



(22) Date de dépôt/Filing Date: 2006/04/04

(41) Mise à la disp. pub./Open to Public Insp.: 2007/09/15

(30) Priorité/Priority: 2006/03/15 (CA2,538,898)

(51) Cl.Int./Int.Cl. *C12N 15/40* (2006.01),
A61K 39/12 (2006.01), *C07K 14/18* (2006.01),
C07K 16/10 (2006.01), *C12N 1/21* (2006.01),
C12N 15/63 (2006.01), *C12N 5/10* (2006.01),
C12N 7/00 (2006.01), *C12N 7/01* (2006.01),
C12Q 1/68 (2006.01), *C12Q 1/70* (2006.01),
G01N 33/569 (2006.01)

(71) Demandeur/Applicant:
INSTITUT PASTEUR, FR

(72) Inventeur/Inventor:
UNKNOWN, (UNKNOWN)

(74) Agent: ROBIC

(54) Titre : NOUVELLES SOUCHES ISOLEES ET PURIFIEES DU VIRUS CHIKUNGUNYA ET SEQUENCES DE
POLYNUCLEOTIDES ET DE POLYPEPTIDES, UTILISATIONS IMMUNOGENES ET DIAGNOSTIQUES CONNEXES

(54) Title: NOVEL ISOLATED AND PURIFIED STRAINS OF CHIKUNGUNYA VIRUS AND POLYNUCLEOTIDES AND
POLYPEPTIDES SEQUENCES, DIAGNOSTIC AND IMMUNOGENICAL USES THEREOF

(57) **Abrégé/Abstract:**

The present invention concerns wild-strains of Chikungunya virus isolated from patients exhibiting severe forms of infection and stemming from a human arbovirolosis epidemic. The present invention also concerns polynucleotide sequences and fragment thereof derived from their genome, the polypeptide encoded same and their use as diagnostic products, as vaccine and/or as immunogenic compositions.



ABSTRACT

The present invention concerns wild-strains of Chikungunya virus isolated from patients exhibiting severe forms of infection and stemming from a human arbovirolosis epidemic. The present invention also concerns polynucleotide sequences and fragment thereof derived from their genome, the polypeptide encoded same and their use as diagnostic products, as vaccine and/or as immunogenic compositions.

DEMANDES OU BREVETS VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.

CECI EST LE TOME _____ DE _____

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.

THIS IS VOLUME 1 OF 2

NOTE: For additional volumes please contact the Canadian Patent Office.

NOVEL ISOLATED AND PURIFIED STRAINS OF CHIKUNGUNYA VIRUS AND POLYNUCLEOTIDES AND POLYPEPTIDES SEQUENCES, DIAGNOSTIC AND IMMUNOGENICAL USES THEREOF

5

FIELD OF THE INVENTION

The present invention concerns wild-strains of Chikungunya virus isolated from
10 patients exhibiting severe forms of infection and stemming from a human
arbovirolosis epidemic. The present invention also concerns polynucleotide
sequences and fragment thereof derived from their genome, the polypeptide
encoded same and their use as diagnostic products, as vaccine and/or as
immunogenic compositions.

15

BACKGROUND OF THE INVENTION

Chikungunya virus (CHIKV) is a mosquito-transmitted *Alphavirus* belonging to
20 family *Togaviridae* [1,2]. It was isolated for the first time from a Tanzanian
outbreak in 1952 [3]. It is responsible for an acute infection of abrupt onset,
characterized by high fever, arthralgia, myalgia, headache and rash [4,5]. Poly-
arthralgia, the pathognomonic sign of the disease, is very painful. Symptoms are
generally self-limiting and last 1 to 10 days. However, arthralgia or arthritic
25 symptoms may persist for months or years. In some patients, minor hemorrhagic
signs such as epistaxis or gingivorrhagia have also been described.

CHIKV is geographically distributed in Africa, India and South East Asia. In Africa,
the virus is maintained through a sylvatic transmission cycle between wild
30 primates and mosquitoes such as *Aedes luteocephalus*, *Ae. furcifer* or *Ae. taylori*
[4]. In Asia, CHIKV is mainly transmitted from human to human by *Ae. aegypti*
and to a lesser extent by *Ae. albopictus* through an urban transmission cycle.

Since the 1952 Tanzania outbreak, CHIKV has caused outbreaks in East Africa (Tanzania, Uganda), in Austral Africa (Zimbabwe, South Africa), in West Africa (Senegal, Nigeria) and in Central Africa (Central African Republic, Democratic Republic of the Congo) [4]. The most recent epidemic re-emergence was documented in 1999-2000 in Kinshasa, where an estimated 50,000 persons were infected [6]. Since the first documented Asian outbreak in 1958 in Bangkok, Thailand, outbreaks have been documented in Thailand, Cambodia, Vietnam, Laos, Myanmar, Malaysia, Philippines and Indonesia [4,5]. The most recent epidemic re-emergence was documented in 2001-2003 in Java after 20 years [7]. Either in Africa or Asia, the re-emergence was unpredictable, with intervals of 7-8 years to 20 years between consecutive epidemics.

Since the end of 2004, Chikungunya virus (CHIKV) has emerged in the islands of the south-western Indian Ocean. Between January and March 2005, more than 5,000 cases were reported in Comoros. Later in 2005, the virus has circulated in the other islands, *i.e* Mayotte, Seychelles, Réunion and Mauritius. Starting in December 2005, the rainy season gave rise to a renewed epidemic circulation of the virus. Between January 1st and March 1st, 2006, 2,553, 3,471, and 4,650 cases have been reported in Mauritius, Mayotte and Seychelles (March 12th, 2006). The most affected island is Reunion with an estimated 212,000 cases until March 12th, 2006 (total population: 770,000) (<http://www.invs.sante.fr>). More recently, circulation of the virus has been documented in Madagascar (<http://www.invs.sante.fr>).

In Reunion Island, the first documented cases were patients coming back from Comoros in March 2005. More than 3,000 cases were reported from March to June. The transmission was limited during the winter season of the southern hemisphere and a major upsurge has been observed since mid-December, with an estimated 210,000 cases between January and March 2006 [8]. Since March 2005, 85 patients with a confirmed CHIKV infection have developed severe clinical signs (meningoencephalitis or fulminant hepatitis) which justified hospitalization in an intensive care unit. Several cases of meningo-encephalitis

and major algic syndrome have been associated with vertical transmission of the virus 9 (<http://www.invs.sante.fr>).

To date, two CHIKV complete nucleotide sequences have been determined, for
5 the strains Ross (accession no: AF490259) and S27 [9], both isolated from
patients during the 1952 Tanzania outbreak. Another complete nucleotide
sequence has been determined for a strain isolated in *Ae. furcifer* during the
Senegal 1983 outbreak (accession no AY726732). Khan and coworkers [9]
10 showed that the S27 genome was similar in its structure to that of other
alphaviruses and that O'nyong-nyong virus (ONN) was the closest relative to
CHIKV. In addition, phylogenetic analyses based on partial E1 sequences from
African and Asian isolates revealed the existence of three distinct CHIKV
phylogroups, one containing all isolates from West Africa, one containing isolates
from Asia, and one corresponding to Eastern, Central and Southern African
15 isolates [10]. Strains isolated in 1999-2000 in the Democratic Republic of the
Congo belonged to the latter phylogroup [6].

In the present study, the inventors determined the nearly complete nucleotide
sequences of viruses isolated from six patients originating from Reunion and
20 Seychelles Islands. The present invention allows to determine the genome
structure as well as the unique molecular features of the Indian Ocean outbreak
isolates, which distinguish them from other reported CHIKV and alphavirus
sequences. In addition, the phylogenetic origin and the diversity and
microevolution of the CHIKV strains responsible for the Indian Ocean outbreak
25 were investigated.

SUMMARY OF THE INVENTION

30 Section à compléter à la suite de la réception de votre approbation du jeu de
revendications.

BRIEF DESCRIPTION OF THE FIGURES

5 **Figure 1: Localization of the E1 changes on the 3D structure modeled from the crystal structure of SFV E1 [43] [19].**

A) Ribbon diagram of E1, with domain I colored red, domain II yellow and domain III, blue. Green tubes mark the disulfide bonds. The fusion peptide, at the tip of the molecule (in domain II) is colored orange and labeled. The N-terminus and
10 the C-terminus observed in the crystal (which is 30 aa upstream of the trans-membrane region) are also labeled. The 2 unique changes observed in the Indian Ocean isolates are indicated by stars and labeled: positions 226 (white) and 284 (magenta).

B) Partial representation (one octant, slightly extended) of the icosahedral E1
15 scaffold at the surface of the virion, viewed down a 5-fold symmetry axis. One E1 protomer is highlighted in colors, as in A); all the others are represented in grey. The location of the some of the icosahedral symmetry axes are drawn as solid black symbols: pentagon for 5-fold axis, triangle for 3-fold axes, ellipse for 2-fold axes (which in the T=4 lattice of alphaviruses are coincident with quasi 6-fold
20 axes). Open triangles indicate roughly the location of the E2 trimers that interact tightly with E1, covering domain II and the fusion peptide, and presenting the main antigenic sites. The open triangles mark also quasi 3-fold symmetry axes of the T=4 surface icosahedral lattice. A magenta ball marks the location of Glu 284, at an inter-E1 protomer contact site. This contact is propagated 240 times at the
25 surface lattice (note all pink balls drawn on the grey protomers). Note that the fusion peptide, in orange, is pointing up and away from contacts with other E1 protomers. This is more easily seen at the periphery of the virion, where one of them is labeled (FP). In the virion, this region of E1 is not accessible, covered underneath the E2 molecule [19].

Figure 2: Phylogenetic relationships among chikungunya isolates based on partial E1 nucleotide sequences. Isolates from the Indian Ocean outbreak (Reunion, Seychelles, Mayotte, Mauritius, Madagascar) represent a distinct clade within a large East, Central and South African (ECSA) phylogroup. Bootstrap resampling values are indicated at major nodes. The branch leading to West-African phylogroup (of length approx. 15%) was shortened for convenience.

Figure 3: Proposed evolutionary scenario of chikungunya virus isolates from the Indian Ocean outbreak. The scenario is based on six genome sequences determined by direct sequencing of RT-PCR products obtained using RNA extracts as templates; the sequences thus correspond to consensus sequences (Cons. Seq.) of the possible mixture of coexisting genomes (quasispecies). Inset: number of cases of E1-226A and E1-226V at different time intervals in Reunion Island, based on partial E1 sequences. E1-226V was observed in consensus sequences 2, 3 and 4, and therefore most E1-226V isolates genotyped based on partial E1 sequences are likely related to these genotypes. However, the independent appearance of E1-226V in other genotypes cannot be excluded. Int.: intermediate sequence. The location, size and relative position of the Islands and the African border are indicative. Consensus sequence 1 was obtained from a Reunion patient who traveled back from Comoros in March 2005, and from a Reunion Island patient. Sequences 2 to 4 were sampled in Reunion Island; sequence 5 was sampled in the Seychelles.

Figure 4 shows the nucleotide sequence of the genome of a CHIK virus strain according to a preferred embodiment of the invention and more specifically for the preferred strain named 05.115.

Figure 5 shows the nucleotide sequence of the genome of a CHIK virus strain according to a preferred embodiment of the invention and more specifically for the preferred strain named 05.209.

Figure 6 shows the nucleotide sequence of the genome of a CHIK virus strain according to a preferred embodiment of the invention and more specifically for the preferred strain named 06.21.

5 **Figure 7** shows the nucleotide sequence of the genome of a CHIK virus strain according to a preferred embodiment of the invention and more specifically for the preferred strain named 06.27.

10 **Figure 8** shows the nucleotide sequence of the genome of a CHIK virus strain according to a preferred embodiment of the invention and more specifically for the preferred strain named 06.49.

15 **Figure 9** shows the nucleotide sequence of the genome of a CHIK virus strain according to a preferred embodiment of the invention and more specifically for the preferred strain named 05.61.

20 **Figure 10** shows a nucleotide sequence of a fragment of a CHIK virus according to a preferred embodiment of the invention, and more specifically a fragment which codes for the ectodomain of the glycoprotein E2 of the preferred strain named 06.21.

25 **Figure 11** shows a nucleotide sequence of a fragment of a CHIK virus according to a preferred embodiment of the invention, and more specifically a fragment which codes for the ectodomain of the glycoprotein E2 of the preferred strain named 06.27.

30 **Figure 12** shows a nucleotide sequence of a fragment of a CHIK virus according to a preferred embodiment of the invention, and more specifically a fragment which codes for the ectodomain of the glycoprotein E2 of the preferred strain named 06.49.

Figure 13 shows a nucleotide sequence of a fragment of a CHIK virus according to a preferred embodiment of the invention, and more specifically a fragment which codes for the ectodomain of the glycoprotein E2 of the preferred strain named 06.115.

5

Figure S1. Repeat Sequence Elements found in the 3'NTR region

- A.** Alignment of Repeat Sequence Elements found in the 3'NTR region of chikungunya virus genome. All sequences form conserved and stable stem-loop structures in which the less conserved nucleotides around position 20 constitute the loop. Three RSE are found in all chikungunya genomes. The first one (RSE1) is inserted before the internal poly-A sequence of S27 genome [9], whereas the two others are found downstream this motif.
- B.** Predicted secondary structure for RSE1 of isolate 05-115.

15 **Figure S2. Focus phenotype of chikungunya viruses on AP61 cells by Focus Immunoassay**

- Mosquito AP61 cells in 24-well plates were infected with CHIK virus stocks grown on mosquito cells (virus titers $2-5 \times 10^8$ FFU. mL⁻¹) at 0.0001 (top well) or 0.00001 (bottom well) multiplicity of infection. Infected cells were overlaid with CMC in Leibovitz L15 growth medium with 2% FBS for 2 days to allow focus development at 28°C. The cells were fixed with 3% PFA in PBS, permeabilized with Triton X-100 in PBS, and foci of CHIK virus replication were immunostained with mouse anti-CHIK HMAF (dilution 1 :2,000) and peroxidase-conjugated goat anti-mouse Ig (dilution 1 :100).

25

DETAILED DESCRIPTION OF THE INVENTION

- The originality of the present invention is the identification of novel strains of the Chikungunya (CHIK) virus which are distinguished from CHIK virus of the prior art.

30

According to a first aspect, the present invention concerns an isolated and purified wild strain of chikungunya virus (CHIKV) capable of *in vitro* infecting human cells. The present invention also concerns a wild strain of CHIK virus which exhibits the same characteristics than those selected from the group consisting of the isolates 05.115, 05.61, 05.209, 06.21, 06.27 and 06.49. According to a preferred embodiment, the strains that are within the scope of the present invention are characterized in that their genome comprises at least one mutation when compared to the sequence of the genome of the CHIK virus strain S-27 (GenBank AF339485). The genome of the preferred strains according to the present invention comprises a sequence as shown in Figure 4, 5, 6, 7, 8 or 9.

Also within the scope of the invention, is any strain grown or obtained by cell culture from a sample of a preferred CHIK strain of the invention.

According to another aspect, the present invention provides an isolated and purified strain of chikungunya virus (CHIKV) comprising at least one mutation in structural protein E1 and/or in structural protein E2, and more particularly in their ectodomain region. According to a preferred embodiment, the strain of the invention is characterized by the fact that its genome comprises at least one mutation in the E2 protein selected from the group consisting of I23T, G57K, I74M, G79E, N160T, A164T, L181M, I211T, S299N, T312M, A344T, S375T, V386A, as shown in Table S3. According to another preferred embodiment, the strain of the invention is characterized by the fact that its genome comprises at least one mutation in the E1 protein selected from the group consisting of A226V, M269V, D284E and V322A as shown in Table S3. Most preferably, the mutation in the E1 protein is A226V.

Another aspect of the invention concerns an isolated and purified polynucleotide comprising a sequence as shown in Figures 4, 5, 6, 7, 8 or 9.

The present invention also provides, as another aspect, a fragment of the polynucleotide as defined above, wherein it is suited to be obtained by:

- use of restriction enzymes wherein their cleavage sites are present in said polynucleotide;
- amplification with specific primers for said polynucleotide;
- *in vitro* transcription; or
- 5 - chemical synthesis.

A preferred fragment of the invention is characterized by the fact that it codes for the glycoprotein E1 or E2, and more preferably for their ectodomain region. Advantageously, the fragment of the invention when coding for the E2
10 ectodomain, comprises, or more preferably, consists of a nucleotide sequence as shown in Figures 10, 11, 12 or 13.

According to another aspect, the present invention is concerned with an isolated and purified polypeptide encoded by a polynucleotide or by a fragment of the
15 invention. As used herein, the terms "polypeptide" and "protein" are used interchangeably to denote an amino acid polymer or a set of two or more interacting or bound amino acid polymers.

By "isolated" is meant, when referring to a polypeptide, that the indicated
20 molecule is separate and discrete from the whole organism with which the molecule is found in nature or is present in the substantial absence of other biological macro-molecules of the same type. The term "isolated" with respect to a polynucleotide is a nucleic acid molecule devoid, in whole or part, of sequences normally associated with it in nature; or a sequence, as it exists in nature, but
25 having heterologous sequences in association therewith; or a molecule disassociated from the chromosome.

Broadly defined, the terms "purified polypeptide" or "purified polynucleotide" refer to polypeptides or polynucleotides that are sufficiently free of other proteins or polynucleotides, or carbohydrates, and lipids with which they are naturally
30 associated. The polypeptide or polynucleotide may be purified by any process by which the protein or polynucleotide are separated from other elements or

compounds on the basis for instance, of charge, molecular size, or binding affinity.

5 The preferred peptides of the invention comprise at least one amino acid substitution compared with the amino acid sequence of strain S-27 (GenBank AF339485) and are derived from the sequence of a protein coded by a fragment of the invention.

10 The present invention is also concerned with a vector comprising a fragment of the invention. As used herein, the term "vector" refers to a polynucleotide construct designed for transduction/transfection of one or more cell types. Vectors may be, for example, "cloning vectors" which are designed for isolation, propagation and replication of inserted nucleotides, "expression vectors" which are designed for expression of a nucleotide sequence in a host cell, or a "viral
15 vector" which is designed to result in the production of a recombinant virus or virus-like particle, or "shuttle vectors", which comprise the attributes of more than one type of vector. Preferred vector are those deposited at the CNCM on March 15, 2006 under accession numbers I-3587, I-3588, I-3589 and I-3590.

20 In a related aspect, the present invention provides a recombinant cell comprising a vector as defined above. The term "recombinant cell" refers to a cell that has a new combination of nucleic acid segments that are not covalently linked to each other in nature. A new combination of nucleic acid segments can be introduced into an organism using a wide array of nucleic acid manipulation techniques
25 available to those skilled in the art. A recombinant cell can be a single eukaryotic cell, or a single prokaryotic cell, or a mammalian cell. The recombinant cell can harbor a vector that is extragenomic. An extragenomic nucleic acid vector does not insert into the cell's genome. A recombinant cell can further harbor a vector or a portion thereof that is intragenomic. The term intragenomic defines a nucleic
30 acid construct incorporated within the recombinant cell's genome. A preferred recombinant cell of the invention *E. coli* such as the one containing a vector of the

invention and deposited at the CNCM on March 15, 2006 under accession numbers I-3587, I-3588, I-3589 and I-3590.

The present invention is further concerned with a monoclonal antibody or polyclonal antibodies, or fragments thereof, that specifically bind to a polypeptide of the invention. As used herein, the term "specifically binds to" refers to antibodies that bind with a relatively high affinity to one or more epitopes of a protein of the invention, but which do not substantially recognize and bind to molecules other than the one(s) of interest. As used herein, the term "relatively high affinity" means a binding affinity between the antibody and the protein of interest of at least 10^{-6} M, and preferably of at least about 10^{-7} M and even more preferably 10^{-8} M to 10^{-10} M. Determination of such affinity is preferably conducted under standard competitive binding immunoassay conditions which is common knowledge to one skilled in the art.

As used herein, the term "antibody" refers to a glycoprotein produced by lymphoid cells in response to a stimulation with an immunogen. Antibodies possess the ability to react *in vitro* and *in vivo* specifically and selectively with an antigenic determinant or epitope eliciting their production or with an antigenic determinant closely related to the homologous antigen. The term "antibody" is meant to encompass constructions using the binding (variable) region of such an antibody, and other antibody modifications. Thus, an antibody useful in the method of the invention may comprise a whole antibody, an antibody fragment, a polyfunctional antibody aggregate, or in general a substance comprising one or more specific binding sites from an antibody. The antibody fragment may be a fragment such as an Fv, Fab or F(ab')₂ fragment or a derivative thereof, such as a single chain Fv fragment. The antibody or antibody fragment may be non-recombinant, recombinant or humanized. The antibody may be of an immunoglobulin isotype, e.g., IgG, IgM, and so forth. In addition, an aggregate, polymer, derivative and conjugate of an immunoglobulin or a fragment thereof can be used where appropriate.

Another aspect of the invention is the use of an element selected from the group consisting of a strain, a polynucleotide, a fragment, a vector, a recombinant cell, a polypeptide and an antibody of the invention for either the detection of a CHIKV associated to an arbovirosis, or for the preparation of a composition that prevents
5 and/or treats an arbovirosis.

Another aspect of the present invention relates to a composition for treating and/or preventing an arbovirosis. The composition of the present invention advantageously comprises at least one element selected from the group
10 consisting of a strain, a polynucleotide, a fragment, a vector, a recombinant cell, a polypeptide and an antibody of the invention.

As used herein, the term "treating" refers to a process by which the development of an infection from a CHIKV is affected or completely eliminated. As used herein,
15 the term "preventing" refers to a process by which the CHIKV infection is obstructed or delayed.

As used herein, the expression "an acceptable carrier" means a vehicle for containing the components (or elements) of the composition of the invention that can be administered to a animal host without adverse effects. Suitable carriers
20 known in the art include, but are not limited to, gold particles, sterile water, saline, glucose, dextrose, or buffered solutions. Carriers may include auxiliary agents including, but not limited to, diluents, stabilizers (i. e., sugars and amino acids), preservatives, wetting agents, emulsifying agents, pH buffering agents, viscosity enhancing additives, colors and the like.

25 The amount of components of the composition of the invention is preferably a therapeutically effective amount. A therapeutically effective amount of components of the composition of the invention is the amount necessary to allow the same to perform their preventing and/or treating role against a CHIKV infection without causing overly negative effects in the host to which the
30 composition is administered. The exact amount of components to be used and

the composition to be administered will vary according to factors such as the mode of administration, as well as the other ingredients in the composition.

The composition of the invention may be given to a host (such as a human) through various routes of administration. For instance, the composition may be administered in the form of sterile injectable preparations, such as sterile injectable aqueous or oleaginous suspensions. These suspensions may be formulated according to techniques known in the art using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparations may also be sterile injectable solutions or suspensions in non-toxic parenterally-acceptable diluents or solvents. They may be given parenterally, for example intravenously, intramuscularly or sub-cutaneously by injection, by infusion or *per os*. Suitable dosages will vary, depending upon factors such as the amount of each of the components in the composition, the desired effect (short or long term), the route of administration, the age and the weight of the host to be treated. Any other methods well known in the art may be used for administering the composition of the invention.

Yet another aspect of the invention is to provide a kit for the detection of a CHIKV associated to an arbovirosis, comprising at least one element selected from the group consisting of a strain, a polynucleotide, a fragment, a vector, a recombinant cell, a polypeptide and an antibody of the invention. Kits according to this embodiment of the invention may comprise packages, each containing one or more of the above mentioned elements (typically in concentrated form) which are required to perform the respective diagnostic tests.

EXAMPLES

The inventors (as sometimes referred therein as "we") report the nearly complete genome sequence of six selected clinical isolates, along with partial sequences of glycoprotein E1 from a total of 60 patients from Réunion, Seychelles, Mauritius,

Madagascar and Mayotte Islands. The present results indicate that the outbreak was initiated by a strain related to East-African isolates, from which viral variants have evolved following a traceable microevolution history. Unique molecular features of the outbreak isolates were identified. Notably, in the region coding for the non-structural proteins, ten amino acid changes were found, three of which being located in alphavirus conserved positions of nsP2 (which contains helicase, protease and RNA triphosphatase activities) and of the polymerase nsP4. The sole isolate obtained from the cerebrospinal fluid of a patient showed unique changes in nsP1 (T301I), nsP2 (Y642N) and nsP3 (E460 deletion). In the structural protein region, two noteworthy changes (A226V and D284E) were observed in the membrane fusion glycoprotein E1. Homology 3D modelling allowed mapping of these two changes to regions that are important for virion assembly and for membrane fusion. Change E1-A226V was absent in the initial strains but was observed in >85% of subsequent viral sequences from Reunion, denoting evolutionary success possibly due to adaptation to the mosquito vector.

Material and Methods

Patients. The 60 patients for whom partial or complete CHIKV nucleotide sequences were determined originated from Reunion (N=43), Seychelles (N=3), Madagascar (N=7), Mayotte (N=4) and Mauritius (N=3). Characteristics of the patients and biological samples are listed in **Table 1**.

Virus isolation and RNA extraction. Viruses were isolated either from serum or cerebrospinal fluid (CSF) (**Table 1**). Briefly, C6-36 *Aedes albopictus* cells were inoculated with 1 ml of serum or CSF diluted 1:10 in L15 medium (Gibco). The cells were grown at 28°C in L15 supplemented with 5% foetal bovine serum and 10% tryptose-phosphate. Cells and supernatants were harvested after the first passage (5 days) and the second passage (7 days). The virus isolates were identified as CHIKV by indirect immunofluorescence, using CHIKV hyperimmune ascitic fluid. In the case of isolates 05.115, 06.21, 06.27 and 06.49 whose

genomes were sequenced, absence of yellow fever, dengue and West Nile viruses was confirmed by indirect immunofluorescence using specific sera. RNA was extracted using the QIAamp Viral Minikit (Qiagen, France).

5 **Nucleotide sequencing.** Primers (**Table S1**) were designed based on the nucleotide sequence 20 of the S27 strain. RT-PCR was performed using the Titan One Tube RT-PCR kit (Roche, France). RT-PCR fragments were purified by ultrafiltration prior to sequencing (Millipore, France). Sequencing reactions were performed using the BigDye Terminator v1.1 cycle sequencing kit (Applied Biosystems, USA) and purified by ethanol precipitation. Sequence chromatograms were obtained on automated sequence analysers ABI3100 or ABI3700 (Applied Biosystems). All amplicons were sequenced on both strands.

