

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
18 October 2001 (18.10.2001)

PCT

(10) International Publication Number
WO 01/77347 A2

- (51) International Patent Classification⁷: C12N 15/74, 1/21
- (21) International Application Number: PCT/US01/11059
- (22) International Filing Date: 5 April 2001 (05.04.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/194,625 5 April 2000 (05.04.2000) US
- (71) Applicant: ARCHER-DANIELS-MIDLAND COMPANY [US/US]; 4666 Faries Parkway, Box 1470, Decatur, IL 62525 (US).
- (71) Applicant and
(72) Inventor: D'ELIA, John [US/US]; 1903 Filipov Street, Champaign, IL 61821 (US).
- (74) Agents: LUDWIG, Steven, R. et al.; Sterne, Kessler, Goldstein & Fox P.L.L.C., Suite 600, 1100 New York Avenue, N.W., Washington, DC 20005-3934 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 01/77347 A2

(54) Title: KETOGULONIGENIUM SHUTTLE VECTORS

(57) Abstract: The present invention relates, in general, to vectors comprising *Ketogulonigenium* replicons. More specifically, the present invention relates to vectors comprising a *Ketogulonigenium* replicon found on the endogenous plasmid contained in Deposit No. NRRL B-30035.

KETOGULONIGENIUM SHUTTLE VECTORS

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates, in general, to vectors comprising a *Ketogulonigenium* replicon. More specifically, the present invention relates to vectors comprising a *Ketogulonigenium* replicon found on the endogenous plasmid contained in Deposit No. NRRL B-30035 (Deposited July 21, 1998, Agriculture Research Culture Collection (NRRL), 1815 N. University Street, Peoria, Illinois 61604, U.S.A.).

Background Information

The exploitation of microorganisms to synthesize vitamin C or its chemical pathway intermediates has both economic and ecological advantages. One key intermediate in vitamin C synthesis is 2-keto-L-gulonic acid (2-KLG), which is easily converted chemically to L-ascorbic acid (vitamin C) by esterification followed by lactonization (Delic, V. *et al.*, "Microbial reactions for the synthesis of vitamin C (L-ascorbic acid," in *Biotechnology of Vitamins, Pigments and Growth Factors*, Vandamme, E.J., ed., Elsevier Applied Science (London & New York) pp. 299-336 (1989)). Members of a number of bacterial genera have been identified that produce 2-KLG from the oxidation of sorbitol. Such 2-KLG producing genera include the acidogenic, alpha-proteobacteria *Gluconobacter* and *Acetobacter*, the gamma-proteobacteria *Pseudomonas*, *Escherichia*, *Klebsiella*, *Serratia* and *Xanthomonas*, the Gram positive *Bacillus*, *Micrococcus* and the unofficial genus *Pseudogluconobacter* (Imai, K. *et al.*, U.S. Patent No. 4,933,289 (1990), Sugisawa, H. *et al.*, "Microbial production of 2-keto-L-gulonic acid from L-sorbose and D-sorbitol by *Gluconobacter melanogenus*," *Agric. Biol. Chem.* 54:1201-1209 (1990), Yin, G. *et al.*, U.S.

Patent No. 4,935,359 (1990) and Nogami, I. *et al.*, U.S. Patent No. 5,474,924 (1995)).

To aid in increasing the yield of bacterial products, attempts have been made to exploit endogenous plasmids within microorganism strains. For example, shuttle vectors derived from endogenous plasmids have been utilized in an attempt increase production of bacterially generated compounds in *Gluconobacter* and *Acetobacter* (Beppu, T. *et al.*, U.S. Patent No. 5,580,782 (1996), Fujiwara, A. *et al.*, U.S. Patent No. 5,399,496 (1995), Tonouchi *et al.*, U.S. Patent No. 6,127,174 (2000), Hoshino, T. *et al.*, U.S. Patent No. 6,127,156, (2000)).

SUMMARY OF THE INVENTION

The present invention provides an isolated or purified nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a nucleotide sequence of a *Ketogulonigenium* plasmid replicon found on the endogenous plasmid contained in Deposit No. NRRL B-30035. The invention further provides host cells transformed with the vector of the present invention. The invention also provides a method of producing polypeptides and/or antisense transcripts by culturing host cells transformed with the vector of the present invention.

Further advantages of the present invention will be clear from the description that follows.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the nucleotide (SEQ ID NO:1) sequence of a replicon of the endogenous plasmid contained in *Ketogulonigenium* strain NRRL Deposit No. B-30035. The nucleotide has a sequence of about 2112 nucleic acid residues.

Figure 2 shows the nucleotide (SEQ ID NO:2) sequence of the endogenous plasmid determined by sequencing of the endogenous plasmid contained in NRRL Deposit No. B-30035. The nucleotide has a sequence of about 8509 nucleic acid residues.

Figure 3 shows the nucleotide (SEQ ID NO:3) sequence of a shuttle vector plasmid contained in NRRL Deposit No. B-30434 (Deposited March 27, 2001, Agriculture Research Culture Collection (NRRL), 1815 N. University Street, Peoria, Illinois 61604, U.S.A.). The nucleotide has a sequence of about 5859 nucleic acid residues.

Figure 4 shows the nucleotide (SEQ ID NO:4) sequence comprising a region that supports plasmid vector replication in *Ketogulonigenium* host species. The nucleotide has a sequence of about 2517 nucleic acid residues.

DETAILED DESCRIPTION OF THE INVENTION

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer (such as the ABI Prism 3700). Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule.

Unless otherwise indicated, each "nucleotide sequence" set forth herein is presented as a sequence of deoxyribonucleotides (abbreviated A, G, C and T). However, by "nucleotide sequence" of a nucleic acid molecule or polynucleotide is intended, for a DNA molecule or polynucleotide, a sequence of deoxyribonucleotides, and for an RNA molecule or polynucleotide, the corresponding sequence of ribonucleotides (A, G, C and U) where each thymidine deoxynucleotide (T) in the specified deoxynucleotide sequence is replaced by the ribonucleotide uridine (U). For instance, reference to an RNA molecule having the sequence of SEQ ID NO:1 set forth using deoxyribonucleotide abbreviations is intended to indicate an RNA molecule having a sequence in which each deoxynucleotide A, G or C of SEQ ID NO:1 has been replaced by the corresponding ribonucleotide A, G or C, and each deoxynucleotide T has been replaced by a ribonucleotide U.

As indicated, nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced synthetically. The DNA may be double-stranded or single-stranded. Single-stranded DNA or RNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

By "isolated" nucleic acid molecule(s) is intended a nucleic acid molecule, DNA or RNA, which has been removed from its native environment. For example, recombinant DNA molecules contained in a vector are considered isolated for the purposes of the present invention. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include in vivo or in vitro RNA transcripts of the DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

In another aspect, the invention provides an isolated nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization

conditions to a portion of the polynucleotide in a nucleic acid molecule of the invention described above, for instance, in the endogenous plasmids contained in NRRL Deposit No. B-30035. By "stringent hybridization conditions" is intended overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150 mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C. By a polynucleotide which hybridizes to a "portion" of a polynucleotide is intended a polynucleotide (either DNA or RNA) hybridizing to at least about 15 nucleotides (nt), and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably about 30-70 nt of the reference polynucleotide. These are useful as diagnostic probes and primers.

Of course, polynucleotides hybridizing to a larger portion of the reference polynucleotide (e.g., the deposited endogenous plasmid), for instance, a portion 50-750 nt in length, or even to the entire length of the reference polynucleotide, also useful as probes according to the present invention, as are polynucleotides corresponding to most, if not all, of the nucleotide sequence of the deposited DNA or the nucleotide sequence as shown in Figure 1 (SEQ ID NO:1). By a portion of a polynucleotide of "at least 20 nt in length," for example, is intended 20 or more contiguous nucleotides from the nucleotide sequence of the reference polynucleotide, (e.g., the deposited DNA or the nucleotide sequence as shown in Figure 1 (SEQ ID NO:1)). As indicated, such portions are useful diagnostically either as a probe according to conventional DNA hybridization techniques or as primers for amplification of a target sequence by the polymerase chain reaction (PCR), as described, for instance, in *Molecular Cloning, A Laboratory Manual*, 2nd. edition, edited by Sambrook, J., Fritsch, E. F. and Maniatis, T., (1989), Cold Spring Harbor Laboratory Press, the entire disclosure of which is hereby incorporated herein by reference.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence is intended that the nucleotide

-6-

sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the pyruvate carboxylase polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence.

As a practical matter, whether any particular nucleic acid molecule is at least 90%, 95%, 97%, 98% or 99% identical to, for instance, the nucleotide sequence shown in Figure 1 or to the nucleotide sequence of the deposited endogenous plasmid can be determined conventionally using known computer programs such as the FastA program. FastA does a Pearson and Lipman search for similarity between a query sequence and a group of sequences of the same type nucleic acid. Professor William Pearson of the University of Virginia Department of Biochemistry wrote the FASTA program family (FastA, TFastA, FastX, TFastX and SSearch). In collaboration with Dr. Pearson, the programs were modified and documented for distribution with GCG Version 6.1 by Mary Schultz and Irv Edelman, and for Versions 8 through 10 by Sue Olson.

The present invention provides an isolated or purified vector comprising a nucleic acid molecule comprising a nucleotide sequence of a *Ketogulonigenium* replicon found on the endogenous plasmid contained in Deposit No. NRRL B-30035 (ADM 291-19), pADM291, and at least one exogenous nucleotide sequence. In an additional embodiment of the invention, the *Ketogulonigenium* replicon comprises the nucleotide sequence in SEQ ID NO:1.

As used herein, an exogenous nucleotide sequence is a nucleotide sequence which is not present in the native or wild type endogenous plasmid contained in Deposit No. NRRL B-30035. However, an exogenous nucleotide sequence also includes a sequence endogenous to the plasmid contained in Deposit No. NRRL B-30035, but mutated or non-natively regulated.

Ketogulonigenium (Ke.to.gu.lo.ni.gen'i.um. M.L. n. *acidum ketogulonicum* ketogulonic acid; Gr. v. *gennaio* to produce; M.L. n. *ketogulonigenium* ketogulonic acid producing) is gram negative, facultatively anaerobic, motile or non-motile, has ovoid to rod-shaped cells, 0.8-1.3 μm long, 0.5-0.7 μm in diameter, with tapered ends, occurring as single cells, pairs and occasionally short chains. Some strains form elongated cells (up to 30 μm in length) on TSB. Flagella and fimbriae have been observed. Colonies are tan colored, smooth, circular, entire, raised to convex, 1-2 mm in diameter with a diffusible brown pigment after 48 hrs incubation. Oxidase and catalase reactions are positive. Optimum temperature range is 27 to 31 °C, optimum pH range is 7.2 to 8.5 and optimum Na^+ concentration is 117-459 mM. Chemoorganotrophic. Carbon sources utilized include arabinose, cellobiose, fructose, glucose, glycerol, inositol, lactose, maltose, mannitol, mannose, rhamnose, sorbitol, sorbose, sucrose, trehalose, pyruvate and succinate. Favored carbon sources are inositol, mannitol, glycerol, sorbitol, lactose and arabinose. All strains examined produce 2-keto-L-gulonic acid from L-sorbose. Major cellular fatty acids are 16:0 and 18:1 ω 7c/ ω 9t/ ω 12t and the mol% DNA G+C is 52.1 to 54.0 percent. Small subunit rDNA sequence analysis place this genus in the alpha subgroup of the Proteobacteria. All strains isolated in the present study group originated in soil. DNA reassociation studies divide the genus into two species. *K. vulgarae* and *K. robustum*, of which *K. vulgarae* is the designated type species. A group of bacteria having the above-mentioned properties does not belong to any known genera as described in Bergey's Manual of Systematic Bacteriology, and therefore belongs to a new genus.

The present invention further provides vectors which have a replicon functional in *Escherichia coli* (*E. coli*) and in *Ketogulonigenium*. *E. coli* is known to be an efficient host for amplification of a vector DNA and manipulation of recombinant DNA by simple and rapid methods. On the other hand, *Ketogulonigenium* can be used as a host for expression of *Ketogulonigenium* genes. Since the vectors of the present invention are such functional constructs,

they enable cloning of certain genes of *Ketogulonigenium* in *E. coli* and thereafter the effective expression of the genes in *Ketogulonigenium*. Furthermore, it is favorable that such functional constructs also contain a DNA region necessary for conjugal transfer (mob site). Hence the vectors of the present invention can first be assembled in *E. coli* and then directly introduced into *Ketogulonigenium* by conjugal mating without isolation of plasmid DNA from *E. coli*.

Polynucleotides of interest may be joined to a vector containing a selectable marker for propagation in the host. A plasmid vector can be introduced by transformation mediated by calcium chloride or other cation salts, electroporation and other transformation methods, viral or phage transduction or other methods. Such methods are described in many standard laboratory manuals, such as Davis *et al.*, "Basic Methods in Molecular Biology," (1986), and, J. Sambrook, E.F. Fritsch and T. Maniatis (1989) "Molecular Cloning: A Laboratory Manual", 2nd Ed..

Preferred are vectors comprising cis-acting control regions to a polynucleotide of interest. Appropriate *trans* acting factors may be supplied by the host, supplied by a complementing vector, or supplied by the vector itself upon introduction into the host.

In certain preferred embodiments in this regard, the vectors provide for specific expression, which may be inducible, mutant-specific and/or condition-specific. Particularly preferred among such vectors are those inducible by environmental factors that are easy to manipulate, such as temperature, nutrient additives or chemical additives. Other suitable environmental factors will be readily apparent to the skilled artisan.

Expression vectors useful in the present invention include chromosomal-, episomal-vectors e.g., vectors derived from plasmids, bacteriophage, and vectors derived from combinations thereof, such as cosmids and phagemids.

A DNA insert of interest should be operatively linked to an appropriate promoter which is preferably a host-derived promoter. Preferably, the host-derived promoter is positioned in front of a polylinker insert such that

transcripts initiated at the promoter cause expression of genes cloned into the polylinker. The expression constructs can further contain sites for transcription initiation, termination and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs can include a translation initiating codon appropriate for the host at the beginning and a termination codon appropriately positioned at the end of the coding sequence to be translated.

The invention further provides the vectors of the present invention comprising a peptide fusion cloned gene expression system whereby a fused peptide serves as a tag for detecting and/or isolating the cloned gene product. The invention also provides a fused short peptide containing a biocytin residue (lysine covalently modified with a biotin), so that avidin-sepharose columns can be used to isolate the protein, and avidin-conjugated alkaline phosphatase can be used to detect the protein fusion on Western Blots.

In an additional embodiment, the DNA insert of interest additionally comprises a nucleotide sequence encoding a His-Tag sequence downstream of the promoter and polylinker, such that genes cloned into the polylinker can form translational fusions with the His-Tag. In this way, isolation of cloned proteins is facilitated because the fused His-Tag sequence allows rapid purification. A His-Tag is a polyhistidine sequence that binds to Nickel-Agarose columns, thus allowing rapid purification of the fusion protein from crude cell extracts. (Chen, B.P and Hai, T., *Gene* 139:73-79 (1994)). Removal of the His-Tag from the cloned polypeptide is achieved by treating the purified protein preparation with factor Xa or other suitable proteases.

In an additional embodiment, the DNA insert of interest additionally comprises nucleotides encoding a polypeptide. The polypeptide can be a small, immunogenic peptide sequence of about 5 to 30 amino acids, and preferably from about 10 to 20 amino acids. Ideally, this peptide sequence is not expressed natively in *Ketogulonigenium*. The DNA insert is adjacent to and translationally fused to the polylinker on either side, such that genes cloned into the polylinker

are expressed as protein fusions with the immunogenic peptide. The fused peptide-antibody interaction can be used in protein isolation schemes and for detecting expression of the cloned gene fusion by immunoblotting of the expressed protein. (Enomoto, S. *et al.*, *Biotechniques* 24:782-788 (1998)).

The invention further provides a functional gene expression vector for a periplasmic, outer membrane, or exported protein. (Payne, M.S. and Jackson, E.N., *J. Bacteriol.* 173:2278-2282 (1991)). Specifically, the invention provides a DNA insert in the vector of the insert further comprising a suitable N-terminal signal sequence from a periplasmic or exported protein gene (derived from *Ketogulonigenium* or other bacterial genera) inserted downstream of the promoter and upstream of the polylinker. In this way a periplasmic expression system is combined with easy recovery and purification of the cloned protein fusion.

The invention further provides a vector comprising a nucleotide molecule comprising a nucleotide sequence of a *Ketogulonigenium* replicon found on the endogenous plasmid of NRRL B-30035 (ADM 291-19) additionally comprising a sequence encoding a periplasmic thioredoxin-like function similar to that of the DsbA or *Pseudomonas* or *Bradyrhizobium* TlpA. The *dsbA* gene in *E. coli* encodes a protein that is able to catalyze formation and isomerization of disulfide bonds (Wunderlich, M. and Glockshuber, R., *J. Biol. Chem.* 268:24547-24550 (1993)). Co-expression of *dsbA*, *tlpA* or the genes of functionally analogous proteins is sometimes necessary for functional expression of cloned periplasmic enzymes.

The invention further provides a vector comprising a nucleotide molecule comprising a nucleotide sequence of a *Ketogulonigenium* replicon found on the endogenous plasmid of NRRL B-30035 (ADM 291-19) additionally comprising a DNA sequence that encodes an easily isolatable protein "handle" inserted adjacent to, and in translational fusion with either side of the polylinker, such that genes cloned into the polylinker are expressed as protein fusions between the cloned protein and the vector protein handle. In this way, purification of the protein products of the cloned gene are facilitated by applying well tested

purification procedures for the "protein handle." More specifically, the cloned protein follows the handle protein during the purification process. After purification, the fusion protein is then separated into its two native protein products through the application of factor Xa or other suitable protease.

In one embodiment, the vectors of the present invention comprise a cosmid site which allows for making a DNA library with about 30 kilobase inserts.

In an additional embodiment, the vectors of the present invention further comprises temperature-sensitive plasmid replication functions and regions of DNA homology to the host chromosome. The temperature sensitive replication functions allow control of integration into the homologous regions of the host chromosome through manipulation of temperature.

In one embodiment, the vectors of the present invention include two different antibiotic resistance genes inserted into two different restriction sites in the vector. One or both of the antibiotic resistance genes can carry a restriction site for gene cloning. In this way, one resistance marker is used to select transformants, and the other is used to screen for cloned DNA insertions by insertional inactivation. Suitable marker genes for the vectors of the present invention are all antibiotic resistance genes which are expressed in *E. coli* or *Ketogulonigenium*. Preferred antibiotic resistance genes include amikacin, ampicillin, chloramphenicol, erythromycin, gentamicin, kanamycin, penicillin, spectinomycin, streptomycin or tetracycline resistance genes. Particularly preferred markers include ampicillin, chloramphenicol, erythromycin, kanamycin, penicillin, spectinomycin, streptomycin, and/or tetracycline. Other suitable markers will be readily apparent to the skilled artisan.

