



US 2010050301A1

(19) **United States**

(12) **Patent Application Publication**
MENDEZ et al.

(10) **Pub. No.: US 2010/0050301 A1**
(43) **Pub. Date:** **Feb. 25, 2010**

(54) **SYSTEM FOR CAPTURING AND
MODIFYING LARGE PIECES OF GENOMIC
DNA AND CONSTRUCTING VASCULAR
PLANTS WITH SYNTHETIC CHLOROPLAST
GENOMES**

(75) Inventors: **MICHAEL MENDEZ, SAN
DIEGO, CA (US); BRYAN
O'NEILL, SAN DIEGO, CA (US);
KARI MIKKELSON, SAN
DIEGO, CA (US)**

Correspondence Address:

**Sapphire Energy
3115 Merryfield Row
San Diego, CA 92121 (US)**

(73) Assignee: **SAPPHIRE ENERGY, INC., SAN
DIEGO, CA (US)**

(21) Appl. No.: **12/560,341**

(22) Filed: **Sep. 15, 2009**

Related U.S. Application Data

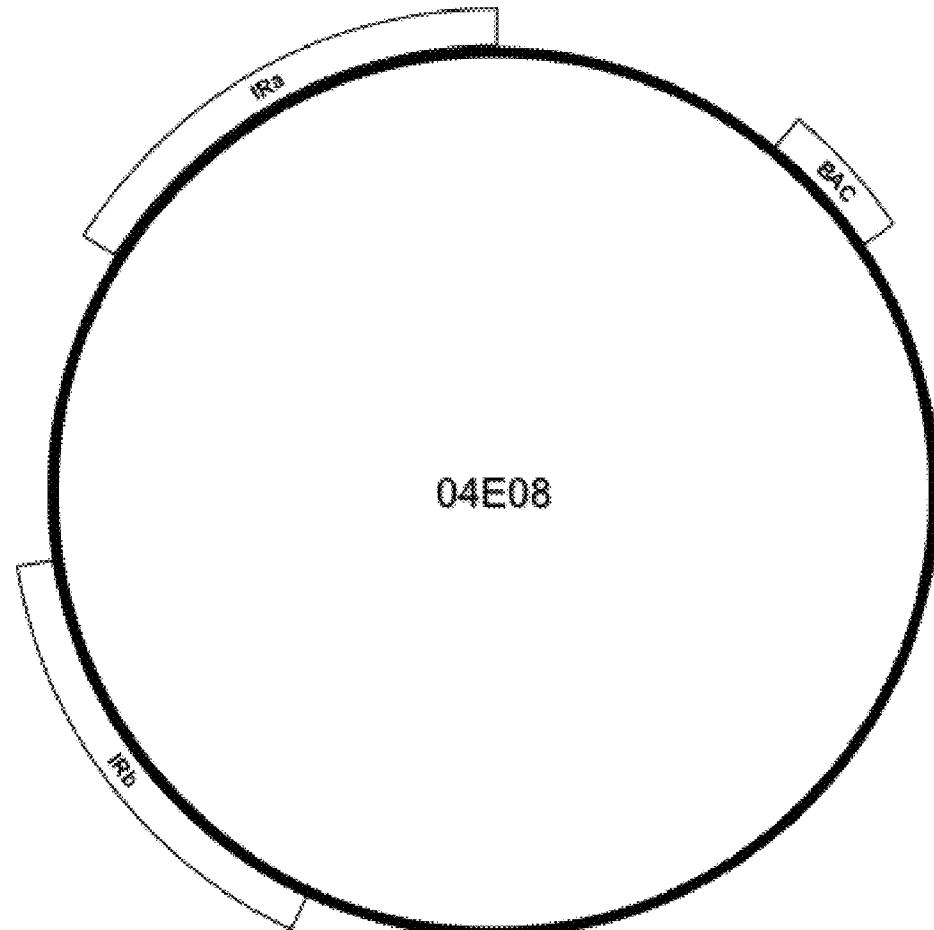
- (63) Continuation-in-part of application No. 12/384,893, filed on Apr. 8, 2009, which is a continuation-in-part of application No. 12/287,230, filed on Oct. 6, 2008.
(60) Provisional application No. 60/978,024, filed on Oct. 5, 2007.

Publication Classification

- (51) **Int. Cl.**
A01H 5/00 (2006.01)
C12N 15/74 (2006.01)
C12N 15/64 (2006.01)
C12N 1/21 (2006.01)
C12N 1/19 (2006.01)
C07H 21/04 (2006.01)
(52) **U.S. Cl.** **800/298; 435/320.1; 435/91.42;**
435/252.3; 435/254.2; 536/23.6

(57) **ABSTRACT**

Vectors capable of stable replication in yeast and bacteria and comprising all essential genes of vascular plant plastids are provided as well as the use of such vectors to construct an recombinant plastid genome and host cells transformed with said vectors and recombinant plastid genomes.