Assembly of genome sequences and sequence analysis. Contig assembly was performed independently by distinct operators and software, using either BioNumerics version 4.5 (Applied-Maths, Sint-Martens-Latem, Belgium) or PhredPhrap / Consed [11]. Both analyses yielded exactly the same consensus sequence for all strains. A single contig of 11,601 nt was obtained for five isolates, whereas for strain 05.61, a sequence portion was missing, between S27 positions 5,246 to 5,649 (positions 390 to 524 of nsP3). Sequence alignments and computation of substitution tables were performed using programs BioNumerics, DNASP version 4.10 [12] and DAMBE version 4.2.13 [13]. Alignments of nucleotide and amino acid sequences against selected alphavirus sequences 1 es were performed with the ClustalW1.7 software [14]. Sequence identities were computed with the Phylip package [15]. RNA secondary structure was predicted with the Vienna RNA secondary structure server [16]. Neighbor-joining trees were constructed using MEGA version 3.1 [17] with the Kimura-2 parameter corrections of multiple substitutions. Reliability of nodes was assessed by bootstrap resampling with 1,000 replicates. Amounts of synonymous substitutions per synonymous site (Ks) and of non synonymous substitutions per non synonymous site (Ka) were estimated using DNASP. RDP2 [18] was used to detect putative mosaic sequences.

3D structure modeling. The crystallographic structure of the ectodomain of the glycoprotein E1 of Semliki Forest Virus (SFV) at neutral pH [19]; Protein Data Bank code 2ALA) was used as a template to model and analyze the two amino acid mutations of the Indian Ocean isolates. The figure 2 was prepared using the program RIBBONS [20].

Detection of viral foci by immunological staining. *Aedes pseudoscutellaris* AP61 cells were grown in a 24-well tissue culture plates in Leibovitz L-15 growth medium with 10% heat inactivated fetal calf serum (FCS) for 24 h. Mosquito cell monolayers were washed once with Leibovitz L-15 and 0.2 ml Leibovitz L-15/2% FCS were added. Cells were infected with CHIK virus in 0.2 ml of Leibovitz L-15/2% FCS and incubated at 28°C for 1 h. Overlay medium consisting of 0.4 ml of Leibovitz L-15/2% FBS and carboxymethylcellulose (CMC) (1.6%) was then added and the tissue culture plates were incubated at 28°C for 2 days. Foci of infected cells were visualized by focus immunoassay (FIA). The cells were washed with PBS, fixed with 3% paraformaldehyde (PFA) in PBS for 20 min, and permeabilized with 0.5% Triton X-100 in PBS for 4 min at room temperature. The fixed cells were incubated for 20 min at 37°C with 1:2,000 dilution of hyperimmune mouse ascitic fluid (HMAF) directed against CHIKV. Goat anti-mouse IgG, horseradish peroxidase conjugated was used as the second antibody (1:100 dilution) at 37°C for 20 min. Foci were visualized with DAB Peroxidase Substrate (Sigma).

1. Genome structure and molecular signatures of the Indian Ocean outbreak chikungunya viruses

Genome organization. We determined the nearly complete genome sequences of six CHIKV isolates (05.115, 05.61, 05.209, 06.21, 06.27 and 06.49) representing distinct geographic origins, time points and clinical forms (**Table 1**) of the Indian Ocean outbreak of chikungunya virus. 11,601 nucleotides were determined, corresponding to positions 52 (5'NTR) to 11,667 (3'NTR, end of third Repeat Sequence Element) in the nucleotide sequence of the 1952 Tanzanian isolate

S27 (total length 11,826 nt). There were three insertion/deletion events between S27 and Réunion isolates, two of which were observed in the 3'NTR. First, the internal poly-A stretch of 14 nucleotides observed in S27 (11,440-11,443) and corresponding to a probable internal poly-A site [9] was replaced by a stretch of
 5 only 5 A in Indian Ocean isolates, similar to what was observed in other chikungunya viruses, e.g. the Ross strain (accession no.: AF490259). Second, one A was missing in Indian Ocean isolates in a 5-A stretch at S27 position 11,625. Finally, one codon was missing in isolate 06.27, corresponding to nsP3 codon 460, at which all other Indian Ocean isolates analyzed and available
 10 alphavirus sequences are GAA, coding for Glu.

The genome sequences of the six isolates presented therein was similar to those previously reported for alphaviruses [9,21,22]. Coding sequences consisted of two large open reading frames (ORF) of 7,422 nt and 3,744 nt encoding the non-
 15 structural polyprotein (2,474 amino-acids) and the structural polyprotein (1,248 amino-acids), respectively. The non structural polyprotein is the precursor of proteins nsP1 (535 aa), nsP2 (798 aa), nsP3 (530 aa) and nsP4 (611 aa), and the structural polyprotein is the precursor of proteins C (261 aa), p62 (487 aa, precursor to E3 - 64 aa – and E2 - 423 aa), 6K (61 aa), and E1 (439 aa).
 20 Cleavage sites characteristic of the alphavirus family in the non-structural and structural polyproteins were conserved. Glycosylation sites in E3, E2 and E1 were also conserved. A 65 nt junction sequence was identified between the stop codon (TAG, 7499-7501) of the non-structural ORF and the start codon (7567-7569) of the structural ORF. The 5' non-translated region (5'NTR) ended at
 25 position 76. The 3'NTR region started at position 11,314 and contained three repeat sequence elements (RSE) with predicted secondary structures (**Fig. S1**) that were consistent with previous work [9].

Differences between Indian Ocean outbreak isolates and strain S27. Compared
 30 to strain S27, Réunion isolate 05.115 showed 28 aa changes (1.13%) in the non-structural proteins (**Table S2**), with the highest proportion in nsP3 (2.26%) and the lowest in nsP2 (0.6%). Ten out of 12 amino acid changes in nsP3 were

concentrated between positions 326 and 524 (5.0% variation), similar to findings in ONN viruses [23]. One important difference with S27 was that the Indian Ocean isolates exhibited an opal stop codon (UGA) at nsP3 codon 524, instead of Arg (CGA) in S27. This opal codon was observed in related alphaviruses [9,22,23], and is believed to regulate the expression of nsP4, the putative RNA polymerase, by a read-through mechanism [21,24].

Compared to S27, the structural proteins showed 21 (1.68%, for 05.115) to 22 (1.76%, for other isolates) amino-acid substitutions in Indian Ocean isolates (**Table S3**). Notably, envelope protein E2 showed the highest variation, with 14 (3.3%) aa changes, higher than envelope protein E1 (0.68%) and the capsid protein (0.38%). The ratio of rates of evolution of synonymous and non-synonymous sites (Ks/Ka) between S27 and 05.115 isolates was 11.0 for the whole polyprotein, whereas it was only 6.12 for protein E2, probably indicative of a positive selection in favor of amino-acid changes in this immunogenic protein. By comparison, Ks/Ka was 18.75 for the non-structural polyprotein.

Indian Ocean outbreak molecular signatures in non-structural proteins and phenotypic variation. Ten positions (excluding polymorphic positions) had aa that were unique to the non-structural proteins of outbreak isolates, when compared to other CHIKV sequences (**Table 2**). First, nsP2-54 was Asn in Indian Ocean isolates and in SFV, but was Ser in all other sequences. Second, nsP2-374 was Tyr in Indian Ocean isolates, but was His or Asn in other alphavirus sequences (Table 2). Third, position 500 in nsP4 was Leu in the Indian Ocean sequences instead of Gln in the four other reported CHIKV sequences. Interestingly, this position, which is about 30 aa from the catalytic “GDD” motif, is a strictly conserved Glu in all other alphaviruses. The remaining seven changes took place in relatively variable regions.

Additional specific changes were observed in isolates 05.209 (S358P) and 06.27 (nsP1-T301I, nsP2-Y642N, and nsP3-460del). Notably, our phenotypic assays conducted in parallel showed differences for strain 06.27. Focus immunoassay

showed that CHIKV stocks 05.115, 06.21, 06.27 and 06.49 formed mixtures of foci with different sizes on *Ae. Albopictus* C6/36 (data not shown) and *Ae. pseudoscuterallis* AP61 cells (**Fig. S2**). Interestingly, only isolate 06-27 formed medium foci, whereas others formed minutes and small foci. The particular
 5 phenotype of 06-27 could be linked to the observed aa differences in the non structural proteins, which are involved in the viral replication [21].

Indian Ocean molecular signatures in structural proteins and 3D modelling. When analyzing the aa sequences of the structural proteins, seven positions (four in E2,
 10 one in 6K and two in E1) were found to be unique to isolates from the Indian Ocean outbreak (**Table 2**). Two of these were located in the E2 ectodomain, with Thr 164 and Met 312 being identified in our isolates instead of Ala and Thr, respectively, in all other available CHIKV sequences (**Table 2**). The first of these
 15 two positions is variable in alphaviruses; it lies in a region defined previously as containing neutralizing epitopes [5,25]. At position 312, Thr is present in other CHIKV, in ONNV and in SFV, but varies in other alphaviruses; it lies in a region identified as important for E1-E2 oligomerization [5,25].

In E1, two crucial substitutions were observed, one at residue 226, specific to
 20 Indian Ocean isolates, and one at residue 284, present in 3 out of 6 Indian isolates (06.21, 06.27 and 06.49). Both mutations were mapped on the 3D structure (modeled from the crystal structure of SFV E1) in **Fig. 1**. Interestingly, residue 226 is Ala in all reported CHIKV sequences (**Table 2**), and was also Ala in the first of our Indian Ocean isolates sequenced here (05.61 and 05.115,
 25 obtained at the beginning of the outbreak). All subsequent isolates (obtained from patients collected in November and December 2005) displayed a Val residue at this position. Although position 226 is relatively variable among alphaviruses, it was observed that a single mutation at this position (Pro to Ser) allowed SFV to adapt to growth in cholesterol-depleted insect cells [26,27].

30 The other unique aa observed in E1 from Indian Ocean isolates was Glu 284. This is a highly conserved position in E1, which displays an Asp in the majority of

alphaviruses or an Asn in SIN (**Table 2**). This amino acid is located at the interface between E1 protomers at the surface of the virion, participating in contacts that make up the icosahedral E1 scaffold (**Fig. 1**).

5 **2. Phylogenetic analysis**

Previous work based on E1 protein sequences showed strong phylogeographic structure of the chikungunya virus species [6,10]. In order to determine the progenitor phylogroup from which the Indian Ocean outbreak isolates emerged, we compared a 1,044 nt region within the E1 coding sequence (positions 271 to
10 1314, *i.e.*, codons 91 to 438) from 63 biological specimens from 60 patients from Reunion, Seychelles, Madagascar, Mayotte and Comoros (**Table 1**) with 29 other available chikungunya sequences (**Table S4**). Phylogenetic analysis (**Fig. 2**) clearly demonstrated that the current Indian Ocean isolates represent a homogeneous clade within a broad group (group ECSA) comprising isolates from
15 East, Central and South Africa (ECSA, **Fig. 2**). The isolates from an outbreak in Democratic Republic of the Congo [6] also formed a homogeneous clade within group ECSA. There was no ECSA group member showing a significantly closer relationship with the Indian Ocean isolates. Asian isolates were less related to Indian Ocean isolates and constituted the sister group of group ECSA, whereas
20 West-African isolates were even more divergent. Inclusion of other alphaviruses, including the closest relative ONN, placed the root of the chikungunya isolates on the branch leading to the West-African phylogroup (data not shown).

Comparison of the sequences of Indian Ocean outbreak isolates to the S27
25 sequence revealed 316 (2.7%) nucleotide substitutions in isolate 05.115 (**Table S5**). The Asian clade Nagpur strain showed 5.1% average nucleotide divergence from 05.115, whereas the West-African clade Senegal strain 37997 displayed 15% difference (**Table S5**). Interestingly, the latter strain showed complete conservation of a 87 nucleotides portion (9,958 – 10,045, at the junction between
30 structural proteins 6K and E1) with East-African and Indian Ocean outbreak isolates. Sequence identity in this portion may reflect a past event of genetic recombination between West-African and East/Central-African strains. Differently,

we did not find statistical support ($P > 7E-2$) for sequence mosaicism or recombination since the split between S27 and Reunion isolates, although some genomic regions differed in their density of nucleotide polymorphisms.

5 **3. Genotypic and phenotypic variation among Indian Ocean outbreak isolates and microevolutionary scenario**

Specific aa changes in the non-structural proteins were observed in the isolates 05.209 (S358P) and 06.27 (nsP1-T301I, nsP2-Y642N, and nsP3-460del). In the structural proteins, change E1-A226V was observed in isolates 06.21, 06.27 and
10 06.49, and change E2-Q146R in the Seychelles isolate 05.209. In addition to these non-synonymous changes, there were 8 silent substitutions, observed in 05.209, 06.27 and 06.49 (**Table 3**).

A history of probable sequence evolution that occurred during the outbreak
15 (**Fig. 3**) was deduced from the 14 amino-acid variations observed among the six complete genomes (**Table 3**). Isolate 05.61 was initially selected for genome analysis because it was isolated in March 2005, at the onset of the outbreak, from a Reunion patient returning 1 turning from Comoros Island, where the outbreak had been going on since January 2005. Remarkably, the isolates 05.61 and
20 05.115 (which was the second earliest isolate analyzed), the African isolate S27 and previous unrelated chikungunya isolates from Africa and Asia were identical at all 14 polymorphic sites. Therefore, the consensus sequence of isolates 05.61 and 05.115 (consensus sequence 1) likely represents the ancestral genotype of the Réunion outbreak. Distribution of the 14 polymorphisms suggested that this
25 founder gave rise to three consensus sequences that likely evolved in four steps. First, substitution at genome position 10,670 (causing the E1 A226V change) gave rise to consensus sequence 2, represented by the late-November 2005 isolate 06.21. Second, a G to A synonymous substitution at position 6,547 (nsP4) led to an intermediate sequence, which itself gave rise to two late sequences:
30 consensus sequence 3 (isolate 06.27), following four additional substitutions and one codon deletion (**Table 3**), and consensus sequence 4 (06.49), which arose after three distinct synonymous substitutions (**Table 3**). A fifth consensus

sequence was represented by the Seychelles isolate 05.209 alone, which exhibited four substitutions (two of them causing aa changes in nsP3 - S358P and in E2 - Q146R) compared to consensus sequence 1 (**Fig. 3**).

- 5 Since Reunion isolates had E1-226A at the beginning of the outbreak and E1-266V A at the beginning of the outbreak and E1-266V later in the epidemics, we compared residue 226 in 57 additional sequences (57 sequences from 54 sera and 3 CSF) from the Indian Ocean epidemic. Remarkably, the nature of E1-226 differed totally on Reunion Island before and after the winter season. Five
10 sequences from patients sampled from March to June 2005 (including the sequence originating from a traveller back from Comoros) had E1-226A. Between September and end December 2005, 21 sequences showed E1-226V. Among 17 Reunion sequences from 2006, E1-226V was observed 12 times and E1-226A 5 times (Table 1). On Madagascar and Seychelles sequences, for which the
15 samples were collected when the first clinical cases were suspected (i.e probably at the beginning of the outbreaks), only the E1-226 Ala was observed. On Mayotte 2006 sequences, only the E1-226 V was observed. On Mauritius 2006 sequences, both E1-226 Ala and Val were observed.
- 20 CHIKV is currently causing one of the largest chikungunya fever outbreaks reported in the last forty years [4-6,28]. The magnitude of the epidemics has surprised the population, policy makers and public health specialists, but chikungunya is not uncommon in nearby regions of this part of the world. Our phylogenetic analyses based on partial glycoprotein E1 sequences indicate that
25 the Indian Ocean outbreak was caused by the same strain on Reunion, Mayotte, Madagascar, Seychelles and Mauritius Islands, and show that the outbreak strain is related to East, Central and South African isolates. Although to our knowledge no outbreak was reported recently in East Africa, this scenario is compatible with the human population exchanges between East Africa and Comoros, where the
30 outbreak is believed to have started. Whether the epidemic strain results from the evolution of an enzootic strain, as previously described for VEE [29], cannot be excluded. Sequencing of additional isolates from recent outbreaks in Africa and

from possible local reservoirs, should define more precisely the origin of the Indian Ocean outbreak.

5 This study represents the first survey of intra-outbreak analysis of CHIKV nucleotide variation on a genomic scale. The availability of multiple nearly complete genome sequences allowed us to deduce the probable history of successive molecular evolutionary changes that may have occurred while the outbreak was still ongoing. This evolutionary scenario is the most likely based on the six consensus sequences obtained, but additional isolates and determination
10 of quasispecies heterogeneity are needed to obtain a more precise picture of viral evolution during the outbreak.

Whereas genotype E1-226 Ala was exclusively observed during the first period of the Reunion outbreak (March-June 2005), our data show the emergence and
15 predominance of genotype E1-226 Val, which was observed from beginning of September 2005 and experienced a spectacular rise in frequency, associated with the explosive increase in the number of cases observed in Reunion since mid-December 2005. Since the presence of cholesterol and sphingolipids is necessary for the fusogenic protein E1 to bind to target membranes to induce
20 membrane fusion, it is possible that such a mutation provides a selective advantage to the virus in mosquitoes, which are cholesterol auxotrophs. Indeed, in the SFV E1 crystal structure, position 226 is located in the ij loop, in contact with the fusion peptide, in a region of the protein that is predicted to interact with the target membrane. A single mutation at this position (Pro to Ser) allowed SFV
25 to adapt to growth in cholesterol-depleted insect cells [26,27]. Further mutagenesis studies in this region of SINV [30] also indicate that this loop controls cholesterol dependence, although in SINV, which is more distant from CHIKV than is SFV, a single amino acid change was not enough to obtain a cholesterol independence effect. This data suggest the possibility that the
30 evolutionary success of E1-A226V may have been a consequence of a selective advantage conferred directly by the A226V change. Further experiments comparing the growth kinetics of our different CHIKV isolates in cholesterol-

depleted insect cells are needed to test this hypothesis. Indeed, this change may possibly be selectively neutral, and other evolutionary factors such as genetic drift or a founder effect could have favored E1-226V by chance alone. Genotyping of other parts of the genome will be needed to control that the A226V change did not appear independently in other lineages. Also, more sampling is needed to determine whether an E1-226 Ala to Val change has occurred independently in other islands, and if the E1-226V experienced an evolutionary success comparable to that observed on Reunion Island.

To date, only CHIKV laboratory strains, passaged many times on mosquito or mammalian cells, had been entirely sequenced [9]. We provide for the first time nearly complete nucleotide sequences of six clinical isolates passaged *in-vitro* only once or twice (see M&M section). The presence in infected patients of a mixed viral population, called quasispecies [31-33], with genotypes co-existing in an equilibrium governed by a balance between mutation and natural selection. The presence in S27 of an Arg codon instead of the opal stop codon in Indian Ocean isolates is probably explained by numerous *in-vitro* passages of S27, as evolution of opal to Arg was observed experimentally in ONN viruses [23]. Whereas it may be advantageous for viral quasispecies to maintain the opal codon *in-vivo*, an Arg codon probably confers a selective advantage *in-vitro*, as observed for the closely related Semliki Forest virus [34]. Chikungunya virus quasispecies situation *in-vivo* could also explain the nsP1-T301I polymorphism observed for the LCR isolate 06.27. Indeed, it is likely that selection for a subset of genotypes harboring this change may be associated with invasion of the LCR [33]. These results underscore that the genome sequence of laboratory "reference" strains may not accurately reflect the natural situation, as the genotypic complexity of quasispecies *in-vivo* is subject to erosion by *in-vitro* selection. Since the Indian Ocean isolates sequenced here were subjected to *in-vitro* selection for only a few generations, they probably correspond more closely to the *in-vivo* genotypes than previously sequenced chikungunya strains.

The aa differences detected among the outbreak 1 isolates may relate to biological or pathogenic characteristics of the virus. Although our viral culture results are preliminary, they clearly show phenotypic differences between the unique isolate from CSF (06.27), isolated from a neonatal encephalopathy case, and three other isolates, associated with either the classical form of the disease or encephalopathy. The larger foci observed in culture with 06.27 could reflect a higher replication rate of the virus and be linked to the specific amino acid changes identified in nsP1, nsP2 and nsP3. Single amino-acid changes in nsP1, including a Thr/Ile change (residue 538 of Sindbis virus) [35,36] and a 18-nt deletion in nsP3 have previously been shown to affect neurovirulence in other alphaviruses [35-37]. However, in the absence of nsP1 structural data, it is difficult to predict the structural or functional impact of the I301T change observed in 06.27 isolate. It should also be noted that all the viral sequences determined from either the serum or the isolates from three neonatal encephalopathy cases and an adult meningo-encephalitis case had E1-226 Val. However, as this genotype is observed also in classical forms of the disease, a potential link of E1-226 Val with neuropathogenesis needs further studies. Host factors have to be considered in the occurrence of neurological forms of the disease. For example, the blood-brain crossing may be favoured by young age or hypertension.

Unique molecular signatures of the Indian Ocean outbreak genomes were identified when they were compared to all other reported alphavirus sequences. These features represent interesting targets for future functional studies, as well as for epidemiological follow-up. One particularly interesting feature was the E1-226 Val residue (see above). Another interesting molecular signature of Indian Ocean outbreak genomes was E1-284 Asp. Although pseudo-atomic model of the scaffold used is of modest resolution (the resolution of the crystal structure is limited - approaching 3Å - and the model results of fitting this structure into a 9Å resolution cryo-electron microscopy reconstruction), it appears that the side-chain of Asp 284 interacts with the main chain of an adjacent E1 polypeptide in the virion. Indeed, it is in a position compatible with acceptance of a hydrogen bond from main chain amide 379 from the neighboring E1 protomer. Because the

packing is very tight (see **Fig. 1B**), it is possible that the longer Glutamic acid side chain (which has an extra CH₂ group compared to Asp or Asn) may introduce a slight distortion at the contact sites, an effect that is propagated by the icosahedral T=4 symmetry of the virion. Thus, a cooperative effect due to this change at position Asp 284 may play a role in either allowing a less efficient assembly of new particles in infected cells, or a more efficient particle disassembly process during invasion of a new cell, or a combination of both. This information can guide new site-directed mutagenesis studies, using reverse genetics, to test the effect of the Asp/Glu replacement on the virus cycle.

The magnitude and high political profile of the outbreak has underlined the critical lack of knowledge on the biology of chikungunya virus, contrasting with related model alphaviruses such as Sindbis, Semliki Forest and Ross River. This situation reflects the fact that chikungunya infection, despite infecting millions of people since its discovery, has been neglected. However, chikungunya disease is clearly responsible for invalidating and persistent arthralgia, although it remains unresolved whether the symptoms are due to persistence of the virus or inappropriate immune response [38,39]. In addition, during the Indian Ocean CHIKV outbreak, a small proportion of the patients (about 70 out of 200,000 infected) developed severe clinical signs such as meningo-encephalitis or hepatitis. Whereas neurovirulence and neuroinvasiveness are established for several alphaviruses such as EEV and VEEV, only two CHIKV strains had previously been isolated from children with clinical signs suggestive of encephalitis and meningitis [40,41]. Whether the CHIKV Indian Ocean strain has acquired a higher neurovirulence or neuroinvasiveness certainly deserves investigation.

DEMANDES OU BREVETS VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.

CECI EST LE TOME _____ DE _____

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.

THIS IS VOLUME 1 OF 2

NOTE: For additional volumes please contact the Canadian Patent Office.

CLAIMS

1. An isolated and purified wild strain of chikungunya virus (CHIKV) capable
5 of *in vitro* infecting human cells.
2. The strain according to claim 1, selected from the group consisting of the
isolates 05.115, 05.61, 05.209, 06.21, 06.27 and 06.49.
- 10 3. The strain according to claim 1 or 2, wherein its genome comprises a
sequence as shown in Figure 4.
4. The strain according to claim 1 or 2, wherein its genome comprises a
sequence as shown in Figure 5.
- 15 5. The strain according to claim 1 or 2, wherein its genome comprises a
sequence as shown in Figure 6.
6. The strain according to claim 1 or 2, wherein its genome comprises a
20 sequence as shown in Figure 7.
7. The strain according to claim 1 or 2, wherein its genome comprises a
sequence as shown in Figure 8.
- 25 8. The strain according to claim 1 or 2, wherein its genome comprises a
sequence as shown in Figure 9.
9. An isolated and purified strain of chikungunya virus (CHIKV) comprising at
least one mutation in structural protein E1 and/or structural protein E2.
- 30 10. The strain of claim 9, wherein the at least one mutation is located in the
ectodomain of protein E1 and/or E2.

11. The strain of claim 9 or 10, wherein the at least one mutation in the E2 protein is selected from the group consisting of:

I23T; G57K; I74M; G79E; N160T; A164T; L181M; I211T; S299N; T312M; A344T;

5 S375T; V386A as shown in Table S3.

12. The strain of claim 9 or 10, wherein the at least one mutation in the E1 protein is selected from the group consisting of: A226V, M269V, D284E and V322A as shown in Table S3.

10

13. The strain of claim 12, wherein the mutation is A226V.

14. An isolated and purified polynucleotide comprising a sequence as shown in Figures 4, 5, 6, 7, 8 or 9.

15

15. Fragment of the polynucleotide as defined in claim 14, wherein it is suited to be obtained by:

- use of restriction enzymes wherein their cleavage sites are present in said polynucleotide;
- 20 - amplification with specific primers for said polynucleotide;
- *in vitro* transcription; or
- chemical synthesis.

25

16. The fragment of claim 15, wherein it codes for the ectodomain of glycoprotein E2.

17. The fragment of claim 16, wherein it comprises a nucleotide sequence as shown in Figures 10, 11, 12 or 13.

30

18. The fragment of claim 15, wherein it codes for the ectodomain of glycoprotein E1.

19. The fragment of claim 16, wherein it consists of a nucleotide sequence as shown in Figures 10, 11, 12 or 13.

20. A vector comprising a fragment as defined in any one of claims 15 to 19.

5

21. The vector according to claim 20, selected from the vector deposited at the CNCM on March 15, 2006 under accession number I-3587, I-3588, I-3589 and I-3590.

10 22. A recombinant cell comprising a vector as defined in claim 20 or 21.

23. An isolated and purified polypeptide encoded by a polynucleotide as defined in claim 14, or by a fragment as defined in any one of claims 15 to 19.

15 24. Monoclonal or polyclonal antibodies, or fragment thereof, that specifically bind to a polypeptide as defined in claim 23.

