOR SEQ ID NO: 969:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 511 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 969:

TGCTCGCGAA CAAAGAGCTG CCACACAGAA GGAGTTGTTA GAAACAGCAC CAGAAGCTAG 60

TCTCTTATCT GCCACAATGA ATATTCCTGG AGAAGTAAAA AATTCGCCCA CCTTAACGGC 120

TGTTTTTTTA GAAGTAATTG ACGAAtCGAG CAACAATTGC TGGATCAAGT ACCGATTGTT 180

AATTTTTACC GGAATGAAAA AACAGGTCCT GAATATTATT TGGCGGTTTC GTTAGCCCCA 240

CAAGAATTAA AACAAAGAAT GGTCAAAATT GAAGAAACAC ATCCTTATGG TCGTTTAGTG 300

GACCTAGATG TTTTATGGGG AAtGAGGAGC TAAAAAGCCT TCATCGAGGA GATTTAGGCT 360

kGCCACCACG GCGCTGTTTT AtTkGCCAAG aAGTGGCcGA AAaGTCTGTn GaCGAAATCG 420

asGTCACAGT CTtGaAGCAt GCCAgaAgAA ATaCCggAAT AATTTgaCcC GaAAGGGcac 480

CAAGTGGGTA AAAAAATCCG TTTAACGGAA A 511

(2) INFORMATION FOR SEQ ID NO: 970:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 522 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 970:

CGTTTGATGT CTTATCCGAT GAGGTCGTGA TGCAGTTGCC TTCTGCACCT AATTATCCTT 60

CAATTTATAG TTCTCTAAAC AAATTATTAA CCTTTGAGTC CTACAAATTA GAAGAGTTAA 120

CCCCGAATCG AATTTATAAT AAAGTTGTTG TTGCGATTGA TGAAGCGTAT TTAAATGAAC 180

GCATCAAGGA AATTCCAGCC TCTTTTTATG AGCGCTTTGA AATTATTAAA ACGCGAAATA 240

ACTTATTAGA ATTTATGCCA AAAGGGATTA CGAAAGCCTA TGGTATTTCC TTGTTAGCCA 300

AAGATTTAGG AATCCGTGCA GAAGAAATCA TGACCCTAGG GGACGArGAA AACgATTTaC 360

CGATGATCGA ATACGCTGGG CTTAGGGTGt CGCCATGGGC AAATGCGatT CCaTTAGGtT 420

AAAgAAGCCG CaGATGTCgT CACaGACmCg GATGrCCCAg AttGGCGkTG cCAAAGCGGG 480

TnGAAAATnT AnTTTAACCC CGTTGGGAAG GGGGCCATAA GA 522

(2) INFORMATION FOR SEQ ID NO: 971:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 470 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 971:

GGCGGACGAG CCTTCCCCAG GAAACCTTAG TCATACGGTG GACAGGATTC TCACCTGTCT 60

TTCGCTACTC ATACCGGCAT TCTCACTTCT AAGCGCTCCA GCCGTCCTCA CGATCGACCT 120

TCAACGCCCT TAGAACGCTC TCCTACCACT ACACCTAATG GTGTAGTTCC ACAGCTTCGG 180

TAATATGTTT AGCCCCGGTA CATTTTCGGC GCAGGtCACT CGACTAGTGA GCTATTACGC 240

ACTCTTTAAA TGGTGGCTGC TTCTAAGCCA ACATCCTAGT TGTCTGTGcA ACCCACATCC 300

TTTTCCACTT AACATATATT TTGGGGACCT TAGCTGGTGG GTCTGGGGCT GGTTTTCCCT 360

TTCGGACTAC GGGATCTTAA TCACTCGGCn GTCTGACTCC GGGATATAAA TGAATGGGCA 420

TCGGAGTTTA nCTGAATCGG TnACCGAGAT GGGGCCCAAG GTCAAAAGnG 470

(2) INFORMATION FOR SEQ ID NO: 972:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 943 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 972:

CGCGTAGAGG ATCCCGCCTG ATGATCCAAA ACCAACCAAA GCCGTTCATA ACAAAAAAGG 60

GGAAGACATT AATCATGGAA AAGTGGCTCG TGGTGATGTT CTTTCTTATG AAATGACTTG 120

GGACTTAAAA GGGTACGATA AGGACTTTGC CTTTGATACA GTCGATCTTG CGACAGGCGT 180

TTCTTTCTTC GATGATTACG ATGAAACGAA GGTGACACCA ATCAAAGACT TACTTCGTGT 240

CAAAGATTCT AAAGGGGAAG ACATTACGAA CCAGTTCACG ATCTCTTGGG ATGATGCCAA 300

AGGcACGGtG ACGATTTCTG CCAAAGAcCC ACAAGCCTTT ATTTTGGCGC ATGGTGGGCA 360

AGAATTACGT GTAACTTTAC CAACAAAAGT TAAAGCCAAT GTTTCTGGTG ATGTGTATAA 420

TTTAGCGGAA CAAAATACAT TTGGTCmACG AATTAAAACC AwTACCGTTG TCAACCATAT 480

TCCAAAAGTG rACCCTAAAA AAGACGTGGT TATTAAAGTC GGTGATAAAC AAAGTCAAAA 540

TGGTGCCACA ATCAAATTAG GGGAGAAATT CTTCTATGAA TTTACAAGTA GTGACATTCC 600

TGCAGAATAC GCTGGTATTG TGGAAGAATG GTCGATTAGC GATAAACTAG ACGTCAAACA 660

TGACaAATTT AGTGGCCAAT GGTCTGTGTT TGCCAATTCT ACGTTTGTCT TaGCAGACGG 720

AACCAAAGTG AATAAAGGGG ACGACATTTC GAAACTATTC ACGATGACCT TTGAACaAGG 780

GGTAGTGAAA ATCACAGCCA GTCAAGCCTT TTTaGATGCG ATGaATCTAA AaGAAAaCCA 840

AAACGTTGcA CACTCATGGa AGCGTcATTG GTGTAGACGA tTGCGGcAGa GcGTtcACaC 900

aTCGAGAtCT TTCaCaTGGG aGTTaACyaT ccGGTTTGnC nCT 943

(2) INFORMATION FOR SEQ ID NO: 973:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 475 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 973:

nCCAGGTACT AAGAkGTTTC AGTyCTCkGC GTCTACCTCT AATCAGCTAT GTATTCACTG 60

AAAAGTAATA TCCTATAAAA GATATTGGGT TCCCCCATTC GGAAATCTCT GGATCATAGC 120

TTACTTACAG CTCCCCAAAG CATATCGGTG TTAGTCCCGT CCTTCATCGG CTCCTAGTGC 180

CAAGGCATCC ACCGTGCGCC CTTATTCACT TAACCTTATC AACcTTACGG TTGGGTCTTG 240

ATCACTTCTT CTAGCGATAG AAGGTTAATC AATAAATAAG CAATTGAACT TATTAAAAAA 300

CTCATTCAAC GCGGTGTTCT CGGTTTGTTT TGATTCTTTT TACTTCAATA TCCAGTTTTC 360

AATGAACGAA TGTTTTGAGA GTAGACCTCT CAAAACTGAA CAAAGTAAAG ACGAAATGTG 420

TATTTCCGTA ATATCCCTTA GAAAGGAGGT GATCCAGCCG CACCTTCCGA TACGG 475

(2) INFORMATION FOR SEQ ID NO: 974:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 382 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 974:

TTGAAGACAG CCGATTCAGG TTACTTwACA CGTCGTTTAG TTGACGTTGC CCAAGACGTA 60

ATCATtCGTG AAGACGACTG TGGCACTGAC CGTGGTCTTG AAATCGAAGC AATTCGTGAA 120

GGTAACGAAA TTATCGAACC ATTAGACGAA CGTCTATTAG GTCGTTATAC ACGTAAATCA 180

GTTGTTCATC CAGAAACTGG TGCAATCATC ATTGGTGCGG ATCAATTAAT CACTGAAGAT 240

CTTGCCAGAG AAATCGTGGA CGCTGGTATT GAAAAAGTAA CCATCCGTTC TGTCTTTACA 300

TGTAACACAA AACATGGTGT ATGTAAGCAC TGTTACGGAC GTAACCTTGC AACTGGTTCT 360

GACGTTGAAG TCGGAGAAGC AG 382

(2) INFORMATION FOR SEQ ID NO: 975:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 496 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 975:

AACATAATAA TCTTTCGGAT ACAAATAATA AGAAATGTTA TTTTTAAAAA TAAAACAATA 60

TTTTTTAAGT AATATTAAGA ATGTGAGAGT GTAAAATATT TTGTGTAAAT GAAAAAATCC 120

ATACAAAAAA GGAAGTCGCT TCTGTAGAAT AAAGTTAACG ACAACCAATT CACAGArAAG 180

AGGACTTCCC TATGAATGAT TTTACTACAG AAATTGTGCA AACTCTAGTC ACTAAAGGCG 240

ATTTAAATGA ATTATTCCGT TCGCACTTAG AAAAAGCGAT AAACACACTC CTACGGACTG 300

AATTAACGGC TTTTTTAGAT TACGAAAAAT ATGATCGCAC TGGTTTTAAT TCAGGTAATT 360

CGAGAAACGG TTCTTACTTT CGATCAATCA AAACCGAATA TGGTGAATTA ACATTGGAAA 420

TACCTAGAGA TCGTAATGGT GAGTTTAAAC ACCAAACTTT ACCAGCCTCC AAAGACAACG 480

AnCATTGGAA CCCTAT 496

(2) INFORMATION FOR SEQ ID NO: 976:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 455 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 976:

CTTGCATAGT AAACTTCATA ATATCCCCAG TAATCGAAAC TTTCGCTGGA TCAATAAAAC 60

GACCACTTTC ATCTAAAAAC GACATTTTGT ACGTCTTCTC TGGAATATCG AATGTCTGCA 120

CTGCTCGCTG CGCACGCAtG ATTTCTGGTG TATTAACTAT TATTGGACTC GTCACTGACG 180

CTTCAACTTG CCCGACTACT TGTTCAGTCT GTTCTGGTAG tCTTCATCAG CTGGCGGTAC 240

cTtACTTCCG TATCTGCCTC TGGAAAAGAA ACGGGCGCCC GTCCTTGgAA CCCCCTCTGg 300

AcGGGgACTA CTGtTTGGGC GGGTCCTCCC CTTGGAACAT CTGGGGGnCG CCATGGGTTT 360

TCAGCAATTG GCTGGGGGAT GGCCATTTTA ATGGAACTGG ATTAGGTnAA GGAACCACAC 420

AAAGCGGATT GGCnAATAAA CGGGTTTCCT TTTTT 455

(2) INFORMATION FOR SEQ ID NO: 977:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 489 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 977:

CTTGTTGTGC CTCTAAGTCT CTTAATACAC CTTGGTACTG CTGTAGCTCT AACTTCGCAT 60

TAGAAAGGTC TTCTTGTGCT TGTGCATAAG CAAGTGTAGC AACGGCTTGT TGTTCTTTAA 120

GTTCCTCCAA TGATTTCATA GAATTTTCAT AGGCTTCTTG TGCTTGAACC ACTGCCGCTT 180

CTGCTTTGTT TAACTCGTTT AATGCTTTTT CGTATGTTGG TTGAGCATTT TCTAAACGAG 240

ACAATGAACG TGTACTTGTT TTAAGCGCTA CTTGTTTTTG ATTTAAGTTT TCTTTTGCTA 300

AATCCAATAC TTCCTTCaAG CTATTCaATG TGtTTTCyTC TGCTTTTTGT GCnTCyTGtT 360

TTCTTTCAGC ACTTtAGCAG ACGtTGkTCm ATCGCTGgtG CnTCCGmACG tTGgTTTcGg 420

CTTTyCaAAA CAGCAGCCgA TCcGaATAAC cCATGATGnG CATGACCTTG TTGTGAAGCT 480

TTCCGCGTG 489

(2) INFORMATION FOR SEQ ID NO: 978:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 473 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 978:

AATTAATTAC ACAGTTATTT AGTCAATTTT GTTTAGGAAA ATAAGACAAG AGAGGAATTT 60

TTTATGAATC mATATCAAGC AGAATCATAT GATGTTATTG TCGTCGGTGC TGGACATGCT 120

GGCTCTGAAG CAGCCTTAGC AGCCGCACGA ATGGGCGTTA AAACGTTACT TTTAACGATT 180

AATTTAGaTA TGGTTGCTTT aTGCCATGTA ACCCTTCTGT TGGTGGACCT GCCAAAGGAG 240

TCGTTGTCCG AGAAATCGtG CATTAGGCGG CGAAATGGGT AAAAACATgA TAAAACCTAC 300

ATCCAAATGC GGATGCTAAA TACTGGTAAA GGTCCAGCAG gCGTGCGTgC GCGCGCAAcT 360

GgACAACATG CCAAGCGCCA nAATGAACAT ACCATGAAAA AGAGAATTTA CCCTCGCAAG 420

GATGTnGAGA TTAACGGGAA ATGCGTTTGC GGGGTGACTC ACAGGCCCAA CGn 473

(2) INFORMATION FOR SEQ ID NO: 979:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 483 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 979:

ACAACGTTCC AAACACCAGC GGATGAAGCA ACCGTAACCG CACCAACAAT CACAGGAGTG 60

ACAGGTAATT CAACAGCAGG TTACGAATTA AAGGAACTGC CGATGCCAAT GCCACGGTTG 120

AGATCCGAAA TGCAGGAGGT GCCGTGATAG GTACAGGTAC TGCTGATGGG ACAGGGGCAT 180

TTACAGTTAC CATTCCCGCA GTGAAGCAGG TGCGAATGAA ACGTTAACCG CCGTAGCGAA 240

AAACGCCAGC GGTACAGAAA GTACGCCAAC AACGTTCCaA ACGCCArCGG rTCCTaATTA 300

CGCCCGTGGG GACGCCAATT TGTTTGAGAC TTGTTAACCG GkTAGTTACC AcCAAAGGsT 360

TATgAGGkTC CAAGGGGACT TGCTTGAAGT TTTgGGCACC aCCATTTnGA GGGTTTCCGC 420

GATTTCAnCT TTGGCACGGT nCCTTTGGTT ACTTTGCAAC AACTTTGGAA CTTTACGGGA 480

AnT 483

(2) INFORMATION FOR SEQ ID NO: 980:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 537 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 980:

CCTnTGGGCC GCCCACGGAA TTGCCCCGGC GGAAATTGCC GCTATTAATT CACGTCCATT 60

CCTAAGGATT TGAACCCCCn GnCTTGACAA GCCGTTGGTG GAGCGAAAAA GGnTCCAAAC 120

TTTACCGTCG TGGAAAACCC AACTGGAGGA CTTAGGAATG TTTCTTCTAA AGATTTAGCT 180

GCAAArGAAA AAGAAGTAGA CCAACTACAA AAAGAACAAG CCCAAAAGAT TGCCCAACAA 240

GCAGCTGAAT TAAAAGCCAA AAATGAAAAA ATTGCCAAAG AAAATGCAGA AATTGCGGCA 300

AAAAATAAAG CGGAAAAAGA GCGCTACGAA AAAGAAGTGG CGGAATACAA CAAACACAAA 360

AATGACAAAG GCTATGTGAA TGAAGCAATC AGCAAAGACT TAGTTTTTGA TTCAAGCATT 420

GTGACTAAGG ATaCTAAAAT AGACaAGATT TACAGGTGGC CAGTTTATTA AAGCTTCCGA 480

TTTTAACAAG TAAACCCAAG GGACAATCCA nAGATATTTT TTACAAAATT TAAGTAA 537

(2) INFORMATION FOR SEQ ID NO: 981:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 325 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 981:

TGTGGGTAGC GGCGAAATTC CAAACGAACT TGGAGATAGC TGGTTCTCTC CGAAATAGCT 60

TTAGGGnTAG CCTCGGAATT GAGaATGATG GAGGTAGAGC ACTGTTTGGA CTAGGGGCCC 120

ATCTCGGGTT ACCGAATTCA GATAAACTCC GrATGCCATT CATTTATaTC CGGGAGTCAG 180

ACTGCGAGTG ATAAGATCCG TAGTCGAAAG GGAAACAGCC CAGACCACCA GCTAAGGTCC 240

CnAAATATAT GTTAAGTGGA AAAGGATGTG GGGTTGCACA GACAACTAGG ATGTTGGCTT 300

AGAAGCAGCC ACCATTTAAA GAGTG 325

(2) INFORMATION FOR SEQ ID NO: 982:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 327 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 982:

GTAATATCCT ATAAAAGATA TTGGGTTCCC CCATTCGGAA ATCTCTGGAT CATAGCTTAC 60

TTACAGsTCC CCAAAGCATA TCGGTGTTAG TCCCGTCCTT CATCGGCTCC TAgTGCCAAG 120

GCATCCACCG TGCGcCCTTA TTCACTTAAc CTTATCAACC TTAcGGTTGG GTCTTGATCA 180

CTTCTTCTAG CGATAGAAGG TTAATCAATA AATAAGCAAT TGaACTaTTA AAAAACTCAT 240

TCAACGCGGk GTTCCgGTTt GGTTTgAAyC TtTtACTCmA AAATCCAGTT TTCAATGAAC 300

GAATGTTTgA GAGTAGACCC CCCCAAA 327

# INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

**A.** The indications made below relate to the microorganism referred to in the description
on page 8 , line 27

**B. IDENTIFICATION OF DEPOSIT** Further deposits are identified on an additional sheet ☐

Name of depositary institution American Type Culture Collection

Address of depositary institution (*including postal code and country*)

10801 University Boulevard
Manasas, Virginia 20110-2209
United States of America

Date of depositMay 2, 1997

Accession Number 55969

**C. ADDITIONAL INDICATIONS** *(leave blank if not applicable)* This information is continued on an additional sheet ☐

**D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE** *(if the indications are not for all designated States)*

**E. SEPARATE FURNISHING OF INDICATIONS** *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

For receiving Office use only

☑ This sheet was received with the international application

Authorized officer

For International Bureau use only

☐ This sheet was received by the International Bureau on:

Authorized officer

**What Is Claimed Is:**

1. Computer readable medium having recorded thereon the nucleotide sequence depicted in SEQ ID NOS:1-982, a representative fragment thereof or a nucleotide sequence at least 95% identical to a nucleotide sequence depicted in SEQ ID NOS:1-982.

2. The computer readable medium of claim 1 having recorded thereon any one of the fragments of SEQ ID NOS:1-982 depicted in Tables 2 and 3 or a degenerate variant thereof.

3. The computer readable medium of claim 1, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.

4. The computer readable medium of claim 3, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.

5. A computer-based system for identifying fragments of the *Enterococcus faecalis* genome of commercial importance comprising the following elements:

a) a data storage means comprising the nucleotide sequence of SEQ ID NOS:1-982, a representative fragment thereof, or a nucleotide sequence at least 95% identical to a nucleotide sequence of SEQ ID NOS:1-982;

b) search means for comparing a target sequence to the nucleotide sequence of the data storage means of step (a) to identify homologous sequence(s), and

c) retrieval means for obtaining said homologous sequence(s) of step (b).