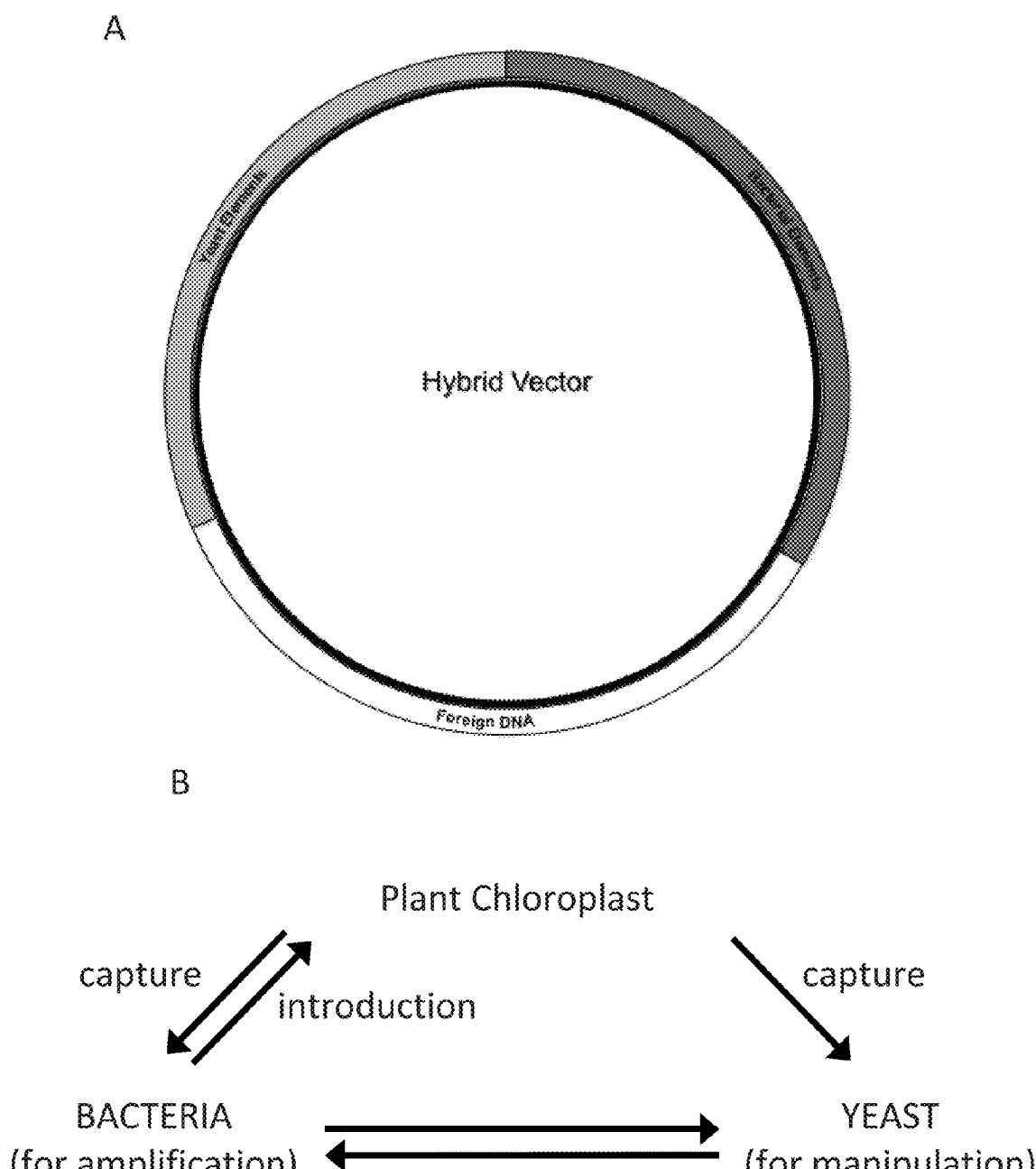


Figure 1

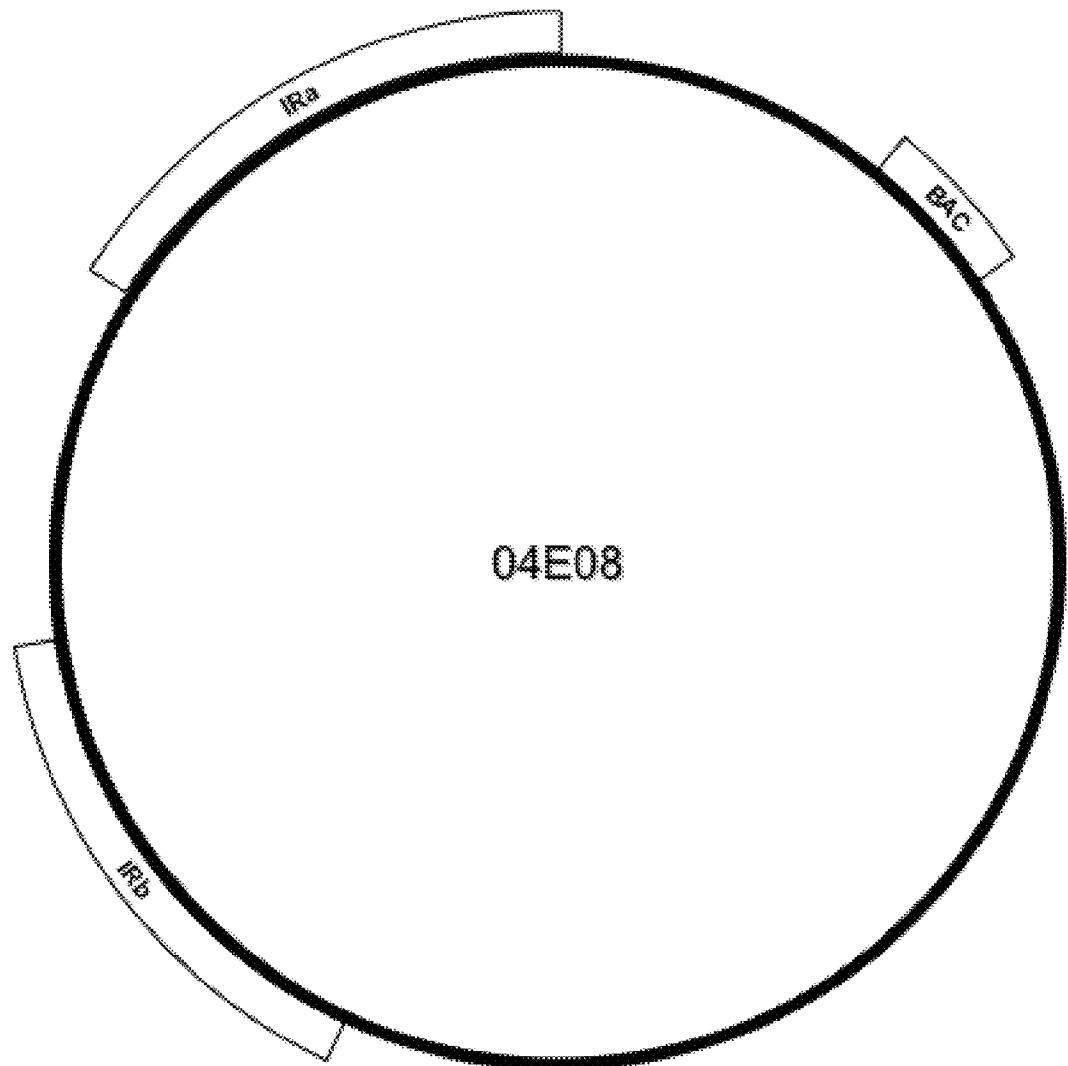


Figure 2

Figure 3A

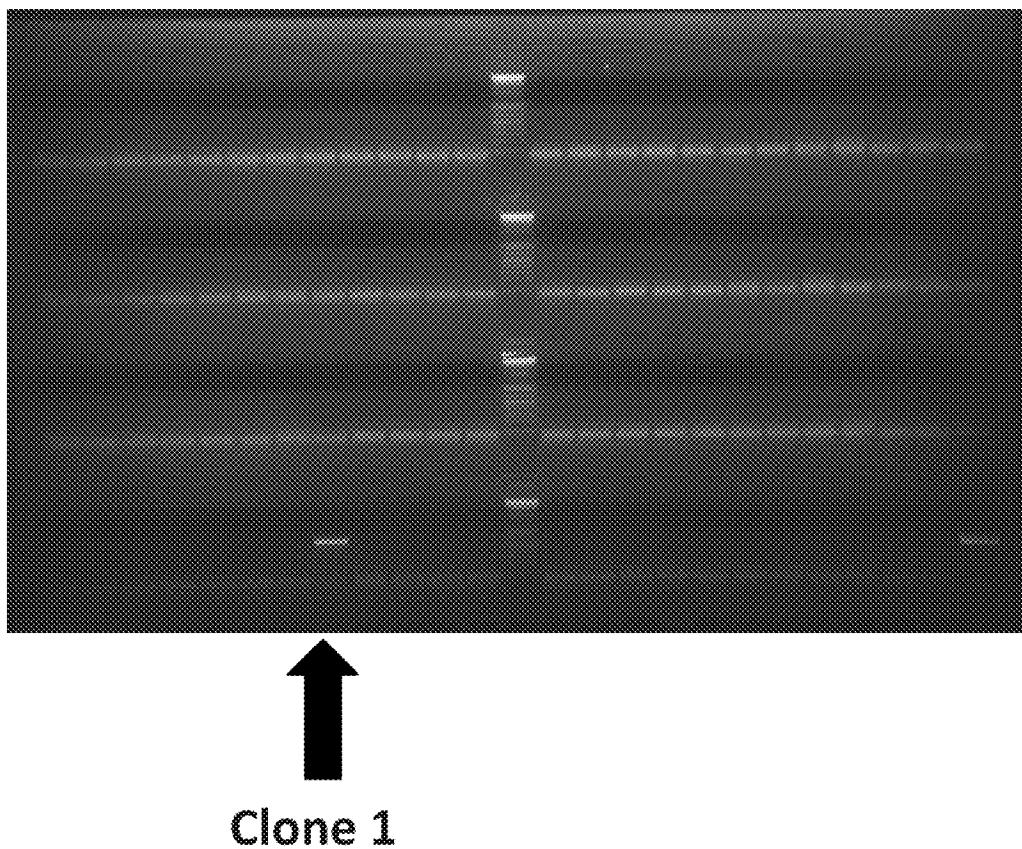


Figure 3B

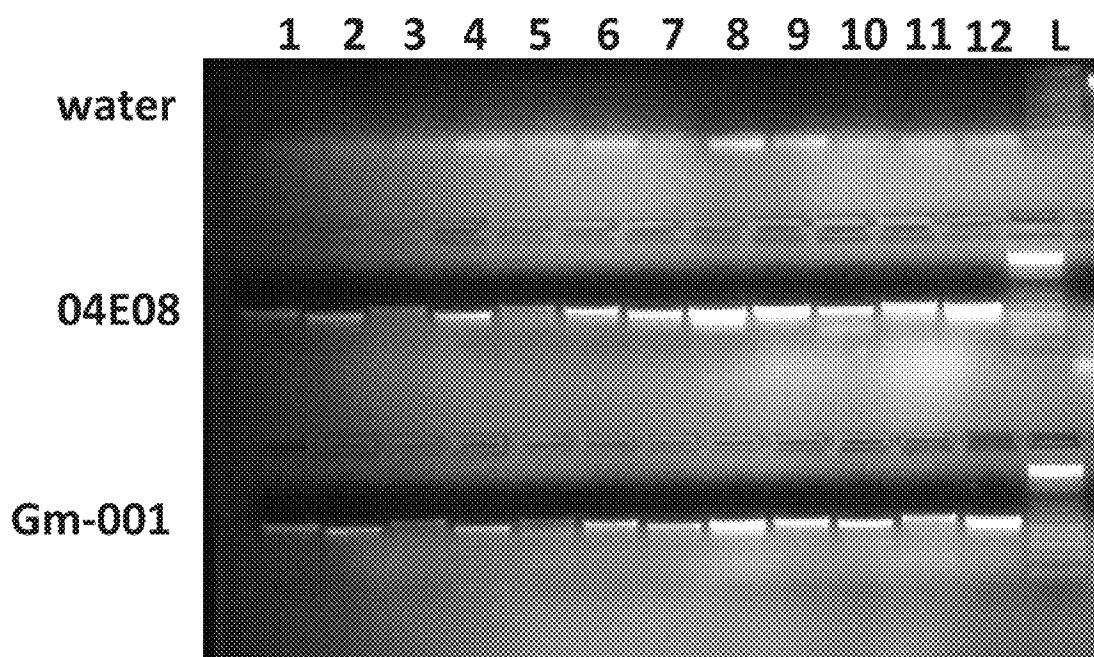


Figure 3C

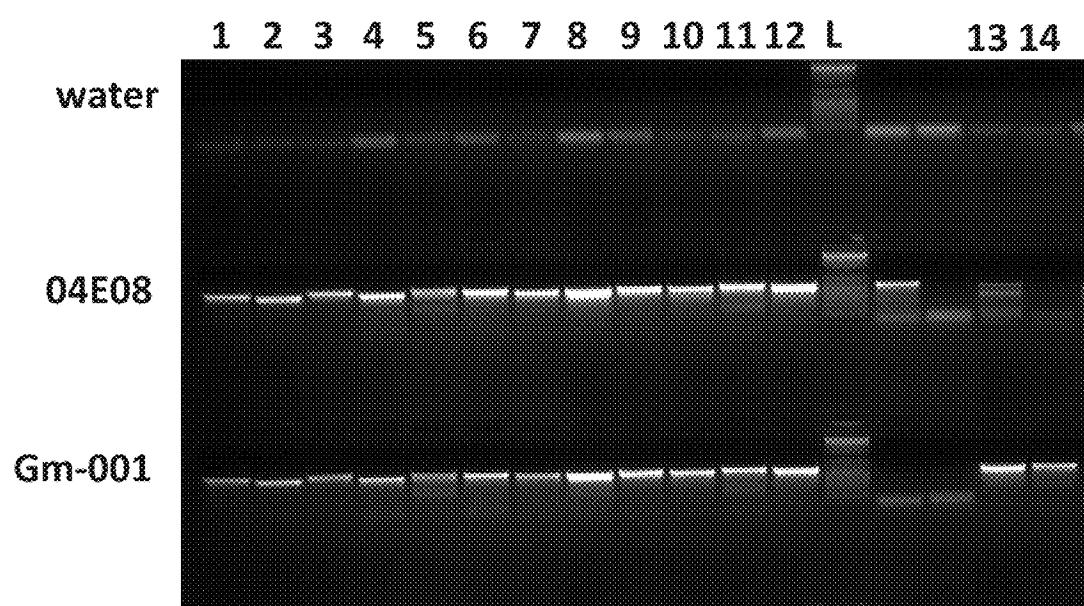
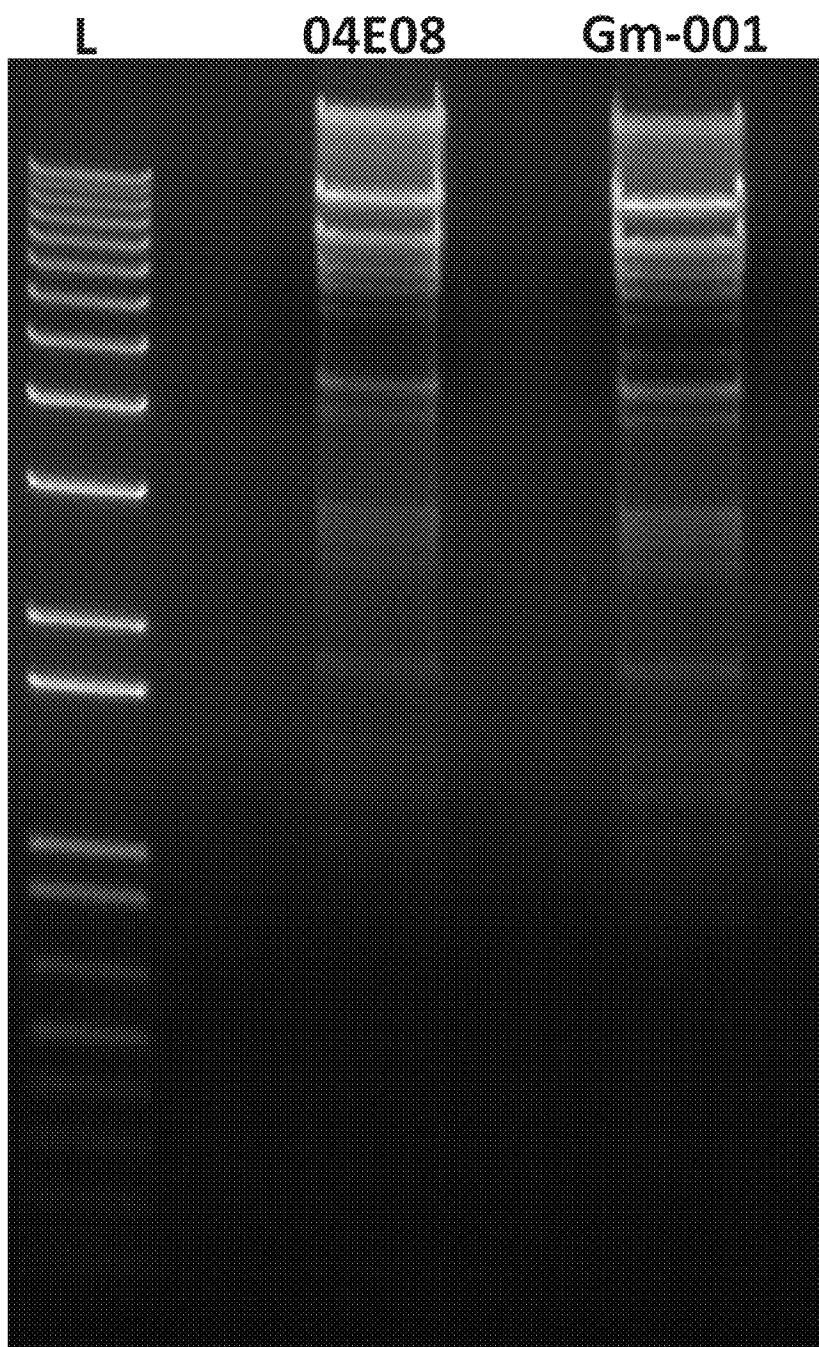


Figure 3D



**SYSTEM FOR CAPTURING AND
MODIFYING LARGE PIECES OF GENOMIC
DNA AND CONSTRUCTING VASCULAR
PLANTS WITH SYNTHETIC CHLOROPLAST
GENOMES**

**CROSS REFERENCE TO RELATED
APPLICATIONS**

[0001] The present application is a continuation in part of copending U.S. patent application Ser. No. 12/384,893 filed Apr. 8, 2009, which is a continuation in part of copending U.S. patent application Ser. No. 12/287,230 filed Oct. 6, 2008, which claims the benefit of U.S. Provisional Patent Application 60/978,024 filed Oct. 5, 2007, now abandoned; each of which is incorporated by reference in its entirety for all purposes.

BACKGROUND

[0002] For the functional analysis of many genes, investigators need to isolate and manipulate large DNA fragments. The advent of genomics and the study of genomic regions of DNA have generated a need for vectors capable of carrying large DNA regions.

[0003] In general, two types of yeast vector systems are presently available. The first type of vector is one capable of transferring small insert DNA between yeast and bacteria. A second type of vector is a fragmenting vector which creates interstitial or terminal deletions in yeast artificial chromosomes (YACs). The small insert shuttle vectors are able to recombine with and recover homologous sequences. They are centromere-based and replicate stably and autonomously in yeast, but also contain a high-copy origin of replication for maintenance as bacterial plasmids. However, these vectors are limited by their small insert capacity. The second type of vector (also known as fragmenting vectors) has recombinogenic sequences, but is unable to transfer the recovered insert DNA to bacteria for large preparations of DNA.

[0004] Researchers use fragmentation techniques to narrow down the region of interest in YACs. Isolating sufficient quantities of YAC DNA from agarose gels for microinjection or electroporation, however, remains cumbersome. Purification remains a problem when the YAC comigrates with an endogenous chromosome. In addition, YACs may be chimeric or contain additional DNA regions that are not required for the particular functional study.

[0005] Types of vectors available for cloning large fragments in bacteria are cosmids, P1s and bacterial artificial chromosomes (BACs). These vectors are limited to bacteria and cannot be shuttled to yeast for modification by homologous recombination. Bacterial vectors are also limited in their use for transforming plants. For example, although chloroplasts are thought to originate from the endosymbiosis of photosynthetic bacteria into eukaryotic hosts, translation in chloroplasts is more complex. Adding to the complexity of genetically engineering plants is the presence of multiple chloroplasts with multiple copies of the chloroplast genome. Thus, there exists a need for developing a method to express proteins from large fragments of DNA in the chloroplasts of plants.

SUMMARY

[0006] Disclosed herein are compositions and methods of isolating, characterizing, and/or modifying large DNA,

including entire genomes of chloroplasts. The compositions include shuttle vectors into which target DNA may be inserted. The methods include modifying or manipulating target DNA by removing, adding or rearranging portions and introducing the modified DNA into a host, in particular a vascular plant.

[0007] One aspect provides an isolated vector comprising a yeast element, a bacterial origin of replication, and at least 90% of the plastid genome of a vascular plant. In some vectors, the yeast element is a yeast centromere, a yeast autonomous replicating sequence, yeast auxotrophic marker, or a combination thereof. The plastid genome may be from, for example, soybean (*G. max*), *arabidopsis* (*A. thaliana*), corn (*Z. mays*), rice (*O. sativa*) or wheat (*T. aestivum*). In some embodiments, the plastid genomic DNA is modified, for example by insertion of a heterologous or homologous polynucleotide, deletion of one or more nucleic acid bases, mutation of one or more nucleic acid bases, rearrangement of one or more polynucleotides, or a combination thereof. In some instances, the modification is synthetic. Vectors of the present disclosure, when transformed into a plant host cell, may result in production of a product not naturally produced by the host plant cell. The vectors disclosed herein may further comprise one or more selection markers, for example, a yeast marker, a yeast antibiotic resistance marker, a yeast auxotrophic marker, a bacterial marker, a bacterial antibiotic resistance marker, a bacterial auxotrophic marker or any combination thereof. Vectors may also contain chloroplast genomic DNA which comprises 1) 1-200 genes; 2) all essential chloroplast genes; 3) at least 90% of a chloroplast genome; 4) at least 135 kb; or 5) at least 150 kb.

[0008] Also described is a plant host cell comprising the vectors described herein. Exemplary host cells may be monocots or dicots and include, but are not limited to, soybeans, tomatoes, potatoes, wheat, rice, corn, barley, rye, and cotton.

[0009] Another aspect provides method for producing a vector where the method involves inserting targeting DNA into a vector—where the vector comprises a yeast centromere, a yeast autonomous replicating sequence, and a bacterial origin of replication, transforming an organism, for example yeast, with the vector and capturing a portion of a chloroplast genome, thus producing a vector with a portion of or a complete chloroplast genome. In some instances, the targeting DNA is chloroplast genomic DNA. This method may be used to capture a portion of a genome which is at least 135 kb in length, at least 150 kb or that comprises at least 90% of a chloroplast genome. In some instances, the capturing step occurs by recombination. The captured portion of a chloroplast genome may be co-transformed into an organism with a vector, thus the recombination step may occur *in vivo*. Vascular plants used to practice methods disclosed herein may be monocots or dicots and include the major agricultural crops. In some embodiments, an additional step of modifying a portion of a chloroplast genome is utilized. A modification may be achieved through homologous recombination. Such recombination may occur in an organism, for example yeast. In embodiments with a modification step, the step may comprise addition of a polynucleotide, deletion of one or more nucleic acid bases, mutation of one or more nucleic acid bases, rearrangement of a polynucleotide, or any combination thereof.

[0010] Further disclosed herein is an isolated vector comprising essential chloroplast genes, a selectable marker and a manipulation in one or more nucleic acids in the vector. In

some instances, essential chloroplast genes are cloned from a vascular plant such as soybean (*G. max*), *arabidopsis* (*A. thaliana*), corn (*Z. mays*), rice (*O. sativa*) or wheat (*T. aestivum*). Essential chloroplast genes for use in the vectors described herein may be synthetic. The vectors described herein may further comprise an expression cassette, which may further comprise a region for integration into target DNA, for example plastid DNA. The vectors described herein may also contain one or more selection markers, for example, an auxotrophic marker, an antibiotic resistance marker, a chloroplast marker, or any combination thereof. In some instances, the essential chloroplast genes are those required for chloroplast function, photosynthesis, carbon fixation, production of one or more products, or any combination thereof. Essential chloroplast genes may comprise up to 200 genes and/or consist of up to 400 kb. In some of the vectors described herein, a manipulation in one or more nucleic acids is an addition, deletion, mutation, or rearrangement. In some instances, expression of the vector in a host cell produces a product not naturally produced by said host cell. In other instances, expression of a vector of the present invention results in an increase production of a product naturally produced by said host cell.

[0011] One aspect provides an isolated chloroplast comprising a vector described herein. In another aspect, a host cell comprising a vector described herein is provided. Host cells useful in the present include those obtained from any vascular plant. Examples of plants useful for the present disclosure include soybeans, tomatoes, potatoes, wheat, rice, corn, barley, rye, and cotton.

[0012] In another aspect, a method for transforming a cell or organism is provided where the method comprises inserting into said cell a vector comprising all essential chloroplast genes or 90% of a chloroplast genome. Optionally, the vector can comprise one or more genes not naturally occurring in said cell or organism. In some embodiments, the method further comprises the step of eliminating substantially all chloroplast genomes in said cell or organism or disabling the photosynthetic capability of the cell. A cell or organism useful for this method may be photosynthetic, non-photosynthetic and/or eukaryotic. In some instances, the vector for use in this method may also comprise an expression cassette and the expression cassette may be capable of integrating into non-nuclear DNA. In one embodiment the one or more genes not naturally occurring in the cell or organism is a gene in the isoprenoid pathway, MVA pathway, or MEP pathway. In another embodiment, the essential chloroplast genes are those that are required for chloroplast function, photosynthesis, carbon fixation, production of one or more hydrocarbons, or a combination thereof. In still another embodiment, the genes not naturally occurring in the cell or organism confer herbicide, insect and/or disease resistance. In another embodiment, the one or more genes not naturally occurring in the cell or organism produce a therapeutic protein. In another embodiment, the one or more genes not naturally occurring in the cell or organism increase production of a lipid, fatty acid or phytosterol.

[0013] Further provided herein is a method for modifying an organism comprising the steps of transforming the organism, and in particular a vascular plant, with a vector disclosed herein. In some instances, a vector useful for this method further comprises a sequence for production and/or secretion of a compound from said organism. In other instances, the vector comprises all essential chloroplast genes. In other

instances, the vector comprises at least 90% of a chloroplast genome. In still other instances, the essential chloroplast genes are rearranged or mutated. An organism useful for some embodiments comprises essentially no chloroplast genome or a chloroplast genome incapable of photosynthesis prior to transformation.

[0014] Yet another method provided herein is a method for making a product from an organism comprising the step of transforming said organism with a vector described herein and further comprising one or more of the following: (i) a gene not naturally occurring in said organism; (ii) a deletion in a gene naturally occurring in said organism; (iii) a rearrangement of genes naturally occurring in said organism; and (iv) a mutation in a gene naturally occurring in said organism. In some instances, the organism is naturally photosynthetic. In other instances, the additional genes encode enzymes in the isoprenoid pathway, MVA pathway, or MEP pathway. In other instances, the additional genes allow for the production of therapeutic products, such as therapeutic proteins and phytosterols. In still another embodiment, the present disclosure provides a method for transforming a cell or organism comprising inserting into said cell or organism a chloroplast and a vector comprising all essential chloroplast genes.

[0015] The present disclosure also provides a method of producing an artificial chloroplast genome comprising the steps of: (a) providing a vector comprising one or more essential chloroplast genes; (b) adding to said vector a DNA fragment; (c) transforming a cell or organism with the vector produced by step (b); and (d) determining whether chloroplast function exists with said added DNA fragment. In some instances, the added DNA fragments comprise one or more coding regions for an enzyme in the isoprenoid, MVA or MEP pathway.

[0016] The present disclosure also provides a shuttle vector comprising at least 90% of a chloroplast genome of a vascular plant. In some instances, the genome may be modified. The shuttle vector comprises at least one selection marker, at least one yeast element, at least one bacterial origin of replication and at least one bacterial selection marker. The shuttle vector is capable of stable replication in yeast and bacterial cells. In certain embodiments, the yeast elements are a yeast centromere sequence, a yeast autonomously replicating nucleotide sequence or both. In other embodiments the bacterial origin of replication is a P1 or F' origin of replication. Also provided herein is a vector comprising an isolated, functional chloroplast genome. A chloroplast genome useful in such a vector may be modified.