Suitable replicons functional in *E. coli* for the vectors of the present invention can be selected from the group comprising DNA fragments containing a replicon of *E. coli*, or of any plasmid or phage which can autonomously replicate in *E. coli*. Such replicons may be isolated from the group comprising plasmids RP4, RSF1010, and pSUP301 (U.S. Patent No. 5,399,496) lambda

phages, such as phage lambda ATCC 10798, P1 phages such as P1 ATCC 25404-B1 or T-coliphages, such as coliphage T4 ATCC 11303-B4. Other suitable plasmids and replication origins will be readily apparent to the skilled artisan.

A suitable replicon functional in *Ketogulonigenium* for the vectors of the present invention is any DNA fragment which can support autonomous replication in *Ketogulonigenium*. Preferred are DNA fragments selected from the group comprising a replicon of *Ketogulonigenium* or of any plasmid, endogenous or otherwise, or phage which can autonomously replicate in *Ketogulonigenium*.

The mob site includes the origin of transfer (*oriT*) and acts as a recognition site for certain *trans* active plasmid transfer functions (R. Simon *et al.*, *Bio/Technology* 1:784-791 (1983)). A mob region is usually a *mob* gene and an *oriT* from any conjugation plasmid. The mob site can be obtained from a conjugative plasmid, e.g., plasmids RK2, RP4, RSF1010, or plasmids belonging to incompatibility groups IncP, IncQ, IncC, IncB, IncF, IncG, IncI, IncK, IncM, IncN, IncPa, IncPb, IncW, IncX, and IncZ or their derivatives. Other suitable conjugative plasmids will be readily apparent to the skilled artisan. The mob site containing plasmid can be transferred from its original host to another host with the help of *tra* genes by using bi-parental conjugal or tri-parental conjugal mating. The *tra* genes are well known as transfer genes of the broad-host-range conjugative plasmids. In the tri-parental conjugal mating, a donor strain harboring mob site-containing plasmids is mixed with a strain harboring plasmids containing *tra* genes, such as RP4 and RK2, and with a recipient strain. Conjugation provides an alternative method, for example, compared to transformation, of placing vectors and other genetic material into the cell. Other suitable methods will be readily apparent to the skilled artisan.

The vectors of the present invention may also comprise one or more further nucleotide molecules (also called "inserts"), for example DNA sequences having multicloning sites (also called "polylinker inserts"), expression control sequences, cos sites, terminator sequences, ribosome binding sites, DNA

sequences encoding signal peptides and/or proteins, to add further desirable functions to the shuttle vector. Other suitable inserts will be readily apparent to the skilled artisan.

In more detail, the vectors of the present invention may comprise DNA sequences including one or more multicloning sites (Messing *et al.*, *Methods in Enzymology* 101:20 (1983) derived from a variety of plasmids and phages, or from synthetic DNA sequences for convenient cloning. Suitable sources include pUC18 (Boehringer Mannheim), pUC19 (Boehringer Mannheim), M13mp8 (Boehringer Mannheim), and pBluescript. Other suitable multicloning sites will be readily apparent to the skilled artisan.

Furthermore, the vectors of the present invention may contain a wide variety of expression control sequences, such as the *E. coli lac*, *trp*, *tac* or beta-lactamase expression control system, control sequences of phage origin, such as the phage Lambda P_L promoter, or expression control sequences derived from *Ketogulonigenium* strains, to name a few. Other expression control systems will be known to the skilled artisan. In addition, the shuttle vectors may contain cos sites for *in vitro* packaging. Furthermore, the expression constructs will further optionally contain terminator sequences for transcription, initiation, termination, natural or synthetic ribosome binding sites for effective translation, DNA sequences encoding signal peptides for efficient localization of the cloned protein(s) and structural genes of marker proteins all of which are used to construct the efficient *Ketogulonigenium* host-vector systems of the present invention.

In another embodiment, the vector of the present invention may contain an *E. coli*, *Ketogulonigenium*, or other promoter from a suitable gene, in front of a promoterless reporter gene, such as β -galactosidase, β -glucuronidase, green fluorescent protein or α -amylase that has a polylinker inside the reporter gene. In this way, cloned nucleic acid inserts may be screened for by inactivation of the reporter function. Other suitable reporter genes will be readily apparent to the skilled artisan.

In another embodiment, pADM291 can contain a promoterless reporter gene such as β -galactosidase, β -glucuronidase, luciferase, green fluorescent protein, α -amylase, uroporphyrinogen III methyltransferase (cobA) from *Propionibacterium freudenreichii* inserted just adjacent to an inserted polylinker, such that DNA fragments cloned into the polylinker and comprising a transcriptional promoter can cause transcription and expression of the vector reporter function in the host. In this way, cloned inserts can be screened for those containing at least a transcriptional promoter that is active in the host organism.

In another embodiment, pADM291 can contain a reporter gene, such as β -galactosidase, β -glucuronidase, green fluorescent protein or α -amylase that lacks a promoter and a translation initiation site, inserted adjacent to a polylinker such that, nucleotide fragments cloned into the polylinker and comprising a promoter and translation initiation site can form a translational fusion with the reporter function and cause expression of the reporter function in the host. In this way, cloned inserts can be screened for those containing a transcriptional promoter and translation initiation site which are active in the host organism.

The vectors of the present invention can be obtained by the following steps using the materials as described in the present description and by using recombinant DNA techniques as described in the art and known to one of ordinary skill:

- (a) Preparing a nucleotide sequence containing a marker gene;
- (b) Preparing a nucleotide sequence containing a replicon functional in *E. coli*;
- (c) Preparing a nucleotide sequence containing a mob site;
- (d) Combining the nucleotides described in (a) through (c) along with the nucleotide of the present invention by digesting the said nucleotide sequences with an appropriate restriction enzyme and ligating them to obtain the recombinant shuttle vectors of the present invention.

By these steps, the vectors containing marker genes, a replicon functional in *E. coli*, a replicon functional in *Ketogulonigenium* and a functional mob site can be constructed.

In another embodiment the pADM291 nucleotide may be fused with another vector which replicates in *E. coli*, *Acetobacter*, *Corynebacterium*, *Bacillus*, *Rhodobacter*, *Paracoccus*, *Pseudomonas*, *Roseobacter*, *Pseudogluconobacter*, *Gluconobacter*, *Serriatia*, *Mycobacterium*, or *Streptomyces*. Preferred is fusion with a vector which replicates in *E. coli*, such as the pET vectors, pUC18, or pUC19. Other suitable vectors will be readily apparent to the skilled artisan.

The vectors of the present invention can be transferred from *E. coli* to *Ketogulonigenium* with a very high frequency by a conjugal mating without the isolation and purification of the vector DNA. With these vectors, a genomic library constructed in *E. coli* can be transferred into *Ketogulonigenium* in one experiment. Thus, the vectors of the present invention are highly efficient in view of their simplicity in cloning experiments.

Among vectors with which the the nucleotide sequence of a *Ketogulonigenium* replicon found on the endogenous plasmid contained in Deposit No. NRRL B-30035 may be combined include pA2, pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; pUC18, available from Takara Shuzo Co., Ltd.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

The vectors of the present invention incorporate a DNA sequence which confers upon the vector the capacity for autonomous replication in host *E. coli* cells.

The vector of the present invention may be combined with an *E. coli*-derived plasmid to create a vector which is able to replicate in both *Ketogulonigenium* and *E. coli*.

The present invention also provides a vector comprising a replicon functional in an organism selected from the genera comprising *Acetobacter*, *Corynebacterium*, *Bacillus*, *Rhodobacter*, *Paracoccus*, *Pseudomonas*, *Roseobacter*, *Pseudogluconobacter*, *Gluconobacter*, *Serratia*, *Mycobacterium*, and *Streptomyces*. Other suitable genera will be readily apparent to the skilled artisan.

The plasmid may optionally contain its native expression vector and/or expression vectors which include chromosomal-, episomal- and virus-derived vectors, e.g., vectors derived from bacterial endogenous plasmids, bacteriophage, yeast episomes, yeast chromosomal elements, viruses such as baculoviruses, papova viruses, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as cosmids and phagemids.

Among known bacterial promoters suitable for use in the present invention include the *E. coli lacI* and *lacZ* promoters, the T3 and T7 promoters, the *gpt* promoter, the lambda P_R and P_L promoters and the *trp* promoter. Other suitable promoters will be readily apparent to the skilled artisan.

The invention also provides a shuttle vector comprising a *Ketogulonigenium* replicon found on the endogenous plasmid contained in Deposit No. B-30035 and at least one exogenous nucleotide sequence, wherein said shuttle vector autonomously replicates in *Ketogulonigenium* and at least one organism selected from the genera comprising *Acetobacter*, *Corynebacterium*, *Bacillus*, *Rhodobacter*, *Paracoccus*, *Roseobacter*, *Pseudomonas*, *Pseudogluconobacter*, *Gluconobacter*, *Serratia*, *Mycobacterium*, and *Streptomyces*.

The invention also provides a method of transforming a host cell with the vectors of the present invention to obtain a stably transformed host cell. Such transformation can be effected by conjugation, including both biparental and triparental mating, transformation mediated by calcium chloride or other cation salts, electroporation and other transformation methods, transduction, or other

methods. Such methods are described in many standard laboratory manuals, such as Davis *et al.*, "Basic Methods in Molecular Biology," (1986) and J. Sambrook, E.F. Fritsch and T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). Preferred methods are conjugation and electroporation. Conjugation is a process by which bacterial unilaterally transfer DNA from a donor to a recipient cell through cell to cell contact.

Suitable host cells comprise organisms selected from the genera comprising *E. coli*, *Ketogulonigenium*, *Acetobacter*, *Corynebacterium*, *Bacillus*, *Rhodobacter*, *Paracoccus*, *Roseobacter*, *Pseudomonas*, *Pseudogluconobacter*, *Gluconobacter*, *Serriatia*, *Mycobacterium*, and *Streptomyces*.

A stably transformed cell is a cell wherein a transgene (recombinant DNA) is transmitted to every successive generation. The vector transformed into the above-mentioned cells can optionally comprise a transgene. As used herein, a transgene is defined as a transplanted nucleotide sequence which is exogenous, or non-native, to the host. An exogenous nucleotide sequence, as used in the current context, is a nucleotide sequence which is not found in Deposit No. NRRL B-30035. Thus, the term exogenous nucleotide sequence is meant to encompass a nucleotide sequence that is foreign to Deposit No. NRRL B-30035, as well as a nucleotide sequence endogenous, or native, to Deposit No. NRRL B-30035 that has been modified. Modification of the endogenous nucleotide sequence may include, for instance, mutation of the native nucleotide sequence or any of its regulatory elements. As used herein, mutation is defined as any change in the wild-type sequence of genomic or plasmid DNA. An additional form of modification may also include fusion of the endogenous nucleotide sequence to a nucleotide sequence that is normally not present, in relation to the endogenous nucleotide sequence. The transgene may be regulated by its normal promoter, or more commonly, by a promoter that normally regulates a different gene. The invention also provides a method for producing transformed *Ketogulonigenium*, comprising transforming *Ketogulonigenium* with a transgene,

comprising, an endogenous *Ketogulonigenium* replicon. Preferably, the endogenous *Ketogulonigenium* replicon is contained in Deposit No. NRRL B-30035. The term replicon as used herein is meant to encompass a DNA sequence comprising those genes and gene expression control elements such as promoters and terminators, other DNA sequence features such as short sequence repeats (iterons), origins of plasmid replication (*ori* or *oriV* sites), or other DNA sequence features that are required to support the autonomous replication of a circular DNA molecule in a bacterial host (Chapter 122, pp. 2295-2324, in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd Edition, Frederick C. Neidhardt, Ed., ASM Press (1996)). The requirements of a replicon can vary from as little as a short *ori* sequence in the case of plasmids that do not require their own replication proteins, to larger sequences comprising one or more plasmid-borne replication genes. The definition of a transformed cell, as used herein, is a cell where DNA has been inserted into a bacterial cell. The transformation of *Ketogulonigenium* may be transient or stable.

In an additional embodiment, the invention provides a method for producing a polypeptide comprising culturing a host cell comprising vectors of the present invention under conditions such that the polypeptide is expressed, and recovering the polypeptide. The translated polypeptide encoded by the DNA in the plasmid may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals, but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification, or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification.

The translated protein encoded by the DNA contained in the plasmid can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography,

hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Methods used and described herein are well known in the art and are more particularly described, for example, in R.F. Schleif and P.C. Wensink, *Practical Methods in Molecular Biology*, Springer-Verlag (1981); J.H. Miller, *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1972); J.H. Miller, *A Short Course in Bacterial Genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1992); M. Singer and P. Berg, *Genes & Genomes*, University Science Books, Mill Valley, California (1991); J. Sambrook, E.F. Fritsch and T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989); P.B. Kaufman *et al.*, *Handbook of Molecular and Cellular Methods in Biology and Medicine*, CRC Press, Boca Raton, Florida (1995); *Methods in Plant Molecular Biology and Biotechnology*, B.R. Glick and J.E. Thompson, eds., CRC Press, Boca Raton, Florida (1993); P.F. Smith-Keary, *Molecular Genetics of Escherichia coli*, The Guilford Press, New York, NY (1989); *Plasmids: A Practical Approach*, 2nd Edition, Hardy, K.D., ed., Oxford University Press, New York, NY (1993); *Vectors: Essential Data*, Gacesa, P., and Ramji, D.P., eds., John Wiley & Sons Pub., New York, NY (1994); *Guide to Electroporation and electrofusions*, Chang, D., *et al.*, eds., Academic Press, San Diego, CA (1992); *Promiscuous Plasmids of Gram-Negative Bacteria*, Thomas, C.M., ed., Academic Press, London (1989); *The Biology of Plasmids*, Summers, D.K., Blackwell Science, Cambridge, MA (1996); *Understanding DNA and Gene Cloning: A Guide for the Curious*, Drlica, K., ed., John Wiley and Sons Pub., New York, NY (1997); *Vectors: A Survey of Molecular Cloning Vectors and Their Uses*, Rodriguez, R.L., *et al.*, eds., Butterworth, Boston, MA (1988); *Bacterial Conjugation*, Clewell, D.B., ed., Plenum Press, New York, NY (1993); Del Solar, G., *et al.*, "Replication and control of circular bacterial plasmids," *Microbiol. Mol. Biol. Rev.* 62:434-464 (1998); Meijer, W.J., *et al.*, "Rolling-circle

plasmids from *Bacillus subtilis*: complete nucleotide sequences and analyses of genes of pTA1015, pTA1040, pTA1050 and pTA1060, and comparisons with related plasmids from gram-positive bacteria," *FEMS Microbiol. Rev.* 21:337-368 (1998); Khan, S.A., "Rolling-circle replication of bacterial plasmids," *Microbiol. Mol. Biol. Rev.* 61:442-455 (1997); Baker, R.L., "Protein expression using ubiquitin fusion and cleavage," *Curr. Opin. Biotechnol.* 7:541-546 (1996); Makrides, S.C., "Strategies for achieving high-level expression of genes in *Escherichia coli*," *Microbiol. Rev.* 60:512-538 (1996); Alonso, J.C., *et al.*, "Site-specific recombination in gram-positive theta-replicating plasmids," *FEMS Microbiol. Lett.* 142:1-10 (1996); Miroux, B., *et al.*, "Over-production of protein in *Escherichia coli*: mutant hosts that allow synthesis of some membrane protein and globular protein at high levels," *J. Mol. Biol.* 260:289-298 (1996); Kurland, C.G., and Dong, H., "Bacterial growth inhibited by overproduction of protein," *Mol. Microbiol.* 21:1-4 (1996); Saki, H., and Komano, T., "DNA replication of IncQ broad-host-range plasmids in gram-negative bacteria," *Biosci. Biotechnol. Biochem.* 60:377-382 (1996); Deb, J.K., and Nath, N., "Plasmids of corynebacteria," *FEMS Microbiol. Lett.* 175:11-20 (1999); Smith, G.P., "Filamentous phages as cloning vectors," *Biotechnol.* 10:61-83 (1988); Espinosa, M., *et al.*, "Plasmid rolling circle replication and its control," *FEMS Microbiol. Lett.* 130:111-120 (1995); Lanka, E., and Wilkins, B.M., "DNA processing reaction in bacterial conjugation," *Ann. Rev. Biochem.* 64:141-169 (1995); Dreiseikelmann, B., "Translocation of DNA across bacterial membranes," *Microbiol. Rev.* 58:293-316 (1994); Nordstrom, K., and Wagner, E.G., "Kinetic aspects of control of plasmid replication by antisense RNA," *Trends Biochem. Sci.* 19:294-300 (1994); Frost, L.S., *et al.*, "Analysis of the sequence gene products of the transfer region of the F sex factor," *Microbiol. Rev.* 58:162-210 (1994); Drury, L., "Transformation of bacteria by electroporation," *Methods Mol. Biol.* 58:249-256 (1996); Dower, W.J., "Electroporation of bacteria: a general approach to genetic transformation," *Genet. Eng.* 12:275-295 (1990); Na, S., *et al.*, "The factors affecting transformation efficiency of coryneform bacteria by

electroporation," *Chin. J. Biotechnol.* 11:193-198 (1995); Pansegrau, W., "Covalent association of the *traI* gene product of plasmid RP4 with the 5'-terminal nucleotide at the relaxation nick site," *J. Biol. Chem.* 265:10637-10644 (1990); and Bailey, J.E., "Host-vector interactions in *Escherichia coli*," *Adv. Biochem. Eng. Biotechnol.* 48:29-52 (1993).

The following examples are illustrative only and are not intended to limit the scope of the invention as defined by the appended claims. It will be apparent to those skilled in the art that various modifications and variations can be made in the methods of the present invention without departing from the spirit and scope of the invention. Thus, it is intended that the present invention cover the modifications and variations of this invention provided they come within the scope of the appended claims and their equivalents.

All patents and publications referred to herein are expressly incorporated by reference.

EXAMPLES

EXAMPLE 1

Isolation of plasmid DNA from *Ketogulonigenium* and *E. coli* strains.