6. A method for identifying commercially important nucleic acid fragments of the *Enterococcus faecalis* genome comprising the step of comparing a database comprising the nucleotide sequences depicted in SEQ ID NOS:1-982, a representative fragment thereof, or a nucleotide sequence at least 95% identical to a nucleotide sequence of SEQ ID NOS:1-982 with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence is not randomly selected.

7. A method for identifying an expression modulating fragment of *Enterococcus faecalis* genome comprising the step of comparing a database comprising the nucleotide sequences depicted in SEQ ID NOS:1-982, a representative fragment thereof, or a nucleotide sequence at least 95% identical to the nucleotide sequence of SEQ ID NOS:1-

982 with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence comprises sequences known to regulate gene expression.

8. An isolated protein-encoding nucleic acid fragment of the *Enterococcus faecalis* genome, wherein said fragment consists of the nucleotide sequence of any one of the fragments of SEQ ID NOS:1-982 depicted in Tables 2 and 3, or a degenerate variant thereof.

9. A vector comprising any one of the fragments of the *Enterococcus faecalis* genome of claim 8.

10. An isolated fragment of the *Enterococcus faecalis* genome, wherein said fragment modulates the expression of an operably linked open reading frame, wherein said fragment consists of the nucleotide sequence from about 10 to 200 bases in length which is 5' to any one of the open reading of claim 8.

11. A vector comprising any one of the fragments of the *Enterococcus faecalis* genome of claim 8.

12. An organism which has been altered to contain any one of the fragments of the *Enterococcus faecalis* genome of claim 8.

13. An organism which has been altered to contain any one of the fragments of the *Enterococcus faecalis* genome of claim 10.

14. A method for regulating the expression of a nucleic acid molecule comprising the step of covalently attaching to said nucleic acid molecule to a a nucleic acid molecule of claim 10.

15. An isolated polypeptide encoded by any of the fragments of the *Enterococcus faecalis* genome of claim 8.

16. An isolated polynucleotide molecule encoding any one of the polypeptides of claim 15.

17. An antibody which selectively binds to any one of the polypeptides of claim 15.

18. A method for producing a polypeptide in a host cell comprising the steps of:

a) incubating a host containing a heterologous nucleic acid molecule whose nucleotide sequence consists of any one of the fragments of the *Enterococcus faecalis* genome of claim 8, under conditions where said heterologous nucleic acid molecule is expressed to produce said protein, and

b) isolating said protein.