[0017] Further provided herein is a method of producing an artificial chloroplast genome comprising the steps of: (a) providing a vector comprising all essential chloroplast genes; and (b) removing, adding, mutating, or rearranging DNA from the chloroplast genome. Such a method may further comprise the steps of transforming a redacted genome into a host organism; and (d) determining chloroplast function in the host organism. In some instances, steps (b), (c), and (d) are repeated. In still other instances, the chloroplast genome is from an organism selected from the group consisting of: soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes (e.g., peas, beans, lentils, alfalfa, etc.). For some embodiments, the method may further comprise the step of removing redundant DNA from a chloroplast genome. In other embodiments, the vector comprises all or substantially all of a chloroplast genome, for example at least 85%, at least 87%, at least 90%,

at least 92% or at least 95% of a chloroplast genome. A chloroplast genome useful in the present invention may be cloned from a photosynthetic organism or may be a synthetic chloroplast genome. In some instances, the vector further comprises a gene not naturally occurring in the host organism.

[0018] Yet another method provided herein is a method of producing an artificial chloroplast genome comprising the steps of: (a) providing a vector comprising an entire chloroplast genome; (b) deleting a portion of said entire chloroplast genome; and (c) determining whether chloroplast function exists without said deleted portion. In another aspect of the present invention, a composition comprising an isolated and functional chloroplast genome is provided. In some instances, a composition comprises a modification to said chloroplast genome.

[0019] Further provided herein is an ex vivo vector comprising a nucleic acid comprising at least about 85%, at least about 87%, at least about 90%, at least about 92%, at least about 95%, or all essential genes of a chloroplast genome and a manipulation in one or more nucleic acids in the vector. In some instances, the vector comprises at least about 135 kb or at least about 150 kb of chloroplast genomic DNA. In some instances, the nucleic acid is cloned from a vascular plant, for example soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes. In some instances, the nucleic acid is synthetic. A vector of the present disclosure may further comprise an expression cassette and an expression cassette may further comprise a region for integration into target DNA. In some instances, the target DNA is organelle DNA. A vector useful in the present disclosure may further comprise one or more selection markers, for example an auxotrophic marker, an antibiotic resistance marker, a chloroplast marker, or combinations thereof. In some embodiments, a manipulation in one or more nucleic acids in a vector may be an addition, deletion, mutation, or rearrangement. Expression of the vector may result in production of a product not naturally produced by a host cell and/or an increase production of a product naturally produced by a host cell. Examples of some products of the present invention include a terpene, terpenoid, fatty acid, a biomass degrading enzyme, a therapeutic protein and/or a phytosterol.

[0020] Further provided herein is a method of producing a vector containing a reconstructed genome, comprising: introducing two or more vectors into a host cell, wherein the vectors comprise fragments of a genome, recombining the vectors into a single vector comprising at least about 85%, at least about 87%, at least about 90%, at least about 92% or at least about 95% of a plastid genome, thereby producing a vector containing a reconstructed genome. In some instances, the host cell is eukaryotic, for example, a plant cell. The plastid may be a chloroplast, for example a chloroplast from a vascular plant such as soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes. In some instances, the two or more vectors comprise a selectable marker. In other instances, at least one of said fragments is synthetic. In still other instances, a further step comprising modifying a portion of the genome is useful in this method. Such a modification may comprise an addition, deletion, mutation, or rearrangement. In other embodiments, the modification is the addition of an exogenous nucleic acid

which results in the production or increased production of a terpene, terpenoid, fatty acid or biomass degrading enzyme.

INCORPORATION BY REFERENCE

[0021] All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0023] FIG. 1 provides a general description of a hybrid vector of the present invention. FIG. 1A) Vector schematic. FIG. 1B) DNA shuttling between organisms.

[0024] FIG. 2 is a schematic showing clone 04E08.

[0025] FIG. 3 shows analysis of the *G. max* chloroplast genome cloned into the hybrid vector system. FIG. 3A) PCR screen data of isolated yeast transformants using primers 6113 and 6114. FIG. 3B) PCR screen data of Gm-001 in its isolated yeast clone using 12 different PCR primer pairs (lane 1, 6113 and 6114; lane 2, 6115 and 6116; lane 3, 6117 and 6118; lane 4, 6119 and 6120; lane 5, 6121 and 6122; lane 6, 6123 and 6124; lane 7, 6125 and 6126; lane 8, 6127 and 6128; lane 9, 6129 and 6130; lane 10, 6131 and 6132; lane 11, 6133 and 6134; and lane 12, 6135 and 6136). FIG. 3C) PCR screen data of Gm-001 in bacteria using 14 different PCR primer pairs (lane 1, 6113 and 6114; lane 2, 6115 and 6116; lane 3, 6117 and 6118; lane 4, 6119 and 6120; lane 5, 6121 and 6122; lane 6, 6123 and 6124; lane 7, 6125 and 6126; lane 8, 6127 and 6128; lane 9, 6129 and 6130; lane 10, 6131 and 6132; lane 11, 6133 and 6134; lane 12, 6135 and 6136, lane 13, 6105 and 6106; and lane 14, 6107 and 6108). FIG. 3D shows the stability of the isolated clone in the hybrid system.

DETAILED DESCRIPTION OF THE INVENTION

[0026] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

[0027] Technical and scientific terms used herein have the meanings commonly understood by one of ordinary skill in the art to which the instant invention pertains, unless otherwise defined. Reference is made herein to various materials and methodologies known to those of skill in the art. Standard reference works setting forth the general principles of recombinant DNA technology include Sambrook et al., "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Plainview, N.Y., 1989; Kaufman et al., eds.,

"Handbook of Molecular and Cellular Methods in Biology and Medicine", CRC Press, Boca Raton, 1995; and McPhereson, ed., "Directed Mutagenesis: A Practical Approach", IRL Press, Oxford, 1991. Standard reference literature teaching general methodologies and principles of yeast genetics useful for selected aspects of the invention include: Sherman et al. "Laboratory Course Manual Methods in Yeast Genetics", Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1986 and Guthrie et al. "Guide to Yeast Genetics and Molecular Biology", Academic, New York, 1991.

[0028] Any suitable materials and/or methods known to those of skill can be utilized in carrying out the instant invention. Materials and/or methods for practicing the instant invention are described. Materials, reagents and the like to which reference is made in the following description and examples are obtainable from commercial sources, unless otherwise noted. This invention teaches methods and describes tools for capturing and modifying large pieces of DNA. It is especially useful for modifying genomic DNA, including the whole genome of an organism or organelle, or a part thereof. Novel prophetic uses of the invention are also described. The disclosure relates to the manipulation and delivery of large nucleic acids. The disclosure further relates to recombinational cloning vectors and systems and to methods of using the same.

[0029] Contemporary methods for genetically engineering genomes (e.g., chloroplast genomes) are time intensive (>1 month) and allow for only a limited number of manipulations at a time. If multiple modifications to a target genome are desired, the process must be iterated, further increasing the time required to generate a desired strain. Because metabolic engineering and/or synthetic biology require numerous modifications to a genome, these technologies are not amenable to rapid introduction of modifications to a genome. Thus, a new technology that allows for multiple modification of the chloroplast genome in a short amount of time will enable the application of metabolic engineering and/or synthetic biology to chloroplast genomes. The disclosure herein describes such technology.

[0030] The instant invention provides a versatile, recombinational approach to the capture, cloning, and manipulation of large nucleic acids from target cells and organelles (e.g., chloroplasts). One aspect of the present disclosure provides a recombinational cloning system. More specifically, the disclosure provides vectors, which in some embodiments, rely on homologous recombination technologies to mediate the isolation and manipulation of large nucleic acid segments. Another aspect of the present disclosure provides methods for using such recombinational cloning vectors to clone, to manipulate and to deliver large nucleic acids to target cells and/or organelles such as chloroplasts.

[0031] In one embodiment, homologous recombination is performed *in vitro*. In another embodiment, homologous recombination is performed *in vivo*. In still another embodiment, homologous recombination occurs in a yeast cell. In one embodiment, homologous recombination occurs in *Saccharomyces cerevisiae* or *Schizosaccharomyces pombe*. In yeast, the combination of efficient recombination processes and the availability of numerous selectable markers provides for rapid and complex engineering of target DNA sequences. Once all of the modifications are made to an ex vivo vector containing chloroplast genome DNA, the entire vector can be introduced into a chloroplast in a single transformation step. Thus, employing yeast technology enables the application of

metabolic engineering and/or synthetic biology to chloroplast genomes. One aspect of the present disclosure provides an isolated vector comprising a yeast element, a bacterial origin of replication, and at least about 135 kb or at least about 150 kb of plastid genomic DNA from a vascular plant. In some vectors, the yeast element is a yeast centromere, a yeast autonomous replicating sequence, or a combination thereof. The DNA may be from a vascular photosynthetic plant, for example soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes. In some embodiments, the genomic plastid DNA is modified, for example by insertion of a heterologous or homologous polynucleotide, deletion of one or more nucleic acid bases, mutation of one or more nucleic acid bases, rearrangement of one or more polynucleotides, or a combination thereof. In some instances, the modification is synthetic. Vectors of the present disclosure, when transformed into a host cell, may result in production of a product not naturally produced by the host cell. Some examples of such products include biomass-degrading enzymes, fatty acids, terpenes, terpenoids, therapeutic proteins and/or phytosterols. In some host cells, expression of the vector results in increased production of a product naturally produced by said host cell, for example, a biomass-degrading enzyme, a terpene, a terpenoid, a therapeutic protein and/or a phytosterol. The vectors of the present invention may further comprise one or more selection markers, for example, a yeast marker, a yeast antibiotic resistance marker, a bacterial marker, a bacterial antibiotic resistance marker, a plant marker, a plant antibiotic resistance marker or a combination thereof. Vectors of the present invention may also contain chloroplast genomic DNA which comprises 1) 1-200 genes; 2) all essential chloroplast genes; and/or 3) at least about 85%, 87%, 90%, 92%, or 95% of a chloroplast genome.

[0032] Also described herein is a host cell comprising the vectors described herein. Exemplary host cells may be naturally non-photosynthetic or photosynthetic and include, for example, *Saccharomyces cerevisiae*, *Escherichia coli*, soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes.

[0033] In another aspect, a method for producing a vector is provided where the method involves inserting targeting DNA into a vector—where the vector comprises a yeast centromere, a yeast autonomous replicating sequence, and a bacterial origin of replication, transforming an organism with the vector and capturing a portion of a chloroplast genome, thus producing a vector with a portion of a chloroplast genome. In some instances, the targeting DNA is chloroplast genomic DNA. This method may be used to capture a portion of a genome which is 10-400 kb in length. In some instances, the capturing step occurs by recombination. The captured portion of a chloroplast genome may be co-transformed into an organism with a vector, thus the recombination step may occur *in vivo*. Organisms used to practice methods disclosed herein may be eukaryotic and photosynthetic. In some instances, the organism is a vascular photosynthetic plant, for example soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes. Organisms used to practice methods disclosed herein may also be non-photosynthetic, for example yeast. In some instances, a non-photosynthetic organism may contain exogenous chloroplast DNA. In some embodiments, an additional step of modifying a portion of a chloroplast genome is utilized. A modification may be achieved through homologous

recombination. Such recombination may occur in an organism, for example a eukaryotic and/or photosynthetic organism. In some instances, the organism is a vascular plant, for example soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes. In other instances, the organism may be non-photosynthetic, for example a yeast. In embodiments with a modification step, the step may comprise addition of a polynucleotide, deletion of one or more nucleic acid bases, mutation of one or more nucleic acid bases, rearrangement or a polynucleotide, or combination thereof.

[0034] Further disclosed herein is an isolated vector comprising essential chloroplast genes, a selectable marker and a manipulation in one or more nucleic acids in the vector. In some instances, essential chloroplast genes are cloned from a vascular photosynthetic organism such as macroalgae, microalgae, *Ch. vulgaris*, *C. reinhardtii*, *D. salina*, *S. quadrangularis* or *H. pluvialis*. In other instances essential chloroplast genes are cloned from a vascular plant, for example, soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes. Essential chloroplast genes for use in the vectors described herein may be synthetic. The vectors described herein may further comprise an expression cassette, which may further comprise a region for integration into target DNA, for example organelle DNA. The vectors described herein may also contain one or more selection markers, for example, an auxotrophic marker, an antibiotic resistance marker, a chloroplast marker, or combinations thereof. In some instances, the essential chloroplast genes are those required for chloroplast function, photosynthesis, carbon fixation, production of one or more hydrocarbons, or a combination thereof. In other instances essential chloroplast genes are those necessary to render an organism photoautotrophic. Essential chloroplast genes may comprise up to 200 genes and/or consist of up to 400 kb. In some of the vectors described herein a manipulation in one or more nucleic acids is an addition, deletion, mutation, or rearrangement. In some instances, expression of the vector in a host cell produces a product not naturally produced by said host cell. In other instances, expression of a vector of the present invention results in an increase production of a product naturally produced by said host cell. Examples of such products are biomass degrading enzymes, fatty acids, terpenes, terpenoids, therapeutic proteins and/or phytosterols.

[0035] As described herein, one aspect provides a chloroplast comprising a vector of the present disclosure. In one aspect the chloroplast is an isolated chloroplast. In another aspect, a host cell comprising a vector of the present disclosure is provided. Host cells useful in the present disclosure may be naturally non-photosynthetic or naturally photosynthetic. Examples of organisms useful for the present invention include *Saccharomyces cerevisiae*, *Escherichia coli*, soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes.

[0036] In another aspect a method for transforming a plant or cell therefrom is provided where the method comprises inserting into said plant or cell therefrom with a vector comprising all essential chloroplast genes and optionally one or more genes not naturally occurring in said cell or organism. In some embodiments, the method further comprises the step of eliminating substantially all chloroplast genomes or disabling the photosynthetic capacity of said cell or organism prior to transformation. A cell or organism useful for this method may be photosynthetic, non-photosynthetic and/or

eukaryotic. In some instances, the vector for use in this method may also comprise an expression cassette and the expression cassette may be capable of integrating into non-nuclear DNA. In one embodiment the one or more genes not naturally occurring in the cell or organism is a gene in the isoprenoid pathway, MVA pathway, or MEP pathway. In another embodiment, the essential chloroplast genes are those that are required for chloroplast function, photosynthesis, carbon fixation, production of one or more hydrocarbons, or a combination thereof.

[0037] Further provided herein is a method for modifying an organism comprising the steps of transforming the organism, for example a vascular plant, with a vector comprising one or more polynucleotides sufficient to perform chloroplast function. In some instances, a vector useful for this method further comprises a sequence for production or secretion of a compound from said organism. In other instances, the vector comprises all essential chloroplast genes. In still other instances, the essential chloroplast genes are rearranged or mutated. An organism useful for some embodiments comprises essentially no chloroplast genome prior to transformation.

[0038] Yet another method provided herein is a method for making a product from an organism comprising the step of transforming said organism with a vector comprising at least 135 kb of genomic plastid DNA and one or more of the following: (i) a gene not naturally occurring in said organism; (ii) a deletion in a gene naturally occurring in said organism; (iii) a rearrangement of genes naturally occurring in said organism; and (iv) a mutation in a gene naturally occurring in said organism. In some instances, the organism is naturally photosynthetic. In other instances, the additional genes encode enzymes in the isoprenoid pathway, MVA pathway, or MEP pathway. In still another embodiment, the present disclosure provides a method for transforming a cell or organism comprising inserting into said cell or organism a chloroplast and a vector comprising all essential chloroplast genes.

[0039] The present disclosure also provides a method of producing an artificial chloroplast genome comprising the steps of: (a) providing a vector comprising one or more essential chloroplast genes; (b) adding to said vector a DNA fragment; (c) transforming a cell or organism with the vector produced by step (b); and (d) determining whether chloroplast function exists with said added DNA fragment. In some instances, the added DNA fragments comprises one or more coding regions for an enzyme in the isoprenoid, MVA or MEP pathway.

[0040] The present disclosure also provides a shuttle vector comprising at least about 85%, at least about 87%, at least about 90%, at least about 92%, or at least about 95% of a chloroplast genome. The genome may be modified by the insertion, deletion and/or rearrangement of one or more nucleotide sequences. Also provided herein is a vector comprising an isolated, functional chloroplast genome. A chloroplast genome useful in such a vector may be modified by the insertion, deletion and/or rearrangement of one or more nucleotide sequences.