A fresh culture of *Ketogulonigenium* strain ADM291-19 (NRRL B-30035) was grown in 10 ml X6L medium (2% Mannitol, 1% Soytone, 1% Yeast Extract, 0.5% Malt Extract, 0.5% NaCl, 0.25% K₂HPO₄, pH 7.8). A fresh culture of *E. coli* strain DH5 α MCR harboring plasmid pUC19 DNA was grown in 10 ml of Luria Broth (10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl) with 100 μ g/ml ampicillin. Both cultures were incubated overnight with orbital shaking at 250 rpm at 30 degrees C for *Ketogulonigenium* and 37 degrees C for *E. coli*. DNA was isolated using the Promega Wizard *Plus* Midipreps DNA Purification System (Madison, WI). The culture was centrifuged at 10,000 x g for 10 minutes at 4°C in a Sorval RC-5B centrifuge using the SS-34 rotor. The pellet was

-22-

suspend in 3 ml of 50mM Tris-HCl, pH 7.5, 10mM EDTA, 100(g/ml RnaseA. Three ml of cell lysis solution (0.2M NaOH, 1%SDS) was added and mixed by inverting the tube, then three ml of neutralization solution (1.32M potassium acetate, pH4.8) was added and mixed by inverting the tube. The lysate was centrifuged at 14,000 x g for 15 minutes at 4(C in a SS-34 rotor, then the supernatant was carefully decanted to a new centrifuge tube. Ten ml of resuspension resin (40% isopropanol, 4.2M guanidine hydrochloride) was added to the supernatant fluid and mixed by swirling, then the mixture was into the Promega Wizard Midicolumn, which was connected a vacuum manifold. Vacuum was applied to pull the resin/DNA mixture completely into the midicolumn. The column was washed twice with 15 ml of column wash solution (95% ethanol, 80mM potassium acetate, 8.3mM Tris-HCl, pH 7.5, 40MEDTA. The reservoir was removed from the midicolumn with a scissors, then the column was placed in a microcentrifuge tube and centrifuged at 10,000 x g for 2 minutes to remove any residual solution. The midicolumn was transferred to a new microcentrifuge tube and 300 (l of sterile dH₂O was applied. The tube was microcentrifuged at 10,000 x g for 20 seconds to elute to the DNA into solution. About 2.5 ug of pADM291 plasmid DNA was recovered. A similar procedure would be employed for other plasmids from *Ketogulonigenium* or *E. coli*, except that choice and concentration of selective antibiotics would be altered as suitable for the plasmid being isolated.

EXAMPLE 2

Isolation of strains ADM29101 and ADMX6L01, Nalidixic acid resistant mutants of *Ketogulonigenium* strains ADM291-19 (NRRL B-30035) and ADMX6L(NRRL B-21627), respectively.

Three ml of X6L medium (2% Mannitol, 1% Soytone, 1% Yeast Extract, 0.5% Malt Extract, 0.5% NaCl, 0.25% K₂HPO₄, pH 7.8) was inoculated with

-23-

strain ADM291-19 (NRRL B-30035) and the culture was grown overnight in an orbital shaker at 30 degrees C, 250 rpm. 50 uL of the overnight culture was transferred to 3ml of X6L medium to which 1 μ l of a 1% aqueous solution of Ethidium bromide had been added. This culture was incubated overnight as before. 200 μ l of the second overnight culture was spread on a plate of X6L medium containing 1.3% agar plus 50 μ g/ml of nalidixic acid. The plate was incubated 2-3 days at 30°C until colonies appeared. Colonies growing on these plates were picked and naladixic acid resistant strains were purified by additional cycles of streak purification on X6L medium plates containing 50 μ g/ml of nalidixic acid. A naladixic acid resistant strain was preserved as strain ADM 29101. The same procedure was used to create ADMX6L01 from ADMX6L.

EXAMPLE 3

Construction of *Ketogulonigenium* suicide plasmid vector pJND1000.

To invent an *E. coli*/*Ketogulonigenium* shuttle vector it was desirable that we first identify a subsequence from pADM291 that, when cloned into *E. coli* or other plasmid vectors, would support plasmid replication in *Ketogulonigenium* hosts. In order to do this, a plasmid vector that could replicate in *E. coli* and be transferred efficiently to *Ketogulonigenium*, but which cannot replicate in *Ketogulonigenium* was devised.

pJND1000 was constructed from segments of plasmids pUC19 (GenBank Accession No. M77789), pUC4K (GenBank Accession No. X06404), pFD288 (GenBank Accession No. U30830), and pFC5 (David M. Lonsdale et.al, 1995, "pFC1 to pFC7: A novel family of combinatorial cloning vectors.", Plant Molecular Biology Reporter 13[4]:343-345).

The ampicillin resistance gene (*amp*^R) was removed from pUC19 by digesting pUC19 DNA with restriction enzymes *Dra*I and *Ssp*I, separating the 1748 bp vector fragment from the smaller *amp*^R fragment and other fragments by

gel electrophoresis, then recovering the 1748 bp fragment from a gel slice. A kanamycin resistance gene (*kan^R*) fragment from pUC4K was prepared by digesting pUC4K with restriction enzyme *Pst*I, treating the mixture with Klenow Fragment to produce blunt ends, separating the fragments by gel electrophoresis, then purifying the 1240 bp *kan^R* fragment from a gel slice. The isolated fragments from pUC19 and pUC4K were mixed and ligated with T4 ligase following the protocol of GibcoBRL technical bulletin 15244-2 (Rockville, MD) to produce "intermediate p1", an intermediate plasmid carrying pUC19 features and a *kan^R* gene. Strain *E. coli* DH5(MCR was transformed with the ligation mixture following an established protocol ("Fresh Competent *E. coli* prepared using CaCl_2 ", pp. 1.82-1.84, in J. Sambrook, E.F. Fritsch and T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, 2nd Ed.). Transformants were selected on Luria Broth plates containing 50 $\mu\text{g/ml}$ of kanamycin.

An *oriT* site for conjugative transfer was obtained from plasmid pFD288 by restricting it with *Hae*II, converting the single stranded ends to blunt ends by treating the mixture with Klenow Fragment, separating the fragments by agarose gel electrophoresis, then purifying the 778 bp *oriT* fragment from a gel slice. The intermediate p1 plasmid was opened at a single site with restriction enzyme *Sap*I, then treated with Klenow Fragment to convert the single stranded ends to blunt ends. The *Sap*I digested, blunt ended p1 intermediate was mixed with the purified *oriT* fragment, then treated with T4 ligase as above to create "intermediate p2", a plasmid which carries *oriT* in addition to *kan^R* and other pUC19-derived features. Strain *E. coli* DH5 α MCR was transformed with the ligation mixture as above. Intermediate p2 was confirmed by restriction digestion analysis.

The polylinker in intermediate plasmid p2 was replaced with the polylinker from pFC5. To do this the two plasmids were separately restricted with *Pvu*II, the fragments from each digestion reaction were separated by gel electrophoresis, then the pFC5-derived polylinker fragment (531 bp), and the larger non-polylinker fragment from intermediate p2 were purified from gel

-25-

slices. The recovered p2 fragment and the pFC5-derived polylinker fragment were mixed and joined using T4 ligase as above to make plasmid pJND1000. The structure of pJND1000 was confirmed by restriction digestion analysis. PJND1000 has a functioning kanamycin resistance gene, an RK2-derived *oriT* site to permit conjugative transfer, a polylinker for DNA cloning, forward and reverse M13 primers to facilitate DNA sequencing reactions into the polylinker, replicates in *E. coli* strains but not in *Ketogulonigenium* strains, and permits screening for cloned inserts by inactivation of *lacZ α* using Xgal indicator plates.

EXAMPLE 4

Cloning of DNA sequences from plasmid pADM291 into suicide vector pJND1000 to make *E. coli/Ketogulonigenium* shuttle vector candidates.

To find the functional replicon of pADM291, the Polymerase Chain Reaction (PCR) was used to generate various plasmid pADM291 DNA fragments flanked by restriction endonuclease recognition sites. These were cloned into the *Ketogulonigenium* suicide vector pJND1000 to make *Ketogulonigenium* shuttle vector candidates, which were transferred to *E. coli* hosts by plasmid transformation.

Forward and reverse oligonucleotide primers were designed with which to initiate PCR reactions using purified pADM291 DNA as substrate. The oligonucleotide primers were prepared by a commercial lab (Sigma/Genosys, The Woodlands, TX) using established methods known in the art. The primers had the following features ("F"=forward primer; "R"=reverse primer):

Table I

| <u>Primer Label</u> | <u>DNA Sequence</u> |
|---------------------|-----------------------------------|
| Primer 1F | CGGAATTCGATCATATCATTCCCCAAGCGGAC |
| Primer 1R | GCTCTAGACGATGTACTCCTTGGTGCTCTCGAT |
| Primer 2F | CGGAATTCTGCTTCTTTTCGTTTCGTTCCGCC |
| Primer 2R | GCTCTAGAACATCGCTCATCTGTAGTCGCC |
| Primer 3F | CGG AATTCACAGTCAGGTGGCACATGTTCC |
| Primer 3R | CGG GATCCTGTGAAAAAGTGAGGACAGGCGGG |
| Primer 4F | CGGAATTCGGCAATGGGTTCGAAATTCATAG |
| Primer 4R | CGGGATCCACGTTCCCTATTTTCCTCAATC |
| Primer 5F | CGGAATTCACACCGAAACACCTAACACGCAAG |
| Primer 5R | CGGGATCCAGTGCGGTTACGTCATCAATG |
| Primer 6F | CGGAATTCTGCACTGCCGCTCTCGAAATG |
| Primer 6R | CGGAATTCACAAGATTGACGCAGCTCTTCGC |
| Primer 7F | CGGAATTCGGCAATGGGTTCGAAATTCATAG |
| Primer 7R | CGCGGATCCACTTGTGTTGTCTTTCCC |
| Primer 8F | GCGGAATTCGACGCTGCAAACATCGAAAAAC |
| Primer 8R | CGGGATCCACGTTCCCTATTTTCCTCAATC |

The PCR reactions generated pADM291 DNA fragments having the following endpoints, using the numbering of the DNA sequence in Figure 2 (SEQ ID NO2:).

Table II

| <u>PCR Fragment Label</u> | <u>Length</u> | <u>Fragment Endpoints</u> | <u>End Restriction Sites</u> |
|---------------------------|---------------|---------------------------|-----------------------------------|
| PCR#1 | 2210 bp | 8060 through 1760 | 5' <i>EcoRI</i> – <i>XbaI</i> 3' |
| PCR#2 | 1523 bp | 876 through 2398 | 5' <i>EcoRI</i> – <i>XbaI</i> 3' |
| PCR#3 | 1222 bp | 2017 through 3238 | 5' <i>EcoRI</i> – <i>BamHI</i> 3' |
| PCR#4 | 2517 bp | 2955 through 5471 | 5' <i>EcoRI</i> – <i>BamHI</i> 3' |
| PCR#5 | 1965 bp | 4704 through 6668 | 5' <i>EcoRI</i> – <i>BamHI</i> 3' |
| PCR#6 | 2516 bp | 5908 through 8423 | 5' <i>EcoRI</i> – <i>EcoRI</i> 3' |
| PCR#7 | 520 bp | 2955 through 3474 | 5' <i>EcoRI</i> – <i>BamHI</i> 3' |
| PCR#8 | 1065 bp | 4407 through 5471 | 5' <i>EcoRI</i> – <i>BamHI</i> 3' |

PCR products were generated in a GENEAMP PCR System 9700 thermal cycler from Applied Biosystems using Gibco BRL *Taq* Polymerase (Rockville, MD), using 35 cycles of 94°C for 30 seconds, 60°C for 30 seconds, 72°C for 60 seconds. Each reaction contained 1 μ l of a 50 μ M solution of each primer, 10ng of pADM291 template, 1.5mM MgCl₂ and PCR Buffer supplied with the *Taq* Polymerase. *Taq* polymerase and buffer were removed from the PCR products using the Promega Wizard DNA Clean-up Kit (Madison, WI). The pADM291 DNA fragments from each PCR reaction were prepared for ligation into plasmid pJND1000 by reaction with the appropriate endonuclease enzymes from Gibco BRL (Rockville, MD) as given in Table II (for example the PCR#1 products were digested with *EcoRI* and *XbaI*), followed by removal of enzyme and reaction buffer using Promega Wizard DNA Clean-up Kits. Likewise, for each PCR fragment a sample of purified plasmid pJND1000 DNA vector was digested with the same restriction enzymes that had been used to digest the PCR fragment, followed by enzyme and buffer removal using Promega Wizard DNA Clean-up Kits. Each PCR fragment was ligated into the pJND1000 vector in reactions utilizing T4 DNA ligase as in Example 3. A plasmid transformation was performed with each ligation product using competent *E. coli* strain DH5 α MCR

-28-

as the recipient following the transformation protocol of Example 3. Transformants were spread on Luria Broth agar plates containing 50 ug/ml of kanamycin and 40 ug/ml of Xgal and grown at 37 deg C. Colonies that were white, indicating successful insertion of a PCR fragment into the pJND1000 vector, were picked and preserved, and a sample was transferred to 1 ml of Luria Broth containing 50 ug/ml kanamycin and incubated in an orbital shaker overnight at 37 deg C, 250 rpm. Plasmid DNA was prepared from each of the grown transformant cultures by the alkaline method of H.C. Birnboim and J. Doly (Nucleic Acids Research, 7:1513-1523 (1979)). The plasmids from these cultures, each of which was a candidate for a *Ketogulonigenium* shuttle vector, were named "p291-1" for PCR product #1, "p291-2" for PCR product #2, etc. Each plasmid was confirmed to be carrying the expected PCR insert via restriction digestion and gel electrophoresis.

EXAMPLE 5

Discovery of *Ketogulonigenium* shuttle vector p291-4 by conjugative transfer from *E. coli* hosts into *Ketogulonigenium* hosts and plasmid maintainance in *Ketogulonigenium* and bi-parental mating.

Ketogulonigenium shuttle vector plasmid candidates from Example 4 were introduced into *E. coli* strain S17-1 (ATCC 47055) by transformation, then screened for their ability to replicate in *Ketogulonigenium* hosts after conjugative plasmid transfer from *E. coli* S17-1 to *Ketogulonigenium* strains ADM 29101 and ADM X6L01 using bi-parental matings.

A 5 ml overnight culture of each of the *E. coli* S17-1 transformants was prepared in Luria Broth containing 50 ug/ml of kanamycin with orbital shaking at 250 rpm, 37 deg.C. Concurrently, 5 ml shaking overnight cultures of strains ADM X6L01 and ADM 29101 (*Ketogulonigenium robustum* and *Ketogulonigenium sp.*, respectively), were prepared in X6L medium containing

-29-

50 µg/ml nalidixic acid at 30 deg C. The next day, 50 µl of each *E. coli* culture was transferred into 3 ml of fresh Luria Broth without antibiotic and grown at 37 deg C, 250 rpm until reaching an optical density of 0.1 to 0.4 at 600 nm wavelength. Likewise, 200 µl of each *Ketogulonigenium* culture was transferred to 3 ml of fresh X6L medium without antibiotic and grown at 30 deg C, 250 rpm until reaching an optical density of 0.6 at 600 nm. At this point a bi-parental mating of each *E. coli* culture with each *Ketogulonigenium* culture was prepared. 400 µl of *E. coli* culture was mixed in a sterilized 1.5 ml microcentrifuge tube with 1 ml of *Ketogulonigenium* culture. The mixed cells were pelleted in an Eppendorf microcentrifuge and the supernatant was decanted. Cell pellets were suspended in 100 µl of fresh X6L medium and the entire cell suspension was spotted onto a sterile, 0.45 µm X 25 mm GN-6 metric filter from Gelman Sciences (Product No. 63068), which had been placed onto the surface of a fresh X6L agar medium plate. The petri plate and filter with the cells on it was incubated overnight at 30 deg C. The mated cells were then removed from the filter by suspension in 3 ml of fresh X6L medium. 200 µl of the cell suspension were then plated on fresh X6L medium containing 50 µg/ml kanamycin, 50 µg/ml nalidixic acid, and 1.3% Difco Bacto Agar. The plates were incubated at 30 deg C for two days then inspected.

Matings between *E. coli* S17-1 hosts harboring plasmid p291-4 and either ADM 29101 or ADM X6L01 recipients produced more than 300 *Ketogulonigenium* transconjugant colonies on the X6L-kanamycin-naladixic acid medium plates. Matings involving any of the other shuttle plasmid candidates produced no colonies with ADM X6L01 recipients, and 3-10 colonies with ADM 29101 recipients. The later may be attributed to recombination of the entering plasmid with the endogenous pADM291 plasmid in the case of ADM 29101 recipients, rather than to maintenance of the incoming plasmid as an independently replicating plasmid.

The identity of the p291-4 *E. coli*/*Ketogulonigenium* shuttle plasmid from the ADM 29101 and ADM X6L01 transconjugant strains was confirmed by

restriction digestion and sequencing of plasmid DNA recovered from a transconjugant. The p291-4 shuttle vector was able to be maintained in the *Ketogulonigenium* recipient hosts. The suitability of strains ADM X6L01 and ADM29101 as conjugation recipients demonstrates that both the p291-4 shuttle vector and the transformation method is efficient with multiple species of the genus *Ketogulonigenium*. Methods analogous to those presented in the examples of this invention could be employed to derive additional shuttle vectors for use in *Ketogulonigenium*, for example, from portions of other endogenous plasmids native to other *Ketogulonigenium* strains, including but not limited to endogenous plasmids of the species *Ketogulonigenium robustum* and *Ketogulonigenium vulgariae*. The DNA sequence of PCR product #4, which corresponds to the *Ketogulonigenium* portion of p291-4, is shown in Figure 4 (SEQ ID NO:4).

EXAMPLE 6

Conjugative transfer of shuttle vectors from *E. coli* hosts into *Ketogulonigenium* recipients using tri-parental mating.

The *E. coli* S17-1 strain that was used as the donor host in the bi-parental mating of Example 5 provides mobilization functions from a chromosomally integrated plasmid. In a tri-parental mating, plasmid-encoded mobilization functions can be provided by a third bacterial strain harboring a broad host range plasmid such as, for example, RP4 (GenBank Accession No. L27758). This would enable a broader range of bacterial strains to be used as the donor host of the shuttle plasmid that is to be conjugated into the *Ketogulonigenium* recipient. An example of a suitable third strain to be used as the source of the mobilization function in a tri-parental mating would be *E. coli* strain HB101 (ATCC 33694) transformed with RP4. An example of a non-mobilizing donor strain harboring the *Ketogulonigenium* shuttle vector would be *E. coli* HB101 transformed with

an *E. coli*/*Ketogulonigenium* shuttle vector plasmid. The procedure would be the same as in the previous Example, except for the following modification:

Fresh liquid cultures of *Ketogulonigenium* and the shuttle vector donor strain would be prepared as in the bi-parental mating, except that *E. coli* strain HB101 or another suitable host transformed with the shuttle vector would be used instead of the S17-1 transformant. The third strain, *E. coli* HB101 harboring RP4 for example, would be cultured in suitable medium (Luria Broth containing 5 ug/ml of tetracycline in this example, or a different antibiotic if dictated by the selected broad host range plasmid), to an optical density of 0.1 to 0.4 at 600 nm. To prepare the mating mixture, 200–400 ul of the two non-*Ketogulonigenium* cultures would each be centrifuged to a pellet in microcentrifuge tubes. The supernatants would be decanted and the pellets would be combined and resuspended in 1.0 ml of fresh X6L medium without antibiotics, then recentrifuged to a pellet and the supernatant decanted again. This washing step in X6L medium is to prevent residual tetracycline from entering the mating mixture. Finally, 1.0 ml of the fresh culture of naladixic acid resistant *Ketogulonigenium* is added and the three cell populations are resuspended together, centrifuged to a pellet, then resuspended in 100 ul of fresh X6L medium and spotted onto the mating filter as in the bi-parental mating. The rest of the tri-parental mating is the same as in the bi-parental mating example.