25. Use of an element selected from the group consisting of:

- a strain as defined in any one of claims 1 to 13;
- 20 - a polynucleotide as defined in claim 14;
- a fragment as defined in any one of claims 15 to 19;
- a vector as defined in claim 20 or 21;
- a recombinant cell as defined in claim 22;
- a polypeptide as defined in claim 23; and
- 25 - an antibody as defined in claim 24;

for the detection of a CHIKV associated to an arbovirosis.

26. Use of an element selected from the group consisting of:

- a strain as defined in any one of claims 1 to 13;
- 30 - a polynucleotide as defined in claim 14;
- a fragment as defined in any one of claims 15 to 19;
- a vector as defined in claim 20 or 21;

- a recombinant cell as defined in claim 22;
- a polypeptide as defined in claim 23; and
- an antibody as defined in claim 24;

for the preparation of a composition that prevents and/or treats an arbovirosis.

5

27. A composition comprising at least one element selected from the group consisting of:

- a strain as defined in any one of claims 1 to 13;
- a polynucleotide as defined in claim 14;
- 10 - a fragment as defined in any one of claims 15 to 19;
- a vector as defined in claim 20 or 21;
- a recombinant cell as defined in claim 22;
- a polypeptide as defined in claim 23; and
- an antibody as defined in claim 24;

15 for treating and/or preventing an arbovirosis.

28. A kit for the detection of a CHIKV associated to an arbovirosis, comprising at least one element selected from the group consisting of:

- a strain as defined in any one of claims 1 to 13;
- 20 - a polynucleotide as defined in claim 14;
- a fragment as defined in any one of claims 15 to 19;
- a vector as defined in claim 20 or 21;
- a recombinant cell as defined in claim 22;
- a polypeptide as defined in claim 23; and
- 25 - an antibody as defined in claim 24.

Application number/numéro de demande: 2545597

Figures: 1, 2, 3, SIA, S2

Pages: _____

DRW-IP

Unscannable items
received with this application
(Request original documents in File Prep. Section on the 10th Floor)

Documents reçus avec cette demande ne pouvant être balayés
(Commander les documents originaux dans la section de préparation des dossiers au
10ième étage)

FIGURE 4

caaagcaaga	gattaataaac	ccatcatgga	tccctgtgtac	gtggacatag	acgctgacag	60
cgcctttttg	aaggccctgc	aacgtgcgta	ccccatgttt	gaggtggaac	caaggcaggt	120
cacaccgaat	gaccatgcta	atgctagagc	gttctcgcac	ctagctataa	aactaataga	180
gcaggaaatt	gaccccgact	caaccatcct	ggatatcggc	agtgcgccag	caaggaggat	240
gatgtcggac	aggaagtacc	actgcgtctg	cccgatgcgc	agtgcggaag	atcccgagag	300
actcgccaat	tatgcgagaa	agctagcatc	tgccgcagga	aaagtcctgg	acagaaacat	360
ctctggaaag	atcggggact	tacaagcagt	aatggccgtg	ccagacacgg	agacgccaac	420
attctgctta	cacacagacg	tctcatgtag	acagagagca	gacgtcgcta	tataccaaga	480
cgtctatgct	gtacacgcac	ccacgtcgct	ataccaccag	gcgattaaag	gggtccgagt	540
ggcgtactgg	gttgggttcg	acacaacccc	gttcatgtac	aatgccatgg	cgggtgccta	600
cccctcatac	tcgacaaact	gggcagatga	gcagggtactg	aaggctaaga	acataggatt	660
atgttcaaca	gacctgacgg	aaggtagacg	aggcaagttg	tctattatga	gagggaaaaa	720
gctaaaaccg	tgcgaccgtg	tgctgttctc	agtagggtea	acgctctacc	cggaaagccg	780
caagctactt	aagagctggc	acctgccatc	ggtgttccat	ttaaagggca	aactcagctt	840
cacatgccgc	tgtgatacag	tgggttccgtg	tgagggtctac	gtcgttaaga	gaataacgat	900
gagcccaggc	ctttatggaa	aaaccacagg	gtatgcggta	accaccacag	cagacggatt	960
cctgatgtgc	aagactaccg	acacggttga	cggcgaaaga	gtgtcattct	cgggtgtgcac	1020
atacgtgccg	gcgaccattt	gtgatcaaat	gaccggcatc	cttgctacag	aagtcacgcc	1080
ggaggatgca	cagaagctgt	tgggtggggct	gaaccagaga	atagtggtta	acggcagaaac	1140
gcaacggaat	acgaacacca	tgaaaaatta	tctgcttccc	gtggtcgccc	aagccttcag	1200
taagtgggca	aaggagtgcc	ggaaagacat	ggaagatgaa	aaactcctgg	gggtcagaga	1260
aagaacactg	acctgctgct	gtctatgggc	attcaagaag	cagaaaacac	acacggtcta	1320
caagaggcct	gatacccagt	caattcagaa	ggttcaggcc	gagtttgaca	gcttttgtggt	1380
accgagtctg	tggtcgtccg	ggttgtcaat	ccctttgagg	actagaatca	aatgggttgtt	1440
aagcaagggtg	ccaaaaaccg	acctgatccc	atacagcgga	gacgcccagag	aagcccggga	1500
cgcagaaaaa	gaagcagagg	aagaacgaga	agcagaactg	actcgcgaag	ccctaccacc	1560
tctacaggca	gcacaggaag	atgttcagggt	cgaaatcgac	gtggaacagc	ttgaggacag	1620
agcggggcgca	ggaataatag	agactccgag	aggagctatc	aaagtactg	cccaaccaac	1680
agaccacgtc	gtgggagagt	acctgggtact	ctccccgcag	accgtactac	gtagccagaa	1740
gctcagtctg	attcacgctt	tggcggagca	agtgaagacg	tgcacgcaca	acggacgagc	1800
agggaggtat	gcggtcgaag	cgtacgacgg	ccgagtccta	gtgccctcag	gctatgcaat	1860
ctcgcctgaa	gacttccaga	gtctaagcga	aagcgcacag	atgggtgtata	acgaaagaga	1920
gttcgtaaac	agaaagctac	accatattgc	gatgcacgga	ccagccctga	acaccgacga	1980
agagtcgtat	gagctggtga	gggcagagag	gacagaacac	gagtacgtct	acgacgtgga	2040
tcagagaaga	tgctgtaaga	aggaagaagc	cgcaggactg	gtactggtgg	gcgacttgac	2100
taatccgccc	taccacgaat	tcgcatatga	agggtctaaa	atccgcccctg	cctgcccata	2160
caaaattgca	gtcataggag	tcttcggagt	accgggatct	ggcaagtcag	ctattatcaa	2220
gaacctagtt	accaggcagg	acctggtgac	tagcggaaag	aaagaaaact	gccaagaaat	2280
caccaccgac	gtgatgagac	agagaggtct	agagatatct	gcacgtacgg	ttgactcgct	2340
gctcttgaat	ggatgcaaca	gaccagtcca	cgtgttgtac	gtagacgagg	cgtttgcgtg	2400
ccactctgga	acgctacttg	ctttgatcgc	cttgggtgaga	ccaaggcaga	aagttgtact	2460
ttgtggtgac	ccgaagcagt	gcggcttctt	caatatgatg	cagatgaaag	tcaactataa	2520
tcacaacatc	tgcacccaag	tgtaccacaa	aagtatctcc	aggcgggtgta	cactgcctgt	2580
gaccgccatt	gtgtcatcgt	tgcattacga	aggcaaaatg	cgcactacga	atgagtacaa	2640
caagccgatt	gtagtggaca	ctacaggctc	aacaaaacct	gaccctggag	acctcgtgtt	2700
aacgtgcttc	agaggggtggg	ttaaacaact	gcaaattgac	tatcgtggat	acgaggtcat	2760
gacagcagcc	gcatecccaag	ggttaaccag	aaaaggagtt	tacgcagtta	gacaaaaagt	2820
taatgaaaac	ccgctctatg	catcaacgtc	agagcacgtc	aacgtactcc	taacgcgtac	2880
ggaaggtaaa	ctgggtatgga	agacactttc	cggcgacccg	tggataaaga	cgctgcagaa	2940

FIGURE 4 (con't)

cccaccgaaa	ggaaacttca	aagcaactat	taaggagtgg	gaggtggagc	atgcatcaat	3000
aatggcgggc	atctgcagtc	accaaatagac	cttcgatata	ttccaaaata	aagccaacgt	3060
ttgttgggct	aagagcttgg	tccttatcct	cgaaacagcg	gggataaaac	taaatgatag	3120
gcagtgggtct	cagataattc	aagccttcaa	agaagacaaa	gcatactcac	ctgaagtagc	3180
cctgaatgaa	atatgtacgc	gcattgtatgg	ggtggatcta	gacagcgggc	tatttttctaa	3240
accgttgggtg	tctgtgtatt	acgcggataa	ccactgggat	aataggcctg	gagggaaaat	3300
gttcggattt	aaccccgagg	cagcatccat	tctagaaaga	aagtatccat	tcacaaaagg	3360
gaagtggaaac	atcaacaagc	agatctgcgt	gactaccagg	aggatagaag	actttaaccc	3420
taccaccaac	atcataccgg	ccaacaggag	actaccacac	tcattagtgg	ccgaacaccg	3480
cccagtaaaa	ggggaaagaa	tggaaatggct	ggttaacaag	ataaacggcc	accacgtgct	3540
cctggtcagt	ggctataacc	ttgcactgcc	tactaagaga	gtcacttggg	tagcgccgtt	3600
aggtgtccgc	ggagcggact	acacatacaa	cctagagtgg	ggtctgccag	caacgcttgg	3660
taggtatgac	ctagtgggtca	taaacatcca	cacacctttt	cgcatacacc	attaccaaca	3720
gtgcgtcgac	cacgcaatga	aactgcaaat	gctcgggggt	gactcattga	gactgctcaa	3780
accgggcggc	tctctattga	tcagagcata	tgggtacgca	gatagaacca	gtgaacgagt	3840
catctgcgta	ttgggacgca	agtttagatc	gtctagagcg	ttgaaaccac	catgtgtcac	3900
cagcaacact	gagatgtttt	tcctattcag	caactttgac	aatggcagaa	ggaatttcac	3960
aactcatgtc	atgaacaatc	aactgaatgc	agccttcgta	ggacaggtca	cccagacagg	4020
atgtgcaccg	tcgtaccggg	taaaacgcat	ggacatcgcg	aagaacgatg	aagagtgcgt	4080
agtcaacgcc	gctaaccctc	gcgggttacc	gggtgacggg	gtttgcaagg	cagtatacaa	4140
aaaatggccg	gagtccttta	agaacagtgc	aacaccagtg	ggaaccgcaa	aaacagttat	4200
gtgcggtacg	tatccagtaa	tcacgcgtgt	tggaccaaac	ttctctaatt	attcggagtc	4260
tgaaggggac	cgggaattgg	cagctgccta	tcgagaagtc	gcaaaggaag	taactagggt	4320
gggagtaaat	agtgtagcta	tacctctcct	ctccacaggt	gtatactcag	gagggaaaga	4380
caggctgacc	cagtcactga	accacctctt	tacagccatg	gactcgacgg	atgcagacgt	4440
ggtcatctac	tgccgcgaca	aagaatggga	gaagaaaata	tctgaggcca	tacagatgcg	4500
gacccaagta	gagctgctgg	atgagcacat	ctccatagac	tgcgatattg	ttcgcgtgca	4560
ccctgacagc	agcttggcag	gcagaaaagg	atacagcacc	acggaaggcg	cactgtactc	4620
atatctagaa	gggacccgtt	ttcatcagac	ggctgtggat	atggcggaga	tacatactat	4680
gtggccaaag	caaacagagg	ccaatgagca	agtctgccta	tatgccctgg	gggaaagtat	4740
tgaatcgatc	aggcagaaat	gcccgggtgga	tgatgcagac	gcatcatctc	ccccaaaac	4800
tgtcccggtc	ctttgcccgtt	acgctatgac	tcagaaacgc	gtcaccgggc	ttcgcgatgaa	4860
ccacgtcaca	agcataattg	tgtgtttcttc	gtttcccttc	ccaaagtaca	aaatagaagg	4920
agtgcaaaaa	gtcaaatgct	ctaaggtaat	gctatttgac	cacaacgtgc	catcgcgct	4980
aagtccaagg	gaatatagat	cttcccagga	gtctgcacag	gaggcgagta	caatcacgtc	5040
actgacgcat	agtcaattcg	acctaagcgt	tgatggcgag	atactgcccg	tcccgtcaga	5100
cctggatgct	gacgccccag	ccctagaacc	agcactagac	gacggggcga	cacacacgct	5160
gccatccaca	accggaaacc	ttgcggccgt	gtctgattgg	gtaatgagca	ccgtacctgt	5220
cgcgcgcgcc	agaagaaggc	gagggagaaa	cctgactgtg	acatgtgacg	agagagaagg	5280
gaatataaca	cccatggcta	gcgtccgatt	ctttagggca	gagctgtgtc	cggtcgtaca	5340
agaaacagcg	gagacgcgtg	acacagcaat	gtctcttcag	gcaccaccga	gtaccgccac	5400
ggaaccgaat	catccgccga	tctccttcgg	agcatcaagc	gagacgttcc	ccattacatt	5460
tggggacttc	aacgaaggag	aaatcgaaaag	cttgtcttct	gagctactaa	ctttcggaga	5520
cttcttacca	ggagaagtgg	atgacttgac	agacagcgac	tgggtccacgt	gctcagacac	5580
ggacgacgag	ttatgactag	acagggcagg	tgggtatata	ttctcgtcgg	acaccggtcc	5640
aggtcattta	caacagaagt	cagtacgcca	gtcagtgtg	ccggtgaaca	ccctggagga	5700
agtccacgag	gagaagtgtt	acccacctaa	gctggatgaa	gcaaaggagc	aactattact	5760
taagaaactc	caggagagtg	catccatggc	caacagaagc	aggtatcagt	cgcgcaaagt	5820
agaaaacatg	aaagcagcaa	tcatccagag	actaaagaga	ggctgtagac	tatacttaat	5880

FIGURE 4 (con't)

gtcagagacc	ccaaaagtcc	ctacttaccg	gactacatat	ccggcgccctg	tgtactcgcc	5940
tccgatcaac	gtccgattgt	ccaatcccga	gtccgcagtg	gcagcatgca	atgagttctt	6000
agctagaaac	tatccaactg	tctcatcata	ccaaattacc	gacgagtatg	atgcatatct	6060
agacatgggtg	gacgggtcgg	agagttgcct	ggaccgagcg	acattcaatc	cgtcaaaact	6120
caggagctac	ccgaaacagc	acgcttacca	cgcgcctcc	atcagaagcg	ctgtaccgtc	6180
cccattccag	aacacactac	agaatgtact	ggcagcagcc	acgaaaagaa	actgcaacgt	6240
cacacagatg	aggggaattac	ccactttgga	ctcagcagta	ttcaacgtgg	agtgtttcaa	6300
aaaattcgca	tgcaaccaag	aatactggga	agaatttgct	gccagcccta	ttaggataac	6360
aactgagaat	ttagcaacct	atgttactaa	actaaaaggg	ccaaaagcag	cagcgctatt	6420
cgcaaaaacc	cataatctac	tgccactaca	ggaagtacca	atggataggt	tcacagtaga	6480
tatgaaaagg	gacgtgaagg	tgactcctgg	tacaaagcat	acagaggaaa	gacctaaagg	6540
gcaggttata	caggcggctg	aacccttggc	gacagcatac	ctatgtggga	ttcacagaga	6600
gctgggttagg	aggctgaacg	ccgtcctcct	acccaatgta	catacactat	ttgacatgtc	6660
tgccgaggat	ttcgatgcca	tcatagccgc	acactttaag	ccaggagaca	ctgttttgga	6720
aacggacata	gcctcctttg	ataagagcca	agatgattca	cttgcgctta	ctgctttgat	6780
gctgttagag	gatttagggg	tggatcactc	cctgctggac	ttgatagagg	ctgctttcgg	6840
agagatttcc	agctgtcacc	taccgacagg	tacgcgcttc	aagttcggcg	ccatgatgaa	6900
atcaggatatg	ttcctaactc	tgttcgtcaa	cacattgtta	aacatcacca	tcgccagccg	6960
agtgctggaa	gacgtctga	caaaatccgc	gtgcgcggcc	ttcatcggcg	acgacaacat	7020
aatacatgga	gtcgtctccg	atgaattgat	ggcagccaga	tgtgccactt	ggatgaacat	7080
ggaagtgaag	atcatagatg	cagttgtatc	cttgaaagcc	ccttactttt	gtggagggtt	7140
tatactgcac	gatactgtga	caggaacagc	ttgcagagtg	gcagaccgcg	taaaaaggct	7200
ttttaaaactg	ggcaaaccgc	tagcggcagg	tgacgaacaa	gatgaagata	gaagacgagc	7260
gctggctgac	gaagtgatca	gatggcaacg	aacagggcta	attgatgagc	tggagaaagc	7320
ggtatactct	aggtacgaag	tgcaggggat	atcagttgtg	gtaatgtcca	tggccacctt	7380
tgcaagctcc	agatccaact	tcgagaagct	cagaggaccc	gtcataactt	tgtacggcgg	7440
tcctaaatag	gtacgcacta	cagctaccta	ttttgcagaa	gccgacagca	agtatctaaa	7500
cactaatcag	ctacaatgga	gttcatccca	acccaaactt	tttacaatag	gaggtaccag	7560
cctcgaccct	ggactccgcg	ccctactatc	caagtcatca	ggcccagacc	gcgcccctcag	7620
aggcaagctg	ggcaacttgc	ccagctgatc	tcagcagtta	ataaactgac	aatgcgcgcg	7680
gtaccccaac	agaagccacg	caggaatcgg	agaataaga	agcaaaagca	aaaacaacag	7740
gcgccacaaa	acaacacaaa	tcaaaagaag	cagccaccta	aaaagaaacc	ggctcaaaag	7800
aaaaagaagc	cgggccgcag	agagaggatg	tgcataaaaa	tcgaaaatga	ttgtattttc	7860
gaagtcaagc	acgaaggtaa	ggtaacaggt	tacgcgtgcc	tgggtggggga	caaagtaatg	7920
aaaccagcac	acgtaaaggg	gaccatcgat	aacgcggacc	tggccaaact	ggcctttaag	7980
cggtcactta	agtatgacct	tgaatgcgcg	cagatacccg	tgccacatgaa	gtccgacgct	8040
tcgaagttca	cccatgagaa	accggagggg	tactacaact	ggcaccacgg	agcagtacag	8100
tactcaggag	gccggttcac	catccctaca	ggtgctggca	aaccagggga	cagcggcaga	8160
ccgatcttcg	acaacaaggg	acgcgtgggt	gccatagtct	taggaggagc	taatgaagga	8220
gcccgtacag	ccctctcggt	ggtgacctgg	aataaagaca	ttgtcactaa	aatcaccccc	8280
gagggggccg	aagagtggag	tcttgccatc	ccagttatgt	gcctgttggc	aaacaccacg	8340
ttcccctgct	cccagcccc	ttgcacgccc	tgctgctacg	aaaaggaacc	ggaggaaacc	8400
ctacgcatgc	ttgaggacaa	cgtcatgaga	cctgggtact	atcagctgct	acaagcatcc	8460
ttaacatggt	ctccccaccg	ccagcgacgc	agcaccaagg	acaacttcaa	tgtctataaa	8520
gccacaagac	catacttagc	tactgtccc	gactgtggag	aagggcactc	gtgccatagt	8580
cccgtagcac	tagaacgcac	cagaaatgaa	gcgacagacg	ggacgctgaa	aatccaggtc	8640
tccttgcaaa	tcggaataaa	gacggatgac	agccacgatt	ggaccaagct	gcgttatatg	8700
gacaaccaca	tgccagcaga	cgcagagagg	gcggggctat	ttgtaagaac	atcagcaccg	8760
tgtacgatta	ctggaacaat	gggacacttc	atcctggccc	gatgtccaaa	aggggaaact	8820

FIGURE 4 (con't)

ctgacggtgg	gattcactga	cagtaggaag	attagtcact	catgtacgca	cccatttcac	8880
cacgaccctc	ctgtgatagg	tcgggaaaaa	ttccattccc	gaccgcagca	cggtaaagag	8940
ctaccttgca	gcacgtacgt	gcagagcacc	gccgcaacta	ccgaggagat	agaggtaacac	9000
atgccccag	acaccctga	tcgcacatta	atgtcacaac	agtccggcaa	cgtaaagatc	9060
acagtcaatg	gccagacggt	gcggtacaag	tgtaatgctg	gtggctcaaa	tgaaggacta	9120
acaactacag	acaaagtgat	taataactgc	aagggttgatc	aatgtcatgc	cgcggtcacc	9180
aatcacaaaa	agtggcagta	taactccctt	ctgggtcccg	gtaatgctga	acttggggac	9240
cgaaaaggaa	aaattcacat	cccgtttccg	ctggcaaatg	taacatgcag	ggtgcctaaa	9300
gcaaggaacc	ccaccgtgac	gtacgggaaa	aaccaagtca	tcattgctact	gtatcctgac	9360
caccaaacac	tcctgtccta	ccggaatatg	ggagaagaac	caaactatca	agaagagtgg	9420
gtgatgcata	agaaggaagt	cgtgctaacc	gtgccgactg	aagggtctga	ggtcacgtgg	9480
ggcaacaacg	agccgtataa	gtattggccg	cagttatcta	caaacggtac	agcccatggc	9540
caccgcgatg	agataattct	gtattattat	gagctgtacc	ccactatgac	tgtagtagtt	9600
gtgtcagtgg	ccacgttcat	actcctgtcg	atgggtgggt	tggcagcggg	gatgtgcatg	9660
tgtgcacgac	gcagatgcat	cacaccgtat	gaactgacac	caggagctac	cgcccccttc	9720
ctgcttagcc	taatattgctg	catcagaaca	gctaaagcgg	ccacatacca	agaggctgcg	9780
atatacctgt	ggaacgagca	gcaacctttg	ttttggctac	aagcccttat	tcgctggca	9840
gccctgattg	ttctatgcaa	ctgtctgaga	ctcttaccat	gctgctgtaa	aacgttggct	9900
tttttagccg	taatgagcgt	cgggtgccac	actgtgagcg	cgtacgaaca	cgtaacagtg	9960
atcccgaaca	cgggtgggagt	accgtataag	actctagtca	atagacctgg	ctacagcccc	10020
atgggtattg	agatggaact	actgtcagtc	actttggagc	caacactatc	gcttgattac	10080
atcacgtgcg	agtacaaaac	cgtcatcccg	tctccgtacg	tgaagtgctg	cggtacagca	10140
gagtgcgaagg	acaaaaacct	acctgactac	agctgtaagg	tcttcaccgg	cgtctacca	10200
tttatgtggg	gcgggcgcta	ctgcttctgc	gacgctgaaa	acacgcagtt	gagcgaagca	10260
cacgtggaga	agtccgaatc	atgcaaaaaca	gaatttgcat	cagcatacag	ggctcatacc	10320
gcatctgcat	cagctaagct	ccgcgtcctt	taccaaggaa	ataacatcac	tgtaaactgcc	10380
tatgcaaacg	gcgaccatgc	cgtcacagtt	aaggacgcca	aattcattgt	ggggccaatg	10440
tcttcagcct	ggacaccttt	cgacaacaaa	attgtggtgt	acaaagggtga	cgtctataac	10500
atggactacc	cgcccttttg	cgcaggaaga	ccaggacaat	ttggcgatat	ccaaagtcgc	10560
acacctgaga	gtaaagacgt	ctatgctaata	acacaactgg	tactgcagag	accggctgcg	10620
ggtacggtac	acgtgccata	ctctcaggca	ccatctggct	ttaagtattg	gctaaaagaa	10680
cgcggggctg	cgctgcagca	cacagcacca	tttggctgcc	aaatagcaac	aaacccggta	10740
agagcggtga	actgcgcgct	agggaaacatg	cccatctcca	tcgacatacc	ggaagcggcc	10800
ttcactaggg	tcgtcgacgc	gccctcttta	acggacatgt	cgtgcgaggt	accagcctgc	10860
accatttctt	cagacttttg	gggcgtcgcc	attattaaat	atgcagccag	caagaaaggc	10920
aagtgtgctg	tgcattcgat	gactaacgcc	gtcactattc	gggaagctga	gatagaagtt	10980
gaagggaatt	ctcagctgca	aatctctttc	tcgacggcct	tagccagcgc	cgaattccgc	11040
gtacaagtct	gttctacaca	agtacactgt	gcagccgagt	gccaccccc	gaaggaccac	11100
atagtcaact	acccggcgct	acataccacc	ctcgggggtc	aggacatctc	cgctacggcg	11160
atgtcatggg	tgcagaagat	cacgggaggt	gtgggactgg	ttgttgctgt	tgccgcactg	11220
attctaactg	tggtgctatg	cgtgtcgttc	agcaggcact	aacttgacaa	ttaagtatga	11280
aggtatatgt	gtcccctaag	agacacactg	tacatagcaa	ataatctata	gatcaaaggg	11340
ctacgcaacc	cctgaatagt	aacaaaatac	aaaatcacta	aaaattataa	aaacagaaaa	11400
atacataaat	aggtatacgt	gtcccctaag	agacacattg	tatgtagggt	ataagtatag	11460
atcaaagggc	cgaataaccc	ctgaatagta	acaaaatatg	aaaatcaata	aaaatcataa	11520
aatagaaaaa	ccataaacag	aagtagttca	aagggtcata	aaaccctga	atagtaacaa	11580
aacataaaat	taataaaaaat	c				11601