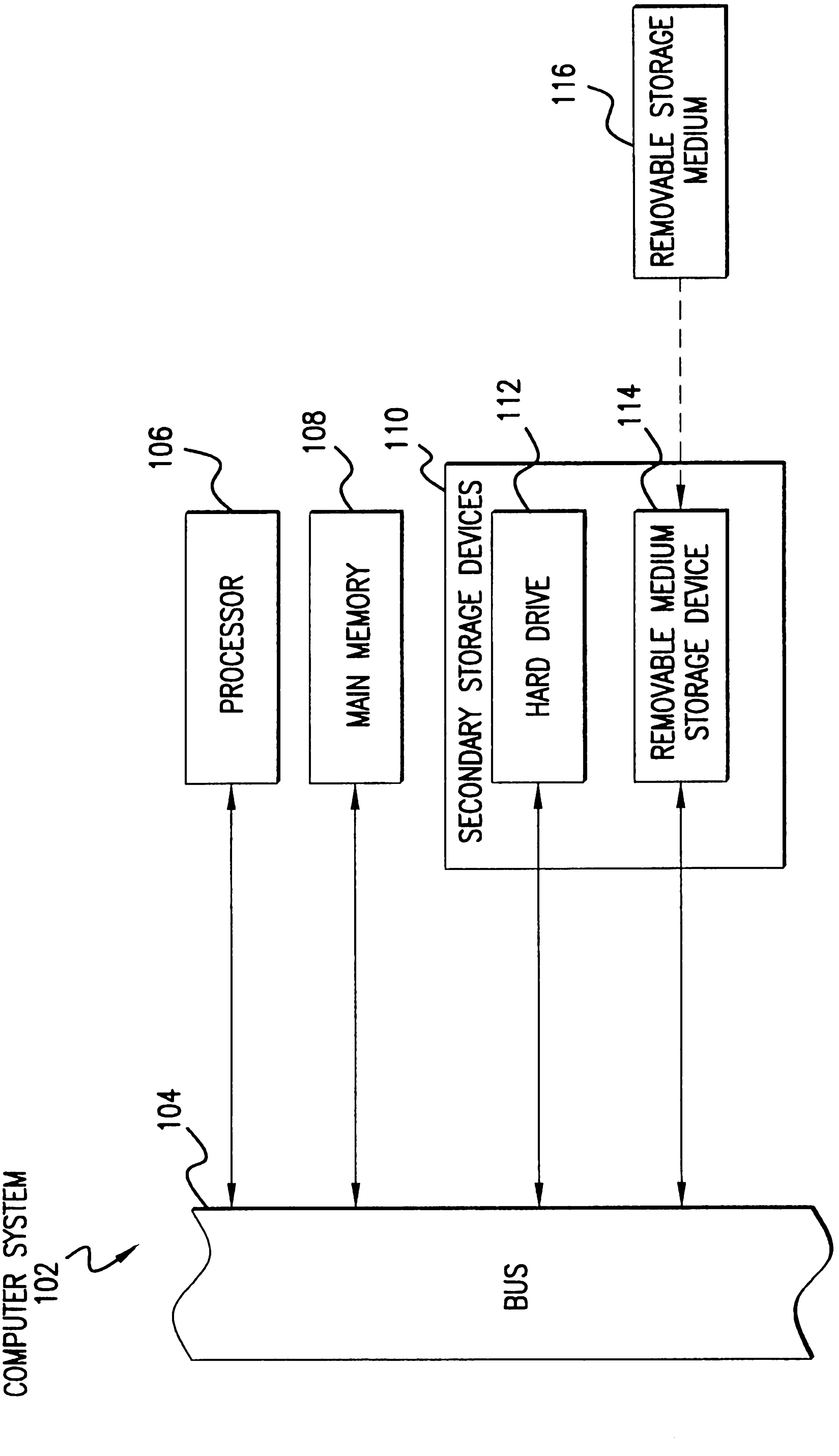

COMPUTER SYSTEM
102
104
BUS
106
PROCESSOR
108
MAIN MEMORY
110
SECONDARY STORAGE DEVICES
112
HARD DRIVE
114
REMOVABLE MEDIUM STORAGE DEVICE
116
REMOVABLE STORAGE MEDIUM