EXAMPLE 7

Subcloning of p291-4 to make p291-4DS and definition of a pADM291 replicon.

Example 4 demonstrated that a 2517 bp DNA sequence of endogenous *Ketogulonigenium* plasmid pADM291, corresponding to the DNA sequence of Figure 4 (SEQ ID NO:4), supports the replication of non-native circular DNA molecules in multiple *Ketogulonigenium* species. Upon examination of this sequence using programs of the "Wisconsin Package" (Genetics Computer Group

[GCG], Madison, WI, version 10.1), several open reading frames (ORFs) with protein encoding potential were discovered to exist. These correspond to Figure 2 (SEQ ID NO:2) bases 499 through 1512 for "ORF1", encoding a 337 aa long protein, and bases 1584 through 1979 for "ORF2", encoding a 131 aa long "protein". The predicted amino acid sequences of these proteins was compared against known proteins in the PIR and SWISS-PROT databases using the GCG FASTA and BLAST programs. The results show both to be probable plasmid replication functions, showing the closest homology to other plasmid rep proteins for ORF1, and plasmid resolvase proteins for ORF2. Potential roles for these proteins include plasmid DNA replication, plasmid partitioning or plasmid mobilization, resolution of plasmid multimers, or other functions. The Figure 4 DNA sequence includes non-coding regions of 498 bp upstream of ORF1, 70 bp between ORF1 and ORF2, and 538 bp downstream of ORF2. These non-coding regions could contain one or more plasmid replication origin sites (*oriV*), iterons, sequences encoding antisense RNA encoding sequences, transcriptional promoters, or other features that may participate in control of plasmid replication, partitioning, or copy number. We therefore sought to determine if the *Ketogulonigenium* functional replicon defined by PCR#4 insert of p291-4 could be made smaller by deleting various regions of p291-4 and reintroducing the smaller plasmids into *Ketogulonigenium* by bi-parental mating as in Example 5.

Plasmid p291-4NX had a 759 bp deletion removing DNA from the *Ketogulonigenium* *NheI* site to the vector polylinker *XbaI* site, leaving only the amino terminus of ORF1 and the noncoding region upstream of ORF1 intact. p291-4ES had a 453 bp deletion removing DNA from the *Ketogulonigenium* *EcoRV* site to the vector *SfoI* site, leaving ORF1 and ORF2 and downstream regions intact but removing DNA upstream of ORF1, perhaps including the ORF1 promoter. The plasmid p291-4DS has a 404 bp deletion removing DNA from the *DraI* site to the *ScaI* site of the *Ketogulonigenium* region, leaving ORF1 and its upstream region and ORF2 intact, but removing some but not all of the DNA downstream of ORF2. p291-4DS is able to be maintained in *E. coli*, be

transferred to *Ketogulonigenium* hosts, and can replicate in *Ketogulonigenium* hosts.

This Example defines a second *E. coli*/*Ketogulonigenium* shuttle plasmid, p291-4DS, having the DNA sequence shown in Figure 3 (SEQ ID NO:3). The DNA sequence of the *Ketogulonigenium* portion of p291-4DS, which comprises a *Ketogulonigenium* plasmid replicon, is shown in Figure 1 (SEQ ID NO:1). It is recognized that additional DNA sequences not present in p291-4DS might participate in some aspects of pADM291 plasmid replication, plasmid multimer resolution or other maintenance functions, or plasmid partitioning in *Ketogulonigenium* hosts under these or other conditions. If so, it is realized that these functions could be localized and defined using similar methods in accordance with the present invention.

EXAMPLE 8

Transformation of a *Ketogulonigenium* host using electroporation.

Ketogulonigenium robustum strain ADM X6L (NRRL B-21627) was transformed with plasmid p291-4DS using the electroporation method. Competent *Ketogulonigenium* cells were prepared by seeding a single colony of ADMX6L into 10 ml of X6L medium (1% soytone, 1% yeast extract, 0.5% malt extract, 0.5% NaCl, 0.25% K₂HPO₄, 2% mannitol, pH 7.8). The culture was shaken at 300 rpm at 30 C until reaching an optical density of 0.8 units at 600 nm wavelength. Five ml of this culture was used to seed 500 ml of fresh X6L medium in a 2 L baffled erlenmeyer flask, which was shook at 300 rpm at 30 C until again reaching an optical density of 0.8 units. The culture was chilled in an ice-water bath 10 minutes, then transferred to a pre-chilled centrifuge bottle and centrifuged at 5,000 rpm in a Sorvall 5C-RB Refrigerated Centrifuge for 15 minutes at 4 degrees C. Cells were maintained at 2-4 degrees C for all steps to follow. The supernatant was decanted and the cell pellet suspended in 5 ml

-34-

ice-cold Milli-Q water, then additional cold Milli-Q water was added to a volume of 500 ml. The cells were centrifuged as before, then rewashed in 500 ml of Milli-Q water as before and recentrifuged. The twice-washed cell pellet was suspended in 40 ml of ice-cold 10% glycerol then centrifuged again as before. The supernatant was decanted, then the cells were suspended in a volume of chilled 10% glycerol approximately equal to the volume of the cell pellet. The competent *Ketogulonigenium* cells were aliquoted to microcentrifuge tubes (40 ul per tube) and stored at -80 degrees C.

A BioRad "Gene Pulser II" electroporator device was set to 1.5 kV, 25 mF. The pulse controller was set to 200 ohms. One ul of p291-4DS DNA prepared as in Example 1 was added to 40 ul of thawed chilled competent *Ketogulonigenium* cells on ice. The cell-DNA mixture was transferred to a pre-chilled electroporation cuvette, which was then transferred to the electroporation device and the pulse was applied. One ml of X6L medium was added to the cuvette, then the mixture was removed, transferred to a 10-ml test tube, and incubated for 2 hours with orbital shaking at 300 rpm and 30 degrees C. The incubated cells (approximately 1.4 ml) were placed in a microcentrifuge tube and spun at 13,000 rpm for 2 minutes, after which 0.9 ml of clear supernatant was carefully removed. The cell pellet was suspended in the remaining supernatant, then the cells were spread onto the surface of an X6L medium agar plate (1.2% Difco Bacto Agar) containing 20 ug/ml of kanamycin, and the plate was incubated for 2 days at 30 degrees C. Colonies that formed on this plate were confirmed to be *Ketogulonigenium* transformed with plasmid vector p291-4DS.

* * * * *

All publications mentioned herein above are hereby incorporated in their entirety by reference. All publications cited in the annotations to the Genbank

accession numbers or in ATCC strain descriptions cited herein are hereby incorporated in their entirety by reference.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be appreciated by one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention and appended claims.

WHAT IS CLAIMED IS:

1. An isolated or purified nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a nucleotide sequence of a *Ketogulonigenium* plasmid replicon found on the endogenous plasmid contained in Deposit No. NRRL B-30035.
2. The nucleic acid molecule of claim 1, wherein said plasmid replicon comprises the nucleic acid sequence in SEQ ID NO:1.
3. The nucleic acid molecule of claim 1, comprising the DNA sequence shown in SEQ ID NO: 3.
4. The nucleic acid molecule of claim 1 comprising the DNA sequence shown in SEQ ID NO:4.
5. The nucleic acid molecule of claim 1, comprising a replicon functional in *E. coli*.
6. The nucleic acid molecule of claim 1, comprising a replicon functional in an organism selected from the genera consisting of *Acetobacter*, *Corynebacterium*, *Bacillus*, *Rhodobacter*, *Paracoccus*, *Roseobacter*, *Pseudomonas*, *Pseudogluconobacter*, *Gluconobacter*, *Serratia*, *Mycobacterium*, and *Streptomyces*.
7. The nucleic acid molecule of claim 1, comprising a mob region.
8. The nucleic acid molecule of claim 7, wherein said mob region comprises a mob gene and an oriT from a conjugation plasmid.

9. The nucleic acid molecule of claim 8, wherein said conjugation plasmid is selected from plasmids which are included within the incompatibility groups consisting of IncP, IncQ, IncC, IncB, IncF, IncG, IncI, IncK, IncM, IncN, IncPa, IncPb, IncW, IncX, and IncZ.

10. The nucleic acid molecule of claim 1, comprising a temperature-sensitive replicon.

11. The nucleic acid molecule of claim 1, comprising at least one marker gene.

12. The nucleic acid molecule of claim 11, wherein said marker gene comprises a nucleotide sequence operative to direct synthesis of a protein conferring antibiotic resistance in a host cell population.

13. The nucleic acid molecule of claim 12, wherein said antibiotic is selected from the group comprising ampicillin, chloramphenicol, erythromycin, kanamycin, spectinomycin, streptomycin and tetracycline.

14. The nucleic acid molecule of claim 1, comprising at least one further nucleic acid sequence, wherein said further nucleic acid sequence is selected from the group consisting of a polylinker insert, an expression control sequence, a cos site, a terminator sequence, a ribosome binding site, a DNA sequence encoding a signal peptide, a DNA sequence encoding a polypeptide and a DNA sequence encoding a polypeptide containing one or more signal peptides.

15. The nucleic acid molecule of claim 14, wherein said expression control sequence comprises an *E. coli*-derived promoter.

16. The nucleic acid molecule of claim 14, wherein said expression control sequence comprises a *Ketogulonigenium*-derived promoter.

17. The nucleic acid molecule of claim 14, further comprising a His-Tag sequence.

18. The nucleic acid molecule of claim 14, further comprising a nucleic acid sequence encoding a polypeptide sequence not present in *Ketogulonigenium*.

19. The nucleic acid molecule of claim 14, further comprising a cosmid site.

20. The nucleic acid molecule of claim 1, further comprising a DNA sequence from an *E. coli*-derived plasmid.

21. The nucleic acid molecule of claim 19, wherein said *E. coli*-derived plasmid is selected from the group comprising pET, pUC18, and pUC19.

22. The nucleic acid molecule of claim 1, further comprising a reporter gene.

23. The nucleic acid molecule of claim 21, wherein said reporter gene encodes a protein selected from the group consisting of β -galactosidase, β -glucuronidase, luciferase, green fluorescent protein α -amylase, and uroporphyrinogen III methyltransferase (cobA) from *Propionibacterium freudenreichii*.

24. The nucleic acid molecule of claim 1, wherein said nucleic acid molecule autonomously replicates in *Ketogulonigenium* and at least one organism selected from the genera consisting of *Acetobacter*, *Corynebacterium*, *Bacillus*, *Rhodobacter*, *Paracoccus*, *Roseobacter*, *Pseudomonas*, *Pseudogluconobacter*, *Gluconobacter*, *Serratia*, *Mycobacterium*, and *Streptomyces*.

25. A transformed *Escherichia coli* cell comprising the nucleic acid molecule of claim 1.

26. A transformed *Ketogulonigenium* cell comprising the nucleic acid molecule of claim 1.

27. A method for producing a polypeptide, comprising culturing a host cell comprising the nucleic acid molecule of claim 14 under conditions such that said polypeptide is expressed, and recovering said polypeptide.

28. A method of transforming a host cell with a nucleic acid comprising:

- (a) transforming a host cell with the nucleic acid of claim 1;
- and
- (b) obtaining a stably transformed host cell.

29. The method of claim 28, wherein said transformation comprises conjugation.

30. The method of claim 28, wherein said transformation comprises electroporation.

SEQ ID NO:1

```

1  GGCAATGGGT CGAAATTCAT AGAATTTTGT GTGAGGTGCG TAGCGGCTCT
51  GACAGGGGTG CTGCGCGGAG ATCTCTGGTC TCAGGTAGGG CGACAATGGA
101 GAGGTGTTAG TTGCCCCCTG TATCGCTCTC TCGGTGGCGC ATTGGGTCAT
151 CCTGCCCGGA CATATGATAT TCCGCTAGAG GATTACTGAT AGTTTCTGCC
201 TGTGCGGCTT GTCGGGCTTG TCGGGCTTGT CGGGCTTGTC GGGCCTGTCC
251 CTCTTGTCCT GCCTGTCCTC ACTTTTTTAC AATCAAAAAA TGGGCGAAGC
301 CCTTCTTGTT CTATAGTTCT TATAGTTCAT ACGAAAAATA CACATAATTA
351 TCAATAGCTT ATTCGCTTAA AAGGGAGTAA TTGGGCCGCA AAAGGGAGTA
401 ATTTGGGCCG AAAAGGGAGT AATTGGGCCG CAAAAGGGAG TAATTGGGCC
451 GATATCGGTT GTTTACATGG GGAGGAATCC CCTTAATCAT TTCTCCCCAT
501 GGGAAAAGACA ACACAAGTGG CCGCAGACCG GGCCTTCGAC CAGACAAAAA
551 CTGTGCTCCC TGCCGAGGTG GCGAGAGGGG TCTATATGCG CAATCCGCCC
601 CGCTGCAGG CGCTCAAGCT CATGCATTTA ATGATAGCCA CTGCGGGCGG
651 CCGCATGGCT GATGATGTGC GCCATGAAAT GCGGCTGGCC GACATTGCGG
701 CAATCGACGG CATGAAAAAC CATGACCGTG AGAGCCTGAC CCCGCTGTTC
751 GAGGAGCTAG CCGCTGCGGT GTTGACCCAT GATGACCCTG CAAAGATGAT
801 CGTGACAGTC GCGGGCTTGG TCGATGAGGC GCGAATAGAC TACCGCCAGG
851 AGGCAAGCGG CGAACTCCTA GTGACGTGGA CCTTCCGGAG TACATTCCGT
901 CGTATGGCGG CGGAGTCGAA CCACTGGGCC ATTCFCGACC GTCAAACGGT
951 ATCCATCTC GGTAGTAAGT ATTCCGTGCT GCTGTTCAG CACGTCTCTA
1001 GTCTCGCCAA TCTTGATCGG ATGAGCGCGA AAACCTTTAC GGTCCCCGAG
1051 TTGCGGGCGC TCCTTGAGT GCCCGAGGGA AAGATGGTTC GTTGAACGA
1101 CGTTAACAGA TTTGCTCTCA AACCTGCACT GGATGAGATC AACCATTTAT
1151 CCGTCTGAC ATTGACGGCA AAGCCGACCA AGATTTGGCC TAGCGTGGCA
1201 AGTGTGACTA TAGGCTGGGA AGTCAAAGAC GACCCAACCG TCGCCAGGGC
1251 CGAGCTGGCG GGTTCCAAGG TCGGTGCGA TGCTCGTCGC AGAGGGGCGAG
1301 CGGAAACGAT AGCCCCCTCC TTCCAGAAG CGGGCGGGAT CACCTACAGT
1351 CCACGTGGC TGGAGCTGAA ACGCTCTGCT GGCAGCAACA AGGACAACGA
1401 TCTGATCGCC TCAGACTTCC GCGTTCCTG TCGGGAGAGA GGCGTGCGTC
1451 TGGACGCTGC AAACATCGAA AACTGTTTT TAGATTTCTG CGCAAAGGTA
1501 GGAAGGTTT GAGTTTTGAG GTATTTACC GCAATAGTGT TAAATGACTT
1551 TCGTGAAACG ATGTGCAATA TAGCGGTAAG ACTATGAAAT ACACGGCTGG
1601 ACAGGCTGCA AAAGCAACGG GTGTGGCGAC CGCAACCATC ACTCGGGCGC
1651 TAAAAAGCGG TAAAATTTCC GGTAAAAAAG ATGAATCTGG GGCATGGGTT
1701 ATAGATCCTG CAGAATTGCA CAGAGTGTTC CTCCCATTT CAAAGAAATA
1751 CACCGAAACA CCTAACACGC AAGTATATGG TAAGCGTGAT GAAACACATG
1801 AAATGACCTC AGAAATCAGC GCATTAGAGC GTGAAGTTCG GACTTTACGC
1851 GATGCTTTAT CTGATGCCAG GGAGGATCGC GACAAATGGC GCGACATGGC
1901 CGAGCGTCTT TCAATTTTAT CACCGATGAG AGAGGAAGAC CGCCCCCTC
1951 AAAAACAAAG ATGGTGAAG ATATTCTGAT CCTGGGCTTC AGGAGCCTTG

```

FIGURE 1A

2001 CCTTACTGG CGGAAAACG CGATATTGAG GCACAGGCC GCACTTTAGA
2051 GCGGAAGCC TATAACGAGT ACCAAAACAC TAGAAGCCAG ATTGAGGAAA
2101 .ATAGGAACG TG