FIGURE 5

caaagcaaga	gattaataac	ccatcatgga	tcctgtgtac	gtggacatag	acgctgacag	60
cgcctttttg	aaggccctgc	aacgtgcgta	cccatgtttt	gaggtggaac	caaggcaggt	120
cacaccgaat	gaccatgcta	atgctagagc	gttctcgcat	ctagctataa	aactaataga	180
gcaggaaatt	gaccccgact	caaccatcct	ggatatcggc	agtgcgccag	caaggaggat	240
gatgtcggac	aggaagtacc	actgcgtctg	cccgatgcgc	agtgcggaag	atcccgagag	300
actcgccaat	tatgcgagaa	agctagcatc	tgccgcagga	aaagtcctgg	acagaaacat	360
ctctggaaag	atcggggact	tacaagcagt	aatggccgtg	ccagacacgg	agacgccaac	420
attctgctta	cacacagacg	tctcatgtag	acagagagca	gacgtcgcta	tataccaaga	480
cgtctatgct	gtacacgcac	ccacgtcgct	ataccaccag	gcgattaaag	gggtccgagt	540
ggcgtactgg	gttgggttcg	acacaacccc	gttcatgtac	aatgccatgg	cgggtgccta	600
cccctcatac	tcgacaaact	gggcagatga	gcagggtactg	aaggctaaga	acataggatt	660
atgttcaaca	gacctgacgg	aaggtagacg	aggcaagtgt	tctattatga	gagggaaaaa	720
gctaaaaccg	tgcgaccgtg	tgctgtttctc	agtaggggtca	acgctctacc	cggaaagccg	780
caagctactt	aagagctggc	acctgccatc	ggtgttccat	ttaaagggca	aactcagctt	840
cacatgccgc	tgtgatacag	tgggttctgtg	tgaggggtac	gtcgttaaga	gaataacgat	900
gagcccaggc	ctttatggaa	aaaccacagg	gtatgcggta	accaccacg	cagacggatt	960
cctgatgtgc	aagactaccg	acacgggttg	cggcgaaaga	gtgtcattct	cgggtgtgcac	1020
atacgtgccg	gcgaccattt	gtgatcaa	gaccggcatc	cttgctacag	aagtcacgcc	1080
ggaggatgca	cagaagctgt	tgggtggggt	gaaccagaga	atagtgggtta	acggcagaa	1140
gcaacggaat	acgaacacca	tgaaaaatta	tctgcttccc	gtgggtcgccc	aagccttcag	1200
taagtgggca	aaggagtgcc	ggaaagacat	ggaagatgaa	aaactcctgg	gggtcagaga	1260
aagaacactg	acctgctgct	gtctatgggc	attcaagaag	cagaaaacac	acacgggtcta	1320
caagaggcct	gataccagct	caattcagaa	ggttcaggcc	gagtttgaca	gctttgtggt	1380
accgagtctg	tggtcgtccg	ggttgtcaat	ccctttgagg	actagaatca	aatggttgtt	1440
aagcaagggtg	ccaaaaaccg	acctgatccc	atacagcgga	gacgcccag	aagcccggga	1500
cgcagaaaaa	gaagcagagg	aagaacgaga	agcagaactg	actcgcgaa	ccctaccacc	1560
tctacaggca	gcacaggaag	atgttcagggt	cgaaatcgac	gtggaacagc	ttgaggacag	1620
agcgggcgca	ggaataatag	agactccgag	aggagctatc	aaagttactg	cccaaccaac	1680
agaccacgtc	gtgggagagt	acctgggtact	ctccccgcag	accgtactac	gtagccagaa	1740
gctcagtctg	attcacgctt	tggcggagca	agtgaagacg	tgacgcaca	acggacgagc	1800
agggagggtat	gcggtcgaag	cgtacgacgg	ccgagtccta	gtgccctcag	gctatgcaat	1860
ctcgctgaa	gacttccaga	gtctaagcga	aagcgcaacg	atgggtgtata	acgaaagaga	1920
gttcgtaaac	agaaagctac	accatattgc	gatgcacgga	ccagccctga	acaccgacga	1980
agagtcgtat	gagctgggtga	gggcagagag	gacagaacac	gagtacgtct	acgacgtgga	2040
tcagagaaga	tgctgtaaga	aggaagaagc	cgcaggactg	gtactgggtg	gcgacttgac	2100
taatccgccc	taccacgaat	tcgcatatga	agggtctaaa	atccgcccctg	cctgcccata	2160
caaaattgca	gtcataggag	tcttcggagt	accgggatct	ggcaagtcag	ctattatcaa	2220
gaacctagtt	accaggcagg	acctgggtgac	tagcggaaag	aaagaaaact	gccaagaaat	2280
caccaccgac	gtgatgagac	agagagggtct	agagatatct	gcacgtacgg	ttgactcgct	2340
gctcttgaat	ggatgcaaca	gaccagtcga	cgtgttgtac	gtagacgagg	cgtttgcgtg	2400
ccactctgga	acgtacttg	ctttgatcgc	cttgggtgaga	ccaaggcaga	aagttgtact	2460
ttgtgggtgac	ccgaagcagt	gcggcttctt	caatatgatg	cagatgaaag	tcaactataa	2520
tcacaacatc	tgcacccaag	tgtaccacaa	aagtatctcc	aggcgggtgta	cactgcctgt	2580
gaccgccatt	gtgtcatcgt	tgcatcaca	aggcaaaatg	cgcactacga	atgagtacaa	2640
caagccgatt	gtagtggaca	ctacaggctc	aacaaaacct	gaccctggag	acctcgtgtt	2700
aacgtgcttc	agagggtggg	ttaaacaact	gcaaattgac	tatcgtggat	acgaggtcat	2760
gacagcagcc	gcacccaag	ggttaaccag	aaaaggagtt	tacgcagtta	gacaaaaagt	2820
taatgaaaac	ccgctctatg	catcaacgtc	agagcacgtc	aacgtactcc	taacgcgtac	2880

FIGURE 5 (con't)

ggaaggtaaa	ctggtatgga	agacactttc	cggcgacccg	tggataaaga	cgctgcagaa	2940
cccaccgaaa	ggaaacttca	aagcaactat	taaggagtgg	gaggtggagc	atgcatcaat	3000
aatggcgggc	atctgcagtc	accaaataac	cttcgatata	ttccaaaata	aagccaacgt	3060
ttgttgggct	aagagcttgg	tccctatcct	cgaaacagcg	gggataaaaac	taaatgatag	3120
gcagtgggtct	cagataattc	aagccttcaa	agaagacaaa	gcatactcac	ctgaagtagc	3180
cctgaatgaa	atatgtacgc	gcatgtatgg	ggtggatcta	gacagcgggc	tattttctaa	3240
accgttgggtg	tctgtgtatt	acgcggataa	ccactgggat	aataggcctg	gagggaaaat	3300
gttcggattt	aaccccgagg	cagcatccat	tctagaaaga	aagtatccat	tcacaaaagg	3360
gaagtggaaac	atcaacaagc	agatctgcgt	gactaccagg	aggatagaag	actttaaccc	3420
taccaccaac	atcataccgg	ccaacaggag	actaccacac	tcattagtgg	ccgaacaccg	3480
cccagtaaaa	ggggaaagaa	tggaaatggct	ggttaacaag	ataaacggcc	accacgtgct	3540
cctggtcagt	ggctataacc	ttgcaactgcc	tactaagaga	gtcacttggg	tagcgccgtt	3600
aggtgtccgc	ggagcggact	acacatacaa	cctagagtgt	ggtctgccag	caacgcttgg	3660
taggtatgac	ctagtgggtca	taaacatcca	cacacctttt	cgcatacacc	attaccaaca	3720
gtgcgtcgac	cacgcaatga	aactgcaaat	gctcgggggt	gactcattga	gactgctcaa	3780
accgggcggc	tctctattga	tcagagcata	tggttacgca	gatagaacca	gtgaacgagt	3840
catctgcgta	ttgggacgca	agtttagatc	gtctagagcg	ttgaaaccac	catgtgtcac	3900
cagcaacact	gagatgtttt	tcctattcag	caactttgac	aatggcagaa	ggaatttcac	3960
aactcatgtc	atgaacaatc	aactgaatgc	agccttcgta	ggacagggtca	cccagcagg	4020
atgtgcaccg	tcgtaccggg	taaaacgcat	ggacatcgcg	agaacgatg	aagagtgcgt	4080
agtcaacgcc	gctaaccctc	gcgggttacc	gggtgacggt	gtttgcaagg	cagtatacaa	4140
aaaatggccg	gagtccttta	agaacagtgc	aacaccagtg	ggaaccgcaa	aaacagttat	4200
gtgcggtacg	tatccagtaa	tcacgcgtgt	tggaccaaac	ttctctaatt	attcggagtc	4260
tgaaggggac	cgggaattgg	cagctgccta	tcgagaagtc	gcaaaggaag	taactaggct	4320
gggagtaaat	agtgtagcta	tacctctcct	ctccacaggt	gtatactcag	gagggaaaga	4380
caggctgacc	cagtcactga	accacctctt	tacagccatg	gactcgacgg	atgcagacgt	4440
ggtcactctac	tgccgcgaca	aagaatggga	gaagaaaata	tctgaggcca	tacagatgcg	4500
gacccaagta	gagctgctgg	atgagcacat	ctccatagac	tgcgatattg	ttcgcgtgca	4560
ccctgacagc	agcttggcag	gcagaaaagg	atacagcacc	acggaaggcg	cactgtactc	4620
atatctagaa	gggacccgtt	ttcatcagac	ggctgtggat	atggcggaga	tacatactat	4680
gtggccaaag	caaacagagg	ccaatgagca	agtctgccta	tatgccctgg	gggaaagtat	4740
tgaatcgatc	aggcagaaat	gcccgggtgga	tgatgcagac	gcatactctc	ccccaaaac	4800
tgtcccgtgc	ctttgccgtt	acgctatgac	tcagaaacgc	gtcaccgggc	ttcgcagtga	4860
ccacgtcaca	agcataattg	tgtgttcttc	gtttcccctc	ccaaagtaca	aaatagaagg	4920
agtgcaaaaa	gtcaaatgct	ctaaggtaat	gctatttgac	cacaacgtgc	catcgcgct	4980
aagtccaagg	gaatatagat	cttcccagga	gtctgcacag	gagggcagta	caatcacgtc	5040
actgacgcat	agtcaattcg	acctaagcgt	tgatggcgag	atactgcccg	tcccgcgaga	5100
cctggatgct	gacgccccag	ccctagaacc	agcactagac	gacggggcga	cacacacgct	5160
gccatccaca	accggaaacc	ttgcggccgt	gtctgatagg	gtaatgagca	ccgtacctgt	5220
cgcgcgcgcc	agaagaaggc	gagggagaaa	cctgactgtg	acatgtgacg	agagagaagg	5280
gaatataaca	cccatggcta	gcgtccgatt	ctttagggca	gagctgtgtc	cggtcgtaca	5340
agaaacagcg	gagacgcgtg	acacagcaat	gtctcttcag	gcaccaccga	gtaccgccac	5400
ggaaccgaat	catccgccga	tctccttcgg	agcatcaagc	gagacgttcc	ccattacatt	5460
tggggacttc	aacgaaggag	aaatcgaaag	cttgtcttct	gagctactaa	ctttcggaga	5520
cttcttacca	ggagaagtgg	atgacttgac	agacagcgac	tgggtccacgt	gctcagacac	5580
ggacgacgag	ttatgactag	acagggcgag	tgggtatata	ttctcgtcgg	acaccggtcc	5640
aggtcattta	caacagaagt	cagtacgcca	gtcagtgtctg	ccggtgaaca	ccctggagga	5700
agtccacgag	gagaagtgtt	acccacctaa	gctggatgaa	gcaaaggagc	aactattact	5760
taagaaactc	caggagagtg	catccatggc	caacagaagc	aggtatcagt	cgcgcaaagt	5820
agaaaacatg	aaagcagcaa	tcatccagag	actaaagaga	ggctgtagac	tatacttaat	5880

FIGURE 5 (con't)

gtcagagacc	ccaaaagtcc	ctacttaccg	gactacatat	ccggcgccctg	tgtactcgcc	5940
tccgatcaac	gtccgattgt	ccaatcccga	gtccgcagtg	gcagcatgca	atgagttctt	6000
agctagaaac	tatccaactg	tctcatcata	ccaaattacc	gacgagtatg	atgcatactt	6060
agacatgggtg	gacgggtcgg	agagttgcct	ggaccgagcg	acattcaatc	cgtcaaaact	6120
caggagctac	ccgaaacagc	acgcttacca	cgcgcctcct	atcagaagcg	ctgtaccgtc	6180
cccattccag	aacacactac	agaatgtact	ggcagcagcc	acgaaaagaa	actgcaacgt	6240
cacacagatg	aggggaattac	ccacttttga	ctcagcagta	ttcaacgtgg	agtgtttcaa	6300
aaaattcgca	tgcaaccaag	aatactggga	agaatttgct	gccagcccta	ttaggataac	6360
aactgagaat	ttagcaacct	atgttactaa	actaaaaggg	ccaaaagcag	cagcgctatt	6420
cgcaaaaacc	cataatctac	tgccactaca	ggaagtacca	atggataggt	tcacagtaga	6480
tatgaaaagg	gacgtgaagg	tgactcctgg	tacaaagcat	acagaggaaa	gacctaaagg	6540
gcaggttata	caggcggtcg	aacccttggc	gacagcatac	ctatgtggga	ttcacagaga	6600
gctgggttagg	aggctgaacg	ccgtcctcct	acccaatgta	catacactat	ttgacatgtc	6660
tgccgaggat	ttcgatgcca	tcatagccgc	acactttaag	ccaggagaca	ctgtttttgga	6720
aacggacata	gcctcctttg	ataagagcca	agatgattca	cttgcgctta	ctgctttgat	6780
gctgttagag	gatttagggg	tggtactctc	cctgctggac	ttgatagagg	ctgctttcgg	6840
agagatttcc	agctgtcacc	taccgacagg	tacgcgcttc	aagttcggcg	ccatgatgaa	6900
atcaggatatg	ttcctaactc	tgttcgtcaa	cacattgtta	aacatcacca	tcgccagccg	6960
agtgtctgaa	gacgtcttga	caaaatccgc	gtgtgcggcc	ttcatcggcg	acgacaacat	7020
aatacatgga	gtcgtctccg	atgaattgat	ggcagccaga	tgtgccactt	ggatgaacat	7080
ggaagtgaag	atcatagatg	cagttgtatc	cttgaaagcc	ccttactttt	gtggaggggt	7140
tatactgcac	gatactgtga	caggaacagc	ttgcagagtg	gcagaccgcg	taaaaaggct	7200
tttttaaactg	ggcaaaccgc	tagcggcagg	tgacgaacaa	gatgaagata	gaagacgagc	7260
gctggctgac	gaagtgatca	gatggcaacg	aacagggcta	attgatgagc	tgagagaaagc	7320
ggtatactct	aggtacgaag	tgcaagggtat	atcagttgtg	gtaatgtcca	tggtccacctt	7380
tgcaagctcc	agatccaact	tcgagaagct	cagaggaccc	gtcataactt	tgtacggcgg	7440
tcctaaatag	gtacgcacta	cagctaccta	ttttgcagaa	gccgacagca	agtatctaaa	7500
cactaatcag	ctacaatgga	gttcatecca	acccaaactt	tttacaatag	gaggtaccag	7560
cctcgaccct	ggactccgcg	ccctactatc	caagtcatca	ggcccagacc	gcgccctcag	7620
aggcaagctg	ggcaacttgc	ccagctgac	tcagcagtta	ataaactgac	aatgcgcgcg	7680
gtaccccaac	agaagccacg	caggaatcgg	agaataaga	agcaaaagca	aaaacaacag	7740
gcgccacaaa	acaacacaaa	tcaaaagaag	cagccaccta	aaaagaaacc	ggctcaaaag	7800
aaaaagaagc	cgggcccgcg	agagaggatg	tgcataaaaa	tcgaaaatga	ttgtattttc	7860
gaagtcaagc	acgaaggtaa	ggtaacaggt	tacgcgtgcc	tggtggggga	caaagtaatg	7920
aaaccagcac	acgtaaaggg	gaccatcgat	aacgcggacc	tggtccaaact	ggcctttaag	7980
cggtcactct	agtatgacct	tgaatgcgcg	cagatacccg	tgccacatgaa	gtccgacgct	8040
tcgaagttca	cccatgagaa	accggagggg	tactacaact	ggcaccacgg	agcagtacag	8100
tactcaggag	gccggttcac	catccctaca	gggtgctggca	aaccagggga	cagcggcaga	8160
ccgatcttcg	acaacaaggg	acgcgtgggt	gccatagtct	taggaggagc	taatgaagga	8220
gcccgtacag	ccctctcggg	ggtgacctgg	aataaagaca	ttgtcactaa	aatcaccccc	8280
gagggggccg	aagagtggag	tcttgccatc	ccagttatgt	gcctgttggc	aaacaccacg	8340
ttcccctgct	cccagccccc	ttgcacgccc	tgctgctacg	aaaaggaacc	ggaggaaacc	8400
ctacgcatgc	ttgaggacaa	cgtcatgaga	cctgggtact	atcagctgct	acaagcatcc	8460
ttaacatggt	ctccccaccg	ccagcgacgc	agcaccaagg	acaacttcaa	tgtctataaa	8520
gccacaagac	catacttagc	tcactgtccc	gactgtggag	aagggcactc	gtgccatagt	8580
cccgtagcac	tagaacgcat	cagaaatgaa	gcgacagacg	ggacgctgaa	aatccaggtc	8640
tccttgcaaa	tcggaataaa	gacggatgac	agccacgatt	ggaccaagct	gcgttatatg	8700
gacaaccaca	tgccagcaga	cgcagagagg	gcgggggctat	ttgtaagaac	atcagcaccg	8760

FIGURE 5 (con't)

tgtacgatta	ctggaacaat	gggacacttc	atcctggccc	gatgtccaaa	aggggaaact	8820
ctgacggtgg	gattcactga	cagtaggaag	attagtcact	catgtacgca	cccatttcac	8880
cacgaccctc	ctgtgatagg	tccggaaaaa	ttccattccc	gaccgcggca	cggtaaagag	8940
ctaccttgca	gcacgtacgt	gcagagcacc	gccgcaacta	ccgaggagat	agaggtagac	9000
atgccccag	acaccctga	tgcacatta	atgtcacaa	agtcgggcaa	cgtaaagatc	9060
acagtcaatg	gccagacggg	gcgggtacaag	tgtaattgcg	gtgggtccaa	tgaaggacta	9120
acaactacag	acaaagtgat	taataactgc	aagggtgatc	aatgtcatgc	cgcggtcacc	9180
aatcacaaaa	agtggcagta	taactcccct	ctgggtccgc	gtaatgctga	acttggggac	9240
cgaaaaggaa	aaattcacat	cccgtttccg	ctggcaaagt	taacatgcag	ggtgcctaaa	9300
gcaaggaacc	ccaccgtgac	gtacgggaaa	aaccaagtca	tcatgctact	gtatcctgac	9360
caccaaacac	tcctgtccta	ccggaatatg	ggagaagaac	caaactatca	agaagagtgg	9420
gtgatgcata	agaaggaagt	cgtgctaacc	gtgccgactg	aagggtcga	ggtcacgtgg	9480
ggcaacaacg	agccgtataa	gtattggccg	cagttatcta	caaacggtag	agcccatggc	9540
caccgcgatg	agataattct	gtattattat	gagctgtacc	ccactatgac	tgtagtagtt	9600
gtgtcagtg	ccacgttcat	actcctgtcg	atgggtgggt	tggcagcggg	gatgtgcatg	9660
tgtgcacgac	gcagatgcat	cacaccgtat	gaactgacac	caggagctac	cgtccctttc	9720
ctgcttagcc	taatattgctg	catcagaaca	gctaaagcgg	ccacatacca	agaggctgcg	9780
atatacctgt	ggaacgagca	gcaacccttg	ttttgggtac	aagcccttat	tcgctggca	9840
gccctgattg	ttctatgcaa	ctgtctgaga	ctcttaccat	gctgctgtaa	aacgttggct	9900
tttttagccg	taatgagcgt	cgggtgccac	actgtgagcg	cgtacgaaca	cgtaacagtg	9960
atccgaaca	cgggtgggagt	accgtataag	actctagtca	atagacctgg	ctacagcccc	10020
atgggtattg	agatggaact	actgtcagtc	actttggagc	caacactatc	gcttgattac	10080
atcacgtgcg	agtacaaaac	cgtcatcccg	tctccgtacg	tgaagtgctg	cggtagacga	10140
gagtgcgaag	acaaaaacct	acctgactac	agctgtaagg	tcttcaccgg	cgtctacca	10200
tttatgtggg	gcgggcgcta	ctgcttctgc	gacgctgaaa	acacgcagtt	gagcgaagca	10260
cacgtggaga	agtccgaatc	atgcaaaaac	gaatttgcat	cagcatacag	ggctcatacc	10320
gcatctgcat	cagctaagct	ccgcgtcctt	taccaaggaa	ataacatcac	tgtaactgcc	10380
tatgcaaacg	gcgaccatgc	cgtcacagtt	aaggacgcca	aattcattgt	ggggccaatg	10440
tcttcagcct	ggacaccttt	cgacaacaaa	attgtggtgt	acaaagggtg	cgtctataac	10500
atggactacc	cgccttttgg	cgcaggaaga	ccaggacaat	ttggcgatat	ccaaagtcgc	10560
acacctgaga	gtaaagacgt	ctatgcta	acacaactgg	tactgcagag	accggctgcg	10620
ggtacggtag	acgtgccata	ctctcaggca	ccatctggct	ttaagtattg	gctaaaagaa	10680
cgcggggcgt	cgtgcagca	cacagcacca	tttggctgcc	aaatagcaac	aaaccgggta	10740
agagcggtag	actgcgcgt	aggggaacatg	cccatctcca	tcgacatacc	ggaagcggcc	10800
ttcactaggg	tcgtcgacgc	gccctcttta	acggacatgt	cgtgcgaggt	accagcctgc	10860
accatttcct	cagactttgg	gggcgtcgcc	attattaaat	atgcagccag	caagaaaggc	10920
aagtgtgcgg	tgcattcgat	gactaacgcc	gtcactattc	gggaagctga	gatagaagtt	10980
gaagggaatt	ctcagctgca	aatctctttc	tcgacggcct	tagccagcgc	cgaattccgc	11040
gtacaagtct	gttctacaca	agtacactgt	gcagccgagt	gccaccccc	gaaggaccac	11100
atagtcaact	acccggcgtc	acataccacc	ctcgggggtcc	aggacatctc	cgtacggcg	11160
atgtcatggg	tgcagaagat	cacgggaggt	gtgggactgg	ttgttgctgt	tgccgcactg	11220
attctaattc	tggtgctatg	cgtatcgttc	agcaggcact	aacttgacaa	ttaagtatga	11280
aggtatatgt	gtcccctaag	agacacactg	tacatagcaa	ataatctata	gatcaaaggg	11340
ctacgcaacc	cctgaatagt	aacaaaatac	aaaatcacta	aaaattataa	aaacagaaaa	11400
atacataaat	aggtatacgt	gtcccctaag	agacacattg	tatgtagggt	ataagtatag	11460
atcaaagggc	cgaataaccc	ctgaatagta	acaaaatatg	aaaatcaata	aaaatcataa	11520
aatagaaaaa	ccataaacag	aagtagttca	aagggtata	aaaccctga	atagtaacaa	11580
aacataaaat	taataaaaat	c				11601

FIGURE 6

caaagcaaga	gattaataac	ccatcatgga	tcctgtgtac	gtggacatag	acgctgacag	60
cgcctttttg	aaggccctgc	aacgtgcgta	ccccatgttt	gaggtggaac	caaggcaggt	120
cacaccgaat	gaccatgcta	atgctagagc	gttctcgcac	ctagctataa	aactaataga	180
gcaggaaatt	gaccccgact	caaccatcct	ggatatcggc	agtgcgccag	caaggaggat	240
gatgtcggac	aggaagtacc	actgcgtctg	cccgatgcgc	agtgcggaag	atcccagagag	300
actcgccaat	tatgcgagaa	agctagcatc	tgccgcagga	aaagtcctgg	acagaaacat	360
ctctggaaag	atcggggact	tacaagcagt	aatggccgtg	ccagacacgg	agacgccaac	420
attctgctta	cacacagacg	tctcatgtag	acagagagca	gacgtcgcta	tataccaaga	480
cgtctatgct	gtacacgcac	ccacgtcgct	ataccaccag	gcgattaaag	gggtccgagt	540
ggcgtactgg	gltgggttcg	acacaacccc	gttcatgtac	aatgccatgg	cgggtgccta	600
cccctcatac	tcgacaaact	gggcagatga	gcaggtactg	aaggctaaga	acataggatt	660
atgttcaaca	gacctgacgg	aaggtagacg	aggcaagttg	tctattatga	gagggaaaaa	720
gctaaaaccg	tgcgaccgtg	tgctgttctc	agtaggggtca	acgctctacc	cggaaagccg	780
caagctactt	aagagctggc	acctgccatc	ggtgttccat	ttaaagggca	aactcagctt	840
cacatgccgc	tgtgatacag	tggtttcgtg	tgagggctac	gtcgtaaga	gaataacgat	900
gagcccagge	ctttatggaa	aaaccacagg	gtatgcggta	accaccacg	cagacggatt	960
cctgatgtgc	aagactaccg	acacggttga	cggcgaaaga	gtgtcattct	cgggtgtgcac	1020
atacgtgccg	gcgaccattt	gtgatcaa	gaccggcatc	cttgctacag	aagtcacgcc	1080
ggaggatgca	cagaagctgt	tgggtggggct	gaaccagaga	atagtggtta	acggcagaac	1140
gcaacggaat	acgaacacca	tgaaaaatta	tctgcttccc	gtggtcgccc	aagccttcag	1200
taagtgggca	aaggagtgcc	ggaaagacat	ggaagatgaa	aaactcctgg	gggtcagaga	1260
aagaacactg	acctgctgct	gtctatgggc	attcaagaag	cagaaaacac	acacgggtcta	1320
caagaggcct	gatacccagt	caattcagaa	ggttcaggcc	gagtttgaca	gctttgtggt	1380
accgagtctg	tggtcgtccg	ggttgtcaat	ccctttgagg	actagaatca	aatgggttgtt	1440
aagcaagggtg	ccaaaaaccg	acctgatccc	atacagcgga	gacgcccag	aagcccggga	1500
cgcagaaaaa	gaagcagagg	aagaacgaga	agcagaactg	actcgcgaag	ccctaccacc	1560
tctacaggca	gcacaggaag	atgttcagggt	cgaaatcgac	gtggaacagc	ttgaggacag	1620
agcggggcgca	ggaataatag	agactccgag	aggagctatc	aaagttactg	cccaaccaac	1680
agaccacgtc	gtgggagagt	acctgggtact	ctccccgcag	accgtactac	gtagccagaa	1740
gctcagtctg	attcacgctt	tggcggagca	agtgaagacg	tgcacgcaca	acggacgagc	1800
agggaggtat	gcggtcgaag	cgtacgacgg	ccgagtccta	gtgccctcag	gctatgcaat	1860
ctcgccctgaa	gacttccaga	gtctaagcga	aagcgcaacg	atgggtgtata	acgaaagaga	1920
gttcgtaaac	agaaagctac	accatattgc	gatgcacgga	ccagccctga	acaccgacga	1980
agagtcgtat	gagctggtga	gggcagagag	gacagaacac	gagtacgtct	acgacgtgga	2040
tcagagaaga	tgctgtaaga	aggaagaagc	cgcaggactg	gtactgggtg	gcgacttgac	2100
taatccgccc	taccacgaat	tcgcatatga	agggctaaaa	atccgcccctg	cctgcccata	2160
caaaattgca	gtcataggag	tcttcggagt	accgggatct	ggcaagtcag	ctattatcaa	2220
gaacctagtt	accaggcagg	acctgggtgac	tagcggaaag	aaagaaaact	gccaagaaat	2280
caccaccgac	gtgatgagac	agagaggtct	agagatatct	gcacgtacgg	ttgactcgct	2340
gctcttgaat	ggatgcaaca	gaccagtcga	cgtgtttgtac	gtagacgagg	cgtttgcgtg	2400
ccactctgga	acgctacttg	ctttgatcgc	cttgggtgaga	ccaaggcaga	aagttgtact	2460
ttgtgggtgac	ccgaagcagt	gcggcttctt	caatatgatg	cagatgaaag	tcaactataa	2520
tcacaacatc	tgcaccaag	tgtaccacaa	aagtatctcc	aggcgggtgta	cactgcctgt	2580
gaccgccatt	gtgtcatcgt	tgcattacga	aggcaaaatg	cgcactacga	atgagtacaa	2640
caagccgatt	gtagtggaca	ctacaggctc	aacaaaacct	gaccctggag	acctcgtgtt	2700
aacgtgcttc	agaggggtggg	ttaaacaact	gcaaattgac	tatcgtggat	acgaggtcat	2760
gacagcagcc	gcatcccaag	ggttaaccag	aaaaggagtt	tacgcagtta	gacaaaaagt	2820
taatgaaaac	ccgctctatg	catcaacgtc	agagcacgtc	aacgtactcc	taacgcgtac	2880
ggaaggtaaa	ctgggtatgga	agacactttc	cggcgacccg	tggataaaga	cgctgcagaa	2940