[0041] Further provided herein is a method of producing an artificial chloroplast genome comprising the steps of: (a) providing a vector comprising all essential chloroplast genes; and (b) removing, adding, mutating, or rearranging DNA from the chloroplast genome. Such a method may further comprise the steps of transforming a redacted genome into a host organism; and (d) determining chloroplast function in

the host organism. In some instances, steps (b), (c), and (d) are repeated. In still other instances, the chloroplast genome is from an organism selected from the group consisting of: soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes. In other instances, the host organism is selected from the group consisting of: soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes. For some embodiments, the method may further comprise the step of removing redundant DNA from a chloroplast genome. In other embodiments, the vector comprises all or substantially all of a chloroplast genome. A chloroplast genome useful in the present disclosure may be cloned from a photosynthetic organism or may be a synthetic chloroplast genome. In some instances, the vector further comprises a gene not naturally occurring in the host organism, for example a gene from the isoprenoid pathway, MVA pathway, or MEP pathway.

[0042] Yet another method provided herein is a method of producing an artificial chloroplast genome comprising the steps of: (a) providing a vector comprising an entire chloroplast genome; (b) deleting a portion of said entire chloroplast genome; and (c) determining whether chloroplast function exists without said deleted portion. In another aspect of the present invention, a composition comprising an isolated and functional chloroplast genome is provided. In some instances, a composition comprises a modification to said chloroplast genome.

[0043] Further provided herein is an ex vivo vector comprising a nucleic acid comprising at least about 85%, at least about 87%, at least about 90%, at least about 92%, or at least about 95% of a chloroplast genome and a manipulation in one or more nucleic acids in the vector. In some instances, the nucleic acid is cloned from a photosynthetic vascular plant such as soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes. In some instances, the nucleic acid is synthetic. A vector may further comprise an expression cassette and an expression cassette may further comprise a region for integration into target DNA. In some instances, the target DNA is organelle DNA. A vector useful in the present invention may further comprise one or more selection markers, for example an auxotrophic marker, an antibiotic resistance marker, a chloroplast marker, or combinations thereof. In some embodiments, a manipulation in one or more nucleic acids in a vector may be an addition, deletion, mutation, or rearrangement. Expression of the vector may result in production of a product not naturally produced by a host cell and/or an increase production of a product naturally produced by a host cell. Examples of some products of the present invention include a terpene, terpenoid, fatty acid, biomass degrading enzyme, a therapeutic protein and/or a phytosterol.

[0044] Also provided herein is an ex vivo vector comprising a nucleic acid comprising at least about 135 kilobases or at least about 150 kilobases of a chloroplast genome and a manipulation in one or more nucleic acids in said vector. In some instances, the nucleic acid is cloned from a photosynthetic vascular plant such as soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes. In some instances, the nucleic acid is synthetic. A vector of the present disclosure may further comprise an expression cassette and an expression cassette may further comprise a region for integration into target DNA. In some instances, the target DNA is organelle DNA. A

vector useful in the present disclosure may further comprise one or more selection markers, for example an auxotrophic marker, an antibiotic resistance marker, a chloroplast marker, or combinations thereof. In some embodiments, a manipulation in one or more nucleic acids in a vector may be an addition, deletion, mutation, or rearrangement. Expression of the vector may result in production of a product not naturally produced by a host cell and/or an increase production of a product naturally produced by a host cell. Examples of some products of the present invention include a terpene, terpenoid, fatty acid, biomass degrading enzyme, a therapeutic protein and/or a phytosterol.

[0045] Further provided herein is a method of producing a vector containing a reconstructed genome, comprising: introducing two or more vectors into a host cell, wherein the vectors comprise fragments of a genome, recombining the vectors into a single vector comprising at least about 85%, at least about 87%, at least about 90%, at least about 92%, or at least about 95% of a genome, thereby producing a vector containing a reconstructed genome. In some instances, the host cell is eukaryotic, for example, a yeast such as *S. cerevisiae*. In other instances, the genome is an organelle genome. The organelle may be a chloroplast, for example a chloroplast from a vascular plant, particularly a plant such as soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes. In some instances, the two or more vectors comprise a selectable marker. In other instances, at least one of said fragments is synthetic. In still other instances, a further step comprising modifying a portion of the genome is useful in this method. Such a modification may comprise an addition, deletion, mutation, or rearrangement. In other embodiments, the modification is the addition of an exogenous nucleic acid which results in the production or increased production of a terpene, terpenoid, fatty acid, a biomass degrading enzyme, a therapeutic protein and/or a phytosterol.

[0046] Large DNA Cloning and Content

[0047] An advantage of this invention is that it provides for the cloning, manipulation, and delivery of a vector containing chloroplast genome DNA consisting of up to all chloroplast genes (or sequences). The chloroplast genome DNA contained in the vector can be obtained by recombination of a hybrid cloning vector with one contiguous fragment of DNA or by recombination of two or more contiguous fragments of DNA.

[0048] The methods and compositions disclosed herein may include captured and/or modified large pieces of DNA may comprise DNA from an organelle, such as mitochondrial DNA or plastid DNA (e.g., chloroplast DNA). The captured and/or modified large pieces of DNA may also comprise the entirety of an organelle's genome, e.g., a chloroplast genome. In other embodiments, the captured and/or modified large pieces of DNA comprise a portion of a chloroplast genome, for example 85%, 87%, 90%, 92%, 95% or more. A chloroplast genome may originate from any vascular or non-vascular plant, including algae, bryophytes (e.g., mosses, ferns), gymnosperms (e.g., conifers), and angiosperms (e.g., flowering plants—trees, grasses, herbs, shrubs). A chloroplast genome, or essential portions thereof, may comprise synthetic DNA, rearranged DNA, deletions, additions, and/or mutations. A chloroplast genome, or portions thereof, may comprise one or more deletions, additions, mutations, and/or rearrangements. The deletions, additions, mutations, and/or rearrangements may be naturally found in an organism, for

example a naturally occurring mutation, or may not be naturally found in nature. The chloroplast or plastid genomes of a number of organisms are widely available, for example, at the public database from the NCBI Organelle Genomes section available at http://www.ncbi.nlm.nih.gov/genomes/static/euk_o.html.

[0049] The target DNA sequence described herein may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more deletions, additions, mutations, and/or rearrangements as compared to a control sequence (naturally occurring sequence). Alternatively, the target DNA sequence described herein may comprise up to 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 deletions, additions, mutations, and/or rearrangements as compared to a control sequence. In some embodiments, the mutations may be functional or nonfunctional. For example, a functional mutation may have an effect on a cellular function when the mutation is present in a host cell as compared to a control cell without the mutation. A non-functional mutation may be silent in function, for example, there is no discernable difference in phenotype of a host cell without the mutation as compared to a cell with the mutation.

[0050] Captured and/or modified large pieces of DNA (e.g., target DNA), may comprise a minimal or minimized chloroplast genome (e.g., the minimum number of genes and/or DNA fragment, required for chloroplast functionality). The captured and/or modified DNA may comprise the essential chloroplast genes, it may comprise a portion or all, or substantially all of the essential chloroplast genes. An essential gene may be a gene that is essential for one or more metabolic processes or biochemical pathways. An essential gene may be a gene required for chloroplast function, such as photosynthesis, carbon fixation, or hydrocarbon production. An essential gene may also be a gene that is essential for gene expression, such as transcription, translation, or other process(es) that affect gene expression. The essential genes may comprise mutations or rearrangements. Essential genes may also comprise a minimally functional set of genes to perform a function. For example, a particular function (e.g., photosynthesis) may be performed inefficiently by a set of genes/gene products, however, this set would still comprise essential genes because the function is still performed. Thus, in one embodiment, substantially all essential chloroplast genes comprises the collection of gene needed to render the organism photoautotrophic.

[0051] Modified DNA may comprise at least 5, 10, 15, 20, 25, 30, 40, or 50 essential genes. Modified DNA may also comprise between 5 and 10, between 10 and 15, between 15 and 20, between 20 and 25, between 25 and 30, between 30 and 40, or between 40 and 50 essential genes. In some embodiments, the DNA may comprise essential chloroplast genomic sequence of up to about 135 kb or up to about 150 kb in length. The DNA may comprise essential chloroplast genes as well as non-essential chloroplast gene sequences. The DNA may be single stranded or double stranded, linear or circular, relaxed or supercoiled. The DNA may also be in the form of an expression cassette. For example, an expression cassette may comprise an essential gene to be expressed in a host cell. The expression cassette may comprise one or more essential genes as well as DNA sequences that promote the expression of the essential genes. The expression cassette may also comprise a region for integration into target DNA of a host. The expression cassette may also comprise one or more essential genes and one or more genes not naturally occurring in a host cell comprising the expression cassette.

One of ordinary skill in the arts will easily ascertain various combinations of the aforementioned aspects of the expression cassettes.

[0052] In other instances, captured and/or modified pieces of DNA may comprise the entire genome of a plastid or organelle. For example, about 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% of a plastid genome. In one embodiment the captured and/or modified large pieces of DNA may comprise 10-100%, 20-100%, 30-100%, 40-100%, 50-100%, 60-100%, 70-100%, 80-100%, or 90-100% of the entire genome of a plastid or cell.

[0053] In still other instances, the captured and/or modified large pieces of DNA may comprise about 10 kb, 11 kb, 12 kb, 13 kb, 14 kb, 15 kb, 16 kb, 17 kb, 18 kb, 19 kb, 20 kb, 21 kb, 22 kb, 23 kb, 24 kb, 25 kb, 26 kb, 27 kb, 28 kb, 29 kb, 30 kb, 31 kb, 32 kb, 33 kb, 34 kb, 35 kb, 36 kb, 37 kb, 38 kb, 39 kb, 40 kb, 41 kb, 42 kb, 43 kb, 44 kb, 45 kb, 46 kb, 47 kb, 48 kb, 49 kb, 50 kb, 51 kb, 52 kb, 53 kb, 54 kb, 55 kb, 56 kb, 57 kb, 58 kb, 59 kb, 60 kb, 61 kb, 62 kb, 63 kb, 64 kb, 65 kb, 66 kb, 67 kb, 68 kb, 69 kb, 70 kb, 71 kb, 72 kb, 73 kb, 74 kb, 75 kb, 76 kb, 77 kb, 78 kb, 79 kb, 80 kb, 81 kb, 82 kb, 83 kb, 84 kb, 85 kb, 86 kb, 87 kb, 88 kb, 89 kb, 90 kb, 91 kb, 92 kb, 93 kb, 94 kb, 95 kb, 96 kb, 97 kb, 98 kb, 99 kb, 100 kb, 101 kb, 102 kb, 103 kb, 104 kb, 105 kb, 106 kb, 107 kb, 108 kb, 109 kb, 110 kb, 111 kb, 112 kb, 113 kb, 114 kb, 115 kb, 116 kb, 117 kb, 118 kb, 119 kb, 120 kb, 121 kb, 122 kb, 123 kb, 124 kb, 125 kb, 126 kb, 127 kb, 128 kb, 129 kb, 130 kb, 131 kb, 132 kb, 133 kb, 134 kb, 135 kb, 136 kb, 137 kb, 138 kb, 139 kb, 140 kb, 141 kb, 142 kb, 143 kb, 144 kb, 145 kb, 146 kb, 147 kb, 148 kb, 149 kb, 150 kb, 151 kb, 152 kb, 153 kb, 154 kb, 155 kb, 156 kb, 157 kb, 158 kb, 159 kb, 160 kb, 161 kb, 162 kb, 163 kb, 164 kb, 165 kb, 166 kb, 167 kb, 168 kb, 169 kb, 170 kb, 171 kb, 172 kb, 173 kb, 174 kb, 175 kb, 176 kb, 177 kb, 178 kb, 179 kb, 180 kb, 181 kb, 182 kb, 183 kb, 184 kb, 185 kb, 186 kb, 187 kb, 188 kb, 189 kb, 190 kb, 191 kb, 192 kb, 193 kb, 194 kb, 195 kb, 196 kb, 197 kb, 198 kb, 199 kb, 200 kb, 201 kb, 202 kb, 203 kb, 204 kb, 205 kb, 206 kb, 207 kb, 208 kb, 209 kb, 210 kb, 211 kb, 212 kb, 213 kb, 214 kb, 215 kb, 216 kb, 217 kb, 218 kb, 219 kb, 220 kb, 221 kb, 222 kb, 223 kb, 224 kb, 225 kb, 226 kb, 227 kb, 228 kb, 229 kb, 230 kb, 231 kb, 232 kb, 233 kb, 234 kb, 235 kb, 236 kb, 237 kb, 238 kb, 239 kb, 240 kb, 241 kb, 242 kb, 243 kb, 244 kb, 245 kb, 246 kb, 247 kb, 248 kb, 249 kb, 50 kb, 51 kb, 252 kb, 253 kb, 254 kb, 255 kb, 256 kb, 257 kb, 258 kb, 259 kb, 260 kb, 261 kb, 262 kb, 263 kb, 264 kb, 265 kb, 266 kb, 267 kb, 268 kb, 269 kb, 270 kb, 271 kb, 272 kb, 273 kb, 274 kb, 275 kb, 276 kb, 277 kb, 278 kb, 279 kb, 280 kb, 281 kb, 282 kb, 283 kb, 284 kb, 285 kb, 286 kb, 287 kb, 288 kb, 289 kb, 290 kb, 291 kb, 292 kb, 293 kb, 294 kb, 295 kb, 296 kb, 297 kb, 298 kb, 299 kb, 300 kb, 301 kb, 302 kb, 303 kb, 304 kb, 305 kb, 306 kb, 307 kb, 308 kb, 309 kb, 310 kb, 311 kb, 312 kb, 313 kb, 314 kb, 315 kb, 316 kb, 317 kb, 318 kb, 319 kb, 320 kb, 321 kb, 322 kb, 323 kb, 324 kb, 325 kb, 326 kb, 327 kb, 328 kb, 329 kb, 330 kb, 331 kb, 332 kb, 333 kb, 334 kb, 335 kb, 336 kb, 337 kb, 338 kb, 339 kb, 340 kb, 341 kb, 342 kb, 343 kb, 344 kb, 345 kb, 346 kb, 347 kb, 348 kb, 349 kb, 350 kb, 351 kb, 352 kb, 353 kb, 354 kb, 355 kb, 356 kb, 357 kb,

358 kb, 359 kb, 360 kb, 361 kb, 362 kb, 363 kb, 364 kb, 365 kb, 366 kb, 367 kb, 368 kb, 369 kb, 370 kb, 371 kb, 372 kb, 373 kb, 374 kb, 375 kb, 376 kb, 377 kb, 378 kb, 379 kb, 380 kb, 381 kb, 382 kb, 383 kb, 384 kb, 385 kb, 386 kb, 387 kb, 388 kb, 389 kb, 390 kb, 391 kb, 392 kb, 393 kb, 394 kb, 395 kb, 396 kb, 397 kb, 398 kb, 399 kb, 400 kb or more genomic (e.g., nuclear or organelle) DNA. In some embodiments the captured and or modified large pieces of DNA may comprise about 10-400 kb, 50-350 kb, 100-300 kb, 100-200 kb, 200-300 kb, 150-200 kb, 200-250 kb genomic DNA.

[0054] In still other instances, the captured and or modified large pieces of DNA may comprise about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 50, 51, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300 or more open reading frames, partial open reading frames, pseudogenes and/or repeating sequences.

[0055] The present disclosure also provides vectors comprising a cassette-able chloroplast genome or portion thereof (e.g., a removable DNA fragment comprising a chloroplast genome or functional portion thereof). A vector of the present invention may comprise functional chloroplast units (e.g., a unit essential for metabolic processes, photosynthesis, gene expression, photosystem I, photosystem II). Vectors of the present invention may comprise a transplantable chloroplast genome or portion thereof. Additionally, the vectors of the present invention may comprise a transferable chloroplast genome or portion thereof. In other embodiments, the vectors comprise: 1) one or more large pieces of modified DNA; 2) all genes necessary to carry out photosynthesis; 3) all genes required for chloroplast survival and/or function; 4) essential chloroplast genes; and/or 5) sufficient naturally occurring or modified chloroplast genes to perform one or more chloroplast functions, such as photosynthesis. A vector may comprise a portion, substantially all, or all of the essential chloroplast genes. A vector may comprise at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100 or more essential genes.