FIGURE 1B

SEQ ID NO:2

1 TGGTGAACGC ATTGGCTTGA TGTTTGAGAA AAGCGAAAAG ACCCGGCCAC
51 AGTTGTGGGT AGAGCGTCGA TATGTGCAAG ACCTGATGCT TGCTGACATC
101 GAACTCCGTG TCTACCTCGC ATCGTCGCTG TATCAGCCTG CTGCGGATGG
151 CGGAAAGCCC GCCTATGGTC GTCACGCAGC CCTTAAGGCG ATGCGCGACT
201 TGGCCCATGC CGATCTGGTG CGTTTCACCA TCGGCCGGAT TACGCAACTG
251 GAGATGATCC TAGAGCGGTT AACCGAGACA TCTGGTTAAC GCCATAAAGG
301 CTGCGGCATG AAAATAGGCG GACAATCTGC GCTTGGCCGC CCCCGTTCCTC
351 AGCCGTGCTT GCTCTCTGCC TGCATGGCAC GACGCAGGAT CGCGTTCATA
401 CGGGTCTGAT ATCCAGACCC GCCCGCCTTG AGCCATGCCA GCACATCGGC
451 ATCAAGCCGC GCGGTGATCT GCTGCTTGAT CGGGCGATAG AAGCGCCAC
501 GCTCGGCGTC TGCCCATTTGG GCTTCGGTCA GCTCGGGAAC ATCGTTGGTG
551 TCGATCTGCT CGGGCGGCAG AGCGTCCAGC CGCGCCAATT TCTTGCGGCG
601 CTCCTCGGTA AGAGCGGGCA GCGTATCGAA GGTGTATTCA ACCATTGGCA
651 TATCTCTTCC TTTCTGCGG TGTAGCGCGG CGAGCCGAAA TGATGCGGAT
701 CGTCTCGACC GGATCGGGGC CAGCCTCGAT GATCAGGTGG GCAACCAGAA
751 GGACGGCAGC GCCATAGATC TGCCCAACGG TTTGCCAGCG GTATTCCCCG
801 CCCTCGATCC TATCTGAAC CGTCAGGTGC AACGGATCGG CGAACACATG
851 CACAGCATCC TCGAACCGGA TGCCATGCTT CTTTTCGTTC GTTTCCGCTT
901 TGGCGGGATC CCAGATAAAC CGCATCTTCA TGGCAGAATT ATAACTACAC
951 ATTTGTAGTT ATTCAATGGC AAGTCGCAGG TTCAAATCAC GCCCCAAAC
1001 CGCAACTGTA TTCGTTCTAC TCACGCGCGC TTTTGAATAG AAGCTTGCAT
1051 GATAACACCC GCCGCTCCT CAACAAAATA AGGCAAATCC GCCGCGCTGG
1101 CGCAATCTGC GCTTTGTCGA TGCAAGGTCT TGTGGTTTCA TACTGCAAGA
1151 GCATGCAAGG AATTGCCCCG GATGAGCACC ACGACGACAC CCACCAAGCC
1201 GGCCTGGAAC AAGGGCCGCG TTGTCGGGAA AAAGCCGCCG CTGACACCTG
1251 ACCAGATTGC CCTGATCCGT CTCATCCTGC GCCAGGAACG GGCGTGGCGG
1301 GATCTGGCTC TGTTCAACGT GGCATCGAC ACCAGTTTGC GCGGCTCGGA
1351 CCTCGTGCGC CTGCGCTCT CGGATGTGGC GACCCCAGCT GGTCTGCGTG
1401 AGATCGTCGA GATCCGCCAG AAGAAGACCG AGGCCCGCAA TGTCCGCCCC
1451 GTACAGGCCC GCCTGTCGGA GGGGACACGC GAGAGCCTGC GGGTCTATCT
1501 CGCGGCCTCT GACAAGCCGC TGCACAGCTG GCTGTTACC GGACAGGGCA
1551 TCCGCTGGTC CCACACCCAC CTTAGCGAGA GCCAGCTGTG GCGCCTGTTC
1601 AAGTCCTGGC TCGAGAAGGC GCGGCTCGAT CCCAGCCTCT ACGGGCTGCA
1651 CTCGCTGCGC CGAACCTTCC CCAGCCACAT CTACCGCGAG ACCGGCAATC
1701 TGCGCGCCGC ACAGCTGCTG CTGGGCCATG CCAGCATCGA GAGACCAAG
1751 GAGTACATCG GCACCGAGCA AGCCGAGGCC CTCGATATCG CACGGAGGTA
1801 TCACCTCTAA CCCATGGAGA CCTATCTCGA GAAGCGCATC CCCGCCAAGA
1851 ACACAGCACG GTTCTACCG ATGGCGGTCC TGCCGAACCT GTTCGGGGAA
1901 TGGACGCTGT ATCGAGAATG GGGCCGATT GGCATCAGCG GCCGCATCCG
1951 GCTCGATTGG TTTGAGAGTG AACAAGATGC CATCGCTGCG ATGCTCGCCA

FIGURE 2A

2001 TCGAGACCGC CAAGCGTCAG CGCGGGTATT GGCTCGAGCC CATCCAGATT
 2051 GACATGTTCC CAGGGGCATA ACAGGCCATC AATGTAAGAG TGCAAGCGGA
 2101 GCAAGCAAAA GCCATTTTAC AGTGAGGTGG CAGATGTTCC TGTTTCACAG
 2151 TGAAAGCGCT GATGCTGTTT CCACGCCACA GACTGATACG ACCAAAGCAA
 2201 CGGGGTCTGC CGCCACAGAC CGGTTCCGCC GCCACCCGCA GAAACGCAGG
 2251 TAAAATGGCG ATTTCCGCAA AAAAACCGTG CAAATGATGG CAAATCACCA
 2301 TCCAGTTTCA TCCTGAAACC CGTCGCTCAA CATGAACGAG CAGGCCATCA
 2351 TCCAAGCCCC AGAAACGCGG TCGCGCGACT ACAGATGAGC GATGTTCTGG
 2401 CTCATAGGCT GCAAGGCCCT GCAACAGTGA TTTCACCGTG AGATTGCAGG
 2451 GTCTTTTGGC TCTCCCGCAA GAGCCACCTC AGGGTGAGCG AGCTAGCCGT
 2501 CTAGGTTTAC AGTGAAATCG CTGAGGAGCG TTGCGGGGCT TATGGTTTGG
 2551 CTGGTCACGT TGGCCATCGG AATGGAGCAT ACGATGGCTT CTACGCAGTC
 2601 GAATCCTGAG GCTTCACGTG GGAAAAATAC GCTCCAAAAA AGCCCTGACC
 2651 AAATCTTGGA AAAATTGCTT GAAAAGTTTG CTCTAAAAA ACTGGGAACG
 2701 AGATATGCAC GAGATCCCTT ACGAGTGCTG TAGGAGTAAT GCAGTGGACA
 2751 AAAACGCCAT TTTTTGCCCC AGTAGGAGTA ATGGAGTGGT TATTTTTTGG
 2801 GAGATTTTGC TTCAGTAGGA GTAACGCGTT GGTAAATTTT GCTTGATTGG
 2851 CGGTTCAAAT CGACCACCGA GCTGCCGTTG GTCGTATTTC ATCTGCCCCG
 2901 CAATTGGGCA CTTGCAGGCC ATCCCCCTGA ACTTCTGGCG ATGACCATTT
 2951 CGAAGGCAAT GGGTCGAAAT TCATAGAATT TTGTGTGAGG TGCGTAGCCG
 3001 CTCTGACAGG GGTGCTGCGC GGAGATCTCT GGTCTCAGGT AGGGCGACAA
 3051 TGGAGAGGTG TTAGTTGCCC CCTGTATCGC TCTCTGCGTG GCGCATTGGG
 3101 TCATCCTGCC CGGACATATG ATATTCCGCT AGAGGATTAC TGATAGTTTC
 3151 TGCCTGTCGG GCTTGTTCGG CTGTGTCGGC TTGTTCGGGCT TGTCGGGCCCT
 3201 GTCCCTCTTG TCCCGCCTGT CCTCACTTTT TCACAATCAA AAAATGGGCG
 3251 AAGCCCTTCT TGTTCTATAG TTCTTATAGT TCATACGAAA ATTACACATA
 3301 ATTATCAATA GCTTATTCGC TTAAGGGGA GTAATTGGGC CGCAAAAGGG
 3351 AGTAATTGGG CCGCAAAAGG GAGTAATTGG GCCGCAAAAG GGAGTAATTG
 3401 GGCCGATATC GGTGTTTAC ATGGGGAGGA ATCCCCTTAA TCATTTCTCC
 3451 CCATGGGAAA GACAACACAA GTGGCCGAG ACCGGGCCCTT CGACCAGACA
 3501 AAAACTGTGC TCCCTGCCGA GGTGGCGAGA GGGGTCTATA TGCGCAATCC
 3551 GCCCCGCTG CAGGCGCTCA AGCTCATGCA TTTAATGATA GCCACTGCGG
 3601 GCGGCCGCAT GGCTGATGAT GTGCGCCATG AAATGCGGCT GGCCGACATT
 3651 CGCGCAATCG ACGGCATGAA AAACCATGAC CGTGAGAGCC TGACCCCGCT
 3701 GTTCGAGGAG CTAGCCGCTG CGGTGTTGAC CCATGATGAC CCTGCAAAGA
 3751 TGATCGTGAC AGTCGGCGGC TTGGTTCGAT AGGCGCGAAT AGACTACCGC
 3801 CAGGAGGCAA GCGGCGAACT CCTAGTGACG TGGACCTTCC GGAGTACATT
 3851 CCGTCGTATG GCGGCGGAGT CGAACCCTG GCCATTCTC GACCGTCAAA
 3901 CGGTATTCCA TCTCGGTAGT AAGTATTCCG TGCTGCTGTT CCAGCACGTC
 3951 TCTAGTCTCG CCAATCTTGA TCGGATGAGC GCGAAAACCT TTACGGTCCC
 4001 CGAGTTGCGG GCGCTCCTTG GAGTGCCCGA GGGAAAAGATG GTTCGTTGGA
 4051 ACGACGTTAA CAGATTTGCT CTCAAACCTG CACTGGATGA GATCAACCAT

FIGURE 2B

4101 TTATCGCGTC TGACATTGAC GGCAAAGCCG ACCAAGATTG GCCGTAGCGT
 4151 GGCAAGTGTG ACTATAGGCT GGGAAAGTGAA AGACGACCCA ACCGTCGCCA
 4201 GCGCGAGCT GCGGGTTCC AAGGTCGGTC GAGATGCTCG TCGCAGAGGG
 4251 GCAGCGGAAA CGATAGCCCC CTCCTTCCCA GAAGCGGGCG GGATCACCTA
 4301 CAGTCCACGT TGGCTGGAGC TGAAACGCTC TGCTGGCAGC AACAAGGACA
 4351 ACGATCTGAT CGCCTCAGAC TTCCGGCGTT TCTGTCGGGA GAGAGGCGTG
 4401 CGTCTGGACG CTGCAAACAT CGAAAACTG TTTT TAGATT TCTGCGCAA
 4451 GGTAGGGAAG GTTTGAGTTT TGAGGTATTT CACCGCAATA GTGTAAATG
 4501 ACTTTCGTGA AACGATGTGC AATATAGCGG TAAGACTATG AAATACACGG
 4551 CTGGACAGGC TGCAAAAGCA ACGGGTGTGG CGACCGCAAC CATCACTCGG
 4601 GCGCTAAAAA GCGGTAAAAT TTCCGGTAAA AAAGATGAAT CTGGGGCATG
 4651 GGTATATAGAT CCTGCAGAAT TGCACAGAGT GTTTCCTCCC ATTTCAAAGA
 4701 AATACACCGA AACACCTAAC ACGCAAGTAT ATGGTAAGCG TGATGAAACA
 4751 CATGAAATGA CCTCAGAAAT CAGCGCATTA GAGCGTGAAG TTCGGACTTT
 4801 ACGCGATGCT TTATCTGATG CCAGGGAGGA TCGCGACAAA TGGCGCGACA
 4851 TGGCCGAGCG TCTTTCAATT TCATCACCGA TGAGAGAGGA AGACCGCCCC
 4901 CCTCAAAAAC AAAGATGGTG GAAGATATTC TGATCCTGGG CTTCAGGAGC
 4951 CTGCGCTTTA AAACCTGAAT CAGCATTCTA GCGATGCTGA TAAGAAGTAA
 5001 ATATAGCCAC AATAGAGCGG CCATTTTCCA TTCACATACA GCTCATCATG
 5051 TGATCAATAT CAAGTATTGA TATTCATCAA TGGAGAAGAA TTTACATGTA
 5101 TCACAGGATC ATCACAGCAT TTGTTTTTGT ATTTCTAAGT GCTAACATAA
 5151 CTATCGCTGG CCCTAAAGAA GATTGTACTA TTGCAGTATC TCACCTTGGG
 5201 TTTGAGACCG ATAATTACAG CTTTGTGCGAA GCCGGTTTTT TTGCCAGAGA
 5251 GAGACACGTT TTTGATGGTG TAATAAACTG CTACGTATCT CATGATGGTA
 5301 ACATACACAG CATCATCCGG GGCAACACAC CTCTTATGGA AGATGGATAT
 5351 TATGGCCCAG AAGTACTGGC GGAAAAACGC GATATTGAGG CACAGGCCCC
 5401 CACTTTAGAG GCGGAAGCCT ATAACGAGTA CAAAACACT AGAAGCCAGA
 5451 TTGAGGAAAA TAGGGAACGT GCCCTCGAGG CGCTGCGGCT AGCTAGCAGT
 5501 CCTTTATTA ATAATGGTAG TACAGAAGAA CAGACAATTA TACAGGCCGC
 5551 AACTCCGACG GCAGATCCTG TTGTATCTGT ACCCGTGGCA TCTCCAGAAT
 5601 CTAACAAGAG TCGAGAGCCG GAACCGGCTG CTGTTCCAGC ATCAGTTTCT
 5651 GTTAGAGAGA TGTGGAGCAC GGCTGACAGA TTGACCACCC GTACATGCCC
 5701 ATCGACTCGA TGCGGAGCAA CTAGCTGGGT AACAGATGGA ACTAAAGTAA
 5751 CAGTTTATGA AGAAAAAGAC GGTGGTCTA GAATCGGAGA GCTACAGTCT
 5801 GCAATGTGCA TAAATGGAAT AAGTGGCGCG GTCGATTCAG GTGAATCTTC
 5851 CTGCAATCCC ACCAATGGTA TCGTTAATGG GCAATTCGCA CCCTGGGTTT
 5901 TCTCGGATTA TCTTACGATC CAAGAGCCAG AAGCTCCCAT ATCCACCCAA
 5951 GAGTGTGCGA ATATGGGGCT CGAGAACTCA GATAATTACC STATCTATTC
 6001 TAGTCAGTTC TGCACTGCCG CTCCTCGAAAT GATCAACGAT AGAGTATGCA
 6051 ATACATCTGA TTTGAGAGAT TTAGCTTGGT TATCTTCTCC TGAAAGAGGA
 6101 CAGGATTACT ACTTCACCTA TTGTGGCGGA TTCAACCTC AAAACAGATG
 6151 GTATTTGAAT GTCAGGACAG GTGAAATCAC CCGCTGATAT TCCACCAAGG

FIGURE 2C

6201 TGAGTCCTGT AGATCAGACT CTCAA^gGGAGT AAACGTTTTA ATCCATCTCC
 6251 ATGAGATCAA CATAGATAGG TGTTCAGTCC CGGCATCTGG TGGATCGGGT
 6301 TTAGGATGAA TCTGTCCGGC TCTTGACATA CCCC^gCGTG AAACCCTGTC
 6351 TTTACAAGAG AAAGTCAGCG GCCTCGAAGC CGCTCTAGCC GATGCCCGGG
 6401 CCAACGGGA TGAGTAGAGC GAACAAGCAA AGCGCCTAGC TATGGCTCTG
 6451 CCCGTCCCCG AAGCTGCAGC CGCAGAATCC GGAAAAAGA AAAAATACAT
 6501 GGCAGCGATT ATTTGGATAG GACACAATCC TTTCTATTA ATATACAACA
 6551 AGATATGGGC ATGCGCCGCG CGTGATCCTC ATTCGATACA ATCCAAATCC
 6601 TGAAAGCTGA CTATGCCCTA CGCATCGCGC ACCATCGGTG CCGTCATTGA
 6651 TGACGTGAAC CGCACCTACC TGCTGCCCGC AATCCAACGC CCCTATGTCT
 6701 GGTCTGCCGG ACAGGTCGTT GCGCTGTTCG ACTCTCTGTT GAAGGGCTAT
 6751 CCGATCAGCA GCTTCATGTT CTGGGCGGTG GACGAGGAGA CCAAGGCAGA
 6801 GCTGCGATGC TACAAATTCA TCGAGAATTA TCGGCCGAA ATGATGAACG
 6851 AGCCGACTAG TGCGGACGGG CGGCAGGTCG TCCTTGTGCT CGACGGACAG
 6901 CAGCGGATGA CCTCACTGTT GATCGGCTTG CGCGGCACAT TCTCTGAGAA
 6951 AGCCAAACAC GCGCGCAACA GCAACGCGGC GGCGTGGTCG GCAAAAACGC
 7001 TATATCTAGA CCTGCTTCGG GACCCGGATC CGAAGAACTC CGATGAAGAC
 7051 GAAGGCA^{atg} AGTTCGGAAT CACTTACGGT CTCTCTTTCC ATGAACGCCG
 7101 CCCGACCAGC AGCCACAGGC ACCACTGGTT CAAGGTGGGA TCGATACTGG
 7151 ATTATCCTAC AGACGAGCAG CTGGAGGGGT TGATTGCCAA GGTGAAGACC
 7201 GAATTT^{catc} ATGGTGTATC GGATTGGGAA AAGGGGCTGG CCGAAGACAC
 7251 CCTGCGCCGG TTGCACCGCG TCATCTGGAA AGACGAGGGC ATCAACTTTT
 7301 TCACTGAACG CGACCAGTCG GTTGATCGGG TGCTGGACAT CTTCGTGCGG
 7351 GCCAATGACG GGGGCACGAA ACTGTCTGAAG GCAGACCTGC TGATGTGAT
 7401 GATCACGTCA AAATGGTCCA GCGGATCGGC CCGCGAGGAA ATCGGCGGCT
 7451 TTGTCGAGCA CATAAACAAA GGTCTCGGTG CGCCGAACAA GATCAGTCGC
 7501 GATCTGGTCC TGAAGGCCTG TCTGGTCGTC TGCGATPATG ATGTGCTCTA
 7551 TAATGTCAGG AACTTTACAA GCGAGGTCAT CGGCAGGATC GAAAGCAACT
 7601 GGGATCGTAT CAAGCAGGCA TTCGAGAACA CGTTCCGCCT GCTGAACAGG
 7651 CATGGCATCA CCGGGGATAA CCTCGGCTCT TTGAACCGG TGCTGCCTCT
 7701 GGTCATATAT ATCTACAACA CGCCGGATTT CGATTTCCGA GGATCGAGCG
 7751 AGTTCGAGCG GGTCAATGCC AGCTCCATGC ACCTCTGGTT GGTGAACAGC
 7801 CTGCTGGTCA GCGCCTTCGT TGGCCATTCG GATCAGACCA TCACCACCGC
 7851 GCGCAATACG ATCCGCGATC ACCTGCGTGT AGGCCGCGAT TTCCCAGTAC
 7901 GAAAGCTGTT CGATGCCATG GCGAAGGGGG GACGGCTATC TCAGGTGGAT
 7951 GAGCGTACCA TCGAAGAATT GCTGGAAATG CAATATGGCA AGCCCCGGAC
 8001 CTTCGTTGCG CTGTCGCTGC TCTATCAGGG CATCGACTGG AACGGATCGA
 8051 CCTGGCATGT CGATCATATC ATTCCCAAG CGGACGCTCA GAAAAATGTG
 8101 CTGCGCGGGC GCAATCTGCC CGAGCATCGC ATTCAGGAAA TCTTGGGCGC
 8151 GGTAAACAGT TTGGGCAACC TGCAACTTTT GCGCGGAGAT GAGAATATCG
 8201 AGAAAGGTGC GCTGCCATTC AGGTCATGGA TTACCGGACG GCGCGTTGAT
 8251 TTCTACGAGC AGCATATGAT CCCGGCGCAC CTTGAACTGT GCGATGTACT