FIGURE 6 (con't)

cccaccgaaa	ggaaacttca	aagcaactat	taaggagtgg	gaggtggagc	atgcatcaat	3000
aatggcgggc	atctgcagtc	accaaatagac	cttcgatata	ttccaaaata	aagccaacgt	3060
ttgttgggct	aagagcttgg	tccctatcct	cgaaacagcg	gggataaaac	taaatagatag	3120
gcagtgggtct	cagataattc	aagccttcaa	agaagacaaa	gcatactcac	ctgaagtagc	3180
cctgaatgaa	atatgtacgc	gcatgtatgg	ggtggatcta	gacagcgggc	tatttttctaa	3240
accgttggtg	tctgtgtatt	acgcggataa	ccactgggat	aataggcctg	gagggaaaaat	3300
gttcggattt	aaccccgagg	cagcatccat	tctagaaaga	aagtatccat	tcacaaaagg	3360
gaagtggaac	atcaacaagc	agatctgogt	gactaccagg	aggatagaag	actttaaccc	3420
taccaccaac	atcataccgg	ccaacaggag	actaccacac	tcattagtgg	ccgaacaccg	3480
cccagtaaaa	ggggaaagaa	tggaaatggct	ggttaacaag	ataaacgggc	accacgtgct	3540
cctggtcagt	ggctataacc	ttgcactgcc	tactaagaga	gtcacttggg	tagcgccgtt	3600
aggtgtccgc	ggagcggact	acacatacaa	cctagagttg	ggtctgccag	caacgcttgg	3660
taggtatgac	ctagtgggtca	taaacaacca	cacacctttt	cgcatacacc	attaccaaca	3720
gtgcgtcgac	cacgcaatga	aactgcaaat	gctcgggggt	gactcattga	gactgctcaa	3780
accgggcggc	tctctattga	tcagagcata	tggttacgca	gatagaacca	gtgaacgagt	3840
catctgcgta	ttgggacgca	agtttagatc	gtctagagcg	ttgaaaccac	catgtgtcac	3900
cagcaacact	gagatgtttt	tcctattcag	caactttgac	aatggcagaa	ggaatttcac	3960
aactcatgtc	atgaacaatc	aactgaatgc	agccttcgta	ggacagggtca	cccagacagg	4020
atgtgcaccg	tcgtaccggg	taaaacgcat	ggacatcgcg	aagaacgatg	aagagtgcgt	4080
agtcaacgcc	gctaaccctc	gcgggttacc	gggtgacggt	gtttgcaagg	cagtatacaa	4140
aaaatggccg	gagtccttta	agaacagtgc	aacaccagtg	ggaaccgcaa	aaacagttat	4200
gtgcggtacg	tatccagtaa	tcacgcgtgt	tggaccaaac	ttctctaat	attcggagtc	4260
tgaaggggac	cggaatttgg	cagctgccta	tcgagaagtc	gcaaaggaag	taactaggct	4320
gggagtaaat	agtgtagcta	tacctctcct	ctccacaggt	gtatactcag	gagggaaaga	4380
caggctgacc	cagtcactga	accacctctt	tacagccatg	gactcgacgg	atgcagacgt	4440
ggtcatctac	tgccgcgaca	aagaatggga	gaagaaaata	tctgaggcca	tacagatgcg	4500
gaccaagta	gagctgctgg	atgagcacat	ctccatagac	tgcgatattg	ttcgcgtgca	4560
ccctgacagc	agcttggcag	gcagaaaagg	atacagcacc	acggaaggcg	cactgtactc	4620
atatctagaa	gggacccgtt	ttcatcagac	ggctgtggat	atggcggaga	tacatactat	4680
gtggccaaag	caaacagagg	ccaatgagca	agtctgccta	tatgccctgg	gggaaagtat	4740
tgaatcgatc	aggcagaaat	gcccgggtgga	tgatgcagac	gcatcatctc	cccccaaac	4800
tgtcccgtgc	ctttgccgtt	acgctatgac	tccagaacgc	gtcaccgggc	ttcgcagtga	4860
ccacgtcaca	agcataattg	tgtgttcttc	gtttcccttc	ccaaagtaca	aaatagaagg	4920
agtgcaaaaa	gtcaaatgct	ctaaggtaat	gctatttgac	cacaacgtgc	catcgcgct	4980
aagtccaagg	gaatatagat	cttcccagga	gtctgcacag	gaggcgagta	caatcacgtc	5040
actgacgcat	agtcaattcg	acctaagcgt	tgatggcgag	atactgcccg	tcccgtcaga	5100
cctggatgct	gacgccccag	ccctagaacc	agcactagac	gacggggcga	cacacacgct	5160
gccatccaca	accggaaacc	ttgcggccgt	gtctgattgg	gtaatgagca	ccgtacctgt	5220
cgcgccgccc	agaagaaggc	gaggagaaaa	cctgactgtg	acatgtgacg	agagagaagg	5280
gaatataaca	cccatggcta	gcgtccgatt	ctttagggca	gagctgtgtc	cggtcgtaca	5340
agaaacagcg	gagacgcgtg	acacagcaat	gtctcttcag	gcaccaccga	gtaccgccac	5400
ggaaccgaat	catccgccga	tctccttcgg	agcatcaagc	gagacgttcc	ccattacatt	5460
tggggacttc	aacgaaggag	aaatcgaaag	cttgtcttct	gagctactaa	ctttcggaga	5520
cttcttacca	ggagaagtgg	atgacttgac	agacagcgac	tgggccacgt	gctcagacac	5580
ggacgacgag	ttatgactag	acagggcgag	tgggtatata	ttctcgtcgg	acaccggtcc	5640
aggtcattta	caacagaagt	cagtacgcca	gtcagtgtctg	ccggtgaaca	ccctggagga	5700
agtccacgag	gagaagtgtt	acccacctaa	gctggatgaa	gcaaaggagc	aactattact	5760
taagaaactc	caggagagtg	catccatggc	caacagaagc	aggtatcagt	cgcgcaaagt	5820
agaaaacatg	aaagcagcaa	tcatccagag	actaaagaga	ggctgtagac	tatacttaat	5880

FIGURE 6 (con't)

gtcagagacc	ccaaaagtcc	ctacttaccg	gactacatat	ccggcgccctg	tgtactcgcc	5940
tccgatcaac	gtccgattgt	ccaatcccga	gtccgcagtg	gcagcatgca	atgagttctt	6000
agctagaaac	tatccaactg	tctcatcata	ccaaattacc	gacgagtatg	atgcatatct	6060
agacatgggtg	gacgggtcgg	agagttgcct	ggaccgagcg	acattcaatc	cgtcaaaact	6120
caggagctac	ccgaaacagc	acgcttacca	cgcgcctcc	atcagaagcg	ctgtaccgtc	6180
cccattccag	aacacactac	agaatgtact	ggcagcagcc	acgaaaagaa	actgcaacgt	6240
cacacagatg	aggggaattac	ccactttgga	ctcagcagta	ttcaacgtgg	agtgtttcaa	6300
aaaatttcgca	tgcaaccaag	aatactggga	agaatttgct	gccagcccta	ttaggataac	6360
aactgagaat	ttagcaacct	atgttactaa	actaaaaggg	ccaaaagcag	cagcgctatt	6420
cgcaaaaacc	cataatctac	tgccactaca	ggaagtacca	atggataggt	tcacagtaga	6480
tatgaaaagg	gacgtgaagg	tgactcctgg	tacaaagcat	acagaggaaa	gacctaaagg	6540
gcaggttata	caggcggtg	aacccttggc	gacagcatac	ctatgtggga	ttcacagaga	6600
gctgggttagg	aggctgaacg	ccgtcctcct	acccaatgta	catacactat	ttgacatgtc	6660
tgccgaggat	ttcgatgcca	tcatagccgc	acactttaag	ccaggagaca	ctgttttgga	6720
aacggacata	gcctcctttg	ataagagcca	agatgattca	cttgcgctta	ctgctttgat	6780
gctgttagag	gatttagggg	tggatcactc	cctgctggac	ttgatagagg	ctgctttcgg	6840
agagatttcc	agctgtcacc	taccgacagg	tacgcgcttc	aagttcggcg	ccatgatgaa	6900
atcagggtatg	ttcctaactc	tgttcgtcaa	cacattgtta	aacatcacca	tcgccagccg	6960
agtgtctgaa	gatcgtctga	caaaatccgc	gtgcgcggcc	ttcatcggcg	acgacaacat	7020
aatacatgga	gtcgtctccg	atgaattgat	ggcagccaga	tgtgccactt	ggatgaacat	7080
ggaagtgaag	atcatagatg	cagttgtatc	cttgaaagcc	ccttactttt	gtggaggggt	7140
tatactgcac	gatactgtga	caggaacagc	ttgcagagtg	gcagacccgc	taaaaaggct	7200
tttttaaactg	ggcaaaccgc	tagcggcagg	tgacgaacaa	gatgaagata	gaagacgagc	7260
gctggctgac	gaagtgatca	gatggcaacg	aacagggcta	attgatgagc	tggagaaagc	7320
ggtatactct	aggtacgaag	tgcaagggtat	atcagttgtg	gtaatgtcca	tggccacctt	7380
tgcaagctcc	agatccaact	tcgagaagct	cagaggaccc	gtcataactt	tgtacggcgg	7440
tcctaaatag	gtacgcacta	cagctaccta	ttttgcagaa	gccgacagca	agtatctaaa	7500
cactaatcag	ctacaatgga	gttcatecca	acccaaactt	tttacaatag	gaggtaccag	7560
cctcgaccct	ggactccgcg	ccctactatc	caagtcatca	ggcccagacc	gcgccctcag	7620
aggcaagctg	ggcaacttgc	ccagctgatc	tcagcagtta	ataaactgac	aatgcgcgcg	7680
gtaccccaac	agaagccacg	caggaatcgg	agaataaga	agcaaaaagca	aaaacaacag	7740
gcgccacaaa	acaacacaaa	tcaaaagaag	cagccaccta	aaaagaaacc	ggctcaaaag	7800
aaaaagaagc	cgggccgcag	agagaggatg	tgcatgaaaa	tcgaaaatga	ttgtattttc	7860
gaagtcaagc	acgaaggtaa	ggtaacaggt	tacgcgtgcc	tggtggggga	caaagtaatg	7920
aaaccagcac	acgtaaaagg	gaccatcgat	aacgcggacc	tggccaaact	ggcctttaag	7980
cggtcactct	agtatgacct	tgaatgcgcg	cagatacccg	tgccatgaa	gtccgacgct	8040
tcgaagttca	cccatgagaa	accggagggg	tactacaact	ggcaccacgg	agcagtacag	8100
tactcaggag	gccggttcac	catccctaca	ggtgctggca	aaccagggga	cagcggcaga	8160
ccgatcttcg	acaacaaggg	acgcgtgggt	gccatagtct	taggaggagc	taatgaagga	8220
gcccgtagag	ccctctcggt	ggtgacctgg	aataaagaca	ttgtcactaa	aatcaccccc	8280
gagggggccg	aagagtggag	tcttgccatc	ccagttatgt	gcctgttggc	aaacaccacg	8340
ttcccttgct	cccagcccc	ttgcacgccc	tgctgctacg	aaaaggaacc	ggaggaaacc	8400
ctacgcatgc	ttgaggacaa	cgcatgaga	cctgggtact	atcagctgct	acaagcatcc	8460
ttaacatgtt	ctccccaccg	ccagcgacgc	agcaccaagg	acaacttcaa	tgtctataaa	8520
gccacaagac	catacttagc	tactgtccc	gactgtggag	aagggcactc	gtgccatagt	8580
cccgtagcac	tagaacgcat	cagaaatgaa	gcgacagacg	ggacgctgaa	aatccaggtc	8640
tccttgcaaa	tcggaataaa	gacggatgac	agccacgatt	ggaccaagct	gcgttatatg	8700
gacaaccaca	tgccagcaga	cgcagagagg	gcggggctat	ttgtaagaac	atcagcaccg	8760

FIGURE 6 (con't)

tgtacgatta	ctggaacaat	gggacacttc	atcctggccc	gatgtccaaa	aggggaaact	8820
ctgacggtgg	gattcactga	cagtaggaag	attagtcact	catgtacgca	cccatttcac	8880
cacgaccctc	ctgtgatagg	tcgggaaaaa	ttccattccc	gaccgcagca	cggtaaagag	8940
ctaccttgca	gcacgtacgt	gcagagcacc	gccgcaacta	ccgaggagat	agaggtagac	9000
atgccccag	acaccctga	tcgcacatta	atgtcacaac	agtccggcaa	cgtaaagatc	9060
acagtcaatg	gccagacggt	gcggtacaag	tgtaatgtcg	gtggctcaaa	tgaaggacta	9120
acaactacag	acaaagtgat	taataactgc	aagggtgatc	aatgtcatgc	cgcggtcacc	9180
aatcacaaaa	agtggcagta	taactccctc	ctgggtccgc	gtaatgctga	acttggggac	9240
cgaaaaggaa	aaattcacat	cccgtttccg	ctggcaaattg	taacatgcag	ggtgcctaaa	9300
gcaaggaacc	ccaccgtgac	gtacgggaaa	aaccaagtca	tcatgctact	gtatcctgac	9360
caccaaacac	tcctgtccta	ccggaatatg	ggagaagaac	caaactatca	agaagagtgg	9420
gtgatgcata	agaaggaagt	cgtgctaacc	gtgccgactg	aagggctcga	ggtcacgtgg	9480
ggcaacaacg	agccgtataa	gtattggccg	cagttatcta	caaacggtac	agcccatggc	9540
caccgcgatg	agataattct	gtattattat	gagctgtacc	ccactatgac	tgtagtagtt	9600
gtgtcagtg	ccacgttcat	actcctgtcg	atgggtgggta	tggcagcggg	gatgtgcatg	9660
tgtgcacgac	gcagatgcat	cacaccgtat	gaactgacac	caggagctac	cgcccccttc	9720
ctgcttagcc	taatatgctg	catcagaaca	gctaaagcgg	ccacatacca	agaggctgcg	9780
atatacctgt	ggaacgagca	gcaacctttg	ttttggctac	aagcccttat	tccgctggca	9840
gccctgattg	ttctatgcaa	ctgtctgaga	ctcttaccat	gctgctgtaa	aacgttggct	9900
tttttagccg	taatgagcgt	cggtgcccac	actgtgagcg	cgtacgaaca	cgtaacagtg	9960
atcccgaaca	cgggtgggagt	accgtataag	actctagtca	atagacctgg	ctacagcccc	10020
atggtatttg	agatggaact	actgtcagtc	actttggagc	caacactatc	gcttgattac	10080
atcacgtgcg	agtacaaaac	cgtcatcccg	tctccgtacg	tgaagtgctg	cggtagacga	10140
gagtgcagg	acaaaaacct	acctgactac	agctgtaagg	tcttcaccgg	cgtctacca	10200
tttatgtggg	gcggcgccta	ctgcttctgc	gacgctgaaa	acacgcagtt	gagcgaagca	10260
cacgtggaga	agtcggaatc	atgcaaaaaca	gaatttgcac	cagcatacag	ggctcatacc	10320
gcactctgcat	cagctaagct	ccgcgtcctt	taccaaggaa	ataacatcac	tgtaaactgcc	10380
tatgcaaacg	gcgaccatgc	cgtcacagtt	aaggacgcca	aattcattgt	ggggccaatg	10440
tcttcagcct	ggacaccttt	cgacaacaaa	attgtgggtg	acaaagggtga	cgtctataac	10500
atggactacc	cgccttttgg	cgcaggaaga	ccaggacaat	ttggcgatat	ccaaagtcgc	10560
acacctgaga	gtaaagacgt	ctatgcta	acacaactgg	tactgcagag	accggctgtg	10620
ggtacggtac	acgtgccata	ctctcaggca	ccatctggct	ttaagtattg	gctaaaagaa	10680
cgcggggctg	cgtgcagca	cacagcacca	tttggtgcc	aaatagcaac	aaaccggtg	10740
agagcggtga	actgcgcgt	agggaaacatg	cccatctcca	tgcacatacc	ggaagcggcc	10800
ttcactaggg	tcgtcgacgc	gccctcttta	acggacatgt	cgtgcgaggt	accagcctgc	10860
acccattcct	cagacttttg	gggcgtcgcc	attattaaat	atgcagccag	caagaaaggc	10920
aagtgtgcgg	tgcattcgat	gactaacgcc	gtcactattc	gggaagctga	gatagaagtt	10980
gaagggaatt	ctcagctgca	aatctctttc	tcgacggcct	tagccagcgc	cgaattccgc	11040
gtacaagtct	gttctacaca	agtacactgt	gcagccgagt	gccaccccc	gaaggaccac	11100
atagtcaact	acccggcgtc	acataccacc	ctcggggctc	aggacatctc	cgctacggcg	11160
atgtcatggg	tgcagaagat	cacgggaggt	gtgggactgg	ttgttgctgt	tgcgcgactg	11220
attctaatacg	tgggtgctatg	cgtgtcggtc	agcaggcact	aacttgacaa	ttaagtatga	11280
aggtatatgt	gtcccctaag	agacacactg	tacatagcaa	ataatctata	gatcaaaggg	11340
ctacgcaacc	cctgaatagt	aacaaaatac	aaaatcacta	aaaattataa	aaacagaaaa	11400
atacataaat	aggtatacgt	gtcccctaag	agacacattg	tatgtagggtg	ataagtatag	11460
atcaaagggc	cgaataaccc	ctgaatagta	acaaaatatg	aaaatcaata	aaaatcataa	11520
aatagaaaaa	ccataaacag	aagtagttca	aagggtctata	aaaccctga	atagtaacaa	11580
aacataaaat	taataaaaaat	c				11601

FIGURE 7

caaagcaaga	gattaataac	ccatcatgga	tccctgtgtac	gtggacatag	acgctgacag	60
cgcctttttg	aaggccctgc	aacgtgcgta	ccccatgttt	gaggtggaac	caaggcaggt	120
cacaccgaat	gaccatgcta	atgctagagc	gttctcgcac	ctagctataa	aactaataga	180
gcaggaaatt	gaccccgact	caaccatcct	ggatatcggc	agtgcgccag	caaggaggat	240
gatgtcggac	aggaagtacc	actgcgtctg	cccgatgcgc	agtgcggaag	atcccgagag	300
actcgccaat	tatgcgagaa	agctagcatc	tgcgcgagga	aaagtccctgg	acagaaacat	360
ctctggaaag	atcggggact	tacaagcagt	aatggccgtg	ccagacacgg	agacgccaac	420
attctgctta	cacacagacg	tctcatgtag	acagagagca	gacgtcgcta	tataccaaga	480
cgtctatgct	gtacacgcac	ccacgtcgct	ataccaccag	gcgattaaag	gggtccgagt	540
ggcgtactgg	gttgggttcg	acacaacccc	gttcatgtac	aatgccatgg	cgggtgccta	600
cccctcatat	tgcacaaact	gggcagatga	gcaggtactg	aaggctaaga	acataggatt	660
atgttcaaca	gacctgacgg	aaggtagacg	aggcaagttg	tctattatga	gagggaaaaa	720
gctaaaaccg	tgcgaccgtg	tgctgtttct	agtaggggtca	acgctctacc	cggaaagccg	780
caagctactt	aagagctggc	acctgccatc	ggtgttccat	ttaaagggca	aactcagctt	840
cacatgccgc	tgtgatacag	tggtttcgtg	tgaggggtac	gtcgttaaga	gaataacgat	900
gagcccaggc	ctttatggaa	aaaccatagg	gtatgcggta	acccaccacg	cagacggatt	960
cctgatgtgc	aagactaccg	acacggttga	cggcgaaaga	gtgtcattct	cgggtgtgcac	1020
atacgtgccg	gcgaccattt	gtgatcaaat	gaccggcatc	cttgctacag	aagtcacgcc	1080
ggaggatgca	cagaagctgt	tgggtggggct	gaaccagaga	atagtggtta	acggcagaac	1140
gcaacggaat	acgaacacca	tgaaaaatta	tctgcttccc	gtggtcgccc	aagccttcag	1200
taagtgggca	aaggagtgcc	ggaaagacat	ggaagatgaa	aaactcctgg	gggtcagaga	1260
aagaacactg	acctgctgct	gtctatgggc	attcaagaag	cagaaaacac	acacggctcta	1320
caagaggcct	gatacccagt	caattcagaa	ggttcaggcc	gagtttgaca	gctttgtggt	1380
accgagtctg	tggtcgtccg	ggttgtcaat	ccctttgagg	actagaatca	aatggttggt	1440
aagcaagggtg	ccaaaaaccg	acctgatccc	atacagcgga	gacgcccag	aagcccggga	1500
cgcagaaaaa	gaagcagagg	aagaacgaga	agcagaactg	actcgcgaag	ccctaccacc	1560
tctacaggca	gcacaggaag	atgttcaggt	cgaaatcgac	gtggaacagc	ttgaggacag	1620
agcgggcgca	ggaataatag	agactccgag	aggagctatc	aaagt tactg	cccaaccaac	1680
agaccacgtc	gtgggagagt	acctggtact	ctccccgcag	accgtactac	gtagccagaa	1740
gctcagtctg	attcacgctt	tggcggagca	agtgaagacg	tgcacgcaca	acggacgagc	1800
agggaggtat	gcggtcgaag	cgtacgacgg	ccgagtccta	gtgccctcag	gctatgcaat	1860
ctcgccctgaa	gacttccaga	gtctaagcga	aagcgcaacg	atggtgtata	acgaaagaga	1920
gttcgtaaac	agaaagctac	accatattgc	gatgcacgga	ccagccctga	acaccgacga	1980
agagtcgtat	gagctggtga	gggcagagag	gacagaacac	gagtacgtct	acgacgtgga	2040
tcagagaaga	tgctgtaaga	aggaagaagc	cgcaggactg	gtactggtgg	gcgacttgac	2100
taatccgccc	taccacgaat	tgcgatatga	agggctaaaa	atccgcccctg	cctgcccata	2160
caaaattgca	gtcataggag	tcttcggagt	accgggatct	ggcaagtcag	ctattatcaa	2220
gaacctagtt	accaggcagg	acctggtgac	tagcggaaag	aaagaaaact	gccaagaaat	2280
caccaccgac	gtgatgagac	agagaggtct	agagatatct	gcacgtacgg	ttgactcgct	2340
gctcttgaat	ggatgcaaca	gaccagtcga	cgtgttgtac	gtagacgagg	cgtttgcgtg	2400
ccactctgga	acgctacttg	ctttgatcgc	cttggtgaga	ccaaggcaga	aagttgtact	2460
ttgtggtgac	ccgaagcagt	gcggcttctt	caatatgatg	cagatgaaag	tcaactataa	2520
tcacaacatc	tgcaccaag	tgtaccacaa	aagtatctcc	aggcgggtga	cactgcctgt	2580
gaccgccatt	gtgtcatcgt	tgcattacga	aggcaaaatg	cgcactacga	atgagtacaa	2640
caagccgatt	gtagtggaca	ctacaggctc	aacaaaacct	gaccctggag	acctcgtggt	2700
aacgtgcttc	agaggggtggg	ttaaacaact	gcaaattgac	tatcgtggat	acgaggtcat	2760
gacagcagcc	gcatcccaag	ggttaaccag	aaaaggagtt	tacgcagtta	gacaaaaagt	2820
taatgaaaac	ccgctctatg	catcaacgtc	agagcacgtc	aacgtactcc	taacgcgtac	2880