[0056] A vector may comprise chloroplast DNA of 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250 kb or more in length. A vector may comprise essential chloroplast genes as well as non-essential chloroplast gene sequences. A vector may comprise one or more, or all, essential chloroplast genes and/or one or more genes not naturally occurring in a host cell comprising a vector. In some embodiments, a vector may

comprise chloroplast genes and genes not naturally occurring in the chloroplast. A vector may comprise one or more essential chloroplast genes and/or one or more DNA sequences or genes involved in chloroplast function, photosynthesis, carbon fixation, and/or hydrocarbon production. For example, a vector may comprise a sequence required for photosynthesis and a sequence involved in the isoprenoid production, MVA, and/or MEP pathways, such as a DNA sequence encoding a terpene synthase, or other polypeptide that produces a hydrocarbon, such as a terpene or isoprenoid. The invention further provides methods for cloning, manipulating and delivering a large target nucleic acid to a cell or particle, such as, for example, yeasts or bacteria. Certain embodiments make use of a hybrid yeast-bacteria cloning system (See, e.g., U.S. Pat. Nos. 5,866,404 and 7,083,971 and Hokanson et al., (2003) *Human Gene Ther.*: 14: 329-339). The vectors herein (e.g., cloning system) are comprised of a shuttle vector that contains elements for function and replication in both yeast and bacteria, allowing them to stably function and replicate in either organism. This composition of functional and replicative elements yields a hybrid system which enjoys the benefits of both genetic engineering systems. The genetics of yeast (e.g., *S. cerevisiae*) are well understood and a powerful assortment of molecular biology tools exists for genetic engineering in bacteria.

[0057] Another aspect produces a gap-filled vector by homologous recombination among the two arms and the target nucleic acid. In still another embodiment, at least one arm further comprises an origin of replication. In another embodiment, each arm further comprises a rare restriction endonuclease recognition site. Homologous recombination may be performed in vitro or in vivo, for example, in a fungal cell (e.g., *S. cerevisiae*, *Sz. pombe* or *U. maydis*). Also provided is a eukaryotic host cell harboring the recombinational cloning system or vector according to the present disclosure, for example, in a fungal cell (e.g., *S. cerevisiae*, *Sz. pombe* or *U. maydis*).

[0058] A gap-filled linear vector may be converted to a circular vector in vitro (e.g. using T4 ligase) or in vivo, for example, in a bacterium. The circular vectors of interest can be amplified, purified, cut and used to recover sufficient amounts of DNA to be introduced either directly into a cell or into a suitable delivery system for subsequent delivery to a target cell. The methodology offers great versatility to clone and to modify any large bacterial or non-bacterial genome, and easily facilitate the use thereof as recombinational vectors. Direct delivery of a gap-filled vector into a cell may be performed by methods well known in the field such as, for example, calcium phosphate transformation methodologies or electroporation (see Sambrook et al., supra).

[0059] Accordingly, provided herein is a method for producing a recombinant delivery unit including the steps of: (a) producing a gap-filled vector containing a target sequence; (b) optionally circularizing the gap-filled vector segments of the invention; (c) propagating the vector; and (d) introducing the gap-filled vector in a delivery unit.

[0060] Bacterial systems are useful for amplifying and purifying DNA, and for functionally testing the genetic modifications and their effect on pathways. One embodiment of the present invention provides a cloning system that will aid in the cloning and modifying of any large genome and easily facilitate the cloning and introduction of pathways. With the

ability to deliver whole pathways, certain embodiments of the present invention allow for a system biological approach to problem solving.

[0061] In general, target DNA (e.g., genomic DNA) may be captured by creating sites allowing for homologous recombination in the vector. For example, such sites may be created by, but not limited to, gap-repair cloning, wherein a gap is created in the vector, usually by restrictive enzyme digestion prior to transformation into the yeast. When the target DNA is mixed with the vector, the target DNA is recombined into the vector. This operation is called "gap filling." This recombination can occur in bacteria, yeast, the original host organism, another organism, or *in vitro*. In some embodiments, recombination is performed in yeast because of the high rate of homologous recombination. Once captured, the target DNA can be modified in many ways including adding, altering, or removing DNA sequences. In some embodiments, the target DNA is genomic DNA. In other embodiments, the target DNA is organelle (e.g., mitochondria or chloroplast) DNA.

[0062] In some embodiments, target DNA is modified by adding, altering or removing genes, coding sequences, partial coding sequences, regulatory elements, positive and/or negative selection markers, recombination sites, restriction sites, and/or codon bias sites. For example, the target DNA sequence may be codon biased for expression in the organism being transformed. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell or organelle in usage of nucleotide codons to specify a given amino acid. Without being bound by theory, by using preferred codons, the rate of translation may be greater. Therefore, when synthesizing a gene for improved expression, it may be desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage. The codons of the present invention may be A/T rich, for example, A/T rich in the third nucleotide position of the codons. In some embodiments, at least 50% of the third nucleotide position of the codons are A or T. In other embodiments, at least 60%, 70%, 80%, 90%, or 99% of the third nucleotide position of the codons are A or T. (see also U.S. Publication No. 2004/0014174 and U.S. Pat. No. 5,545,817).

[0063] Such manipulations are well known in the art and can be performed in numerous ways. In some embodiments, the modifications may be performed using cloned sequences. In other embodiments, the modifications may be performed using synthetic DNA.

[0064] Genetic manipulations include cloning large pieces of target DNA (e.g., chromosomes, genomes) and/or dividing and reorganizing target DNA based on functional relations between genes, such as metabolic pathways or operons. Genetic manipulations also include introducing and removing metabolic pathways, recombining DNA into functional units (e.g., metabolic pathways, synthetic operons), and/or determining sites of instability in large pieces of DNA (e.g., sites where a native or non-native host tends to delete or recombine a sequence of DNA).

[0065] Target DNA may be DNA from a prokaryote. Target DNA may also be genomic DNA, mitochondrial DNA, or chloroplast DNA from a eukaryote. Examples of such organisms from which genomic and/or organelle DNA may serve as target DNA include, but are not limited to vascular plants and more specifically soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes. One of skill in the art will recognize that these organisms are listed only as examples and that the methods

disclosed herein are applicable to the large DNA from any organism, including bacteria, plants, fungi, protists, and animals. Genetic manipulations of the present invention may include stabilizing large pieces of DNA by removing or inserting sequences that force transformed cells to preserve certain sequences of DNA and to stably maintain the sequences in its progeny. Genetic manipulations may also include altering codons of the target DNA, vector DNA, and/or synthetic DNA to reflect any codon bias of the host organism. Additionally, genetic manipulations may include determining the minimal set of genes required for an organism to be viable. In another embodiment, the genetic manipulations include determining the minimal set of genes required for a certain metabolic pathway to be created or maintained.

[0066] The genetic manipulations of the present disclosure may include determining redundant genes both within a genome, and between two genomes (e.g., redundancy between the nuclear and chloroplast genome). Additionally, the genetic manipulations may include determining a functional sequence of DNA that could be artificially synthesized (e.g. the genes in a certain metabolic pathway, the genes of a functional genome). In another embodiment, the genetic manipulations include creating DNA and genomes packaged into cassettes (e.g., sites within a vector where genes can be easily inserted or removed). The genetic manipulations may also include creating a nuclear or organelle genome that is viable in multiple species (e.g. a transplantable chloroplast genome). Furthermore, the genetic manipulations may include a method for testing the viability of any of these manipulations or creations (e.g., transferring a shuttle vector back into a host system and assaying for survival).

[0067] Vectors, Markers and Transformation

[0068] A vector or other recombinant nucleic acid molecule may include a nucleotide sequence encoding a selectable marker. The term or "selectable marker" or "selection marker" refers to a polynucleotide (or encoded polypeptide) that confers a detectable phenotype. A selectable marker generally encodes a detectable polypeptide, for example, a green fluorescent protein or an enzyme such as luciferase, which, when contacted with an appropriate agent (a particular, wavelength of light or luciferin, respectively) generates a signal that can be detected by eye or using appropriate instrumentation (Giacomin, *Plant Sci.* 116:59-72, 1996; Scikantha, *J. Bacteriol.* 178:121, 1996; Gerdes, *FEBS Lett.* 389:44-47, 1996; see, also, Jefferson, *EMBO J.* 6:3901-3907, 1997, fl-glucuronidase). A selectable marker generally is a molecule that, when present or expressed in a cell, provides a selective advantage (or disadvantage) to the cell containing the marker, for example, the ability to grow in the presence of an agent that otherwise would kill the cell.

[0069] A selectable marker can provide a means to obtain prokaryotic cells, yeast cell, plant cells or any combination that express the marker and, therefore, can be useful as a component of a vector of the present disclosure (see, for example, Bock, *supra*, 2001). Examples of selectable markers include, but are not limited to, those that confer antimetabolite resistance, for example, dihydrofolate reductase, which confers resistance to methotrexate (Reiss, *Plant Physiol. (Life Sci. Adv.)* 13:143-149, 1994); neomycin phosphotransferase, which confers resistance to the aminoglycosides neomycin, kanamycin and paromycin (Herrera-Estrella, *EMBO J.* 2:987-995, 1983), hygro, which confers resistance to hygromycin (Marsh, *Gene* 32:481-485, 1984), trpB, which allows cells to utilize indole in place of tryptophan; hisD, which

allows cells to utilize histinol in place of histidine (Hartman, *Proc. Natl. Acad. Sci., USA* 85:8047, 1988); mannose-6-phosphate isomerase which allows cells to utilize mannose (WO 94/20627); ornithine decarboxylase, which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine (DFMO; McConlogue, 1987, In: Current Communications in Molecular Biology, Cold Spring Harbor Laboratory ed.); and deaminase from *Aspergillus terreus*, which confers resistance to Blasticidin S (Tamura, *Biosci. Biotechnol. Biochem.* 59:2336-2338, 1995). Additional selectable markers include those that confer herbicide resistance, for example, phosphinothricin acetyltransferase gene, which confers resistance to phosphinothricin (White et al., *Nucl. Acids Res.* 18:1062, 1990; Spencer et al., *Theor. Appl. Genet.* 79:625-631, 1990), a mutant EPSPV-synthase, which confers glyphosate resistance (Hinchee et al., *BioTechnology* 91:915-922, 1998), a mutant acetolactate synthase, which confers imidazolione or sulfonylurea resistance (Lee et al., *EMBO J.* 7:1241-1248, 1988), a mutant psbA, which confers resistance to atrazine (Smeda et al., *Plant Physiol.* 103:911-917, 1993), or a mutant protoporphyrinogen oxidase (see U.S. Pat. No. 5,767,373), or other markers conferring resistance to an herbicide such as glufosinate. Selectable markers include polynucleotides that confer dihydrofolate reductase (DHFR) or neomycin resistance for eukaryotic cells and tetracycline; ampicillin resistance for prokaryotes such as *E. coli*; and bleomycin, gentamycin, glyphosate, hygromycin, kanamycin, methotrexate, phleomycin, phosphinothricin, spectinomycin, streptomycin, sulfonamide and sulfonylurea resistance in plants (see, for example, Maliga et al., Methods in Plant Molecular Biology, Cold Spring Harbor Laboratory Press, 1995, page 39).

[0070] Methods for nuclear and plastid transformation are routine and well known for introducing a polynucleotide into a plant cell chloroplast (see U.S. Pat. Nos. 5,451,513, 5,545,817, and 5,545,818; WO 95/16783; McBride et al., *Proc. Natl. Acad. Sci., USA* 91:7301-7305, 1994). In some embodiments, chloroplast transformation involves introducing regions of chloroplast DNA flanking a desired nucleotide sequence, allowing for homologous recombination of the exogenous DNA into the target chloroplast genome. In some instances one to 1.5 kb flanking nucleotide sequences of chloroplast genomic DNA may be used. Using this method, point mutations in the chloroplast 16S rRNA and rps12 genes, which confer resistance to spectinomycin and streptomycin, can be utilized as selectable markers for transformation (Svab et al., *Proc. Natl. Acad. Sci., USA* 87:8526-8530, 1990), and can result in stable homoplasmic transformants, at a frequency of approximately one per 100 bombardments of target leaves.

[0071] Microprojectile mediated transformation also can be used to introduce a polynucleotide into a plant cell chloroplast (Klein et al., *Nature* 327:70-73, 1987). This method utilizes microprojectiles such as gold or tungsten, which are coated with the desired polynucleotide by precipitation with calcium chloride, spermidine or polyethylene glycol. The microprojectile particles are accelerated at high speed into a plant tissue using a device such as the BIOLISTIC PDS-1000 particle gun (BioRad; Hercules Calif.). Methods for the transformation using biolistic methods are well known in the art (see, e.g.; Christou, *Trends in Plant Science* 1:423-431, 1996). Microprojectile mediated transformation has been used, for example, to generate a variety of transgenic plant species, including cotton, tobacco, corn, hybrid poplar and

papaya. Important cereal crops such as wheat, oat, barley, sorghum and rice also have been transformed using micro-projectile mediated delivery (Duan et al., *Nature Biotech.* 14:494-498, 1996; Shimamoto, *Curr. Opin. Biotech.* 5:158-162, 1994). The transformation of most dicotyledonous plants is possible with the methods described above. Transformation of monocotyledonous plants also can be transformed using, for example, biolistic methods as described above, protoplast transformation, electroporation of partially permeabilized cells, introduction of DNA using glass fibers, the glass bead agitation method, and the like.

[0072] Transformation frequency may be increased by replacement of recessive rRNA or r-protein antibiotic resistance genes with a dominant selectable marker, including, but not limited to the bacterial aadA gene (Svab and Maliga, *Proc. Natl. Acad. Sci., USA* 90:913-917, 1993). Approximately 15 to 20 cell division cycles following transformation are generally required to reach a homoplasmidic state. It is apparent to one of skill in the art that a chloroplast may contain multiple copies of its genome, and therefore, the term "homoplasmic" or "homoplasmy" refers to the state where all copies of a particular locus of interest are substantially identical. Plastid expression, in which genes are inserted by homologous recombination into all of the several thousand copies of the circular plastid genome present in each plant cell, takes advantage of the enormous copy number advantage over nuclear-expressed genes to permit expression levels that can readily exceed 10% of the total soluble plant protein.

[0073] Any of the nucleotide sequences of target DNA, vector DNA, or synthetic DNA in the vectors disclosed herein can further include codons biased for expression of the nucleotide sequences in the organism transformed. In some instances, codons in the nucleotide sequences are A/T rich in a third nucleotide position of the codons. For example, at least 50% of the third nucleotide position of the codons may be A or T. In other instances, the codons are G/C rich, for example at least 50% of the third nucleotide positions of the codons may be G or C.

[0074] The nucleotide sequences of the shuttle vectors of the present disclosure can be adapted for chloroplast expression. For example, a nucleotide sequence herein can comprise a chloroplast specific promoter or chloroplast specific regulatory control region.

[0075] In embodiments where a vector encodes genes capable of fuel production, fuel products are produced by altering the enzymatic content of the cell to increase the biosynthesis of specific fuel molecules. For example, nucleotides sequences (e.g., an ORF isolated from an exogenous source) encoding biosynthetic enzymes can be introduced into the chloroplast of a photosynthetic organism. Nucleotide sequences encoding fuel biosynthetic enzymes can also be introduced into the nuclear genome of the photosynthetic organisms. Nucleotide sequences introduced into the nuclear genome can direct accumulation of the biosynthetic enzyme in the cytoplasm of the cell, or may direct accumulation of the biosynthetic enzyme in the chloroplast of the photosynthetic organism.

[0076] Any of the nucleotide sequences herein may further comprise a regulatory control sequence. Regulatory control sequences can include one or more of the following: a promoter, an intron, an exon, processing elements, 3' untranslated region, 5' untranslated region, RNA stability elements, or translational enhancers. A promoter may be one or more of the following: a promoter adapted for expression in the organ-

ism, an algal promoter, a chloroplast promoter, and a nuclear promoter, any of which may be a native or synthetic promoters. A regulatory control sequence can be inducible or auto-regulatable. A regulatory control sequence can include autologous and/or heterologous sequences. In some cases, control sequences can be flanked by a first homologous sequence and a second homologous sequence. The first and second homologous sequences can each be at least 500 nucleotides in length. The homologous sequences can allow for either homologous recombination or can act to insulate the heterologous sequence to facilitate gene expression.

[0077] Vectors may also comprise sequences involved in producing products useful as biopharmaceuticals, such as, but not limited to, antibodies (including functional portions thereof), interleukins and other immune modulators, and antibiotics. See, e.g., Mayfield et al., (2003) *Proc. Nat'l Acad. Sci.*: 100 (438-42) and U.S. Pub. No. 2004/0014174.

[0078] Vectors may comprise a cassette-able bacterial genome or portion thereof (e.g., a removable DNA fragment comprising a bacterial genome or functional portion thereof). Additionally, vectors may comprise functional genomic units (e.g., a unit essential for metabolic processes, biochemical pathways, gene expression). Vectors may also comprise a transplantable bacterial genome or portion thereof. Vectors may comprise a transferable bacterial genome or portion thereof.

[0079] In some embodiments, the large piece of target DNA is modified. The modified DNA may comprise all genes necessary to carry out ethanologenesis, all genes required for the Entner-Duodorff pathway, the glucose tolerance pathway, the ethanol tolerance pathway, the carboxylic acid byproduct resistance pathway, the acetic acid tolerance pathway, the sugar transport pathway, sugar fermentation pathways, and/or the cellulose and hemicellulose digestive pathways.