FIGURE 2D

8301 GCATCTGCCC GAGTTCGTGC GCGAACGGGA AAAGGTGATC CGGCGCCGTT
8351 TGATGGAGTT GGTCCGAGCA CGACGCGCAT GAATGAGGTC GTCTTGTCAC
8401 GCGAAGAGCT GCGTCAATCT TGTCTCGACC TGCTTGAAAA ACGCGCTGTC
8451 GAACACCCTG CGGGACACCA AGGCAAGCTC GCCGCCCGCT ATGTTGTGCA
8501 CCGCGACGA

FIGURE 2E

SEQ ID NO:3

1 TCGCGCGTTT CCGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG
51 GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG
101 TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG
151 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
201 CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT
251 CAGGCTGCGC AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT
301 TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA
351 ACGCCAGGGT TTTCCAGTC ACGACGTTGT AAAACGACGG CCAGTGAATT
401 CGGCAATGGG TCGAAATTCA TAGAATTTTG TGTGAGGTGC GTAGCGGCTC
451 TGACAGGGGT GCTGCGCGGA GATCTCTGGT CTCAGGTAGG GCGACAATGG
501 AGAGGTGTTA GTTGCCCCCT GTATCGCTCT CTGCGTGGCG CATTGGGTCA
551 TCCTGCCCCG ACATATGATA TTCCGCTAGA GGATTA CTGA TAGTTTCTGC
601 CTGTGCGGGCT TGTCGGGCTT GTCGGGCTTG TCGGGCTTGT CGGGCCTGTC
651 CCTCTTGTC CGCCTGTCT CACTTTTTTCA CAATCAAAAA ATGGGCGAAG
701 CCCTTCTTGT TCTATAGTTC TTATAGTTCA TACGAAAATT ACACATAATT
751 ATCAATAGCT TATTCGCTTA AAAGGGAGTA ATTGGGCCGC AAAAGGGAGT
801 AATTGGGCCG CAAAAGGGAG TAATTGGGCC GCAAAGGGA GTAATTGGGC
851 CGATATCGGT TGTTTACATG GGGAGGAATC CCTTAATCA TTTCTCCCA
901 TGGGAAAGAC AACACAAGTG GCCGCAGACC GGGCCTTCGA CCAGACAAAA
951 ACTGTGCTCC CTGCCGAGGT GCGGAGAGGG GTCTATATGC GCAATCCGCC
1001 CCGCCTGCAG GCGCTCAAGC TCATGCATTT AATGATAGCC ACTGCGGGCG
1051 GCCGCATGGC TGATGATGTG CGCCATGAAA TGCGGCTGGC CGACATTCGC
1101 GCAATCGACG GCATGAAAAA CCATGACCGT GAGAGCCTGA CCCCCTGTT
1151 CGAGGAGCTA GCCGCTGCGG TGTGACCCA TGATGACCCT GCAAAGATGA
1201 TCGTGACAGT CGGCGGCTTG GTCGATGAGG CGCGAATAGA CTACCGCCAG
1251 GAGGCAAGCG GCGAACTCCT AGTGACGTGG ACCTTCCGGA GTACATTCCG
1301 TCGTATGGCG GCGGAGTCGA ACCACTGGGC CATTCTCGAC CGTCAAACGG
1351 TATTCATCT CGGTAGTAAG TATTCCTGTC TGCTGTTCCA GCACGTCTCT
1401 AGTCTCGCCA ATCTTGATCG GATGAGCGCG AAAACCTTTA CGGTCCCCGA
1451 GTTGCGGGCG CTCCTTGAG TGCCCGAGGG AAAGATGGTT CGTTGGAACG
1501 ACGTTAACAG ATTTGCTCTC AAACCTGCAC TGGATGAGAT CAACCATTTA
1551 TCGCGTCTGA CATTGACGGC AAAGCCGACC AAGATTGGCC GTAGCGTGGC
1601 AAGTGTGACT ATAGGCTGGG AAGTGAAAGA CGACCCAACC GTCGCCAGGC
1651 GCGAGCTGGC GGGTTCCAAG GTCGGTCGAG ATGCTCGTCG CAGAGGGGCA
1701 GCGGAAACGA TAGCCCCCTC CTTCCAGAA GCGGGCGGGA TCACCTACAG
1751 TCCACGTTGG CTGGAGCTGA AACGCTCTGC TGGCAGCAAC AAGGACAACG
1801 ATCTGATCGC CTCAGACTTC CGGCGTTTCT GTCGGGAGAG AGGCGTGCCT
1851 CTGGACGCTG CAAACATCGA AAACTGTTT TTAGATTTCT GCGCAAAGGT
1901 AGGGAAGGTT TGAGTTTTGA GGTATTTTAC CGCAATAGTG TTAAATGACT
1951 TTCGTGAAAC GATGTGCAAT ATAGCGGTAA GACTATGAAA TACACGGCTG

FIGURE 3A

2001 GACAGGCTGC AAAAGCAACG GGTGTGGCGA CCGCAACCAT CACTCGGGCG
 2051 CTAAAAAGCG GTAAAATTC CGGTAAAAAA GATGAATCTG GGGCATGGGT
 2101 TATAGATCCT GCAGAATTGC ACAGAGTGTT TCCTCCCATT TCAAAGAAAT
 2151 ACACCGAAAC ACCTAACACG CAAGTATATG GTAAGCGTGA TGAAACACAT
 2201 GAAATGACCT CAGAAATCAG CGCATTAGAG CGTGAAGTTC GGACTTTACG
 2251 CGATGCTTTA TCTGATGCCA GGGAGGATCG CGACAAATGG CGCGACATGG
 2301 CCGAGCGTCT TTCAATTTCA TCACCGATGA GAGAGGAAGA CCGCCCCCTT
 2351 CAAAAACAAA GATGGTGGAA GATATTCTGA TCCTGGGCTT CAGGAGCCTT
 2401 GCCTTTACTG GCGGAAAAAC GCGATATTGA GGCACAGGCC CGCACTTTAG
 2451 AGGCGGAAGC CTATAACGAG TACCAAAACA CTAGAAGCCA GATTGAGGAA
 2501 AATAGGGAAC GTGGGATCCT CTAGAGTCGA CCTGCAGGCA TGCAAGCTTG
 2551 GCGTAATCAT GGTCATAGCT GTTTCCTGTG TGAAATGTGT ATCCGCTCAC
 2601 AATTCCACAC AACATACGAG CCGGAAGCAT AAAGTGTAAG GCCTGGGGTG
 2651 CCTAATGAGT GAGCTAACTC ACATTAATTG CGTTGCGCTC ACTGCCCGCT
 2701 TTCCAGTCGG GAAACCTGTC GTGCCAGCTG CATTAATGAA TCGGCCAACG
 2751 CGCGGGGAGA GCGGTTTGC GTATTGGGCG CTCTTCCGCG CTCGGTCTTG
 2801 CCTTGCTCGT CCGTGATGTA CTTACCAGC TCCGCGAAGT CGCTCTTCTT
 2851 GATGGAGCGC ATGGGGACGT GCTTGGCAAT CACGCGACC CCCCAGCCGT
 2901 TTTAGCGGCT AAAAAAGTCA TGGCTCTGCC CTCGGGCGGA CCACGCCCAT
 2951 CATGACCTTG CCAAGCTCGT CTTGCTTCTC TTCGATCTTC GCCAGCAGGG
 3001 CGAGGATCGT GGCATCACCG AACCGCGCCG TCGCGGGGTC GTCGGTGAGC
 3051 CAGAGTTTCA GCAGGCCGCC CAGGCGGCC AGGTCGCCAT TGATGCGGGC
 3101 CAGCTCGCGG ACGTGCTCAT AGTCCACGAC GCCCGTGATT TTGTAGCCCT
 3151 GGCCGACGGC CAGCAGGTAG GCCGACAGG TCATGCCGGC CGCCGCCGCC
 3201 TTTTCTCAA TCGCTCTTCG TTCGTCTGGA AGGCAGTACA CCTTGATAGG
 3251 TGGGCTGCCC TTCCTGGTTG GCTTGGTTTC ATCAGCCATC CGCTTGCCCT
 3301 CATCTGTTAC GCCGGCGGTA GCCGCCAGC CTCGCAGAGC AGGATTTCCC
 3351 TTGAGCACCG CCAGGTGCGA ATAAGGGACA GTGAAGAAG AACACCCGCT
 3401 CGCGGGTGGG CCTACTTCAC CTATCCTGCC CGGCTGACGC CGTTGGATAC
 3451 ACCAAGGAAA GTCTACACGA ACCCTTTGGC AAAATCCTGT ATATCGTGCG
 3501 AAAAAAGGATG GATATACCGA AAAAAATCGCT ATAATGACCC CGAAGCAGGG
 3551 TTATGCAGCG GAAAAGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTCGT
 3601 TCGGCTGCGG CGAGCGGTAT CAGCTCACTC AAAGGCGGTA ATACGGTTAT
 3651 CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG
 3701 CAAAAGGCCA GGAACCGTAA AAAGGCCGCG TTGCTGGCGT TTTTCCATAG
 3751 GCTCCGCCCC CCTGACGAGC ATCACAATAA TCGACGCTCA AGTCAGAGGT
 3801 GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC
 3851 TCCCTCGTGC GCTCTCCTGT TCCGACCCTG CCGCTTACCG GATACCTGTC
 3901 CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCATAGC TCACGCTGTA
 3951 GGTATCTCAG TTCGGTGTAG GTCGTTGCTT CCAAGCTGGG CTGTGTGCAC
 4001 GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA ACTATCGTCT
 4051 TGAGTCCAAC CCGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG

FIGURE 3B

4101 GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG
4151 AAGTGGTGGC CTAAC TACGG CTACACTAGA AGGACAGTAT TTGGTATCTG
4201 CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT
4251 CCGGCAAACA AACCACCGCT GGTAGCGGTG GTTTTTTTGT TTGCAAGCAG
4301 CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT TGATCTTTTC
4351 TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTTGG
4401 TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTTT GGGGGGGGGG
4451 GCGCTGAGG TCTGCCTCGT GAAGAAGGTG TTGCTGACTC ATACCAGGCC
4501 TGAATCGCCC CATCATCCAG CCAGAAAGTG AGGGAGCCAC GTTTGATGAG
4551 AGCTTTGTTG TAGGTGGACC AGTTGGTGAT TTTGAACTTT TGCTTTGCCA
4601 CGGAACGGTC TCGTTGTCG GGAAGATGCG TGATCTGATC CTTCAACTCA
4651 GCAAAAAGTTC GATTTATTCA ACAAAGCCGC CGTCCCCTCA AGTCAGCGTA
4701 ATGCTCTGCC AGTGTTACAA CCAATTAACC AATTCTGATT AGAAAACTC
4751 ATCGAGCATC AAATGAACT GCAATTTATT CATATCAGGA TTATCAATAC
4801 CATATTTTTG AAAAAGCCGT TTCTGTAATG AAGGAGAAAA CTCACCGAGG
4851 CAGTTCCATA GGATGGCAAG ATCCTGGTAT CGGTCTGCGA TTCCGACTCG
4901 TCCAACATCA ATACAACCTA TTAATTTCCC CTCGTCAAAA ATAAGGTTAT
4951 CAAGTGAGAA ATCACCATGA GTGACGACTG AATCCGGTGA GAATGGCAAA
5001 AGCTTATGCA TTTCTTTCCA GACTTGTTCA ACAGGCCAGC CATTACGCTC
5051 GTCATCAAAA TCACTCGCAT CAACCAAACC GTTATTTCATT CGTGATTGCG
5101 CCTGAGCGAG ACGAAATACG CGATCGCTGT TAAAAGGACA ATTACAAACA
5151 GGAATCGAAT GCAACCGGCG CAGGAACACT GCCAGCGCAT CAACAATATT
5201 TTCACCTGAA TCAGGATATT CTTCTAATAC CTGGAATGCT GTTTTCCCGG
5251 GGATCGCAGT GGTGAGTAAC CATGCATCAT CAGGAGTACG GATAAAATGC
5301 TTGATGGTCG GAAGAGGCAT AAATCCGTC AGCCAGTTA GTCTGACCAT
5351 CTCATCTGTA ACATCATTGG CAACGCTACC TTTGCCATGT TTCAGAAACA
5401 ACTCTGGCGC ATCGGGCTTC CCATACAATC GATAGATTGT CGCACCTGAT
5451 TGCCCGACAT TATCGCGAGC CCATTTATAC CCATATAAAT CAGCATCCAT
5501 GTTGGAAATTT AATCGCGGCC TCGAGCAAGA CGTTTCCCGT TGAATATGGC
5551 TCATAACACC CCTTGTATTA CTGTTTATGT AAGCAGACAG TTTTATTGTT
5601 CATGATGATA TATTTTTATC TTGTGCAATG TAACATCAGA GATTTTGAGA
5651 CACAACGTGG CTTTCCCCC CCCCCATTA TTGAAGCATT TATCAGGGTT
5701 ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA AAATAAACAA
5751 ATAGGGGTTT CGCGCACATT TCCCCGAAAA GTGCCACCTG ACGTCTAAGA
5801 AACCATTATT ATCATGACAT TAACCTATAA AAATAGGCGT ATCACGAGGC
5851 CCTTTCGTC

FIGURE 3C

SEQ ID NO:4

1 GGCAATGGGT CGAAATTCAT AGAATTTTGT GTGAGGTGCG TAGCGGCTCT
51 GACAGGGGTG CTGCGCGGAG ATCTCTGGTC TCAGGTAGGG CGACAATGGA
101 GAGGTGTTAG TTGCCCCCTG TATCGCTCTC TCGGTGGCGC ATTEGGTCAT
151 CCTGCCCCGA CATATGATAT FCCGCTAGAG GATTACTGAT AGTTTCTGCC
201 TGTGCGGCTT GTCGGGCTTG TCGGGCTTGT CGGGCTTGTC GGGCCTGTCC
251 CTCTTGTCCC GCCTGTCCTC ACTTTTTCAC AATCAAAAAA TGGGCGAAGC
301 CCTTCTTGTT CTATAGTTCT TATAGTTCAT ACGAAAATTA CACATAATTA
351 TCAATAGCTT ATTCGCTTAA AAGGGAGTAA TTGGGCCGCA AAAGGGAGTA
401 ATTGGGCCGC AAAAGGGAGT AATTGGGCCG CAAAAGGGAG TAATTGGGCC
451 GATATCGGTT GTTTACATGG GGAGGAATCC CCTPAATCAT TTCTCCCAT
501 GGGAAAGACA ACACAAGTGG CCGCAGACCG GGCCCTCGAC CAGACAAAAA
551 CTGTGCTCCC TGCCGAGGTG GCGAGAGGGG TCTATATGCG CAATCCGCCC
601 CGCCTGCAGG CGCTCAAGCT CATGCATTTA ATGATAGCCA CTGCGGGCGG
651 CCGCATGGCT GATGATGTGC GCCATGAAAT GCGGCTGGCC GACATTCCGG
701 CAATCGACGG CATGAAAAAC CATGACCGTG AGAGCCTGAC CCCGCTGTTC
751 GAGGAGCTAG CCGCTGCGGT GTTGACCCAT GATGACCCTG CAAAGATGAT
801 CGTGACAGTC GCGGGCTTGG TCGATGAGGC GCGAATAGAC TACCGCCAGG
851 AGGCAAGCGG CGAACTCCTA GTGACGTGGA CCTTCCGGAG TACATTCCGT
901 CGTATGGCGG CGGAGTCGAA CCACTGGGCC ATTCTCGACC GTCAAACGGT
951 ATTCCATCTC GGTAGTAAGT ATTCCGTGCT GCTGTTCCAG CACGTCTCTA
1001 GTCTCGCCAA TCTTGATCGG ATGAGCGCGA AAACCTTTAC GGTCCCCGAG
1051 TTGCGGGCGC TCCTTGAGT GCCCGAGGGA AAGATGGTTC GTTGAACGA
1101 CGTTAACAGA TTTGCTCTCA AACCTGCACT GGATGAGATC AACCATTTAT
1151 CCGTCTGAC ATTGACGGCA AAGCCGACCA AGATTGGCCG TAGCGTGGCA
1201 AGTGTGACTA TAGGCTGGGA AGTGAAGAC GACCCAACCG TCGCCAGGGC
1251 CGAGCTGGCG GGTTCOAAG TCGGTGCGA TGCTCGTCGC AGAGGGGCAG
1301 CGGAAACGAT AGCCCCCTCC TTCCAGAAG CGGGCGGGAT CACCTACAGT
1351 CCACGTTGGC TGGAGCTGAA ACGTCTGCT GGCAGCAACA AGGACAACGA
1401 TCTGATCGCC TCAGACTTCC GCGTTTCTG TCGGGAGAGA GCGGTGCGTC
1451 TGGACGCTGC AAACATCGAA AACTGT'TTT TAGATTTCTG CGCAAAGGTA
1501 GGAAGGTTT GAGTTT'TGAG GTATTT'CACC GCAATAGTGT TAAATGACTT
1551 TCGTGAAACG ATGTGCAATA TAGCGGTAAG ACTATGAAAT ACACGGCTGG
1601 ACAGGCTGCA AAAGCAACGG GTGTGGCGAC CGCAACCATC ACTCGGGCGC
1651 TAAAAAGCGG TAAAATTTCC GGTAAAAAAG ATGAATCTGG GGCATGGGTT
1701 ATAGATCCTG CAGAATTGCA CAGAGTGT'TT CCTCCCATTT CAAAGAAATA
1751 CACCGAAACA CCTAACACGC AAGTATATGG TAAGCGTGAT GAAACACATG
1801 AAATGACCTC AGAAATCAGC GCATTAGAGC GTGAAGTTCG GACTTTACGC
1851 GATGCTTTAT CTGATGCCAG GGAGGATCGC GACAAATGGC GCGACATGGC
1901 CGAGCGTCTT TCAATTTTAT CACCGATGAG AGAGGAAGAC CGCCCCCTC
1951 AAAAACAAAG ATGGTGGAAG ATATTCTGAT CCTGGGCTTC AGGAGCCTTG

FIGURE 4A

12/12

2001 CCTTTAAAAC CTGAATCAGC ATTCTAGCGA TGCTGATAAG AAGTAAATAT
2051 AGCCACAATA GAGCGGCCAT TTTCCATTCA CATAACAGCTC ATCATGTGAT
2101 CAATATCAAG TATTGATATT CATCAATGGA GAAGAATTTA CATGTATCAC
2151 AGGATCATCA CAGCATTTGT TTTTGTATTT CTAAGTGCTA ACATAACTAT
2201 CGCTGGCCCT AAAGAAGATT GTACTATTGC AGTATCTCAC CTTGGGTTTC
2251 AGACCGATAA TTACAGCTTT GTCGAAGCCG GTTTTTTTTGC CAGAGAGAGA
2301 CACGTTTTTG ATGGTGTAAT AAAC TGCTAC GTATCTCATG ATGGTAACAT
2351 ACACAGCATC ATCCGGGGCA ACACACCTCT TATGGAAGAT GGATATTATG
2401 GCCCAGAAGT ACTGGCGGAA AAACGCGATA TTGAGGCACA GGCCCGCACT
2451 TTAGAGGCGG AAGCCTATAA CGAGTACCAA AACACTAGAA GCCAGATTGA
2501 GGAAAATAGG GAACGTG