FIGURE 7 (con't)

ggaaggtaaa	ctggtatgga	agacactttc	cggcgacccg	tggataaaga	cgctgcagaa	2940
cccaccgaaa	ggaaacttca	aagcaactat	taaggagtgg	gaggtggagc	atgcatcaat	3000
aatggcgggc	atctgcagtc	accaaatagac	cttcgatata	ttccaaaata	aagccaacgt	3060
ttgttgggct	aagagcttgg	tccctatcct	cgaaacagcg	gggataaaac	taaatagatag	3120
gcagtgggtct	cagataattc	aagccttcaa	agaagacaaa	gcatactcac	ctgaagtagc	3180
cctgaatgaa	atatgtacgc	gcatgtatgg	ggtggatcta	gacagcgggc	tatttttctaa	3240
accgttgggtg	tctgtgtatt	acgcggataa	ccactgggat	aataggcctg	gagggaaaat	3300
gttcggatth	aaccccgagg	cagcatccat	tctagaaaga	aagtatccat	tcacaaaagg	3360
gaagtggaaac	atcaacaagc	agatctgcgt	gactaccagg	aggatagaag	actttaaccc	3420
taccaccaac	atcataccgg	ccaacaggag	actaccacac	tcattagtgg	ccgaacaccg	3480
cccagtaaaa	ggggaaagaa	tggaaatggct	ggttaacaag	ataaacggcc	accacgtgct	3540
cctggtcagt	ggcaataacc	ttgcactgcc	tactaagaga	gtcacttggg	tagcgccggt	3600
aggtgtccgc	ggagcggact	acacatacaa	cctagagtth	ggtctgccag	caacgcttgg	3660
taggtatgac	ctagtgggtca	taaacatcca	cacaccttht	cgcatacacc	attaccaaca	3720
gtgcgtcgac	cacgcaatga	aactgcaaata	gctcgggggt	gactcattga	gactgctcaa	3780
accggggcggc	tctctattga	tcagagcata	tggttacgca	gatagaacca	gtgaacgagt	3840
catctgcgta	ttgggacgca	agtttagatc	gtctagagcg	ttgaaaccac	catgtgtcac	3900
cagcaacact	gagatgttht	tcctattcag	caactthtgac	aatggcagaa	ggaatttcac	3960
aactcatgtc	atgaacaatc	aactgaatgc	agccttcgta	ggacagggtca	cccgagcagg	4020
atgtgcaccg	tcgtaccggg	taaaacgcat	ggacatcgcg	aagaacgatg	aagagtgcgt	4080
agtcaacgcc	gctaaccctc	gcgggttacc	gggtgacggg	gtttgcaagg	cagtatacaa	4140
aaaatggccg	gagtccttht	agaacagtgc	aacaccagtg	ggaaccgcaa	aaacagttat	4200
gtgcgggtacg	tatccagtaa	tccacgctgt	tggaccaaac	ttctctaatt	attcggagtc	4260
tgaaggggac	cgggaattgg	cagctgccta	tcgagaagtc	gcaaaggaag	taactaggct	4320
gggagtaaat	agtgtagcta	tacctctcct	ctccacagg	gtatactcag	gagggaaaga	4380
caggctgacc	cagtcactga	accacctctt	tacagccatg	gactcgacgg	atgcagacgt	4440
ggtcatctac	tgccgcgaca	aagaatggga	gaagaaaata	tctgaggcca	tacagatgcg	4500
gacccaagta	gagctgctgg	atgagcacat	ctccatagac	tgcgatattg	ttcgcgtgca	4560
ccctgacagc	agcttggcag	gcagaaaagg	atacagcacc	acggaaggcg	cactgtactc	4620
atatctagaa	gggacccgth	ttcatcagac	ggcagtggat	atggcggaga	tacatactat	4680
gtggccaaag	caaacagagg	ccaatgagca	agtctgccta	tatgccctgg	gggaaagtat	4740
tgaatcgatc	aggcagaaat	gcccgggtgga	tgatgcagac	gcatcatctc	cccccaaac	4800
tgtcccgtgc	ctthtgccgth	acgctatgac	tccagaacgc	gtcacccggc	ttcgcgatgaa	4860
ccacgtcaca	agcataattg	tgtgtthcttc	gtthtcccctc	ccaaagtaca	aaatagaagg	4920
agtgcaaaaa	gtcaaatgct	ctaaggtaat	gctattthgac	cacaacgtgc	catcgcgctg	4980
aagtccaagg	gaatatagat	cttcccagga	gtctgcacag	gagggcagta	caatcacgtc	5040
actgacgcat	agtcaattcg	acctaagcgt	tgatggcgag	atactgcccg	tcccgtcaga	5100
cctggatgct	gacgccccag	ccctagaacc	agcactagac	gacggggcga	cacacacgct	5160
gccatccaca	accggaaaacc	ttgcggccgth	gtctgattgg	gtaatgagca	ccgtacctgt	5220
cgcgccgccc	agaagaaggc	gagggagaaa	cctgactgtg	acatgtgacg	agagagaagg	5280
gaatataaca	cccatggcta	gcgtccgatt	ctthtagggca	gagctgtgtc	cggtcgtaca	5340
agaaacagcg	gagacgcgtg	acacagcaat	gtctcttcag	gcaccaccga	gtaccgccac	5400
accgaatcat	ccgccgatct	ccttcggagc	atcaagcgag	acgttccccca	ttacattthg	5460
ggacttcaac	gaaggagaaa	tcgaaaagctt	gtcttctgag	ctactaactt	tcggagactt	5520
cttaccagga	gaagtggatg	acttgacaga	cagcgactgg	tccacgtgct	cagacacgga	5580
cgacgagtha	tgactagaca	gggcaggtgg	gtatatattc	tcgtcggaca	ccggtccagg	5640
tcattttacaa	cagaagtcag	tacgccagtc	agtgtctgcc	gtgaacaccc	tggaggaagt	5700
ccacgaggag	aagtgttacc	cacctaaagct	ggatgaagca	aaggagcaac	tattacttaa	5760

FIGURE 7 (con't)

gaaactccag	gagagtgc	ccatggccaa	cagaagcagg	tatcagtcgc	gcaaagtaga	5820
aaacatgaaa	gcagcaatca	tccagagact	aaagagagggc	tgtagactat	acttaatgtc	5880
agagacccca	aaagtcccta	cttaccggac	tacatatccg	gcgcctgtgt	actcgcctcc	5940
gatcaacgtc	cgattgtcca	atcccagagtc	cgcagtggca	gcatgcaatg	agttcttagc	6000
tagaaactat	ccaactgtct	catcatacca	aattaccgac	gagtatgatg	catatctaga	6060
catggtggac	gggtcggaga	gttgccctgga	ccgagcgcaca	ttcaatccgt	caaaactcag	6120
gagctacccg	aaacagcacg	cttaccacgc	gccctccatc	agaagcgcctg	taccgtcccc	6180
attccagaac	acactacaga	atgtactggc	agcagccacg	aaaagaaact	gcaacgtcac	6240
acagatgagg	gaattaccca	ctttggactc	agcagtattc	aacgtggagt	gtttcaaaaa	6300
attcgcatgc	aaccaagaat	actgggaaga	at ttgctgcc	agccctatta	ggataacaac	6360
tgagaattta	gcaacctatg	ttactaaact	aaaagggcca	aaagcagcag	cgctattcgc	6420
aaaaacccat	aatctactgc	cactacagga	agtaccaatg	gataggttca	cagtagatat	6480
gaaaagggac	gtaaaggtga	ctcctggtag	aaagcataca	gaggaaagac	ctaaggtgca	6540
ggttatacag	gcggctgaac	ccttggcgac	agcataccta	tgtgggattc	acagagagct	6600
ggttaggagg	ctgaacgccg	tcctcctacc	caatgtacat	acactatttg	acatgtctgc	6660
cgaggatttc	gatgccatca	tagccgcaca	ctttaagcca	ggagacactg	ttttggaaac	6720
ggacatagcc	tcctttgata	agagccaaga	tgattcactt	gcgcttactg	ctttgatgct	6780
gttagaggat	ttaggggtgg	atcactccct	gctggacttg	atagaggctg	ctttcggaga	6840
gatttccagc	tgtcacctac	cgacaggtac	gcgcttcaag	ttcggcgcga	tgatgaaatc	6900
aggtatgttc	ctaactctgt	tcgtcaacac	attgttaaac	atcaccatcg	ccagccgagt	6960
gctggaagat	cgtctgacaa	aatccgcgtg	cgcggccttc	atcggcgcag	acaacataat	7020
acatggagtc	gtctccgatg	aattgatggc	agccagatgt	gccacttgga	tgaacatgga	7080
agtgaagatc	atagatgcag	ttgtatcctt	gaaagcccct	tacttttgtg	gagggtttat	7140
actgcacgat	actgtgacag	gaacagcttg	cagagtggca	gacccgctaa	aaaggctttt	7200
taaactgggc	aaaccgctag	cggcaggtga	cgaacaagat	gaagatagaa	gacgagcgct	7260
ggctgacgaa	gtgatcagat	ggcaacgaac	agggctaatt	gatgagctgg	agaaagcggg	7320
atactctagg	tacgaagtgc	agggatatatc	agttgtggta	atgtccatgg	ccacctttgc	7380
aagctccaga	tccaacttcg	agaagctcag	aggacccgtc	ataactttgt	acggcgggtcc	7440
taaataggta	cgcactacag	ctacctat tt	tgcagaagcc	gacagcaagt	atctaaacac	7500
taatcagcta	caatggagtt	catcccaacc	caaacttttt	acaataggag	gtaccagcct	7560
cgaccctgga	ctccgcgccc	tactatccaa	gtcatcaggc	ccagaccgcg	ccctcagagg	7620
caagctgggc	aacttgccca	gctgatctca	gcagttaata	aactgacaat	gcgcgcggta	7680
ccccaacaga	agccacgcag	gaatcggaag	aataagaagc	aaaagcaaaa	acaacaggcg	7740
ccacaaaaca	acacaaatca	aaagaagcag	ccacctaaaa	agaaaccggc	tcaaaaagaaa	7800
aagaagccgg	gccgcagaga	gaggatgtgc	atgaaaatcg	aaaatgattg	tat ttttcgaa	7860
gtcaagcacg	aaggtaaggt	aacaggttac	gcgtgcctgg	tgggggacaa	agtaatgaaa	7920
ccagcacacg	taaaggggac	catcgataac	gcggacctgg	ccaaactggc	ctttaagcgg	7980
tcatctaagt	atgaccttga	atgcgcgcag	ataccctgtc	acatgaagtc	cgacgcttcg	8040
aagttcaccc	atgagaaacc	ggaggggtac	tacaactggc	accacggagc	agtacagtac	8100
tcaggaggcc	ggttcaccat	ccctacaggt	gctggcaaac	caggggacag	cggcagaccg	8160
atcttcgaca	acaagggacg	cgtgggtggc	atagtcttag	gaggagctaa	tgaaggagcc	8220
cgtacagccc	tctcgggtgg	gacctggaat	aaagacattg	tcactaaaat	cacccccgag	8280
ggggccgaag	agtggagtct	tgccatccca	gttatgtgcc	tgttggaaca	caccacgttc	8340
ccctgctccc	agcccccttg	cacgccttgc	tgctacgaaa	aggaaccgga	ggaaacccta	8400
cgcattgctg	aggacaacgt	catgagacct	gggtactatc	agctgctaca	agcctcctta	8460
acatgttctc	cccaccgcca	gcgacgcagc	accaaggaca	acttcaatgt	ctataaagcc	8520
acaagaccat	acttagctca	ctgtcccgac	tgtggagaag	ggcactcgtg	ccatagtccc	8580
gtagcactag	aacgcacacg	aaatgaagcg	acagacggga	cgctgaaaat	ccaggtctcc	8640
ttgcaaatcg	gaataaagac	ggatgacagc	cacgattgga	ccaagctgcg	ttatatggac	8700

FIGURE 7 (con't)

aaccacatgc	cagcagacgc	agagagggcg	gggctatttg	taagaacatc	agcaccgtgt	8760
acgattactg	gaacaatggg	acacttcac	ctggcccgat	gtccaaaagg	ggaaactctg	8820
acgggtgggat	tcactgacag	taggaagatt	agtcactcat	gtacgcaccc	atttcaccac	8880
gaccctcctg	tgataggtcg	ggaaaaattc	cattcccgac	cgcagcacgg	taaagagcta	8940
ccttgcagca	cgtacgtgca	gagcaccgcc	gcaactaccg	aggagataga	ggtacacatg	9000
ccccagaca	cccctgatcg	cacattaatg	tcacaacagt	ccggcaacgt	aaagatcaca	9060
gtcaatggcc	agacgggtcg	gtacaagtgt	aattgcggtg	gctcaaataa	aggactaaca	9120
actacagaca	aagtgattaa	taactgcaag	gttgatcaat	gtcatgccgc	ggtcaccaat	9180
cacaaaaagt	ggcagtataa	ctccccctctg	gtcccgcgta	atgctgaact	tggggaccga	9240
aaaggaaaaa	ttcacatccc	gtttccgctg	gcaaatgtaa	catgcagggt	gcctaaagca	9300
aggaacccca	ccgtgacgta	cgggaaaaaac	caagtcatca	tgctactgta	tcctgaccac	9360
ccaacactcc	tgtcctaccg	gaatatggga	gaagaaccaa	actatcaaga	agagtgggtg	9420
atgcataaga	aggaagtctg	gctaaccgtg	ccgactgaag	ggctcgaggt	cacgtggggc	9480
aacaacgagc	cgtataagta	ttggccgcag	ttatctacaa	acggtacagc	ccatggccac	9540
ccgcatgaga	taattctgta	ttattatgag	ctgtacccca	ctatgactgt	agtagttgtg	9600
tcagtggcca	cgttcatact	cctgtcgatg	gtgggtatgg	cagcggggat	gtgcatgtgt	9660
gcacgacgca	gatgcacac	accgtatgaa	ctgacaccag	gagctaccgt	ccctttcctg	9720
cttagcctaa	tatgctgcat	cagaacagct	aaagcggcca	cataccaaga	ggctgcgata	9780
tacctgtgga	acgagcagca	acctttgttt	tggctacaag	cccttattcc	gctggcagcc	9840
ctgattgttc	tatgcaactg	tctgagactc	ttaccatgct	gctgtaaaac	gttggctttt	9900
ttagccgtaa	tgagcgtcgg	tgcccacact	gtgagcgcgt	acgaacacgt	aacagtgatc	9960
ccgaacacgg	tgggagtacc	gtataagact	ctagtcaata	gacctggcta	cagcccccag	10020
gtattggaga	tggaaactact	gtcagtcact	ttggagccaa	cactatcgct	tgattacatc	10080
acgtgcgagt	acaaaaccgt	catcccgtct	ccgtacgtga	agtgtctcgg	tacagcagag	10140
tgcaaggaca	aaaacctacc	tgactacagc	tgtaaggctc	tcaccggcgt	ctacccattt	10200
atgtggggcg	gcgcctactg	cttctgcgac	gctgaaaaca	cgcagttgag	cgaagcacac	10260
gtggagaagt	ccgaatcatg	caaaacagaa	tttgcatacag	catacagggc	tcataccgca	10320
tctgcatcag	ctaagctccg	cgtcctttac	caaggaaata	acatcactgt	aactgcctat	10380
gcaaacggcg	accatgccgt	cacagttaag	gacgccaaat	tcattgtggg	gccaatgtct	10440
tcagcctgga	cacctttcga	caacaaaatt	gtggtgtaca	aaggtgacgt	ctataacatg	10500
gactacccgc	cctttggcgc	aggaagacca	ggacaatttg	gcgatatcca	aagtcgcaca	10560
cctgagagta	aagacgtcta	tgctaataca	caactgggtac	tgacagagacc	ggctgtgggt	10620
acggtacacg	tgccatactc	tcaggcacca	tctggcttta	agtattgggt	aaaagaacgc	10680
ggggcgctcg	tgacgacac	agcaccattt	ggctgccaaa	tagcaacaaa	cccggtaaga	10740
gcgggtgaact	gcgccgtagg	gaacatgccc	atctccatcg	acataccgga	agcggccttc	10800
actagggctg	tcgacgcgcc	ctcttttaacg	gacatgtcgt	gcgaggtacc	agcctgcacc	10860
cattcctcag	actttggggg	cgtcgccatt	attaaatatg	cagccagcaa	gaaaggcaag	10920
tgtgcgggtg	attcgatgac	taacgccgtc	actattcggg	aagctgagat	agaagttgaa	10980
gggaattctc	agctgcaaata	ctctttctcg	acggccttag	ccagcgccga	attccgcgta	11040
caagtctgtt	ctacacaagt	acactgtgca	gccgagtgcc	accccccgaa	ggaccacata	11100
gtcaactacc	cggcgtcaca	taccaccctc	ggggtccagg	acatctccgc	tacggcgatg	11160
tcattgggtg	agaagatcac	gggaggtgtg	ggactgggtg	ttgctgttgc	cgcactgatt	11220
ctaactcgtg	tgctatgcgt	gtcgttcagc	aggcactaac	ttgacaatta	agtatgaagg	11280
tatatgtgtc	ccctaagaga	cacactgtac	atagcaataa	atctatagat	caaagggtta	11340
cgcaaccctt	gaatagtaac	aaaatacaaa	atcactaaaa	attataaaaa	cagaaaaata	11400
cataaatagg	tatacgtgtc	ccctaagaga	cacattgtat	gtagggtgata	agtatagatc	11460
aaagggccga	ataacccttg	aatagtaaca	aaatatgaaa	atcaataaaa	atcataaaat	11520
agaaaaacca	taaacagaag	tagttcaaag	ggctataaaa	cccctgaata	gtaacaaaac	11580
ataaaaattaa	taaaaatc					11598

FIGURE 8

caaagcaaga	gattaataaac	ccatcatgga	tcctgtgtac	gtggacatag	acgctgacag	60
cgcctttttg	aaggccctgc	aacgtgcgta	cccatgttt	gaggtggaac	caaggcaggt	120
cacaccgaat	gaccatgcta	atgctagagc	gttctcgcat	ctagctataa	aactaataga	180
gcaggaaatt	gaccccgact	caaccatcct	ggatatcggc	agtgcgccag	caaggaggat	240
gatgtcggac	aggaagtacc	actgcgtctg	cccgatgcgc	agtgcggaag	atcccgagag	300
actcgccaat	tatgcgagaa	agctagcatc	tgccgcagga	aaagtcctgg	acagaaacat	360
ctctggaaag	atcggggact	tacaagcagt	aatggccgtg	ccagacacgg	agacgccaac	420
attctgctta	cacacagacg	tctcatgtag	acagagagca	gacgtcgcta	tataccaaga	480
cgtctatgct	gtacacgcac	ccacgtcgct	ataccaccag	gcgattaaag	gggtccgagt	540
ggcgtactgg	gttgggttcg	acacaacccc	gttcatgtac	aatgccatgg	cgggtgccta	600
cccctcatac	tcgacaaact	gggcagatga	gcaggtagctg	aaggctaaga	acataggatt	660
atgttcaaca	gacctgacgg	aaggtagacg	aggcaagttg	tctattatga	gagggaaaaa	720
gctaaaaccg	tgcgaccgtg	tgtgtgtctc	agtaggggtca	acgctctacc	cggaaagccg	780
caagctactt	aagagctggc	acctgccatc	ggtgttccat	ttaaagggca	aactcagctt	840
cacatgccgc	tgtgatacac	tggtttcgtg	tgaggggtac	gtcgttaaga	gaataacgat	900
gagcccaggc	ctttatggaa	aaaccacagg	gtatgcggta	acccaccacg	cagacggatt	960
cctgatgtgc	aagactaccg	acacgggtga	cggcgaaaga	gtgtcattct	cgggtgtgcac	1020
atacgtgccg	gcgaccattt	gtgatcaaata	gaccggcatc	cttgctacag	aagtcacgcc	1080
ggaggatgca	cagaagctgt	tgggtggggct	gaaccagaga	atagtgggtta	acggcagAAC	1140
gcaacggaat	acgaacacca	tgaaaaatta	tctgcttccc	gtggtcgccc	aagccttcag	1200
taagtgggca	aaggagtgcc	ggaaagacat	ggaagatgaa	aaactcctgg	gggtcagaga	1260
aagaacactg	acctgctgct	gtctatgggc	attcaagaag	cagaaaacac	acacggtcta	1320
caagagacct	gataccagct	caattcagaa	gggttcaggcc	gagtttgaca	gctttgtggg	1380
accgagtctg	tggtcgtccg	ggttgtcaat	ccctttgagg	actagaatca	aatgggttgt	1440
aagcaaggtg	ccaaaaaccg	acctgatccc	atacagcgga	gacgcccag	aagcccggga	1500
cgcagaaaaa	gaagcagagg	aagaacgaga	agcagaactg	actcgcgagg	ccctaccacc	1560
tctacaggca	gcacaggaag	atgttcagggt	cgaatcgac	gtggaacagc	ttgaggacag	1620
agcgggcgca	ggaataatag	agactccgag	aggagctatc	aaagttactg	cccaaccaac	1680
agaccacgtc	gtgggagagt	acctggtact	ctccccgcag	accgtactac	gtagccagaa	1740
gctcagtctg	attcacgctt	tggcggagca	agtgaagacg	tgcacgcaca	acggacgagc	1800
agggaggtat	gcggtcgaag	cgtacgacgg	ccgagtccta	gtgccctcag	gctatgcaat	1860
ctcgcctgaa	gacttccaga	gtctaagcga	aagcgcaacg	atgggtgtata	acgaaagaga	1920
gttcgtaaac	agaaagctac	accatattgc	gatgcacgga	ccagccctga	acaccgacga	1980
agagtcgtat	gagctggtga	gggcagagag	gacagaacac	gagtacgtct	acgacgtgga	2040
tcagagaaga	tgtgtgaaga	aggaagaagc	cgcaggactg	gtactgggtg	gcgacttgac	2100
taatccgccc	taccacgaat	tcgcatatga	agggctaaaa	atccgccttg	cctgcccata	2160
caaaattgca	gtcataggag	tcttcggagt	accgggatct	ggcaagtcag	ctattatcaa	2220
gaacctaagt	accaggcagg	acctggtgac	tagcggaaag	aaagaaaact	gccaagaaat	2280
caccaccgac	gtgatgagac	agagaggtct	agagatatct	gcacgtacgg	ttgactcgct	2340
gctcttgaat	ggatgcaaca	gaccagtcga	cgtgttgtac	gtagacgagg	cgtttgcgtg	2400
ccactctgga	acgctacttg	ctttgatcgc	cttgggtgaga	ccaaggcaga	aagttgtact	2460
ttgtgggtgac	ccgaagcagt	gcggcttctt	caatatgatg	cagatgaaag	tcaactataa	2520
tcacaacatc	tgcacccaag	tgtaccacaa	aagtatctcc	aggcgggtga	cactgcctgt	2580
gaccgccatt	gtgtcatcgt	tgcattacga	aggcaaaatg	cgcactacga	atgagtacaa	2640
caagccgatt	gtagtggaca	ctacaggctc	aacaaaacct	gaccctggag	acctcgtgtt	2700
aacgtgcttc	agagggtggg	ttaaacaact	gcaaattgac	tatcgtggat	acgaggtcat	2760
gacagcagcc	gcatcccaag	ggttaaccag	aaaaggagtt	tacgcagtta	gacaaaaagt	2820
taatgaaaac	ccgctctatg	catcaacgtc	agagcacgtc	aacgtactcc	taacgcgtac	2880
ggaaggtaaa	ctgggtatgga	agacactttc	cggcgacccg	tggataaaga	cgctgcagaa	2940

FIGURE 8 (con't)

cccaccgaaa	ggaaacttca	aagcaactat	taaggagtgg	gaggtggagc	atgcatcaat	3000
aatggcgggc	atctgcagtc	accaaatagac	cttcgatata	ttccaaaata	aagccaacgt	3060
ttgttgggct	aagagcttgg	tccttatcct	cgaaacagcg	gggataaaac	taaatgatag	3120
gcagtggctc	cagataattc	aagccttcaa	agaagacaaa	gcatactcac	ctgaagtagc	3180
cctgaatgaa	atatgtacgc	gcattgtatg	ggtggatcta	gacagcgggc	tattttctaa	3240
accgttgggtg	tctgtgtatt	acgcggataa	ccactgggat	aataggcctg	gagggaaaat	3300
gttcggattt	aaccccgagg	cagcatccat	tctagaaaga	aagtatccat	tcacaaaagg	3360
gaagtggaaac	atcaacaagc	agatctgcgt	gactaccagg	aggatagaag	actttaaccc	3420
taccaccaac	atcataccgg	ccaacaggag	actaccacac	tcattagtgg	ccgaacaccg	3480
cccagtaaaa	ggggaaagaa	tggaatggct	ggttaacaag	ataaacggcc	accacgtgct	3540
cctggtcagt	ggctataacc	ttgcactgcc	tactaagaga	gtcacttggg	tagcgccgtt	3600
aggtgtccgc	ggagcggact	acacatacaa	cctagagtgt	ggtctgccag	caacgcttgg	3660
taggtatgac	ctagtgggtc	taaacatcca	cacacctttt	cgcatacacc	attaccaaca	3720
gtgcgtcgac	cacgcaatga	aactgcaaata	gctcgggggt	gactcattga	gactgctcaa	3780
accgggcggc	tctctattga	tcagagcata	tgggttacgca	gatagaacca	gtgaacgagt	3840
catctgcgta	ttgggacgca	agtttagatc	gtctagagcg	ttgaaaccac	catgtgtcac	3900
cagcaaacact	gagatgtttt	tcctattcag	caactttgac	aatggcagaa	ggaatttcac	3960
aactcatgtc	atgaacaatc	aactgaatgc	agccttcgta	ggacagggtc	cccagcagg	4020
atgtgcaccg	tcgtaccggg	taaaacgcata	ggacatcgcg	aagaacgatg	aagagtgcgt	4080
agtcacacgc	gctaaccctc	gcgggttacc	gggtgacggg	gtttgcaagg	cagtatacaa	4140
aaaatggccg	gagtccttta	agaacagtgc	aacaccagtg	ggaaccgcaa	aaacagttat	4200
gtgcggtacg	tatccagtaa	tcacacgctg	tggaccaaac	ttctctaatt	attcggagtc	4260
tgaaggggac	cgggaattgg	cagctgccta	tcgagaagtc	gcaaaggaag	taactaggct	4320
gggagtaaat	agtgtagcta	tacctctcct	ctccacagg	gtatactcag	gagggaaaga	4380
caggctgacc	cagtcactga	accacctctt	tacagccatg	gactcgacgg	atgcagacgt	4440
ggtcactctac	tgccgcgaca	aagaatggga	gaagaaaata	tctgaggcca	tacagatgcg	4500
gacccaagta	gagctgctgg	atgagcacat	ctccatagac	tgcgatattg	ttcgcgtgca	4560
ccctgacagc	agcttggcag	gcagaaaagg	atacagcacc	acggaaggcg	cactgtactc	4620
atatctagaa	gggacccggt	ttcatcagac	ggctgtggat	atggcggaga	tacatactat	4680
gtggccaaag	caaacagagg	ccaatgagca	agtctgccta	tatgccctgg	gggaaagtat	4740
tgaatcgatc	aggcagaaat	gcccgggtgga	tgatgcagac	gcatactctc	cccccaaac	4800
tgteccgtgc	ctttgcccgt	acgctatgac	tcagaacgc	gtcacccggc	ttcgcatgaa	4860
ccacgtcaca	agcataattg	tgtgttcttc	gtttcccttc	ccaaagtaca	aaatagaagg	4920
agtgcaaaaa	gtcaaatagt	ctaaggtaat	gctatttgac	cacaacgtgc	catcgcgctg	4980
aagtccaagg	gaatatagat	cttcccagga	gtctgcacag	gagggcagta	caatcacgtc	5040
actgacgcat	agtcaattcg	acctaagcgt	tgatggcgag	atactgcccg	tcccgtcaga	5100
cctggatgct	gacgccccag	ccctagaacc	agcactagac	gacggggcga	cacacacgct	5160
gccatccaca	accggaaacc	ttgcggccgt	gtctgatttg	gtaatgagca	ccgtacctgt	5220
cgcgccgccc	agaagaaggc	gagggagaaa	cctgactgtg	acatgtgacg	agagagaagg	5280
gaatataaca	cccatggcta	gcgtccgatt	ctttagggca	gagctgtgtc	cggtcgtaca	5340
agaaacagcg	gagacgcgtg	acacagcaat	gtctcttcag	gcaccaccga	gtaccgccac	5400
ggaaccgaat	catccgccga	tctccttcgg	agcatcaagc	gagacgttcc	ccattacatt	5460
tggggacttc	aacgaaggag	aaatcgaaag	cttgtcttct	gagctactaa	ctttcggaga	5520
cttcttacca	ggagaagtgg	atgacttgac	agacagcgac	tgggtccacgt	gctcagacac	5580
ggacgacgag	ttatgactag	acagggcagg	tgggtatata	ttctcgtcgg	acaccggtcc	5640
aggtcattta	caacagaagt	cagtacgcca	gtcagtgtctg	ccggtgaaca	ccctggagga	5700
agtccacgag	gagaagtgtt	acccacctaa	gctggatgaa	gcaaaggagc	aactattact	5760
taagaaactc	caggagagtg	catccatggc	caacagaagc	aggtatcagt	cgcgcaaagt	5820
agaaaacatg	aaagcagcaa	tcatccagag	actaaagaga	ggctgtagac	tatacttaat	5880