[0080] Hybrid cloning systems and methods of the invention combine the high versatility of yeast as a system for the capture and manipulation of a given nucleic acid and the high efficiency of bacterial systems for the amplification of such nucleic acid. Recombinational vectors relying on homologous recombination to mediate the isolation, manipulation and delivery of large nucleic acid fragments were constructed. Also described herein are methods for using such recombinational cloning vectors to clone, to manipulate and to deliver large nucleic acids. Additionally, this disclosure provides methods for using such recombinational cloning systems as potentiaters of biochemical pathway analysis, organelle analysis, and synthetic chloroplast construction.

[0081] The vectors herein may be introduced into yeast. The yeast may be a suitable strain of *Saccharomyces cerevisiae*; however, other yeast models may be utilized. Introduction of vectors into yeast may allow for genetic manipulation of the vectors. Yeast vectors have been described extensively in the literature and methods of manipulating the same also are well known as discussed hereinafter (see e.g., Ketner et al. (1994) *Proc. Natl. Acad. Sci. (USA)* 91:6186 6190).

[0082] Following genetic manipulation, the present system allows for the transition to a bacterial environment, suitable for the preparation of larger quantities of nucleic acids. Also provided is a shuttle vector comprising a yeast selectable marker, a bacterial selectable marker, a telomere, a centromere, a yeast origin of replication, and/or a bacterial origin of replication.

[0083] Shuttle vectors described herein may enable homologous recombination in yeast to capture and to inte-

grate in a vector of interest a target nucleic acid of interest. Shuttle vectors may allow for the manipulation of target DNA in any of the hosts to which the vectors can be introduced. In some embodiments, after desired manipulations, shuttle vector components may be removed, leaving just the modified target DNA. Such extraction of vector sequences can be performed using standard methodologies and may occur in any host cell. The target nucleic acid of interest can be a large nucleic acid, and can include, for example, a vector, such as a viral vector, including the foreign gene of interest contained therein. The target nucleic acid can also be a bacterial (including archaebacteria and eubacteria), viral, fungal, protist, plant or animal genome, or a portion thereof. For example, a target nucleic acid of the present disclosure may comprise the chloroplast genome of a eukaryotic organism.

[0084] Shuttle vectors according to the disclosure may comprise an appropriately oriented DNA that functions as a telomere in yeast and a centromere. Any suitable telomere may be used. Suitable telomeres include without limitation telomeric repeats from many organisms, which can provide telomeric function in yeast. The terminal repeat sequence in humans (TTAGGG)_N, is identical to that in trypanosomes and similar to that in yeast ((TG)₁₋₃)_N and Tetrahymena (TTGGG)_N (Szostak & Blackburn (1982) *Cell* 29:245 255; Brown (1988) *EMBO J.* 7:2377 2385; and Moyzis et al. (1988) *Proc. Natl. Acad. Sci.* 85:6622 6626).

[0085] The term "centromere" is used herein to identify a nucleic acid, which mediates the stable replication and precise partitioning of the vectors at meiosis and at mitosis thereby ensuring proper segregation into daughter cells. Suitable centromeres include, without limitation, the yeast centromere, CEN4, which confers mitotic and meiotic stability on large linear plasmids (Murray & Szostak (1983) *Nature* 305:189 193; Carbon (1984) *Cell* 37:351 353; and Clark et al. (1990) *Nature* 287:504 509)).

[0086] In some embodiments, at least one of the two segments of the circular vector according to the disclosure includes at least one replication system that is functional in a host cell/particle of choice. As it will become apparent hereinafter, one of skill will realize that the manipulation, amplification and/or delivery of a target nucleic acid of choice may entail the use of more than one host cell/particle. Accordingly, more than one replication system functional in each host cell/particle of choice may be included.

[0087] When a host cell(s) is a prokaryote, particularly *E. coli*, replication system(s) include those which are functional in prokaryotes, such as, for example, P1 plasmid replicon, ori, P1 lytic replicon, ColE1, BAC, single copy plasmid F factors and the like. Either one or both segments, and/or the circular vector, may further include a yeast origin of replication capable of supporting the replication of large nucleic acids. Non-limiting examples of replication regions according to the invention include the autonomously replicating sequence or "ARS element." ARS elements were identified as yeast sequences that conferred high-frequency transformation. *Tetrahymena* DNA termini have been used as ARS elements in yeast along with ARS1 and ARS4 (Kiss et al. (1981) *Mol. Cell. Biol.* 1:535 543; Stinchcomb et al. (1979) *Nature* 282:39; and Barton & Smith (1986) *Mol. Cell. Biol.* 6:2354). For each segment (e.g., those corresponding to the yeast and bacterial elements of the gap-filling shuttle vector) there may be two or more origins of replication.

[0088] The first and/or the second segment according to an aspect of the disclosure may be joined in a circularized vector

form (e.g., plasmid form). Circularization can occur in vivo or in vitro using the segment of interest. Alternatively, a circular vector of interest can be used. As used herein, the term "vector" designates a plasmid or phage DNA or other nucleic acid into which DNA or other nucleic acid may be cloned. The vector may replicate autonomously in a host cell and may be characterized further by one or a small number of restriction endonuclease recognition sites at which such nucleic acids may be cut in a determinable fashion and into which nucleic acid fragments may be inserted. The vector further may contain one or more selectable markers suitable for the identification of cells transformed with the vector.

[0089] Target nucleic acids of the invention may vary considerably in complexity. The target nucleic acid may include viral, prokaryotic or eukaryotic DNA, cDNA, exonic (coding), and/or intronic (noncoding) sequences. Hence, the target nucleic acid may include one or more genes. A target nucleic acid may be a chromosome, genome, or operon and/or a portion of a chromosome, genome or operon. A target nucleic acid may comprise coding sequences for all the genes in a pathway, the minimum complement of genes necessary for survival of an organelle, and/or the minimum complement of genes necessary for survival of an organism. A target nucleic acid may comprise *Zymomonas mobilis* DNA sequence, including, but not limited to genomic DNA and/or cDNA. A target nucleic acid may comprise eukaryotic chloroplast DNA sequence, including but not limited to, chloroplast genome DNA and/or cDNA. A target nucleic acid may comprise cyanobacteria DNA, including but not limited to genomic DNA and/or cDNA. The target nucleic acid also may be of any origin and/or nature.

[0090] It may be desirable for the gene to also comprise a promoter operably linked to the coding sequence in order to effectively promote transcription. Enhancers, repressors and other regulatory sequences may also be included in order to modulate activity of the gene, as is well known in the art. A gene as provided herein can refer to a gene that is found in the genome of the individual host cell (i.e., endogenous) or to a gene that is not found in the genome of the individual host cell (i.e., exogenous or a "foreign gene"). Foreign genes may be from the same species as the host or from different species. For transfection of a cell using DNA containing a gene with the intent that the gene will be expressed in the cell, the DNA may contain any control sequences necessary for expression of the gene in the required orientation for expression. The term "intron" as used herein, refers to a DNA sequence present in a given gene which is not translated into protein and is generally found between exons.

[0091] Genetic elements, or polynucleotides comprising a region that encodes a polypeptide or a region that regulates transcription or translation or other processes important to expression of the polypeptide in a host cell, or a polynucleotide comprising both a region that encodes a polypeptide and a region operably linked thereto that regulates expression. Genetic elements may be comprised within a vector that replicates as an episomal element; that is, as a molecule physically independent of the host cell genome. They may be comprised within mini-chromosomes, such as those that arise during amplification of transfected DNA by methotrexate selection in eukaryotic cells. Genetic elements also may be comprised within a host cell genome; not in their natural state but, rather, following manipulation such as isolation, cloning and introduction into a host cell in the form of purified DNA or in a vector, among others.

[0092] Vectors of the present disclosure may contain sufficient linear identity or similarity (homology) to have the ability to hybridize to a portion of a target nucleic acid made or which is single-stranded, such as a gene, a transcriptional control element or intragenic DNA. Without being bound to theory, such hybridization is ordinarily the result of base-specific hydrogen bonding between complementary strands, preferably to form Watson-Crick base pairs. As a practical matter, such homology can be inferred from the observation of a homologous recombination event. In some embodiments, such homology is from about 8 to about 1000 bases of the linear nucleic acid. In other embodiments, such homology is from about 12 to about 500 bases. One skilled in the art will appreciate that homology may extend over longer stretches of nucleic acids.

[0093] Homologous recombination is a type of genetic recombination, a process of physical rearrangement occurring between two strands of DNA. Homologous recombination involves the alignment of similar sequences, a crossover between the aligned DNA strands, and breaking and repair of the DNA to produce an exchange of material between the strands. The process homologous recombination naturally occurs in organisms and is also utilized as a molecular biology technique for introducing genetic changes into organism.

[0094] The vectors disclosed herein may be modified further to include functional entities other than the target sequence which may find use in the preparation of the construct(s), amplification, transformation or transfection of a host cell, and—if applicable—for integration in a host cell. For example, the vector may comprise regions for integration into host DNA. Integration may be into nuclear DNA of a host cell. In some embodiments, integration may be into non-nuclear DNA, such as chloroplast DNA. Other functional entities of the vectors may include, but are not limited to, markers, linkers and restriction sites.

[0095] A target nucleic acid may include a regulatory nucleic acid. This refers to any sequence or nucleic acid which modulates (either directly or indirectly, and either up or down) the replication, transcription and/or expression of a nucleic acid controlled thereby. Control by such regulatory nucleic acid may make a nucleic acid constitutively or inducibly transcribed and/or translated. Any of the nucleotide sequences herein may further comprise a regulatory control sequence. Examples of regulatory control sequences can include, without limitation, one or more of the following: a promoter, an intron, an exon, processing elements, 3' untranslated region, 5' untranslated region, RNA stability elements, or translational enhancers. A promoter may be one or more of the following: a promoter adapted for expression in the organism (e.g., bacterial, fungal, viral, plant, mammalian, or protist), an algal promoter, a chloroplast (or other plastid) promoter, a mitochondrial promoter, and a nuclear promoter, any of which may be a native or synthetic promoters. A regulatory control sequence can be inducible or autoregulatable. A regulatory control sequence can include autologous and/or heterologous sequences. In some cases, control sequences can be flanked by a first homologous sequence and a second homologous sequence. The first and second homologous sequences can each be at least 500 nucleotides in length. The homologous sequences can allow for either homologous recombination or can act to insulate the heterologous sequence to facilitate gene expression.

[0096] In some instances, target DNA, vector DNA or other DNA present in a shuttle vector of the present disclosure does

not result in production of a polypeptide product but rather allows for secretion of the product from the cell. In these cases, the nucleotide sequence may encode a protein that enhances or initiates or increases the rate of secretion of a product from an organism to the external environment.

[0097] One embodiment provides a method of producing a gap-filled vector. A gap-filled vector may undergo homologous recombination and insertion of a target nucleic acid according to the invention by filling in the region (gap) between the sequences homologous to the 5' and the 3' regions of the target nucleic acid. Hence, in some embodiments, one would contact the instant cloning system with a target nucleic acid under conditions that allow homologous recombination.

[0098] Another embodiment combines: (i) a first segment including a first nucleic acid homologous to the 5' terminus of a target nucleic acid, a first selectable marker and a first cyclization element; (ii) a target nucleic acid; and (iii) a second segment including a second nucleic acid homologous to the 3' terminus of a target nucleic acid, a second selectable marker and a second cyclization element, under conditions which allow homologous recombination. One embodiment produces a gap-filled vector by homologous recombination between the two arms and the target nucleic acid. The exchange between the homologous regions found in the arms and the target nucleic acid is effected by homologous recombination at any point between the homologous nucleic acids. With respect to a circular vector of the present invention, the "gap filling" essentially is insertion (i.e., subcloning) of the target sequence into the vector.

[0099] Homologous recombination may be effected in vitro according to methodologies well known in the art. For example, the methods described herein can be practiced using yeast lysate preparations. Homologous recombination may take place in vivo. Hence, the method of the present disclosure may be practiced using any host cell capable of supporting homologous recombination events such as, for example, bacteria, yeast, plant and mammalian cells. One skilled in the art will appreciate that the choice of a suitable host depends on the particular combination of selectable markers used in the cloning system of the method.

[0100] One methodology makes use of a "gene gun" approach. The gene gun is part of a method called the biostatic (also known as bioballistic) method, and under certain conditions, DNA (or RNA) become "sticky," adhering to biologically inert particles such as metal atoms (usually tungsten or gold). By accelerating this DNA-particle complex in a partial vacuum and placing the target tissue within the acceleration path, DNA is effectively introduced. Uncoated metal particles could also be shot through a solution containing DNA surrounding the cell thus picking up the genetic material and proceeding into the living cell. A perforated plate stops the shell cartridge but allows the slivers of metal to pass through and into the living cells on the other side. The cells that take up the desired DNA, identified through the use of a marker gene (in plants the use of GUS is most common), are then cultured to replicate the gene and possibly cloned. Different methods have been used to accelerate the particles: these include pneumatic devices; instruments utilizing a mechanical impulse or macroprojectile; centripetal, magnetic or electrostatic forces; spray or vaccination guns; and apparatus based on acceleration by shock wave, such as electric discharge (for example, see Christou and McCabe, 1992,

Agracetus, Inc. Particle Gun Transformation of Crop Plants Using Electric Discharge (ACCELL™ Technology)).

[0101] Transformation can be performed, for example, according to the method of Cohen et al. (Proc. Natl. Acad. Sci. USA, 69:2110 (1972)), the protoplast method (Mol. Gen. Genet., 168:111 (1979)), or the competent method (J. Mol. Biol., 56:209 (1971)) when the hosts are bacteria (*E. coli*, *Bacillus subtilis*, and such), the method of Hinnen et al. (Proc. Natl. Acad. Sci. USA, 75:1927 (1978)), or the lithium method (J. Bacteriol., 153:163 (1983)) when the host is *S. cerevisiae*. Typically, following a transformation event, potential transformants are plated on nutrient media for selection and/or cultivation.

[0102] The nutrient media preferably comprises a carbon source, an inorganic nitrogen source, or an organic nitrogen source necessary for the growth of host cells (transformants). Examples of the carbon source are glucose, dextran, soluble starch, and sucrose, and examples of the inorganic or organic nitrogen source are ammonium salts, nitrates, amino acids, corn steep liquor, peptone, casein, meat extract, soy bean cake, and potato extract. If desired, the media may comprise other nutrients (for example, an inorganic salt (for example, calcium chloride, sodium dihydrogenphosphate, and magnesium chloride), vitamins, antibiotics (for example, tetracycline, neomycin, ampicillin, kanamycin, etc.). Media for some photosynthetic organisms may not require a carbon source as such organisms may be photoautotrophs and, thus, can produce their own carbon sources.

[0103] Cultivation and/or selection are performed by methods known in the art. Cultivation and selection conditions such as temperature, pH of the media, and cultivation time are selected appropriately for the vectors, host cells and methods of the present invention. One of skill in the art will recognize that there are numerous specific media and cultivation/selection conditions which can be used depending on the type of host cell (transformant) and the nature of the vector (e.g., which selectable markers are present). The media herein are merely described by way of example and are not limiting.

[0104] When the hosts are bacteria, actinomycetes, yeast, or filamentous fungi, media comprising the nutrient source(s) mentioned above are appropriate. When the host is *E. coli*, examples of preferable media are LB media, M9 media (Miller et al. Exp. Mol. Genet., Cold Spring Harbor Laboratory, p. 431 (1972)), and so on. When the host is yeast, an example of medium is Burkholder minimal medium (Bostian, Proc. Natl. Acad. Sci. USA, 77:4505 (1980)).

[0105] The selection of vectors in yeast may be accomplished by the use of yeast selectable markers. Examples include, but are not limited to, TRP1, MET2, MAZF, ADE2, ADE6, URA3, URA3, ARG1, ARG2, ARG3, HIS1, HIS2, HIS3, HIS5, HIS6, and LEU2. In certain embodiments, the HIS3, TRP1, URA3, LEU2 and ADE2 markers are used. In some embodiments, a vector or segment thereof may comprise two or more selectable markers. Thus, in one embodiment, a segment of a vector of the present invention may comprise an ADE marker to be lost upon homologous recombination with the target nucleic acid, and a HIS3 marker. The other segment may comprise a TRP1 marker. Selection is achieved by growing transformed cells on a suitable drop-out selection media (see e.g., Watson et al. (1992) Recombinant DNA, 2.sup.nd ed., Freeman and Co., New York, N.Y.). For example, HIS3 allows for selection of cells containing the first segment. TRP1 allows for selection of cells containing the second segment. ADE allows screening and selection of

clones in which homologous recombination took place. ADE enables color selection (red). In some embodiments, pairs of selection markers comprising the URA3 gene in combination with a second marker are utilized. URA3 is used in each pair because it allows for both positive and negative selection. The URA3 gene is then coupled with a second marker such as the LEU2, HIS3, LYS2 or kanMX6 marker. Either member of the pair can be used for selection when a single modification is introduced or just the non-URA3 marker in the case of two or more modifications.