FIG. 1

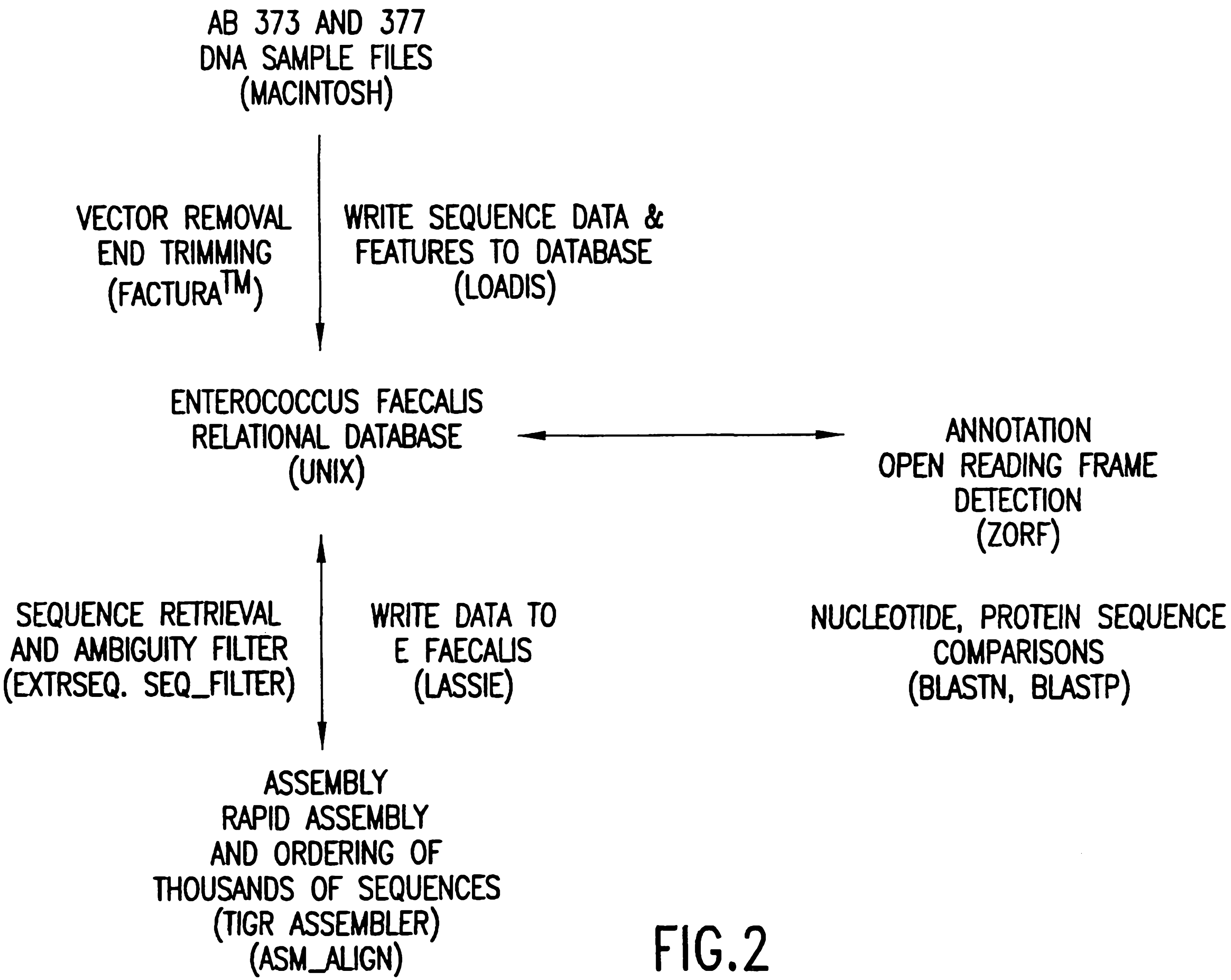

AB 373 AND 377
DNA SAMPLE FILES
(MACINTOSH)
VECTOR REMOVAL
END TRIMMING
(FACTURA™)
WRITE SEQUENCE DATA &
FEATURES TO DATABASE
(LOADIS)
ENTEROCOCCUS FAECALIS
RELATIONAL DATABASE
(UNIX)
ANNOTATION
OPEN READING FRAME
DETECTION
(ZORF)
NUCLEOTIDE, PROTEIN SEQUENCE
COMPARISONS
(BLASTN, BLASTP)
SEQUENCE RETRIEVAL
AND AMBIGUITY FILTER
(EXTRSEQ. SEQ_FILTER)
WRITE DATA TO
E FAECALIS
(LASSIE)
ASSEMBLY
RAPID ASSEMBLY
AND ORDERING OF
THOUSANDS OF SEQUENCES
(TIGR ASSEMBLER)
(ASM_ALIGN)

FIG.2