FIGURE 4B

-1-

SEQUENCE LISTING

<110> Archer-Daniels-Midland Company
D'Elia, John

<120> *Ketogulonigenium* Shuttle Vectors

<130> 1533.110PC01

<150> US 60/194,625

<151> 2000-04-05

<160> 4

<170> PatentIn version 3.0

<210> 1

<211> 2112

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<223> replicon from pADM291

<400> 1
ggcaatgggt cgaaattcat agaattttgt gtgaggtgcg tagcggctct gacaggggtg 60
ctgcgcggag atctctggtc tcaggtaggg cgacaatgga gaggtgtag ttgccccctg 120
tatcgctctc tgcgtggcgc attgggtcat cctgcccgga catatgatat tccgctagag 180
gattactgat agtttctgcc tgcggggctt gtcgggcttg tcgggcttgt cgggcttgtc 240
gggcctgtcc ctcttgtccc gctgtcctc actttttcac aatcaaaaaa tgggcgaagc 300
ccttcttgtt ctatagttct tatagttcat acgaaaatta cacataatta tcaatagctt 360

-2-

attcgcttaa aagggagtaa ttgggcccga aaagggagta attgggcccgc aaaagggagt 420
 aattgggccc caaaagggag taattgggcc gatatcggtt gtttacatgg ggaggaatcc 480
 ccttaatcat ttctcccat gggaaagaca acacaagtgg cgcagaccg ggccttcgac 540
 cagacaaaaa ctgtgctccc tgccgaggtg gcgagagggg tctatatgcg caatccgccc 600
 cgcctgcagg cgctcaagct catgcattta atgatagcca ctgcgggccc cgcgatggct 660
 gatgatgtgc gccatgaaat gcggctggcc gacattcgcg caatcgaccg catgaaaaac 720
 catgaccgtg agagcctgac cccgctgttc gaggagctag ccgctgcggt gttgacccat 780
 gatgaccctg caaagatgat cgtgacagtc ggcggcttgg tcgatgaggc gcgaatagac 840
 .taccgccagg aggcaagcgg cgaactccta gtgacgtgga ccttccggag tacattccgt 900
 cgtatggcgg cggagtcgaa ccaactggcc attctcgacc gtcaaacggt attccatctc 960
 ggtagtaagt attccgtgct gctgttccag cacgtctcta gtctcgccaa tcttgatcgg 1020
 atgagcgcga aaacctttac ggtccccgag ttgcccggcg tccttgaggt gcccgagggg 1080
 aagatggttc gttggaacga cgttaacaga tttgctctca aacctgact ggatgagatc 1140
 aaccatttat cgcgtctgac attgacggca aagccgacca agattggccg tagcgtggca 1200
 agtgtgacta taggctggga agtgaaagac gaccaaccg tcgccaggcg cgagctggcg 1260
 ggttccaagg tcggtcgaga tgctcgtcgc agaggggagc cggaaacgat agccccctcc 1320
 ttcccagaag cgggcccggat cacctacagt ccacgttggc tggagctgaa acgctctgct 1380
 ggagcaaca aggacaacga totgatcgcc tcagacttcc ggcgtttctg tcgggagaga 1440
 ggcgtgcgtc tggacgctgc aaacatcgaa aaactgtttt tagatttctg cgcaaaggta 1500
 gggaaaggttt gagttttgag gtatttcacc gcaatagtgt taaatgactt tcgtgaaacg 1560
 atgtgcaata tagcggtaag actatgaaat acacggctgg acaggctgca aaagcaacgg 1620
 gtgtggcgac cgcaaccatc actcgggccc taaaagcgg taaaatttcc ggtaaaaaag 1680
 atgaatctgg ggcatggggt atagatcctg cagaattgca cagagtgttt cctccattt 1740
 caaagaaata caccgaaaca cctaacacgc aagtatatgg taagcgtgat gaaacacatg 1800
 aatgacctc agaaatcagc gcattagagc gtgaagtctg gactttacgc gatgctttat 1860
 ctgatgccag ggaggatcgc gacaaatggc gcgacatggc cgagcgtctt tcaatttcat 1920
 caccgatgag agaggaagac cgcctccctc aaaaacaaag atgggtggaag atattctgat 1980
 cctgggcttc aggagccttg cctttactgg cggaaaaacg cgatattgag gcacaggccc 2040
 gcactttaga ggcggaagcc tataacgagt accaaaacac tagaagccag attgaggaaa 2100
 ataggaacg tg 2112

<210> 2

<211> 8509

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<223> pADM291

<400> 2

| | |
|--|------|
| tgggtgaacgc attggcttga tgtttgagaa aagogaaaag acccggccac agttgtgggt | 60 |
| agagcgtcga tatgtgcaag acctgatgct tgctgacatc gaactccgtg tctacctcgc | 120 |
| atcgtcgcctg tatcagcctg ctgcggatgg cggaaagccc gcctatggtc gtcacgcagc | 180 |
| ccttaaggcg atgcgcgact tggcccatgc cgatctggtg cgtttcacca tcggccggat | 240 |
| tacgcaactg gagatgatcc tagagcgggt aaccgagaca tctggttaac gccataaagg | 300 |
| ctgcggcatg aaaataggcg gacaatctgc gcttggccgc ccccgttctc agccgtgctt | 360 |
| gctctctgcc tgcattggcac gacgcaggat cgcgttcata cgggtctgat atccagacct | 420 |
| gcccgccttg agccatgcca gcaatcggc atcaagccgc gcggtgatct gctgcttgat | 480 |
| cgggcgatag aagcgcacc gctcggcgtc tgcccattgg gcttcgggtca gctcgggaac | 540 |
| atcgttggtg tcgatctgct cgggcggcag agcgtccagc cgcgccaatt tcttgcggcg | 600 |
| ctcctcggta agagcgggca gcgtatcgaa ggtgtattca accattggca tatctcttcc | 660 |
| tttctcggcg tgtagcgcgg cgagccgaaa tgatgcggat cgtctcgacc ggatcggggc | 720 |
| cagcctcgat gatcagggtg gcaaccagaa ggacggcagc gccatagatc tgcccaacgg | 780 |
| tttgccagcg gtattccccg cctcogatcc tatcctgaac cgtcagggtc aacggatcgg | 840 |
| cgaacacatg cacagcatcc tcgaaccgga tgccatgctt cttttcgttc gtttcgcct | 900 |
| tggcgggatc ccagataaac cgcattctca tggcagaatt ataactacac atttgtagtt | 960 |
| attcaatggc aagtcgcagg ttcaaatcac gccccaaaac cgcaactgta ttcgcttctac | 1020 |
| tcacgcgcgc ttttgaatag aagcttgcatt gataacacct gccgcgtcct caacaaaata | 1080 |
| aggcaaatcc gccgcgctgg cgcaatctgc gctttgtcga tgcaaggctt tgtggtttca | 1140 |
| tactgcaaga gcatgcaagg aattgccccg gatgagcacc acgacgacac ccaccaagcc | 1200 |
| ggcctggaac aagggccgcg ttgtcgggaa aaagccgcgg ctgacacctg accagattgc | 1260 |
| cctgatccgt ctcatcctgc gccaggaacg ggcgtagcgg gatctggctc tgttcaacgt | 1320 |
| ggcgatcgac accagtttgc gcggctcggg cctcgtgcgc ctgcgcgtct cggatgtggc | 1380 |

gaccccagct ggtctgctg agatcgtoga gatccgccag aagaagaccg aggcccgcaa 1440
tgtccgcccc gtacaggccc gcctgtcggg ggggacacgc gagagcctgc gggctctatct 1500
cgcggcctct gacaagccgc tgcacagctg gctgttcacc ggacagggca tccgctggtc 1560
ccacaccac cttagcgaga gccagctgtg ggcctgttc aagtcttggc tcgagaaggc 1620
gcggtctgat cccagcctct acgggctgca ctgctgctgc cgaaccttc ccagccacat 1680
ctaccgcgag accggcaatc tgcgcgcgc acagctgctg ctgggccatg ccagcatoga 1740
gagaccaag gagtacatcg gcacogagca agccgaggcc ctcgatatcg cacggaggta 1800
tcacctctaa ccatggaga cctatctoga gaagcgcac cccgccaaga acacagcagc 1860
gttctaccgc atggcggctc tgcggaacct gttcggggaa tggacgctgt atcgagaatg 1920
gggccgcatt ggcattcagc gccgcattcg gctcgattgg tttgagagt aacaagatgc 1980
catcgtctgc atgctcgcca tcgagaccgc caagcgtcag cgcggtatt ggctcgagcc 2040
catccagatt gacatgttcc caggggcata acagggcctc aatgtaagag tgcaagcggg 2100
gcaagcaaaa gccatttcac agtgaggctg cagatgttcc tgtttcacag tgaaagcgtc 2160
gatgctgttt ccacgccaca gactgatacg accaaagcaa cggggtctgc cgccacagac 2220
cggttcgcgc gccacccgca gaaacgcagg taaaatggcg atttccgcaa aaaaaccgtg 2280
caaatgatgg caaatcacca tccagtttca toctgaaacc cgtcgtctca catgaacgag 2340
caggccatca tccaagcccc agaaacgcgg tgcggcgact acagatgagc gatgttctgg 2400
ctcataggct gcaaggccct gcaacagtga tttcacctgc agattgcagg gtcttttggc 2460
tctcccgcaa gagccacctc agggtgagcg agctagccgt ctaggttcac agtgaaatcg 2520
ctgaggagcg ttgcggggct tatggtttgg ctggtcacgt tggccatcgg aatggagcat 2580
acgatggctt ctacgcagtc gaatcctgag gcttcacgtg ggaaaaatac gctccaaaaa 2640
agccctgacc aaatcttggg aaaattgctt gaaaagtgtt cttctaaaaa actgggaacg 2700
agatatgcac gagatccctt acgagtgtg taggagtaat gcagtggaca aaaacgcat 2760
tttttgcccc agtaggagta atggagtggg ttttttttgg gagattttgc ttcagtagga 2820
gtaacgcgtt ggttaaattt gcttgattgg cggttcaaat cgaccaccga gctgccgttg 2880
gtcgtattcg atctgccccg caattgggca cttgcaggcc atccccctga acttctggcg 2940
atgaccattt cgaaggcaat gggtcgaaat tcatagaatt ttgtgtgagg tgcgtagcgg 3000
ctctgacagg ggtgctgctg ggagatctct ggtctcaggt agggcgacaa tggagaggtg 3060
ttagttgcc cctgtatcgc tctctgcgtg gcgcattggg tcatcctgcc cggacatatg 3120
atattccgct agaggattac tgatagtttc tgctgtcgg gcttgtcggg cttgtcgggc 3180
ttgtcgggct tgtcgggcct gtccctcttg tccgcctgt cctcactttt tcacaatcaa 3240

-5-

aaaatgggcg aagcccttct tgttctatag ttcttatagt tcatacghaa attacacata 3300
attatcaata gcttattcgc ttaaaagggg gtaattgggc cgcaaaaggg agtaattggg 3360
ccgcaaaagg gagtaattgg gcgcgcaaaag ggagtaattg ggccgatata ggttgtttac 3420
atggggagga atccccctaa tcattttctcc ccatgggaaa gacaacacaa gtggccgcag 3480
accgggcctt cgaccagaca aaaactgtgc tccctgccga ggtggcgaga ggggtctata 3540
tgcgcaatcc gccccgcctg caggcgctca agctcatgca tttaatgata gccactgcgg 3600
gcgccgcgat ggctgatgat gtgcgccatg aaatgctggc ggccgacatt cgcgcaatcg 3660
acggcatgaa aaacctatgac cgtgagagcc tgaccccgct gttcgaggag ctaccgctg 3720
cgggtgtgac ccatgatgac cctgcaaaga tgatcgtgac agtcggcggc ttggtcgatg 3780
aggcgcaat agactaccgc caggaggcaa gcggcgaaact cctagtgcag tggaccttcc 3840
ggagtacatt ccgtcgtatg gcggcggagt cgaacctatg ggccattctc gacctcaaa 3900
cggattcca tctcggtagt aagtattcgc tgctgctgtt ccagcacgtc tctagtctcg 3960
ccaatcttga tcggatgagc gcgaaaacct ttacggctcc cgagttgcgg gcgctccttg 4020
gagtgcccg gggaaagatg gttcgttggg acgacgttaa cagatttgct ctcaaacctg 4080
cactggatga gatcaacct ttatcgcgtc tgacattgac ggcaaagccg accaagattg 4140
gccgtagcgt ggcaagtgtg actataggct gggaaagtga agacgacca accgtcgcca 4200
ggcgcgagct ggcggttcc aaggctcggc gagatgctcg tcgcagaggg gcagcgaaa 4260
cgatagcccc ctcttccca gaagcgggcg ggatcaccta cagtccacgt tggctggagc 4320
tgaaacgctc tgctggcagc aacaaggaca acgatctgat cgcctcagac ttccggcgtt 4380
tctgtcggga gagaggcgtg cgtctggacg ctgcaaacat cgaaaaactg tttttagatt 4440
tctgcgcaa ggtagggaag gtttgagttt tgaggatttt caccgcaata gtgttaaattg 4500
actttcgtga aacgatgtgc aatatagcgg taagactatg aaatacacgg ctggacaggc 4560
tgcaaaagca acgggtgtg gcaccgcaac catcactcgg gcgctaaaaa gcggtaaaat 4620
ttccggtaaa aaagatgaat ctggggcatg ggttatagat cctgcagaat tgcaagagt 4680
gtttcctccc atttcaaaga aatacaccga aacacctaac acgcaagtat atggtaagcg 4740
tgatgaaaca catgaaatga cctcagaaat cagcgcatta gagcgtgaag ttcggacttt 4800
acgcatgct ttatctgat ccaggaggga tcgcgacaaa tggcgcgaca tggccgagcg 4860
tctttcaatt tcatcaccga tgagagagga agaccgcccc cctcaaaaac aaagatggtg 4920
gaagatattc tgatcctggg cttcaggagc cttgccttta aaacctgaat cagcattcta 4980
gcgatgctga taagaagtaa atatagccac aatagagcgg ccattttcca ttcacataca 5040
gctcatcatg tgatcaatat caagtattga tattcatcaa tggagaagaa tttacatgta 5100

-6-

tcacaggatc atcacagcat ttgtttttgt atttctaagt gctaacataa ctatcgctgg 5160
ccctaaagaa gattgtacta ttgcagtatc tcaccttggg tttcagaccg ataattacag 5220
ctttgtcgaa gccggttttt ttgccagaga gagacacggt tttgatggtg taataaactg 5280
ctacgtatct catgatggtg acatacacag catcatccgg ggcaacacac ctcttatgga 5340
agatggatat tatggcccag aagtactggc ggaaaaacgc gatattgagg cacaggcccg 5400
cactttagag gcggaagcct ataacgagta ccaaaacact agaagccaga ttgaggaaaa 5460
tagggaacgt gccctcgagg cgctgoggct agctagcagt cttttatta ataatggtag 5520
tacagaagaa cagacaatta tacaggccgc aactccgacg gcagatcctg ttgtatctgt 5580
accctggca tctccagaat ctaaacaag tcgagagccg gaaccggctg ctgttccagc 5640
atcagtttct gttagagaga tgtggagcac ggctgacaga ttgaccacc gtacatgcc 5700
atcgactcga tgcggagcaa ctagctgggt aacagatgga actaaagtaa cagtttatga 5760
agaaaaagac ggttggtcta gaatcggaga gctacagtct gcaatgtgca taaatggaat 5820
aagtggcgcg gtcgattcag gtgaatcttc ctgcaatccc accaatggta tcgttaatgg 5880
gcaattcgca ccctgggttt tctcggatta tcttacgac caagagccag aagctcccat 5940
atccaccaa gagtgtcgaa atatggggct cgagaactca gataattacc gtatctattc 6000
tagtcagttc tgcactgcg ctctcgaaat gatcaacgat agagtatgca atacatctga 6060
ttcagagat ttagcttgggt tatcttctcc tgaagagga caggattact acttcaccta 6120
ttgtggcggg tttcaacctc aaaacagatg gtatttgaat gtcaggacag gtgaaatcac 6180
ccgctgatat tccaccaagg tgagtcctgt agatcagact ctcaaggagt aaacgtttta 6240
atccatctcc atgagatcaa catagatagg tgttcagtc ccgcatctgg tggatcgggt 6300
ttaggatgaa tctgtccggc tcttgacata ccccgcgctg aaaccctgtc tttacaagag 6360
aaagtcagcg gcctcgaagc cgctctagcc gatgcccggg cccaacggga tgagtagagc 6420
gaacaagcaa agcgcctagc tatggctctg cccgtcccgg aagctgcagc cgcagaatcc 6480
ggaaaaaaga aaaaatacat ggcagcgatt atttggatag gacacaatcc ttttctatta 6540
atatacaaca agatatgggc atgcgcccgc cgtgatcctc attcgataca atccaaatcc 6600
tgaaagctga ctatgcccta cgcctcggc accatcgggt ccgtcattga tgacgtgaac 6660
cgcacctacc tgctgcccgc aatccaacgc ccctatgtct ggtctgcccg acaggtcgtt 6720
gcgctgttcg actctctggt gaagggtat ccgatcagca gcttcatggt ctgggcccgtg 6780
gacgaggaga ccaaggcaga gctgcgatgc taaaaattca tcgagaatta tcggcccga 6840
atgatgaacg agccgactag tgcggacggg cggcaggtcg tccttgtgct cgacggacag 6900
cagcggatga cctcactggt gatcggcttg cgcggcacat tctctgagaa agccaaacac 6960