[0106] Recombinant yeast cells may be selected using the selectable markers described herein according to methods well known in the art. Hence, one skilled in the art will appreciate that recombinant yeast cells harboring a gap-filled vector of the invention may be selected on the basis of the selectable markers included therein. For example, recombinant vectors carrying HIS3 and TRP1 may be selected by growing transformed yeast cells in the presence of drop-out selection media lacking histidine and tryptophan. Isolated positive clones may be purified further and analyzed to ascertain the presence and structure of the recombinant vector of the invention by, e.g., restriction analysis, electrophoresis, Southern blot analysis, polymerase chain reaction or the like. Also provided are gap-filled vectors engineered according to the methods described herein. Such a vector is the product of homologous recombination between the segments or vectors of the present disclosure and a target nucleic acid of choice. Also provided is a prokaryotic cell and/or a eukaryotic host cell harboring the cloning system or vector according to the present disclosure. The organism can be unicellular or multicellular. The organism may be naturally photosynthetic or naturally non-photosynthetic. Other examples of organisms that can be transformed include vascular and non-vascular organisms. When hosts, such as plant, yeast, or algal cells are used, a vector may contain, at least, a promoter, an initiation codon, the polynucleotide encoding a protein, and a termination codon. The vectors may also contain, if required, a polynucleotide for gene amplification (marker) that is usually used.

Products

[0107] The vectors described herein may comprise sequences that result in production of a product naturally, or not naturally, produced in the organism comprising the vector. In some instances the product encoded by one or more sequences on a vector is a polypeptide, for example an enzyme. Enzymes utilized may be encoded by nucleotide sequences derived from any organism, including bacteria, plants, fungi and animals. Vectors may also comprise nucleotide sequences that affect the production or secretion of a product from the organism. In some instances, such nucleotide sequence(s) encode one or more enzymes that function in isoprenoid biosynthetic pathway. Examples of polypeptides in the isoprenoid biosynthetic pathway include synthases such as C5, C10, C15, C20, C30, and C40 synthases. In some instances, the enzymes are isoprenoid producing enzymes. In some instances, an isoprenoid producing enzyme produces isoprenoids with two phosphate groups (e.g., GPP synthase, FPP synthase, DMAPP synthase). In other instances, isoprenoid producing enzymes produce isoprenoids with zero, one, three or more phosphates or may produce isoprenoids with other functional groups. Polynucleotides encoding enzymes and other proteins useful in the

present invention may be isolated and/or synthesized by any means known in the art, including, but not limited to cloning, sub-cloning, and PCR.

[0108] An isoprenoid producing enzyme may also be botryococcene synthase, β -caryophyllene synthase, germacrene A synthase, 8-epicedrol synthase, valencene synthase, (+)- δ -cadinene synthase, germacrene C synthase, (E)- β -farnesene synthase, casbene synthase, vetispiradiene synthase, 5-epiaristolochene synthase, aristolchene synthase, α -humulene, (E,E)- α -farnesene synthase, (-)- β -pinene synthase, γ -terpinene synthase, limonene cyclase, linalool synthase, (+)-bornyl diphosphate synthase, levopimaradiene synthase, isopimaradiene synthase, (E)- γ -bisabolene synthase, copalyl pyrophosphate synthase, kaurene synthase, longifolene synthase, γ -humulene synthase, δ -selinene synthase, β -phellandrene synthase, terpinolene synthase, (+)-3-carene synthase, syn-copalyd diphosphate synthase, α -terpineol synthase, syn-pimara-7,15-diene synthase, ent-sandaaracopimaradiene synthase, sterner-13-ene synthase, E- β -ocimene, S-linalool synthase, geraniol synthase, γ -terpinene synthase, linalool synthase, E- β -ocimene synthase, epi-cedrol synthase, α -zingerene synthase, guaiadiene synthase, cascarilladiene synthase, cis-muroladiene synthase, aphidicolan-16b-ol synthase, elisabethatriene synthase, sandalol synthase, patchoulol synthase, zinanol synthase, cedrol synthase, scareol synthase, copalol synthase, or manool synthase.

[0109] Other enzymes which may be produced by vectors of the present disclosure include biomass-degrading enzymes. Non-limiting examples of biomass-degrading enzymes include: cellulolytic enzymes, hemicellulolytic enzymes, pectinolytic enzymes, xylanases, ligninolytic enzymes, cellulases, cellobiases, softening enzymes (e.g., endopolygalacturonase), amylases, lipases, proteases, RNAses, DNAses, inulinase, lysing enzymes, phospholipases, pectinase, pullulanase, glucose isomerase, endoxylanase, beta-xylosidase, alpha-L-arabinofuranosidase, alpha-glucuronidase, alpha-galactosidase, acetylxyran esterase, and feruloyl esterase. Examples of genes that encode such enzymes include, but are not limited to, amylases, cellulases, hemicellulases, (e.g., β -glucosidase, endocellulase, exocellulase), exo- β -glucanase, endo- β -glucanase and xylanse (endoxylanase and exoxylanse). Examples of ligninolytic enzymes include, but are not limited to, lignin peroxidase and manganese peroxidase from *Phanerochaete chrysosporium*. One of skill in the art will recognize that these enzymes are only a partial list of enzymes which could be used.

[0110] The present disclosure contemplates making enzymes that contribute to the production of fatty acids, lipids or oils by transforming host cells and/or organisms comprising host cells with nucleic acids encoding one or more different enzymes. In some embodiments the enzymes that contribute to the production of fatty acids, lipids or oils are anabolic enzymes. Some examples of anabolic enzymes that contribute to the synthesis of fatty acids include, but are not limited to, acetyl-CoA carboxylase, ketoreductase, thioesterase, malonyltransferase, dehydratase, acyl-CoA ligase, ketoacyl synthase, enoylreductase and a desaturase. In some embodiments the enzymes are catabolic or biodegrading enzymes. In some embodiments, a single enzyme is produced.

[0111] Some host cells may be transformed with multiple genes encoding one or more enzymes. For example, a single transformed cell may contain exogenous nucleic acids encoding enzymes that make up an entire fatty acid synthesis pathway. One example of a pathway might include genes encod-

ing an acetyl CoA carboxylase, a malonyltransferase, a ketoacylsynthase, and a thioesterase. Cells transformed with entire pathways and/or enzymes extracted from them, can synthesize complete fatty acids or intermediates of the fatty acid synthesis pathway. In some embodiments constructs may contain multiple copies of the same gene, and/or multiple genes encoding the same enzyme from different organisms, and/or multiple genes with mutations in one or more parts of the coding sequences.

[0112] In some instances, a product (e.g. fuel, fragrance, insecticide) is a hydrocarbon-rich molecule, e.g. a terpene. A terpene (classified by the number of isoprene units) can be a hemiterpene, monoterpane, sesquiterpene, diterpene, triterpene, or tetraterpene. In specific embodiments the terpene is a terpenoid (aka isoprenoid), such as a steroid or carotenoid. Subclasses of carotenoids include carotenes and xanthophylls. In specific embodiments, a fuel product is limonene, 1,8-cineole, α -pinene, camphene, (+)-sabinene, myrcene, abietadiene, taxadiene, farnesyl pyrophosphate, amorphaadiene, (E)- α -bisabolene, beta carotene, alpha carotene, lycopene, or diaphytoene. Some of these terpenes are pure hydrocarbons (e.g. limonene) and others are hydrocarbon derivatives (e.g. cineole).

[0113] Examples of fuel products include petrochemical products and their precursors and all other substances that may be useful in the petrochemical industry. Fuel products include, for example, petroleum products, and precursors of petroleum, as well as petrochemicals and precursors thereof. The fuel product may be used for generating substances, or materials, useful in the petrochemical industry, including petroleum products and petrochemicals. The fuel or fuel products may be used in a combustor such as a boiler, kiln, dryer or furnace. Other examples of combustors are internal combustion engines such as vehicle engines or generators, including gasoline engines, diesel engines, jet engines, and others. Fuel products may also be used to produce plastics, resins, fibers, elastomers, lubricants, and gels.

[0114] Examples of products contemplated herein include hydrocarbon products and hydrocarbon derivative products. A hydrocarbon product is one that consists of only hydrogen molecules and carbon molecules. A hydrocarbon derivative product is a hydrocarbon product with one or more heteroatoms, wherein the heteroatom is any atom that is not hydrogen or carbon. Examples of heteroatoms include, but not limited to, nitrogen, oxygen, sulfur, and phosphorus. Some products are hydrocarbon-rich, wherein at least 50%, 60%, 70%, 80%, 90%, or 95% of the product by weight is made up carbon and hydrogen.

[0115] Fuel products, such as hydrocarbons, may be precursors or products conventionally derived from crude oil, or petroleum, such as, but not limited to, liquid petroleum gas, naptha (ligroin), gasoline, kerosene, diesel, lubricating oil, heavy gas, coke, asphalt, tar, and waxes. For example, fuel products may include small alkanes (for example, 5 to approximately 4 carbons) such as methane, ethane, propane, or butane, which may be used for heating (such as in cooking) or making plastics. Fuel products may also include molecules with a carbon backbone of approximately 5 to approximately 9 carbon atoms, such as naptha or ligroin, or their precursors. Other fuel products may be about 5 to about 12 carbon atoms or cycloalkanes used as gasoline or motor fuel. Molecules and aromatics of approximately 10 to approximately 18 carbons, such as kerosene, or its precursors, may also be fuel products. Fuel products may also include molecules, or their precur-

sors, with more than 12 carbons, such as used for lubricating oil. Other fuel products include heavy gas or fuel oil, or their precursors, typically containing alkanes, cycloalkanes, and aromatics of approximately 20 to approximately 70 carbons. Fuel products also includes other residuals from crude oil, such as coke, asphalt, tar, and waxes, generally containing multiple rings with about 70 or more carbons, and their precursors.

[0116] In one embodiment, the vector comprises a nucleotide sequence encoding an enzyme utilized in the production of plant sterols or phytosterols. As used herein the term "phytosterols" includes in addition to phytosterols, phytosterol esters, phytostanols and phytostanol esters. Phytosterols have been shown in clinical trials to reduce absorption of cholesterol and are approved by the U.S. Food and Drug Administration for use as a food additive. In other embodiments, the vectors comprises nucleotide sequences encoding therapeutic proteins. Numerous therapeutic proteins are known in the art and include, but are not limited to, hormones such as insulin and erythropoietin, antibodies, vaccines, albumins and interferons.

[0117] Host Cells and Organisms

[0118] Examples of organisms that can be transformed using the vectors and methods herein include vascular and non-vascular organisms. The organism can be prokaryotic or eukaryotic. The organism can be unicellular or multicellular.

[0119] Eukaryotic cells, such as a fungal cell (e.g., *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* or *Ustilago maydis*) may be transformed using the methods and compositions disclosed herein. Methods for introducing nucleic acids in a fungal/yeast cells are well known in the art. Hence, such a step may be accomplished by conventional transformation methodologies. Non-limiting examples of suitable methodologies include electroporation, alkali cations protocols and spheroplast transformation.

[0120] Examples of non-vascular photosynthetic organisms include bryophytes, such as marchantiophytes or anthocerotophytes. In some instances the organism is a cyanobacteria. In some instances, the organism is algae (e.g., macroalgae or microalgae). The algae can be unicellular or multicellular algae. In some instances the organism is a rhodophyte, chlorophyte, heterokontophyte, tribophyte, glaucophyte, chlorarachniophyte, euglenoid, haptophyte, cryptomonad, dinoflagellum, or phytoplankton.

[0121] The use of microalgae to express a polypeptide or protein complex according to a method disclosed herein provides the advantage that large populations of the microalgae can be grown, including commercially (Cyanotech Corp.; Kailua-Kona Hi.), thus allowing for production and, if desired, isolation of large amounts of a desired product. However, the ability to express, for example, functional polypeptides, including protein complexes, in the chloroplasts of any plant and/or modify the chloroplasts or any plant allows for production of crops of such plants and, therefore, the ability to conveniently produce large amounts of the polypeptides. Accordingly, the methods of the invention can be practiced using any plant having chloroplasts, including, for example, macroalgae, for example, marine algae and seaweeds, as well as plants that grow in soil, for example, corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum*

miliaceum), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Coffea spp.*), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus spp.*), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa spp.*), avocado (*Persea americana*), fig (*Ficus carica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugar cane (*Saccharum spp.*), oats, duckweed (*Lemna*), barley, tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus spp.*), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals such as azalea (*Rhododendron spp.*), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosa-sinensis*), roses (*Rosa spp.*), tulips (*Tulipa spp.*), daffodils (*Narcissus spp.*), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum are also included. Additional ornamentals useful for practicing a method of the invention include impatiens, Begonia, Pelargonium, Viola, Cyclamen, Verbena, Vinca, Tagetes, Primula, Saint Paulia, Agertum, Amaranthus, Antihirrinum, Aquilegia, Cineraria, Clover, Cosmo, Cowpea, Dahlia, Datura, Delphinium, Gerbera, Gladiolus, Gloxinia, Hippeastrum, Mesembryanthemum, Salpiglossis, and Zinnia. Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliottii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*), Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga heterophylla*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*).

[0122] Leguminous plants that may be used include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mung bean, lima bean, fava bean, lentils, chickpea, etc. Legumes include, but are not limited to, *Arachis*, e.g., peanuts, *Vicia*, e.g., crown vetch, hairy vetch, adzuki bean, mung bean, and chickpea, *Lupinus*, e.g., lupine, *trifolium*, *Phaseolus*, e.g., common bean and lima bean, *Pisum*, e.g., field bean, *Melilotus*, e.g., clover, *Medicago*, e.g., alfalfa, *Lotus*, e.g., trefoil, lens, e.g., lentil, and false indigo. Exemplary forage and turf grass include alfalfa, orchard grass, tall fescue, perennial ryegrass, creeping bent grass, and redtop. Other plants useful in the invention include *Acacia*, aneth, artichoke, arugula, blackberry, canola, cilantro, clementines, escarole, eucalyptus, fennel, grapefruit, honey dew, jicama, kiwifruit, lemon, lime, mushroom, nut, okra, orange, parsley, persimmon, plantain, pomegranate, poplar, radicchio, Southern pine, sweetgum, tangerine, triticale, vine, yams, apple, pear, quince, cherry, apricot, melon, hemp, buckwheat, grape, raspberry, *chenopodium*, blueberry, nectarine, peach, plum, strawberry, watermelon, eggplant, pepper, cauliflower, *Brassica*, e.g., broccoli,

cabbage, ultilan sprouts, onion, carrot, leek, beet, broad bean, celery, radish, pumpkin, endive, gourd, garlic, snapbean, spinach, squash, turnip, ultilane, chicory, groundnut and zucchini. Thus, the compositions contemplated herein include host organisms comprising any of the above vectors. The host organism can be any chloroplast-containing organism.

[0123] The term "plant" is used broadly herein to refer to a eukaryotic organism containing plastids, particularly chloroplasts, and includes any such organism at any stage of development, or to part of a plant, including a plant cutting, a plant cell, a plant cell culture, a plant organ, a plant seed, and a plantlet. A plant cell is the structural and physiological unit of the plant, comprising a protoplast and a cell wall. A plant cell can be in the form of an isolated single cell or a cultured cell, or can be part of higher organized unit, for example, a plant tissue, plant organ, or plant. Thus, a plant cell can be a protoplast, a gamete producing cell, or a cell or collection of cells that can regenerate into a whole plant. As such, a seed, which comprises multiple plant cells and is capable of regenerating into a whole plant, is considered plant cell for purposes of this disclosure. A plant tissue or plant organ can be a seed, protoplast, callus, or any other groups of plant cells that is organized into a structural or functional unit. Particularly useful parts of a plant include harvestable parts and parts useful for propagation of progeny plants. A harvestable part of a plant can be any useful part of a plant, for example, flowers, pollen, seedlings, tubers, leaves, stems, fruit, seeds, roots, and the like. A part of a plant useful for propagation includes, for example, seeds, fruits, cuttings, seedlings, tubers, rootstocks, and the like.

[0124] Eukaryotic host cells may be a fungal cell (e.g., *S. cerevisiae*, *Sz. pombe* or *U. maydis*). Examples of prokaryotic host cells include *E. coli* and *B. subtilis*, cyanobacteria and photosynthetic bacteria (e.g. species of the genus *Synechocystis* or the genus *Synechococcus* or the genus *Athrobacteria*). Examples of non-vascular plants which may be a host organism (or the source of target DNA) include bryophytes, such as marchantiophytes or anthocerotophytes. In some instances, the organism is algae (e.g., macroalgae or microalgae, such as *Chlamydomonas reinhardtii*, *Chorella vulgaris*, *Dunaliella salina*, *Haematococcus pluvialis*, *Scenedesmus* spp.). The algae can be unicellular or multicellular algae. In some instances the organism is a rhodophyte, chlorophyte, heterokontophyte, tribophyte, glaucophyte, chlorarachniophyte, euglenoid, haptophyte, cryptomonad, dinoflagellum, or phytoplankton. In other instances one of skill in the art will recognize that these organisms are given merely as examples and other organisms may be substituted where appropriate positive and negative selectable markers are available.