FIGURE 8 (con't)

gtcagagacc	ccaaaagtcc	ctacttaccg	gactacatat	ccggcgcttg	tgtactcgcc	5940
tccgatcaac	gtccgattgt	ccaatcccga	gtccgcagtg	gcagcatgca	atgagttctt	6000
agctagaaac	tatccaactg	tctcatcata	ccaaattacc	gacgagtatg	atgcatactt	6060
agacatgggtg	gacgggtcgg	agagttgcct	ggaccgagcg	acattcaatc	cgtcaaaact	6120
caggagctac	ccgaaacagc	acgcttacca	cgcgccctcc	atcagaagcg	ctgtaccgtc	6180
cccattccag	aacacactac	agaatgtact	ggcagcagcc	acgaaaagaa	actgcaacgt	6240
cacacagatg	aggggaattac	ccacttttga	ctcagcagta	ttcaacgtgg	agtgtttcaa	6300
aaaatttcgca	tgcaaccaag	aatactggga	agaatttgct	gccagcccta	ttaggataac	6360
aactgagaat	ttagcaacct	atgttactaa	actaaaaggg	ccaaaagcag	cagcgctatt	6420
cgcaaaaacc	cataatctac	tgccactaca	ggaagtacca	atggataggt	tcacagtaga	6480
tatgaaaagg	gacgtaaagg	tgactcctgg	tacaaagcat	acagaggaaa	gacctaaagg	6540
gcaggttata	caggcggtcg	aacccttggc	gacagcatac	ctatgtggga	ttcacagaga	6600
gctgggttagg	aggctgaacg	ccgtcctcct	acccaatgta	catacactat	ttgacatgtc	6660
tgccgaggat	ttcgatgcca	tcatagccgc	acactttaag	ccaggagaca	ctgtttttga	6720
aacggacata	gcctcctttg	ataagagcca	agatgattca	cttgccgctta	ctgctttgat	6780
gctgttagag	gatttagggg	tggatcactc	cctgctggac	ttgatagagg	ctgctttcgg	6840
agagatttcc	agctgtcacc	taccgacagg	tacgcgcttc	aagttcggcg	ccatgatgaa	6900
atcaggtatg	ttcctaactc	tgttcgtcaa	cacattgtta	aacatcacca	tcgccagccg	6960
agtgtctgaa	gacgtcttga	caaaatccgc	gtgcgcggcc	ttcatcggcg	acgacaacat	7020
aatacatgga	gtcgtctccg	atgaattgat	ggcagccaga	tgtgccactt	ggatgaacat	7080
ggaagtgaag	atcatagatg	cagttgtatc	cttgaaagcc	ccttactttt	gtggaggggt	7140
tatactgcac	gatactgtga	caggaacagc	ttgcagagtg	gcagaccgcg	taaaaaggct	7200
ttttaaactg	ggcaaaccgc	tagcggcagg	tgacgaacaa	gatgaagata	gaagacgagc	7260
gctggctgac	gaagtgatca	gatggcaacg	aacagggcta	attgatgagc	tgagaaaagc	7320
ggtatactct	aggtacgaag	tgcaagggtat	atcagttgtg	gtaatgtcca	tggtccacct	7380
tgcaagctcc	agatccaact	tcgagaagct	cagaggaccc	gtcataactt	tgtacggcgg	7440
tcctaaatag	gtacgcacta	cagctaccta	ttttgcagaa	gccgacagca	agtatctaaa	7500
cactaatcag	ctacaatgga	gttcatecca	acccaaactt	tttacaatag	gaggtaccag	7560
cctcgaccct	ggactccgcg	ccctactatc	caagtcatca	ggcccagacc	gcgccctcag	7620
aggcaagctg	ggcaacttgc	ccagctgatc	tcagcagtta	ataaactgac	aatgcgcgcg	7680
gtaccccaac	agaagccacg	caggaatcgg	agaataaga	agcaaaagca	aaaacaacag	7740
gcgccacaaa	acaacacaaa	tcaaaagaag	cagccaccta	aaaagaaacc	ggctcaaaag	7800
aaaaagaagc	cgggccgcag	agagaggatg	tgcatgaaaa	tcgaaaatga	ttgtattttc	7860
gaagtcaagc	acgaaggtaa	ggtaacaggt	tacgcgtgcc	tggtggggga	caaagtaatg	7920
aaaccagcac	acgtaaaggg	gaccatcgat	aacgcggacc	tggtccaaact	ggcctttaag	7980
cggtcactta	agtatgacct	tgaatgcgcg	cagatacccg	tgccatgaa	gtccgacgct	8040
tcgaagttca	cccatgagaa	accggagggg	tactacaact	ggcaccacgg	agcagtacag	8100
tactcaggag	gccggttcac	catccctaca	ggtgctggca	aaccagggga	cagcggcaga	8160
ccgatcttcg	acaacaaggg	acgcgtgggt	gccatagttt	taggaggagc	taatgaagga	8220
gcccgtagac	ccctctcggt	ggtgacctgg	aataaagaca	ttgtcactaa	aatcaccccc	8280
gagggggccg	aagagtggag	tcttgccatc	ccagttatgt	gcctgttggc	aaacaccacg	8340
ttccccctgct	cccagccccc	ttgcacgccc	tgctgctacg	aaaaggaacc	ggaggaaacc	8400
ctacgcattgc	ttgaggacaa	cgtcatgaga	cctgggtact	atcagctgct	acaagcatcc	8460
ttaacatggt	ctccccaccg	ccagcgacgc	agcaccaagg	acaacttcaa	tgtctataaa	8520
gccacaagac	catacttagc	tcactgtccc	gactgtggag	aagggcactc	gtgccatagt	8580
cccgtagcac	tagaacgcat	cagaaatgaa	gcgacagacg	ggacgctgaa	aatccagggtc	8640
tccttgcaaa	tcggaataaa	gacggatgac	agccacgatt	ggaccaagct	gcgttatatg	8700
gacaaccaca	tgccagcaga	cgcagagagg	gcggggctat	ttgtaaagaa	atcagcaccg	8760
tgtacgatta	ctggaacaat	gggacacttc	atcctggccc	gatgtccaaa	aggggaaact	8820

FIGURE 8 (con't)

ctgacgggtgg	gattcactga	cagtaggaag	attagtcact	catgtacgca	cccatttcac	8880
cacgaccctc	ctgtgatagg	tccgggaaaa	ttccattccc	gaccgcagca	cggtaaagag	8940
ctaccttgca	gcacgtacgt	gcagagcacc	gccgcaacta	ccgaggagat	agaggtagac	9000
atgccccag	acacccctga	tccgacatta	atgtcacaa	agtccggcaa	cgtaaagatc	9060
acagtcaatg	gccagacggg	gcggtacaag	tgttaattgcg	gtgggtcaaa	tgaaggacta	9120
acaactacag	acaaagtgat	taataactgc	aagggttgatc	aatgtcatgc	cgcggtcacc	9180
aatcacaaaa	agtggcagta	taactccctc	ctgggtccgc	gtaatgctga	acttggggac	9240
cgaaaaggaa	aaattcacat	cccgtttccg	ctggcaaagt	taacatgcag	gggtgcctaaa	9300
gcaaggaacc	ccaccgtgac	gtacgggaaa	aaccaagtca	tcatgctact	gtatcctgac	9360
cacccaacac	tcctgtccta	ccggaatatg	ggagaagaac	caaactatca	agaagagtgg	9420
gtgatgcata	agaaggaagt	cgtgctaacc	gtgccgactg	aagggctcga	ggtcacgtgg	9480
ggcaacaacg	agccgtataa	gtattggccg	cagttatcta	caaacggtag	agcccatggc	9540
cacccgcacg	agataattct	gtattattat	gagctgtacc	ccactatgac	tgtagtagtt	9600
gtgtcagtg	ccacgttcat	actcctgtcg	atgggtgggt	tggcagcggg	gatgtgcatg	9660
tgtgcacgac	gcagatgcat	cacaccgtat	gaactgacac	caggagctac	cgtccctttc	9720
ctgcttagcc	taatatgctg	catcagaaca	gctaaagcgg	ccacatacca	agaggctgcg	9780
atatacctgt	ggaacgagca	gcaacctttg	ttttggctac	aagcccttat	tccgctggca	9840
gccctgattg	ttctatgcaa	ctgtctgaga	ctcttaccat	gctgctgtaa	aacgttggct	9900
tttttagccg	taatgagcgt	cgggtgccac	actgtgagcg	cgtacgaaca	cgtaacagtg	9960
atcccgaaca	cgggtgggagt	accgtataag	actctagtca	atagacctgg	ctacagcccc	10020
atgggtattg	agatggaact	actgtcagtc	actttggagc	caacactatc	gcttgattac	10080
atcacgtgcg	agtacaaaac	cgtcatcccg	tctccgtacg	tgaagtgctg	cggtagacga	10140
gagtgcgaag	acaaaaacct	acctgactac	agctgtaagg	tcttcaccgg	cgtctacca	10200
tttatgtggg	gcgggcgccta	ctgcttctgc	gacgctgaaa	acacgcagtt	gagcgaagca	10260
cacgtggaga	agtcogaatc	atgcaaaaca	gaatttgcat	cagcatacag	ggctcatacc	10320
gcattctgcat	cagctaagct	ccgcgtcctt	taccaaggaa	ataacatcac	tgttaactgcc	10380
tatgcaaacg	gcgaccatgc	cgtcacagtt	aaggacgcca	aattcattgt	ggggccaatg	10440
tcttcagcct	ggacaccttt	cgacaacaaa	attgtgggtg	acaaagggtg	cgtctataac	10500
atggactacc	cgcccttttg	cgcaggaaga	ccaggacaat	ttggcgatat	ccaaagtccg	10560
acacctgaga	gtaaagacgt	ctatgcta	acacaactgg	tactgcagag	accggctgtg	10620
ggtagcgtac	acgtgccata	ctctcaggca	ccatctggct	ttaagtattg	gctaaaagaa	10680
cgcggggcgt	cgtgcagca	cacagcacca	tttggtgccc	aaatagcaac	aaacccggta	10740
agagcgggtg	actgcgcctg	agggaaacatg	cccattctcca	tcgacatacc	ggaagcggcc	10800
ttcactaggg	tgcgcagcgc	gccctcttta	acggacatgt	cgtgcgaggt	accagcctgc	10860
accatttct	cagacttttg	gggcgtcgcc	attattaaat	atgcagccag	caagaaaggc	10920
aagtgtgcgg	tgcattcgat	gactaacgcc	gtcactattc	gggaagctga	gatagaagtt	10980
gaagggaatt	ctcagctgca	aatctctttc	tcgacggcct	tagccagcgc	cgaattccgc	11040
gtacaagtct	gttctacaca	agtacactgt	gcagccgagt	gccaccccc	gaaggaccac	11100
atagtcaact	acccggcgtc	acataccacc	ctcgggggtcc	aggacatctc	cgtacggcg	11160
atgtcatggg	tgcagaagat	cacggggaggt	gtgggactgg	ttgttgctgt	tgcgcgactg	11220
attctaactg	tgggtgctatg	cgtgtcggtc	agcaggcact	aacttgacaa	ttaagtatga	11280
aggtatatgt	gtcccctaag	agacacactg	tacatagcaa	ataatctata	gatcaaaggg	11340
ctacgcaacc	cctgaatagt	aacaaaatat	aaaatcacta	aaaattataa	aaacagaaaa	11400
atacataaat	aggtatacgt	gtcccctaag	agacacattg	tatgtagggtg	ataagtatag	11460
atcaaagggc	cgaataaccc	ctgaatagta	acaaaatatg	aaaatcaata	aaaatcataa	11520
aatagaaaaa	ccataaacag	aagtagttca	aagggtctata	aaaccctga	atagtaacaa	11580
aacataaaat	taataaaaat	c				11601

FIGURE 9

cagtttctta	ctgctctact	ctgcaaagca	agagattaat	aacccatcat	ggatcctgtg	60
tacgtggaca	tagacgctga	cagcgccttt	ttgaaggccc	tgcaacgtgc	gtaccccatg	120
tttgaggtgg	aaccaaggca	ggtcacaccg	aatgaccatg	ctaatagctag	agcgttctcg	180
catctagcta	taaaactaat	agagcaggaa	attgaccccg	actcaaccat	cctggatatc	240
ggcagtgcgc	cagcaaggag	gatgatgtcg	gacaggaagt	accactgcgt	ctgcccgatg	300
cgcagtgcgg	aagatcccga	gagactcgcc	aattatgcga	gaaagctagc	atctgccgca	360
ggaaaagtcc	tggacagaaa	catctctgga	aagatcgggg	acttacaagc	agtaatggcc	420
gtgccagaca	cggagacgcc	aacattctgc	ttacacacag	acgtctcatg	tagacagaga	480
gcagacgtcg	ctatatacca	agacgtctat	gctgtacacg	caccacgtc	gctataccac	540
caggcgatta	aaggggtccg	agtggcgtag	tgggttgggt	tgcacacaac	cccgttcatg	600
tacaatgcca	tggcgggtgc	ctacccctca	tactcgacaa	actgggcaga	tgagcaggta	660
ctgaaggcta	agaacatagg	attatgttca	acagacctga	cggaaggtag	acgaggcaag	720
ttgtctatta	tgagagggaa	aaagctaaaa	ccgtgcgacc	gtgtgctgtt	ctcagtaggg	780
tcaacgctct	acccggaaag	ccgcaagcta	cttaagagct	ggcacctgcc	atcgggtgtc	840
cattttaaagg	gcaaactcag	cttcacatgc	cgtgtgtgata	cagtggtttc	gtgtgagggc	900
tacgtcgtta	agagaataac	gatgagccca	ggcctttatg	gaaaaaccac	agggtatgcg	960
gtaacccacc	acgcagacgg	attcctgatg	tgcaagacta	ccgacacggt	tgacggcgaa	1020
agagtgtcat	tctcgggtgtg	cacatacgtg	ccggcgacca	tttgtgatca	aatgaccggc	1080
atccttgcta	cagaagtcac	gccggaggat	gcacagaagc	tgttggtggg	gctgaaccag	1140
agaatagtgg	ttaacggcag	aacgcaacgg	aatacgaaca	ccatgaaaaa	ttatctgctt	1200
cccgtggctg	cccaagcctt	cagtaagtgg	gcaaaggagt	gccggaaaga	catggaagat	1260
gaaaaactcc	tgggggtcag	agaaagaaca	ctgacctgct	gctgtctatg	ggcattcaag	1320
aagcagaaaa	cacacacggt	ctacaagagg	cctgataccc	agtcaattca	gaaggttcag	1380
gccgagtttg	acagcttttg	ggtaccgagt	ctgtggtcgt	ccgggttgtc	aatccctttg	1440
aggactagaa	tcaaattggtt	gttaagcaag	gtgccaaaaa	ccgacctgat	cccatacagc	1500
ggagacgccc	gagaagcccg	ggacgcagaa	aaagaagcag	aggaagaacg	agaagcagaa	1560
ctgactcgcg	aagccctacc	acctctacag	gcagcacagg	aagatgttca	ggtcgaaatc	1620
gacgtggaac	agcttgagga	cagagcgggc	gcaggaataa	tagagactcc	gagaggagct	1680
atcaaagtta	ctgcccaccc	aacagaccac	gtcgtgggag	agtacctggt	actctccccg	1740
cagaccgtac	tacgtagcca	gaagctcagt	ctgattcacg	ctttggcgga	gcaagtgaag	1800
acgtgcacgc	acaacggacg	agcagggagg	tatgcggtcg	aagcgtacga	cggccgagtc	1860
ctagtgcctt	caggctatgc	aatctcgcct	gaagacttcc	agagtctaag	cgaaagcgca	1920
acgatgggtg	ataacgaaag	agagttcgta	aacagaaagc	tacaccatat	tgcgatgcac	1980
ggaccagccc	tgaacaccga	cgaagagtcg	tatgagctgg	tgagggcaga	gaggacagaa	2040
cacgagtacg	tctacgacgt	ggatcagaga	agatgctgta	agaaggaaga	agccgcagga	2100
ctgggtactgg	tgggcgactt	gactaatccg	ccctaccacg	aattcgcata	tgaagggcta	2160
aaaatccgcc	ctgcctgccc	atacaaaatt	gcagtcatag	gagtcttcgg	agtaccggga	2220
tctggcaagt	cagctattat	caagaacctt	gttaccaggc	aggacctggt	gactagcgga	2280
aagaaagaaa	actgccaaga	aatcaccacc	gacgtgatga	gacagagagg	tctagagata	2340
tctgcacgta	cggttgactc	gctgctcttg	aatggatgca	acagaccagt	cgacgtgttg	2400
tacgtagacg	aggcgtttgc	gtgccactct	ggaacgctac	ttgctttgat	cgccttggtg	2460
agaccaaggc	agaaagtgtg	actttgtggg	gaccogaagc	agtgcggctt	cttcaatatg	2520
atgcagatga	aagtcaacta	taatcacaac	atctgcaccc	aagtgtacca	caaaagtatc	2580
tccaggcggt	gtacactgcc	tgtgaccgcc	attgtgtcat	cgttgcatca	cgaaggcaaa	2640
atgcgcacta	cgaatgagta	caacaagccg	attgtagtgg	acactacagg	ctcaacaaaa	2700
cctgaccctg	gagacctcgt	gttaacgtgc	ttcagagggg	gggttaaaaa	actgcaaatt	2760
gactatcggt	gatacgaggt	catgacagca	gccgcacccc	aagggttaac	cagaaaagga	2820
gtttacgcag	ttagacaaaa	agttaatgaa	aaccgcgtct	atgcatcaac	gtcagagcac	2880

FIGURE 9 (con't)

gtcaacgtac	tcctaacgcg	tacggaaggt	aaactgggtat	ggaagacact	ttccggcgac	2940
ccgtggataa	agacgctgca	gaacccaccg	aaaggaaact	tcaaagcaac	tattaaggag	3000
tgggaggtgg	agcatgcatc	aataatggcg	ggcatctgca	gtcaccaaact	gaccttcgat	3060
acattccaaa	ataaagccaa	cgtttgttgg	gctaagagct	tggtccctat	cctcgaaaca	3120
gcgggggataa	aactaaatga	taggcagtgg	tctcagataa	ttcaagcctt	caaagaagac	3180
aaagcatact	cacctgaagt	agccctgaat	gaaatatgta	cgcgcagtga	tgggggtggat	3240
ctagacagcg	ggctattttc	taaaccgttg	gtgtctgtgt	attacgcgga	taaccactgg	3300
gataataggc	ctggagggaa	aatgttcgga	tttaaccccg	aggcagcatc	cattctagaa	3360
agaaagtatc	cattcacaaa	agggaaagtgg	aacatcaaca	agcagatctg	cgtgactacc	3420
aggaggatag	aagactttta	ccctaccacc	aacatcatac	cggccaacag	gagactacca	3480
cactcattag	tggccgaaca	ccgccagta	aaaggggaaa	gaatggaatg	gctgggttaac	3540
aagataaacg	gccaccacgt	gctcctggtc	agtggctata	accttgcaact	gcctactaag	3600
agagtcactt	gggtagcgcc	gttaggtgtc	cgcggagcgg	actacacata	caacctagag	3660
ttgggtctgc	cagcaacgct	tggtaggtat	gacctagtgg	tcataaacat	ccacacacct	3720
tttcgcatac	accattacca	acagtgcgtc	gaccacgcaa	tgaaactgca	aatgctcggg	3780
ggtgactcat	tgagactgct	caaaccgggc	ggctctctat	tgatcagagc	atatggttac	3840
gcagatagaa	ccagtgaacg	agtcactctgc	gtattgggac	gcaagtttag	atcgtctaga	3900
gcgttgaaac	caccatgtgt	caccagcaac	actgagatgt	ttttcctatt	cagcaacttt	3960
gacaatggca	gaaggaattt	cacaactcat	gtcatgaaca	atcaactgaa	tgcagccttc	4020
gtaggacagg	tcacccgagc	aggatgtgca	ccgtcgtacc	gggtaaaacg	catggacatc	4080
gcgaagaacg	atgaagagtg	cgtagtcaac	gccgctaacc	ctcgcggggt	accgggtgac	4140
ggtgttttgc	aggcagtata	caaaaaatgg	ccggagtcct	ttaagaacag	tgcaacacca	4200
gtgggaaccg	caaaaacagt	tatgtgcggt	acgtatccag	taatccacgc	tgttggacca	4260
aacttctcta	attattcgga	gtctgaaggg	gaccgggaat	tggcagctgc	ctatcgagaa	4320
gtcgcaaagg	aagtaactag	gctgggagta	aatagtgtag	ctatacctct	cctctccaca	4380
ggtgtatact	caggagggaa	agacaggctg	accagtcac	tgaaccacct	ctttacagcc	4440
atggactcga	cggatgcaga	cgtggtcatc	tactgccgcg	acaaagaatg	ggagaagaaa	4500
atatctgagg	ccatacagat	gcggacccaa	gtagagctgc	tggatgagca	catctccata	4560
gactgcgata	ttgttcgcgt	gcaccctgac	agcagcttgg	caggcagaaa	aggatacagc	4620
accacggaag	gcgcactgta	ctcatatcta	gaagggaccc	gttttcatca	gacggctgtg	4680
gatatggcgg	agatacatat	tatgtggcca	aagcaaacag	aggccaatga	gcaagtctgc	4740
ctatatgccc	tgggggaaaag	tattgaatcg	atcaggcaga	aatgcccggg	ggatgatgca	4800
gacgcatcat	ctccccccaa	aactgtcccg	tgcctttgcc	gttacgctat	gactccagaa	4860
cgcgtcacc	ggcttcgcat	gaaccacgtc	acaagcataa	ttgtgtgttc	ttcgtttccc	4920
ctcccaaagt	acaaaataga	aggagtgcaa	aaagtcaaact	gctctaaggt	aatgctattt	4980
gaccacaacg	tgccatcgcg	cgtaagtcca	agggaatata	gatctttcca	ggagtctgca	5040
caggaggcga	gtacaatcac	gtcactgacg	catagtcaat	tcgacctaa	cgttgatggc	5100
gagatactgc	ccgtcccgtc	agacctggat	gctgacgccc	cagccctaga	accagcacta	5160
gacgacgggg	cgacacacac	gctgccatcc	acaaccggaa	accttgccgc	cgtgtctaga	5220
cagggcaggt	gggtatatat	tctcgtcgga	caccgggtcca	ggtcattttac	aacagaagtc	5280
agtacgccag	tcagtgtctgc	cgggtgaacac	cctggaggaa	gtccacgagg	agaagtgtta	5340
cccacctaag	ctggatgaag	caaaggagca	actattactt	aagaaactcc	aggagagtgc	5400
atccatggcc	aacagaagca	ggtatcagtc	gcgcaaagta	gaaaacatga	aagcagcaat	5460
catccagaga	ctaaagagag	gctgtagact	atacttaatg	tcagagaccc	caaaagtccc	5520
tacttaccgg	actacatatc	cggcgctgtg	gtactcgctt	cogatcaacg	tccgattgtc	5580
caatcccag	tccgcagtgg	cagcatgcaa	tgagttctta	gctagaaact	atccaactgt	5640
ctcatcatac	caaattaccg	acgagtatga	tgcatactta	gacatgggtg	acgggtcgga	5700
gagttgcctg	gaccgagcga	cattcaatcc	gtcaaaactc	aggagctacc	cgaaacagca	5760
cgttaccac	gcgcctcca	tcagaagcgc	tgtaccgtcc	ccattccaga	acacactaca	5820