-7-

gcgcgcaaca gcaacgcggc ggcgtggtcg gcaaaaacgc tatactaga cctgcttcgg 7020
gacccggatc cgaagaactc cgatgaagac gaaggcaatg agttcggaat cacttacggt 7080
ctctctttcc atgaacgccg cccgaccagc agccacaggc accactgggt caaggtggga 7140
tcgatactgg attatcctac agacgagcag ctggaggggt tgattgcaa ggtgaagacc 7200
gaatttcatac atggtgtatac ggattgggaa aaggggctgg cggaagacac cctgcgccgg 7260
ttgcaccgcg tcatctggaa agacgagggc atcaactttt tcaactgaac cgaccagtcg 7320
gttgatcggg tgctggacat ctctgtgogg gccaatgacg ggggcacgaa actgtcgaag 7380
gcagacctgc tgatgtogat gatcacgtca aatgggtcca gcggatcggc ccgagaggaa 7440
atcggcggct ttgtcgagca cataaaca ggtctcgggt gcccgaaaca gatcagtcgc 7500
gatctggtcc tgaaggcctg tctgggtgct tgcgattatg atgtcgtcta taatgtcagg 7560
aactttacaa gcgaggtcac cggcaggatc gaaagcaact gggatcgtat caagcaggca 7620
ttcgagaaca cgttccgcct gctgaacagg catggcatca ccggggataa cctcggctct 7680
ttgaacgcgg tgctgcctct ggtctattat atctacaaca gcccgattt cgatttccga 7740
ggatcgagcg agttcgagcg ggtcaatgcc agctccatgc acctctgggt ggtgaacagc 7800
ctgctggtca gcgccttcgt tggccattcg gatcagacca tcaccaccgc gcgcaatacg 7860
atccgcgatc acctgcgtgt aggccggat ttcccagtac gaaagctggt cgatgccatg 7920
gcgaaggggg gacggctatac tcaggtggat gagcgtacca tcgaagaatt gctggaaatg 7980
caatatggca agccccggac ctctggtgog ctgtcgtctc tctatcaggg catcgactgg 8040
aacggatcga cctggcatgt cgatcatatac attccccaaag cggacgctca gaaaaatgtg 8100
ctgcgcggggc gcaatctgcc cgagcatcgc attcaggaaa tcttggggcg ggttaacagt 8160
ttgggcaacc tgcaactttt gcgcggagat gagaatatcg agaaagggtgc gctgccattc 8220
aggtcatgga ttaccggacg gcgcggtgat ttctacgagc agcatatgat cccggcgcac 8280
cttgaactgt gcgatgtact gcatctgccc gagttcgtgc gcgaacggga aaaggtgatc 8340
cggcgcctgt tgatggagtt ggtcggagca cgacgcgat gaatgaggtc gtcttgtcac 8400
gcgaagagct gcgtcaatct tgtctogacc tgcttgaaaa acgcgctgtc gaacaccctg 8460
cgggacacca aggcaagctc gccgcccgt atgttgtgca ccgcgacga 8509

<210> 3

<211> 5859

<212> DNA

<213> Artificial Sequence

-8-

<220>

<221> misc_feature

<223> pADM291-4DS

<400> 3

```

tcgcgcggtt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca      60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg      120
ttggcgggtg tcggggctgg cttactatg cggcatcaga gcagattgta ctgagagtgc      180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcggc      240
attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcgggcc tcttcgctat      300
tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt      360
tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt cggcaatggg tcgaaattca      420
tagaattttg tgtgaggtgc gtagcggctc tgacaggggt gctgcgcgga gatctctggt      480
ctcaggtagg gcgacaatgg agaggtgtta gttgccccct gtatcgctct ctgctggtgc      540
cattgggtca tcctgcccgg acatatgata ttccgctaga ggattactga tagtttctgc      600
ctgtcgggct tgtcgggott gtcgggcttg tcgggcttgt cgggcctgtc cctcttgtcc      660
cgcctgtcct cactttttca caatcaaaaa atgggcgaag cccttcttgt tctatagttc      720
ttatagttca tacgaaaatt acacataatt atcaatagct tattcgctta aaagggagta      780
attgggccgc aaaagggagt aattgggccg caaaaggag taattgggcc gcaaaagga      840
gtaattgggc cgatatoggt tgtttacatg gggaggaatc cccttaatca tttctcccca      900
tggaagac aacacaagtg gcgcagacc gggccttca ccagacaaa actgtgtctc      960
ctgccgaggt ggcgagaggg gtctatatgc gcaatccgcc ccgcctgcag gcgctcaagc     1020
tcatgcattt aatgatagcc actgcgggcg gccgcatggc tgatgatgtg cgccatgaaa     1080
tgcggtggc cgacattogc gcaatcgacg gcatgaaaaa ccatgaccgt gagagcctga     1140
ccccgctgtt cgaggagcta gccgctgctg tttgaccca tgatgaccct gcaaagatga     1200
tcgtgacagt cggcggcttg gtogatgagg cgcgaaataga ctaccgccag gaggcaagcg     1260
gcgaaactct agtgacgtgg accttcggga gtacattccg tcgtatggcg gcggagtcca     1320
accactgggc cattctogac cgtcaaacgg tattccatct cggtagtaag tattccgtgc     1380
tgctgttcca gcacgtctct agtctcgcca atcttgatcg gatgagcgcg aaaaccttta     1440
cggccccga gttgcgggcg ctccctggag tgcccggagg aaagatggtt cgttggaaacg     1500
acgttaacag atttgctctc aaacctgcac tggatgagat caaccattta tcgctctga     1560
cattgacggc aaagccgacc aagattggcc gtagcgtggc aagtgtgact ataggctggg     1620

```

-9-

aagtgaaaga cgacccaacc gtcgccagggc gcgagctggc gggttccaag gtcggtcgag 1680
atgctcgtcg cagaggggca gcggaaacga tagccccctc cttcccagaa gcgggcggga 1740
tcacctacag tccacgttgg ctggagctga aacgctctgc tggcagcaac aaggacaacg 1800
atctgatcgc ctcagacttc cggcgtttct gtcgggagag aggcgtgcgt ctggacgctg 1860
caaacatcga aaaactgttt ttagatthct gcgcaaaggt agggaaggth tgagthttga 1920
ggtatthcac cgcaatagtg ttaaatgact ttcgtgaaac gatgtgcaat atagcggtaa 1980
gactatgaaa tacacggctg gacaggctgc aaaagcaacg ggtgtggcga ccgcaaccat 2040
cactcgggcg ctaaaaagcg gtaaaatthc cggtaaaaaa gatgaatctg gggcatgggt 2100
tatagatcct gcagaattgc acagagtgtt tcctcccatt tcaaagaaat acaccgaaac 2160
acctaacacg caagtatatg gtaagcgtga tgaaacacat gaaatgacct cagaaatcag 2220
cgcattagag cgtgaagthc ggactthtac cgatgcttht tctgatgcca gggaggatcg 2280
cgacaaatgg cgcgacatgg ccgagcgtct tthcaatthc taccgatga gagaggaaga 2340
ccgccccct caaaaaaaa gatggtggaa gatattctga tcctgggctt caggagcctt 2400
gcctthactg gcgaaaaaac gcgatattga ggcacaggcc cgcactthag aggcggaagc 2460
ctataacgag taccaaaaaca ctagaagcca gattgaggaa aatagggaa gtgggatcct 2520
ctagagtcga cctgcaggca tgcagcttg gcgtaatcat ggtcatagct gthtctctg 2580
tgaaattht atccgctcac aattccacac aacatacag ccggaagcat aaagtgtaaa 2640
gcctggggtg cctaatgagt gagctaactc acattaatg cgttgcgctc actgcccgt 2700
thccagtcgg gaaacctgth gthgocagctg cattaatgaa tcggccaacg cgcggggaga 2760
ggcgththc gtatthggcg ctctthcccg ctccgthctg ccttgctcgt cggthgatg 2820
ctthaccagc thccggaagt cgotctthct gatggagcgc atggggacgt gctthgcaat 2880
cacgcgacc ccccgccgt thtagcggct aaaaaagtca tggctctgcc ctccggcgga 2940
ccacgcccac catgacctg ocaagctcgt cctgctthc thgatctth gccagcaggg 3000
cgaggatcgt ggcacaccg aaccgcgccg tgcgcgggth gtcggtgagc cagagththc 3060
gcaggccgcc caggcggccc aggtcgccat tgatgcggc cagctcgcg acgtgctcat 3120
agtccacgac gcccgthatt thgtagccct ggccgacggc cagcaggtag gccgacaggc 3180
thcatgccggc cgcgcgcgcc ththctcaa thgctctth thgctthgga aggcagthc 3240
cctthgatag thggctgccc thctgthth gctthgthth atcagccatc cgtthgccc 3300
catctgthac gccggcggth gcggccagc ctgcagagc aggathccc thgagcaccg 3360
ccagthcga ataagggaca gthgagaagg aacaccgct cgcggthgg cthactthc 3420
ctatctgcc cggctgacgc cgtthgthc accaaggaaa gthctacgca accctthggc 3480

-10-

aaaatcctgt atatcgtgcg aaaaaggatg gatataccga aaaaatcgct ataatgacct 3540
 cgaagcaggg ttatgcagcg gaaaagcttc ctcgctcact gactcgctgc gctcggctgt 3600
 tcggctgagg cgagcgggat cagctcactc aaaggcggta atacggttat ccacagaatc 3660
 aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa 3720
 aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc atcacaaaaa 3780
 tcgacgctca agtcagaggt ggcgaaaacc gacaggacta taaagatacc aggcgtttcc 3840
 ccctggaagc tccctcgtgc gctctcctgt tccgacctg ccgcttaccg gatacctgtc 3900
 cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta ggtatctcag 3960
 ttcgggtgtag gtogttcgtt ccaagctggg ctgtgtgcac gaaccccccg ttcagcccga 4020
 ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccgtaagac acgacttatc 4080
 gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac 4140
 agagtctctg aagtggggc ctaactacgg ctacactaga aggacagtat ttggatatctg 4200
 cgctctgctg aagccagtta ccttcggaaa aagagttggg agctcttgat ccggcaaaaa 4260
 aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa 4320
 aggatctcaa gaagatcctt tgatcttttc tacgggggtct gacgctcagt ggaacgaaaa 4380
 ctcacgttaa gggattttgg tcatgagatt atcaaaaagg atcttcacct agatcctttt 4440
 gggggggggg ggcgctgagg tctgcctcgt gaagaagggtg ttgctgactc ataccaggcc 4500
 tgaatcgccc catcatccag ccagaaagtg agggagccac ggttgatgag agctttgttg 4560
 taggtggacc agttgggtgat tttgaaacttt tgctttgcca cggaacggtc tgcgttgctg 4620
 ggaagatgcy tgatctgatc cttcaactca gcaaaagttc gatttattca acaaagccgc 4680
 cgtcccgtca agtcagcgtc atgctctgcc agtggttaca ccaattaacc aattctgatt 4740
 agaaaaactc atcgagcadc aaatgaaact gcaatttatt catatcagga ttatcaatac 4800
 catatttttg aaaaagccgt ttctgtaatg aaggagaaaa ctcaccgagg cagttccata 4860
 ggatggcaag atcctgggat cggctctgoga ttccgactcg tccaacatca atacaaccta 4920
 ttaatttccc ctogtcaaaa ataaggttat caagtgagaa atcaccatga gtgacgactg 4980
 aatccggtga gaatggcaaa agcttatgca tttctttcca gacttgttca acaggccagc 5040
 cattacgctc gtcatcaaaa tcaactcgcac caaccaaac gttattcatt cgtgattgcy 5100
 cctgagcagc acgaaatacg cgatcgcctgt taaaaggaca attacaaca ggaatcgaat 5160
 gcaaccggcg caggaaact gccagcgcac caacaatatt ttcacctgaa tcaggatatt 5220
 cttctaatac ctggaatgct gttttcccg ggatcgcagt ggtgagtaac catgcatcat 5280
 caggagtacg gataaaatgc ttgatggctg gaagaggcat aaattccgct agccagttta 5340

-11-

gtctgacat ctcatctgta acatcattgg caacgctacc tttgcatgt ttcagaaaca 5400
 actctggcgc atcgggcttc ccatacaatc gatagattgt cgcacctgat tgcccacat 5460
 tatcgcgagc ccatttatac ccatataaat cagcatccat gttggaattt aatcgcggcc 5520
 tcgagcaaga cgtttcccgt tgaatatggc tcataacacc cttgtatta ctgtttatgt 5580
 aagcagacag ttttattggt catgatgata tatttttatac ttgtgcaatg taacatcaga 5640
 gattttgaga cacaacgtgg ctttcccccc cccccatta ttgaagcatt tatcagggtt 5700
 attgtctcat gagcggatac atatttgaat gtatttagaa aaataaaca ataggggttc 5760
 cgcgcacatt tccccgaaaa gtgccacctg acgtotaaga aaccattatt atcatgacat 5820
 taacctataa aaataggcgt atcacgaggc cctttcgtc 5859

<210> 4

<211> 2517

<212> DNA

<213> Artificial Sequence

<220>

<221> misc_feature

<223> Ketogulonigenium part of pADM291-4

<400> 4

ggcaatgggt cgaaattcat agaattttgt gtgagggtgcg tagcggctct gacaggggtg 60
 ctgcgcggag atctctggtc tcaggtaggg cgacaatgga gaggtgtag ttgccccctg 120
 tatcgtctc tgcgtggcgc attgggtcat cctgcccgga catatgatat tccgctagag 180
 gattactgat agtttctgcc tgtcgggctt gtcgggcttg tcgggcttgt cgggcttgtc 240
 gggcctgtcc ctcttgtccc gcctgtcctc actttttcac aatcaaaaaa tgggcgaagc 300
 cttcttgtt ctatagttct tatagttcat acgaaaatta cacataatta tcaatagctt 360
 attcgttaa aaggagtaa ttgggcccga aaaggagta attgggcccgc aaaaggaggt 420
 aattgggccc caaaaggag taattgggcc gatatcgggt gtttacctgg ggaggaatcc 480
 ccttaatcat ttctccccat gggaaagaca acacaagtgg ccgcagaccg ggccttcgac 540
 cagacaaaaa ctgtgctccc tgccgagggtg gcgagagggg tctatatgcg caatccgccc 600
 cgctgcagg cgctcaagct catgcattta atgatagcca ctgcgggccc cgcgatggct 660
 gatgatgtgc gccatgaaat gcggctggcc gacattcgcg caatcgacgg catgaaaaac 720

-12-

catgaccgtg agagcctgac cccgctgttc gaggagctag ccgctgcggt gttgacccat 780
gatgaccctg caaagatgat cgtgacagtc ggcggcttgg tcgatgaggc gcgaatagac 840
taccgccagg aggcaagcgg cgaactccta gtgacgtgga ccttccggag tacattccgt 900
cgtatggcgg cggagtcgaa ccaactggcc attctcgacc gtcaaacggt attccatctc 960
ggtagtaagt attccgtgct gctgttccag cacgtctcta gtctcgcaa tcttgatcgg 1020
atgagcgcga aaacctttac ggtccccgag ttgcgggcgc tccttggagt gcccgagga 1080
aagatggttc gttggaacga cgttaacaga tttgctctca aacctgcaact ggatgagatc 1140
aaccatttat cgcgtctgac attgacggca aagccgacca agattggccg tagcgtggca 1200
agtgtgacta taggctggga agtgaaagac gaccaaccg tcgccaggcg cgagctggcg 1260
ggttccaagg tcggtcgaga tgctcgtcgc agaggggcag cggaaacgat agccccctcc 1320
ttcccagaag cgggcgggat cacctacagt ccacgttggc tggagctgaa acgctctgct 1380
ggcagcaaca aggacaacga tctgatcgcc tcgacttcc ggcgtttctg tcgggagaga 1440
ggcgtgcgtc tggacgctgc aaacatcgaa aaactgtttt tagatttctg cgcaaaggta 1500
gggaaggttt gagttttgag gtatttcacc gcaatagtgt taaatgactt tcgtgaaacg 1560
atgtgcaata tagcggtaag actatgaaat acacggctgg acaggctgca aaagcaacgg 1620
gtgtggcgac cgcaaccatc actcggggcg taaaagcgg taaaatttcc ggtaaaaaag 1680
atgaatctgg ggcattgggtt atagatcctg cagaattgca cagagtgttt cctcccattt 1740
caaagaaata caccgaaaca cctaacacgc aagtatatgg taagcgtgat gaaacacatg 1800
aatgacctc agaaatcagc gcattagagc gtgaagtctg gactttacgc gatgctttat 1860
ctgatgccag ggaggatcgc gacaaatggc ggcacatggc cgagcgtctt tcaatttcat 1920
caccgatgag agaggaagac cgccccctc aaaaacaaag atggtggaag atattctgat 1980
cctgggcttc aggagccttg cctttaaaac ctgaatcagc attctagcga tgctgataag 2040
aagtaaataat agccacaata gagcggccat tttcattca catacagctc atcatgtgat 2100
caatatcaag tattgatatt catcaatgga gaagaattta catgtatcac aggatcatca 2160
cagcatttgt ttttgatatt ctaagtgcta acataactat cgctggccct aaagaagatt 2220
gtactattgc agtatctcac cttgggtttc agaccgataa ttacagcttt gtcgaagccg 2280
gtttttttgc cagagagaga cacgtttttg atggtgtaat aaactgctac gtatctcatg 2340
atggtaacat acacagcatc atccggggca acacacctct tatggaagat ggatattatg 2400
gccagaagt actggcggaa aaacgcgata ttgaggcaca ggcccgcact ttagaggcgg 2460
aagcctataa cgagtaccaa aacactagaa gccagattga ggaaaatagg gaacgtg 2517

| | | | |
|--|--------------|-------------------------------|----------------|
| Applicant's or agent's file reference number | 1533.110PC01 | International application No. | To Be Assigned |
|--|--------------|-------------------------------|----------------|

**INDICATIONS RELATING TO DEPOSITED MICROORGANISM
OR OTHER BIOLOGICAL MATERIAL**
(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 1, line 11 and 12.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet

Name of depositary institution
Agricultural Research Culture Collection (NRRL)

Address of depositary institution *(including postal code and country)*

1815 N. University Street
Peoria, Illinois 61604
United States of America

| | |
|----------------------------------|----------------------------------|
| Date of deposit July 21, 1998 | Accession Number NRRL B-30035 |
|----------------------------------|----------------------------------|

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet

Bacteria sp. ADM 291-19


D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the international Bureau later *(specify the general nature of the indications, e.g., "Accession Number of Deposit")*

For receiving Office use only

This sheet was received with the international application

Authorized officer 

For International Bureau use only

This sheet was received by the International Bureau on:

Authorized officer

| | | | |
|--|--------------|-------------------------------|----------------|
| Applicant's or agent's file reference number | 1533.110PC01 | International application No. | To Be Assigned |
|--|--------------|-------------------------------|----------------|

**INDICATIONS RELATING TO DEPOSITED MICROORGANISM
OR OTHER BIOLOGICAL MATERIAL**
(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 3, line 11.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet

Name of depositary institution
Agricultural Research Culture Collection (NRRL)

Address of depositary institution (including postal code and country)
1815 N. University Street
Peoria, Illinois 61604
United States of America

| | |
|-----------------------------------|----------------------------------|
| Date of deposit March 27, 2001 | Accession Number NRRL B-30434 |
|-----------------------------------|----------------------------------|


C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet

Escherichia coli DH5 α MCR p291-4DS

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)

E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)

The indications listed below will be submitted to the international Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")

| | |
|--|--|
| For receiving Office use only | For International Bureau use only |
| <input checked="" type="checkbox"/> This sheet was received with the international application | <input type="checkbox"/> This sheet was received by the International Bureau on: |
| Authorized officer  | Authorized officer |