FIGURE 9 (con't)

gaatgtactg	gcagcagcca	cgaaaagaaa	ctgcaacgtc	acacagatga	gggaattacc	5880
cacttttgac	tcagcagtat	tcaacgtgga	gtgtttcaaa	aaattcgcac	gcaaccaaga	5940
atactgggaa	gaatttgctg	ccagccctat	taggataaca	actgagaatt	tagcaacctc	6000
tgttactaaa	ctaaaagggc	caaaagcagc	agcgctattc	gcaaaaaccc	ataatctact	6060
gccactacag	gaagtaccaa	tggatagggt	cacagtagat	atgaaaaggg	acgtgaaggt	6120
gactcctggt	acaaagcata	cagaggaaag	acctaagggt	cagggttatac	aggcggctga	6180
acccttggcg	acagcatacc	tatgtgggat	tcacagagag	ctggttagga	ggctgaacgc	6240
cgtcctccta	cccaatgtac	atacactatt	tgacatgtct	gccgaggatt	tcgatgccat	6300
catagccgca	cactttaagc	caggagacac	tgttttgga	acggacatag	cctcctttga	6360
taagagccaa	gatgattcac	ttgcgcttac	tgccttgatg	ctgttagagg	atttaggggt	6420
ggatcactcc	ctgctggact	tgatagaggc	tgccttcgga	gagatttcca	gctgtcacct	6480
accgacaggt	acgcgcttca	agttcggcgc	catgatgaaa	tcaggatatgt	tcctaactct	6540
gttcgtcaac	acattgttaa	acatcaccat	cgccagccga	gtgctggaag	atcgtctgac	6600
aaaatccgcg	tgcgcggcct	tcatcggcga	cgacaacata	atacatggag	tcgtctccga	6660
tgaattgatg	gcagccagat	gtgccacttg	gatgaacatg	gaagtgaaga	tcatagatgc	6720
agttgtatcc	ttgaaagccc	cttacttttg	tggaggggtt	atactgcacg	atactgtgac	6780
aggaacagct	tgcagagtgg	cagaccgcgt	aaaaaggcct	tttaaaactgg	gcaaaccgct	6840
agcggcaggt	gacgaacaag	atgaagatag	aagacgagcg	ctggctgacg	aagtgatcag	6900
atggcaacga	acagggctaa	ttgatgagct	ggagaaagcg	gtatactcta	ggtacgaagt	6960
gcaggggtata	tcagttgtgg	taatgtccat	ggccaccttt	gcaagctcca	gatccaactt	7020
cgagaagctc	agaggacccg	tcataacttt	gtacggcggt	cctaaatagg	tacgcactac	7080
agctacctat	tttgcagaag	ccgacagcaa	gtatctaaac	actaatcagc	tacaatggag	7140
ttcatcccaa	cccaaacttt	ttacaatagg	aggtaccagc	ctcgaccctg	gactccgcgc	7200
cctactatcc	aagtcatcag	gcccagaccg	cgccctcaga	ggcaagctgg	gcaacttgcc	7260
cagctgatct	cagcagttaa	taaactgaca	atgcgcgcgg	tacccaaca	gaagccacgc	7320
aggaatcgga	agaataagaa	gcaaaagcaa	aaacaacagg	cgccacaaaa	caacacaaat	7380
caaaagaagc	agccacctaa	aaagaaaccg	gctcaaaaga	aaaagaagcc	gggccgcaga	7440
gagaggatgt	gcatgaaaat	cgaaaatgat	tgtattttcg	aagtcaagca	cgaaggtaag	7500
gtaacagggt	acgcgtgcct	ggtggggggac	aaagtaatga	aaccagcaca	cgtaaagggg	7560
accatcgata	acgcggacct	ggccaaactg	gcctttaagc	ggtcatctaa	gtatgacctt	7620
gaatgcgcgc	agatacccg	gcacatgaag	tcgcagcgtt	cgaagtccac	ccatgagaaa	7680
ccggagggggt	actacaactg	gcaccacgga	gcagtacagt	actcaggagg	ccggttcacc	7740
atccctacag	gtgctggcaa	accagggggac	agcggcagac	cgatcttcga	caacaaggga	7800
cgcggtgggt	ccatagtctt	aggaggagct	aatgaaggag	cccgtacagc	cctctcggtg	7860
gtgacctgga	ataaagacat	tgtcactaaa	atcacccccc	agggggccga	agagtggagt	7920
cttgccatcc	cagttatgtg	cctgttgga	aacaccacgt	tccctgctc	ccagccccct	7980
tgcacgcctt	gctgctacga	aaaggaaccg	gaggaaaccc	tacgcagtgt	tgaggacaac	8040
gtcatgagac	ctgggtacta	tcagctgcta	caagcatcct	taacatgttc	tccccaccgc	8100
cagcgacgca	gcaccaagga	caacttcaat	gtctataaag	ccacaagacc	atacttagct	8160
cactgtcccc	actgtggaga	agggcactcg	tgccatagtc	ccgtagcact	agaacgcac	8220
agaaatgaag	cgacagacgg	gacgctgaaa	atccaggtct	ccttgcaaat	cggaataaag	8280
acggatgaca	gccacgattg	gaccaagctg	cgttatatgg	acaaccacat	gccagcagac	8340
gcagagaggg	cggggctatt	tgtaagaaca	tcagcacccg	gtacgattac	tggaacaatg	8400
ggacacttca	tcctggcccc	atgtccaaaa	ggggaaactc	tgacgggtgg	attcactgac	8460
agtaggaaga	ttagtacttc	atgtacgcac	ccatttcacc	acgaccctcc	tgtgatagg	8520
cgggaaaaat	tccattcccc	accgcagcac	ggtaaagagc	taccttgacg	cacgtacgtg	8580
cagagcaccg	ccgcaactac	cgaggagata	gaggtacaca	tgcccccaga	caccctgat	8640
cgcacatata	tgtcacaaca	gtccggcaac	gtaaagatca	cagtcaatgg	ccagacgggtg	8700
cggtacaagt	gtaattgcgg	tggctcaaat	gaaggactaa	caactacaga	caaagtgatt	8760

FIGURE 9 (con't)

aataactgca	aggttgatca	atgtcatgcc	gcggtcacca	atcacaaaaa	gtggcagtat	8820
aactcccctc	tggteccgcg	taatgctgaa	cttgggggacc	gaaaaggaaa	aattcacatc	8880
ccgtttccgc	tggcaaatgt	aacatgcagg	gtgcctaaag	caaggaaccc	caccgtgacg	8940
tacgggaaaa	accaagtcac	catgctactg	tatcctgacc	acccaacact	cctgtcctac	9000
cggaaatatg	gagaagaacc	aaactatcaa	gaagagtggg	tgatgcataa	gaaggaagtc	9060
gtgctaaccg	tgccgactga	agggtctcag	gtcacgtggg	gcaacaacga	gccgtataag	9120
tattggccgc	agttatctac	aaacgggtaca	gcccattggcc	acccgcatga	gataattctg	9180
tattattatg	agctgtaccc	cactatgact	gtagtagttg	tgtcagtggc	cacgttcata	9240
ctcctgtcga	tgggtgggtat	ggcagcgggg	atgtgcatgt	gtgcacgacg	cagatgcac	9300
acaccgtatg	aactgacacc	aggagctacc	gtccctttcc	tgcttagcct	aatatgctgc	9360
atcagaacag	ctaaagcggc	cacataccaa	gaggctgcga	tatacctgtg	gaacgagcag	9420
caacctttgt	tttggctaca	agcccttatt	ccgctggcag	ccctgattgt	tctatgcaac	9480
tgtctgagac	tcttaccatg	ctgctgtaaa	acgttggctt	ttttagccgt	aatgagcgtc	9540
ggtgcccaca	ctgtgagcgc	gtacgaacac	gtaacagtga	tcccgaacac	ggtgggagta	9600
ccgtataaga	ctctagtcaa	tagacctggc	tacagcccca	tggtattgga	gatggaacta	9660
ctgtcagtc	ctttggagcc	aacactatcg	cttgattaca	tcacgtgcga	gtacaaaacc	9720
gtcatcccgt	ctccgtacgt	gaagtgctgc	ggtacagcag	agtgcaagga	caaaaaccta	9780
cctgactaca	gctgtaaggt	cttcaccggc	gtctacccat	ttatgtgggg	cggcgcctac	9840
tgcttctgcg	acgttgaaaa	cacgcagttg	agcgaagcac	acgtggagaa	gtccgaatca	9900
tgcaaaacag	aatttgcatc	agcatacagg	gctcataccg	catctgcac	agctaagctc	9960
cgcgtccttt	accaaggaaa	taacatcact	gtaactgcct	atgcaaacgg	cgaccatgcc	10020
gtcacagtta	aggacgccaa	attcattgtg	gggccaatgt	cttcagcctg	gacacctttc	10080
gacaacaaaa	ttgtggtgta	caaagggtgac	gtctataaca	tggaactacc	gccctttggc	10140
gcaggaagac	caggacaatt	tggcgatatc	caaagtcgca	cacctgagag	taaagacgtc	10200
tatgctaata	cacaactgg	actgcagaga	ccggctgcgg	gtacgggtaca	cgtgccatac	10260
tctcaggcac	catctggctt	taagtattgg	ctaaaagaac	gcggggcgctc	gctgcagcac	10320
acagcaccat	ttggctgcca	aatagcaaca	aaccgcggtaa	gagcgggtgaa	ctgcgcctga	10380
gggaacatgc	ccatctccat	cgacataccg	gaagcggcct	tcactagggt	cgtcgacgcg	10440
ccctctttaa	cggacatgtc	gtgcgaggta	ccagcctgca	cccatctctc	agactttggg	10500
ggcgtcgcca	ttattaaata	tgcagccagc	aagaaaggca	agtgtgcgg	gcattcgatg	10560
actaacgcgc	tcactattcg	ggaagctgag	atagaagttg	aagggaattc	tcagctgcaa	10620
atctctttct	cgacggcctt	agccagcgcc	gaattccgcg	tacaagtctg	ttctacacaa	10680
gtacactgtg	cagccgagtg	ccaccccccg	aaggaccaca	tagtcaacta	cccggcgctc	10740
cataccaccc	tcgggggtcca	ggacatctcc	gctacggcga	tgtcatgggt	gcagaagatc	10800
acgggaggtg	tgggactgg	tgttgctgtt	gccgcactga	ttctaatacgt	ggtgctatgc	10860
gtgtcggttca	gcaggcacta	acttgacaat	taagtatgaa	ggtatatgtg	tcccctaaga	10920
gacacactgt	acatagcaaa	taatctatag	atcaaagggc	tacgcaaccc	ctgaatagta	10980
acaaaataca	aatcactaa	aaattataaa	aacagaaaaa	tacataaata	ggtatacgtg	11040
tcccctaaga	gacacattgt	atgtaggtga	taagtataga	tcaaagggcc	gaataacccc	11100
tgaatagtaa	caaaatatga	aatcaataa	aatcataaa	atagaaaaac	cataaacaga	11160
agtagttcaa	agggtataaa	aaccctgaa	tagtaacaaa	acataaaaatt	aataaaaatc	11220
aatgaatac	catatgg					11237

FIGURE 10

ataccegtgc	acatgaagtc	cgacgcttcg	aagttcaccc	atgagaaacc	ggaggggtac	60
tacaactggc	accacggagc	agtacagtac	tcaggaggcc	ggttcaccat	ccctacaggt	120
gctggcaaac	caggggacag	cggcagaccg	atcttcgaca	acaagggacg	cgtggtggcc	180
atagtcttag	gaggagctaa	tgaaggagcc	cgtacagccc	tctcggtggt	gacctggaat	240
aaagacattg	tcactaaaat	cacccccgag	ggggccgaag	agtggagtct	tgccatccca	300
gttatgtgcc	tgttggcaaa	caccacgttc	ccctgctccc	agcccccttg	cacgccctgc	360
tgctacgaaa	aggaaccgga	ggaaacccta	cgcattgctt	aggacaacgt	catgagacct	420
gggtactatc	agctgctaca	agcatcctta	acatgttctc	cccaccgcca	gcgacgcagc	480
accaaggaca	acttcaatgt	ctataaagcc	acaagaccat	acttagctca	ctgtcccgac	540
tgtggagaag	ggcactcgtg	ccatagtccc	gtagcactag	aacgcattcag	aaatgaagcg	600
acagacggga	cgtgaaaat	ccaggtctcc	ttgcaaatac	gaataaagac	ggatgacagc	660
cacgattgga	ccaagctgcg	ttatatggac	aaccacatgc	cagcagacgc	agagagggcg	720
gggctatttg	taagaacatc	agcaccgtgt	acgattactg	gaacaatggg	acacttcatc	780
ctggccccgat	gtccaaaagg	ggaaactctg	acggtgggat	tcactgacag	taggaagatt	840
agtcactcat	gtacgcaccc	atttcaccac	gacctcctg	tgataggtcg	ggaaaaattc	900
cattccccgac	cgcagcacgg	taaagagcta	ccttgcagca	cgtacgtgca	gagcaccgcc	960
gcaactaccg	aggagataga	ggtacacatg	ccccagaca	cccctgatcg	cacattaatg	1020
tcacaacagt	ccggcaacgt	aaagatcaca	gtcaatggcc	agacggtgcg	gtacaagtgt	1080
aattgcggtg	gctcaaatga	aggactaaca	actacagaca	aagtgattaa	taactgcaag	1140
gttgatcaat	gtcatgccgc	ggtcaccaat	cacaaaaagt	ggcagtataa	ctccccctctg	1200
gtcccgcgta	atgctgaact	tggggaccga	aaaggaaaaa	ttcacatccc	gtttccgctg	1260
gcaaatgtaa	catgcagggt	gcctaaagca	aggaacccca	ccgtgacgta	cgggaaaaaac	1320
caagtcatca	tgtactgta	tcctgaccac	ccaacactcc	tgtcctaccg	gaatatggga	1380
gaagaaccaa	actatcaaga	agagtgggtg	atgcataaga	aggaagtcgt	gctaaccgtg	1440
ccgactgaag	ggctcgaggt	cacgtggggc	aacaacgagc	cgtataagta	ttggccgcag	1500
ttatctacaa	acggtacagc	ccatggccac	ccgcatgaga	taattctgta	ttattatgag	1560
ctgtacccca	ctatgactgt	agtagttgtg	tcagtggcca	cgttcatact	cctgtcgatg	1620
gtgggtatgg	cagcggggat	gtgcatgcgt	gcacgacgca	gatgcatcac	accgtatgaa	1680
ctgacaccag	gagctaccgt	ccctttcctg	cttagcctaa	tatgctgcat	cagaacagct	1740
aaagcggcca	cataccaaga	ggctgcgata	tacctgtgga	acgagcagca	accttta	1797

FIGURE 11

atacccgtgc	acatgaagtc	cgacgcttcg	aagttcaccc	atgagaaacc	ggaggggtac	60
tacaactggc	accacggagc	agtacagtac	tcaggaggcc	ggttcaccat	ccctacaggt	120
gctggcaaac	caggggacag	cggcagaccg	atcttcgaca	acaagggacg	cgtggtggcc	180
atagtcttag	gaggagctaa	tgaaggagcc	cgtacagccc	tctcgggtgt	gacctggaat	240
aaagacattg	tcactaaaat	cacccccgag	ggggccgaag	agtggagtct	tgccatccca	300
gttatgtgcc	tggtggcaaa	caccacgttc	ccctgctccc	agcccccttg	cacgccctgc	360
tgctacgaaa	aggaaccgga	ggaaacccta	cgcattgctt	aggacaacgt	catgagacct	420
gggtactatc	agctgctaca	agcatcctta	acatgtttct	cccaccgcca	gcgacgcagc	480
accaaggaca	acttcaatgt	ctataaagcc	acaagaccat	acttagctca	ctgtcccagc	540
tgtggagaag	ggcactcgtg	ccatagtccc	gtagcactag	aacgcattcg	aaatgaagcg	600
acagacggga	cgctgaaaat	ccaggtctcc	ttgcaaatac	gaataaagac	ggatgacagc	660
cacgattgga	ccaagctgcg	ttatatggac	aaccacatgc	cagcagacgc	agagagggcg	720
gggctatatt	taagaacatc	agcaccgtgt	acgattactg	gaacaatggg	acacttcata	780
ctggccccgat	gtccaaaagg	ggaaactctg	acggtgggat	tcactgacag	taggaagatt	840
agtcactcat	gtacgcaccc	atcttcaccac	gacctctctg	tgataggtcg	ggaaaaattc	900
cattcccagc	cgcagcacgg	taaagagcta	ccttgacgca	cgtacgtgca	gagcacccgc	960
gcaactaccg	aggagataga	ggtacacatg	ccccagaca	cccctgatcg	cacattaatg	1020
tcacaacagt	ccggcaacgt	aaagatcaca	gtcaatggcc	agacggtgcg	gtacaagtgt	1080
aattgcggtg	gctcaaatga	aggactaaca	actacagaca	aagtgattaa	taactgcaag	1140
gttgatcaat	gtcatgccgc	ggtcaccaat	cacaaaaagt	ggcagtataa	ctccccctct	1200
gtcccgcgta	atgctgaact	tggggaccga	aaaggaaaaa	ttcacatccc	gtttccgctg	1260
gcaaatgtaa	catgcagggt	gcctaaagca	aggaacccca	ccgtgacgta	cgggaaaaac	1320
caagtcatac	tgctactgta	tcctgaccac	ccaacactcc	tgctcctacc	gaatatggga	1380
gaagaaccaa	actatcaaga	agagtgggtg	atgcataaga	aggaagtcgt	gctaaccgtg	1440
ccgactgaag	ggctcgaggt	cacgtggggc	aacaacgagc	cgtataagta	ttggccgcag	1500
ttatctacaa	acggtacagc	ccatggccac	ccgcatgaga	taattctgta	ttattatgag	1560
ctgtacccca	ctatgactgt	agtagttgtg	tcagtggcca	cgttcatact	cctgtcgatg	1620
gtgggtatgg	cagcggggat	gtgcatgtgt	gcacgacgca	gatgcattcac	accgtatgaa	1680
ctgacaccag	gagctaccgt	ccctttcctg	cttagcctaa	tatgctgcat	cagaacagct	1740
aaagcggcca	cataccaaga	ggctgcgata	tacctgtgga	acgagcagca	accttta	1797

FIGURE 12

ataccctgtgc	acatgaagtc	cgacgcttcg	aagttcaccc	atgagaaacc	ggaggggtac	60
tacaactggc	accacggagc	agtacagtac	tcaggaggcc	ggttcaccat	ccctacaggt	120
gctggcaaac	caggggacag	cggcagaccg	atcttcgaca	acaagggacg	cgtggtggcc	180
atagtcttag	gaggagctaa	tgaaggagcc	cgtacagccc	tctcggtggt	gacctggaat	240
aaagacattg	tcactaaaat	cacccccgag	ggggccgaag	agtggagtct	tgccatccca	300
gttatgtgcc	tgttggaata	caccacgttc	ccctgctccc	agcccccttg	cacgccccgc	360
tgctacgaaa	aggaaccgga	ggaaacccta	cgcattgctt	aggacaacgt	catgagacct	420
gggtactatc	agctgctaca	agcatcctta	acatgttctc	cccaccgcca	gcgacgcagc	480
accaaggaca	acttcaatgt	ctataaagcc	acaagaccat	acttagctca	ctgtccccgac	540
tgtggagaag	ggcactcgtg	ccatagtccc	gtagcactag	aacgcattcag	aaatgaagcg	600
acagacggga	cgctgaaaat	ccaggctctc	ttgcaaatcg	gaataaagac	ggatgacagc	660
cacgattgga	ccaagctgcg	ttatatggac	aaccacatgc	cagcagacgc	agagagggcg	720
gggctatatt	taagaacatc	agcaccgtgt	acgattactg	gaacaatggg	acacttcatc	780
ctggccccgat	gtccaaaagg	ggaaactctg	acgggtgggat	tcactgacag	taggaagatt	840
agtcactcat	gtacgcaccc	atttcaccac	gacccctctg	tgataggctc	ggaaaaattc	900
cattccccgac	cgcagcacgg	taaagagcta	ccttgcagca	cgtacgtgca	gagcaccgcc	960
gcaactaccg	aggagataga	ggtacacatg	ccccagaca	cccctgatcg	cacattaatg	1020
tcacaacagt	ccggcaacgt	aaagatcaca	gtcaatggcc	agacgggtgc	gtacaagtgt	1080
aattgcggtg	gctcaaatga	aggactaaca	actacagaca	aagtgattaa	taactgcaag	1140
gttgatcaat	gtcatgccgc	ggtcaccaat	cacaaaaagt	ggcagtataa	ctccccctct	1200
gtcccgcgta	atgctgaact	tggggaccga	aaaggaaaaa	ttcacatccc	gtttccgctg	1260
gcaaatgtaa	catgcagggt	gcctaaagca	aggaacccca	ccgtgacgta	cgggaaaaaac	1320
caagtcatca	tgctactgta	tcctgaccac	ccaacactcc	tgctcctacc	gaatatggga	1380
gaagaaccaa	actatcaaga	agagtgggtg	atgcataaga	aggaagtcgt	gctaaccgtg	1440
ccgactgaag	ggctcgaggt	cacgtggggc	aacaacgagc	cgtataagta	ttggccgcag	1500
ttatctacaa	acgggtacagc	ccatggccac	ccgcacgaga	taattctgta	ttattatgag	1560
ctgtacccca	ctatgactgt	agtagttgtg	tcagtggcca	cgttcatact	cctgtcgatg	1620
gtgggtatgg	cagcggggat	gtgcatgtgt	gcacgacgca	gatgcatcac	accgtatgaa	1680
ctgacaccag	gagctaccgt	ccctttcctg	cttagcctaa	tatgctgcat	cagaacagct	1740
aaagcggcca	cataccaaga	ggctgcgata	tacctgtgga	acgagcagca	accttta	1797

FIGURE 13

atacccgtgc	acatgaagtc	cgacgcttcg	aagttcaccc	atgagaaacc	ggaggggtac	60
tacaactggc	accacggagc	agtacagtac	tcaggaggcc	ggttcaccat	ccctacaggt	120
gctggcaaac	caggggacag	cggcagaccg	atcttcgaca	acaagggacg	cgtggtggcc	180
atagtcttag	gaggagctaa	tgaaggagcc	cgtacagccc	tctcgggtgt	gacctggaat	240
aaagacattg	tcactaaaat	cacccccgag	ggggccgaag	agtggagtct	tgccatccca	300
gttatgtgcc	tgttggcaaa	caccacgttc	ccctgctccc	agcccccttg	cacgccctgc	360
tgctacgaaa	aggaaccgga	ggaaacccta	cgcattgctt	aggacaacgt	catgagacct	420
gggtactatc	agctgctaca	agcatcctta	acatgttctc	cccaccgcca	gcgacgcagc	480
accaaggaca	acttcaatgt	ctataaagcc	acaagaccat	acttagctca	ctgtcccgcg	540
tgtggagaag	ggcactcgtg	ccatagtccc	gtagcactag	aacgcattcag	aaatgaagcg	600
acagacggga	cgctgaaaat	ccagggtctc	ttgcaaatac	gaataaagac	ggatgacagc	660
cacgattgga	ccaagctgcg	ttatatggac	aaccacatgc	cagcagacgc	agagagggcg	720
gggctatatt	taagaacatc	agcaccgtgt	acgattactg	gaacaatggg	acacttcatc	780
ctggccccgat	gtccaaaagg	ggaaactctg	acgggtgggat	tactgacag	taggaagatt	840
agtcactcat	gtacgcaccc	atttcaccac	gacctcctg	tgataggctg	ggaaaaattc	900
cattccccgac	cgcagcacgg	taaagagata	ccttgacgca	cgtacgtgca	gagcaccgcc	960
gcaactaccg	aggagataga	ggtacacatg	ccccagaca	cccctgatcg	cacattaatg	1020
tcacaacagt	ccggcaacgt	aaagatcaca	gtcaatggcc	agacgggtgc	gtacaagtgt	1080
aattgcgggt	gctcaaata	aggactaaca	actacagaca	aagtgattaa	taactgcaag	1140
gttgatcaat	gtcatgccgc	ggtcaccaat	cacaaaaagt	ggcagtataa	ctccccctct	1200
gtcccgcgta	atgctgaact	tggggaccga	aaaggaaaaa	ttcacatccc	gtttccgctg	1260
gcaaatgtaa	catgcagggt	gcctaaagca	aggaacccca	ccgtgacgta	cgggaaaaaac	1320
caagtcatca	tgctactgta	tcctgaccac	ccaacactcc	tgtcctaccg	gaatatggga	1380
gaagaacca	actatcaaga	agagtgggtg	atgcataaga	aggaagtcgt	gctaaccgtg	1440
ccgactgaag	ggctcgaggt	cacgtggggc	aacaacgagc	cgtataagta	ttggccgcag	1500
ttatctacaa	acggtacagc	ccatggccac	ccgcatgaga	taattctgta	ttattatgag	1560
ctgtacccca	ctatgactgt	agtagttgtg	tcagtggcca	cgttcatact	cctgtcgatg	1620
gtgggtatgg	cagcggggat	gtgcatgtgt	gcacgacgca	gatgcatcac	accgtatgaa	1680
ctgacaccag	gagctaccgt	ccctttcctg	cttagcctaa	tatgctgcat	cagaacagct	1740
aaagcggcca	cataccaaga	ggctgcgata	tacctgtgga	acgagcagca	accttta	1797

Fig S1. Repeat Sequence Elements found in the 3'NTR region.

A

			10	20	30	40	50	60
Repl116_ESE1	11327	DATAGATCAAGG-GGTACGGT-AGCTCTGATATTAACAAAATACAAAATC-ATTAAAAAT						
Repl116_KSE2	11456	GGGAT.....TG.....ATCA.....						
Repl116_KSE3	11542	AG...T.....TAA.....C.T.....T.A.....						
SD24fr1_ESE1	11346	AG.....TAT.....						
SD24fr1_ESE2	11351	GAA.....T.....ATT.A.....						
SD24fr1_ESE3	11357	AG...T.....TAA.....C.T.....T.A.....						
SD2793_ESE1	11415	AT.....TAA.....C.....A.....						
SD2793_ESE2	11444	TAT.....C.....A.....						
SD2793_ESE3	11450	AG...T.....TAA.....C.T.....GT.A.....						
SD2894_ESE1	4607	AG...T.....TAA.....C.T.....T.A.....						
SD2894_ESE2	4692	GTGAT.....A...T.....ATAA...G...						
SD2894_ESE3	4696	AGAT-T.....TAA.....T.....T.....T.A.....G...						