[0125] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

EXAMPLES

Example 1

DNA Purification and Analysis

[0126] DNA is isolated and analyzed according to methods known in the art. Various methods for the isolation of nuclear

and plastid DNA from plants have been published and are available to the skilled artisan for example in Murray and Thompson, *Nuc. Acids Res.*, 8:4321, 1980; Bovenberg et al., *Nuc. Acids Res.* 9:503, 1981; Triboush et al., *Plant Molec Biol. Reporter*, 16:183, 1998; and *Methods in Plant Molecular Biology, A Laboratory Course Manual*, Maliga et al, Cold Spring Harbor Laboratory Press, 1995.

[0127] To prepare DNA from plant cells, the DNeasy Plant Mini Kit is used (Qiagen).

[0128] To prepare DNA from yeast to use as a template for PCR, 10^6 yeast cells (from agar plate or liquid culture) are suspended in lysis buffer (6 mM KHPO₄, pH=7.5, 6 mM NaCl, 3% glycerol, 1 U/mL zymolyase) and heated to 37° C. for 30 ml, 95° C. for 10 minutes, then cooled to near 23° C. The solution is added to the PCR mixture directly.

[0129] To prepare plasmid DNA from yeast, desired clones are grown in selective liquid media (e.g., CSM-Trp) to saturation at 30° C. Cells are collected by centrifugation at 3000×g for 10 minutes and resuspended in 150 μ L of lysis buffer (1 M sorbitol, 0.1 M sodium citrate, 0.06 M EDTA pH=7.0, 100 mM beta-mercaptoethanol, and 2.5 mg/mL zymolyase). The solution is incubated for 1 hr at 37° C. 300 μ L of denaturing solution (1% SDS and 0.2N NaOH) is added and solution is incubated at 60° C. for 15 min. 150 μ L of neutralizing solution (3M potassium acetate, pH=4.8) is added and the solution is incubated on ice for 10 min. The solution is centrifuged at 14,000 RPM for 10 min and the supernatant is transferred to another tube. 1 mL of isopropanol is added, the mixture is gently mixed and centrifuged at 14,000 RPM for 10 min. The pellet is washed once with 1 mL of 70% ethanol and centrifuged at 14,000 RPM for 10 min. The DNA pellet is air-dried and resuspended in 60 μ L of resuspension buffer (10 mM Tris pH=7.4, 1 mM EDTA, and 0.1 mg/mL RNase).

[0130] To prepare plasmid DNA from bacteria, cells are grown to saturation at 37° C. in LB containing the appropriate antibiotic (Kan or Amp). If the DNA of interest contains standard replication elements, cells are harvested by centrifugation. If the DNA of interest contains P1 replication elements, saturated cell cultures are diluted 1:20 in LB+Kan+IPTG and grown for 4 hours at 37° C., then harvested. The Plasmid Maxi kit (QIAGEN) is used to prepare plasmid DNA from the cell pellets.

[0131] For illustrative purposes, and without limiting the invention to the specific methods described, DNA samples prepared from algae, yeast, or bacteria (in plugs or in solution) are analyzed by pulse-field gel electrophoresis (PFGE), or digested with the appropriate restriction endonuclease (e.g., SmaI) and analyzed by PFGE, conventional agarose gel electrophoresis, and/or Southern blot. Standard protocols useful for these purposes are fully described in Gemmill et al. (in "Advances in Genome Biology", Vol. 1, "Unfolding The Genome," pp 217 251, edited by Ram S. Verma).

[0132] One of skill will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

Example 2

Transformation Methods

[0133] *E. coli* strains DH 10B or Genehog are made electrocompetent by growing the cells to an OD₆₀₀ of 0.7, then collected and washed twice with ice-cold 10% glycerol, flash frozen in a dry-ice ethanol bath and kept at -80° C. Total yeast

or algae DNA is prepared and electroporated into *E. coli* by using, for example, a 0.1 cm cuvette at 1,800 V, 200 ohms and 25 mF in a Bio-Rad Gene Pulsar Electroporator. Cells are allowed to recover and clones are selected on agar growth media containing one or more antibiotics, such as kanamycin (50 μ g/mL), ampicillin (100 μ g/mL), gentamycin (50 μ g/mL), tetracycline (51 μ g/mL), or chloramphenicol (34 μ g/mL).

[0134] Yeast strains YPH857, YPH858 or AB1380 may be transformed by the lithium acetate method as described in Sheistl & Geitz (*Curr. Genet.* 16:339 346, 1989) and Sherman et al., "Laboratory Course Manual Methods in Yeast Genetics" (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986) or a spheroplast method such as the one described by Sipiczki et al., *Curr. Microbiol.*, 12(3):169-173 (1985). Yeast transformants are selected and screened on agar media lacking amino and/or nucleic acids, such as tryptophan, leucine, or uracil. Standard methods for yeast growth and phenotype testing are employed as described by Sherman et al., supra.

[0135] Chloroplasts of plants are transformed based on the method of Maenpaa et al., *Molec. Biotech.*, 13:67, 1999. Tungsten or gold particles are sterilized for use as microcarriers in bombardment experiments. Particles (50 mg) are sterilized with 1 ml of 100% ethanol, and stored at -20° C. or -80° C. Immediately prior to use, particles are sedimented by centrifugation, washed with 2 to 3 washes of 1 ml sterile deionised distilled water, vortexed and centrifuged between each wash. Washed particles are resuspended in 500 μ l 50% glycerol.

[0136] Sterilized particles are coated with DNA for transformation. 25 μ l aliquots of sterilized particles are added to a 1.5 ml microfuge tube, and 5 μ g of DNA of interest is added and mixed by tapping. 35 μ l of a freshly prepared solution of 1.8M CaCl₂ and 30 mM spermidine is added to the particle/DNA mixture, mixed gently, and incubated at room temperature for 20 minutes. The coated particles are sedimented by centrifuging briefly. The particles are washed twice by adding 200 μ l 70% ethanol, mixing gently, and centrifuging briefly. The coated particles are resuspended in 50 μ l of 100% ethanol and mixed gently. Five to ten microliters of coated particles are used for each bombardment.

[0137] Transformation by particle bombardment is carried out using the PDS 1000 Helium gun (Bio Rad, Richmond, Calif.) using a modified protocol described by the manufacturer. Plates containing leaf samples are placed on the second shelf from the bottom of the vacuum chamber and bombarded using the 1100 p.s.i. rupture disk. After bombardment, petri-plates containing the leaf samples are wrapped in plastic bags and incubated at 24° C. for 48 hours.

[0138] After incubation, bombarded leaves are cut into approximately 0.5 cm² pieces and placed abaxial side up on selection medium. After 3 to 4 weeks on the selection medium, small, green shoots will appear on the leaf tissue. These shoots will continue to grow on selection medium and are referred to as putative transformants. When the putative transformants have developed 2 to 3 leaves, 2 small pieces (approximately 0.5 cm²) are cut from each leaf and used for additional rounds of shoot regeneration or further growth.

[0139] Two to four shoots of each positive transformant are selected and transferred to selection medium for generation of roots. Analysis is performed on 2 shoots to confirm homoplasmy. Shoots from homoplasmic events are transferred to the greenhouse for seed production, while transfor-

mants which are not homoplasmic are sent through a second round or regeneration on selection medium to attain homoplasm.

[0140] One of skill will appreciate that many other transformation methods known in the art may be substituted in lieu of the ones specifically described or referenced herein.

Example 3

Hybrid Vector System

[0141] In this example, a system is established using a hybrid vector to support replication of chloroplast DNA from vascular plants in yeast and bacteria (FIG. 1). The hybrid gap filling vector backbone contains yeast elements that allow it to function as a yeast artificial plasmid (YAP) and bacterial elements that allow it to function as a plasmid artificial chromosome (PAC) or a bacterial artificial chromosome (BAC). The yeast elements include a yeast selection marker sequence (e.g. TRP1 or LEU2), a yeast centromere sequence (CEN), and a yeast autonomously replicating nucleotide sequence (ARS). Bacterial elements include the P1 origin of replication or the F' bacterial origin of replication sequences and a bacterial selection marker sequence (e.g. Cam^r or Kan^r). The yeast and bacterial elements may be combined with chloroplast DNA in a single vector by a variety of methods, including, but not limited to, integration of yeast elements into existing BAC and/or PAC clones containing chloroplast DNA (described below), capture of chloroplast DNA fragments into a hybrid vector by homologous recombination in yeast (aka gap-filling), and constructing of large DNA fragment libraries using a vector containing yeast and bacterial elements as the backbone.

[0142] To obtain chloroplast DNA in a hybrid vector by integrating yeast elements into existing BAC clones, homologous recombination in yeast may be used. The region of DNA in pTrp-10-Kan (SEQ ID NO. 1) that contains sequences encoding ARS, CEN and TRP1 is PCR amplified using primers (SEQ ID NOs. 2 and 3) with 5' tails homologous to sequences within pBeloBAC-11 (SEQ ID NO. 4) or its derivatives. The PCR product is then mixed with the desired BAC clone and transformed into yeast. Transformants are identified by growth on selective media. Once confirmed, DNA from desired yeast clones is prepared and transformed into bacteria for amplification and purification.

[0143] To increase recombination efficiency and obtain modified BACs more efficiently, longer regions of homology were used. A 5.0 kb region that contains the yeast replication and selection elements and terminal sequences homologous pBeloBAC-11 was PCR amplified from a BAC modified by the PCR method described above (previous paragraph) using primers (SEQ ID NOs. 5 and 6) that both have SnaBI and XhoI sites at the 5' ends. The PCR product was purified by ethanol precipitation, digested with SnaBI, and isolated using agarose gel electrophoresis and the QIAQUICK gel extraction kit (QIAGEN). The SnaBI digested DNA fragment was then ligated to pUC-SE (SEQ ID NO. 7) that was digested with NotI and treated with Klenow enzyme to create blunt ends, creating pBeloYAP (SEQ ID NO. 8). The integration cassette is liberated from pBeloYAP by digestion with XhoI and isolated using agarose gel electrophoresis and the QIAQUICK gel extraction kit (QIAGEN). The 5.0 kb restriction fragment is mixed with the desired BAC clone and transformed into yeast. Transformants are identified by growth on selective media. Once confirmed, DNA from desired yeast clones is prepared and transformed into bacteria for amplification and purification.

[0144] One of skill will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

Example 4

Vectors to Stabilize and/or Modify Chloroplast Genome DNA in an Exogenous Host

[0145] Often, large pieces of heterologous DNA are unstable in host organisms such as yeast or bacteria. This may be due to multiple factors, including, but not limited to, the presence of toxic gene products or codon bias and/or lack of selective pressure. Therefore, the target DNA within the shuttle vector may be altered within yeast or bacteria. For example, certain portions of a target DNA sequence (e.g., coding regions or promoters) may be deleted or moved by recombination within the host organism. In a similar way, when a shuttle vector carrying the target DNA is transferred back to the organism (or a closely related species) that donated the target DNA, the target DNA can become unstable.

[0146] Pairs of yeast selection markers were constructed so that multiple stabilization sites could be employed simultaneously. Each marker pair contains the URA3 gene (SEQ ID NO. 9), which was PCR amplified from pRS416. Each marker pair also contains the LEU2 gene (SEQ ID NO. 10) amplified from pRS415, the HIS3 gene (SEQ ID NO. 11) amplified from pRS413, the ADE2 gene (SEQ ID NO. 12) amplified from pTrp-AU, the LYS2 gene (SEQ ID NO. 13) amplified from *S. cerevisiae* genomic DNA, or the kanMX6 gene (SEQ ID NO. 14) from pFA6a-kanMX6, which confers resistance to the antifungal agent G418. The primers used for the URA3 gene add the XmaI restriction site to the 5' end (SEQ ID NO. 15) and Sall and SacII to the 3' end (SEQ ID NO. 16). The primers used for the LEU2, HIS3, ADE2, LYS2, and G418^r genes add the XmaI restriction site to the 5' end (SEQ ID NO. 17 for LEU2, SEQ ID NO. 18 for HIS3, SEQ ID NO. 19 for ADE2, SEQ ID NO. 20 for LYS2, and SEQ ID NO. 21 for G418^r) and Sall, FseI, and SpeI sites to the 3' end (SEQ ID NO. 22 for LEU2, SEQ ID NO. 23 for HIS3, SEQ ID NO. 24 for ADE2, SEQ ID NO. 25 for LYS2, and SEQ ID NO. 26 for G418^r). Each PCR product was digested with XmaI and Sall, mixed pairwise, and ligated into Sall-digested pUC1 (SEQ ID NO. 27). URA3 is used in each case because it allows for positive and negative selection. Marker pairs can be introduced based on selection for either gene (in the case of a single modification), the non-URA3 gene in the case of two or more modification. Then the markers can be removed by introducing DNA with terminal sequences homologous to those surrounding the marker pairs and selecting for growth on minimal media containing 5-FOA.

[0147] To promote sequence stability in bacteria, antibiotic resistance markers were cloned into the yeast selection marker pairs. The bacterial stability markers include, but are not limited to, the ampicillin resistance gene (Amp^r, SEQ ID NO. 14) amplified from pET-21a, the tetracycline resistance gene (Tet^r, SEQ ID NO. 27) amplified from pBR322, the chloramphenicol resistance gene (Cam^r, SEQ ID NO. 28) amplified from pETcoco-1, and the gentamycin resistance gene (Gent^r, SEQ ID NO. 29) amplified from pJQ200. For each gene, primer pairs (SEQ ID NOs. 30 and 31 for Amp^r, SEQ ID NOs. 32 and 33 for Tet^r, SEQ ID NOs. 34 and 35 for Cam^r, and SEQ ID NOs. 36 and 37 for Gent^r) that add XmaI sites to both the 5' and 3' ends were used to PCR amplify the antibiotic resistance fragment. Each PCR product was digested with XmaI and ligated into XmaI-digested vectors containing yeast marker pair cassettes.

[0148] The yeast selection marker cassettes (with or without the bacterial antibiotic resistance markers) may be cloned into traditional targeting vectors for integration into the desired target sequence by homologous recombination in yeast. In addition, these same selection marker cassettes may serve as the template for a one-step PCR-mediated technique for introducing selection markers into target regions. Two PCR primers are designed such that the sequence of the first 40-42 nucleotides (5'→3') of each primer are identical to the target sequences, and the final 18-20 nucleotides are identical to sequences within a vector containing a selection marker cassette. Thus, PCR amplification of the selection marker cassette adds flanking sequences that target the selection marker(s) to the desired region.

Example 5

Cloning the *Glycine Max* Chloroplast Genome with a Hybrid Cloning System

[0149] In this example, a system is established using a hybrid vector to support replication of chloroplast DNA from *G. max* in yeast and bacteria (FIG. 1). *G. max* chloroplast genome DNA was derived from BAC clone 04E08 from a BAC library derived from *G. max* total genomic DNA (described in Saski et al., *Plant Mol. Bio.*, 59: 309-322 (2005)) (FIG. 2). Clone 04E08 contains the entire chloroplast genome in a single, circular molecule that stably replicates in bacteria. Incorporation of yeast elements for replication and selection was thus required to adapt clone 04E08 to the hybrid system.

[0150] To incorporate the required yeast elements, we used homologous recombination in yeast. First, pBeloYAP was digested with XbaI, liberating a 5.0 kb fragment containing a yeast ARS and CEN, TRP1 selection marker, and flanking sequences homologous to those found in the pIndigoBAC vector backbone. This fragment was isolated using agarose gel electrophoresis and the QIAQUICK gel extraction kit (QIAGEN). Next, the URA3-LEU2 selection marker cassette was PCR amplified using primers (SEQ ID NOS. 38 and 39) with 5' tails homologous to sequences flanking *G. max* chloroplast genome nucleotide 122,283 (according to the sequence available from NCBI, NC_007942). The PCR product was isolated using agarose gel electrophoresis and the QIAQUICK gel extraction kit (QIAGEN).

[0151] The DNA fragments containing the yeast replication and selection elements were mixed with clone 04E08 and transformed into YPH858 using the spheroplast method. Transformants were identified by growth on CSM-ura-leu agar medium and propagated on CSM-ura-leu-trp agar medium. Yeast clones that grow on both media types were screened by colony PCR using primers (SEQ ID NOS. 40 and 41) that amplify a region within the *G. max* chloroplast genome surrounding nucleotide 000,060 (according to the sequence available from NCBI, NC_007942). Desired clones are those that give rise to a PCR product of expected size. FIG. 3A shows that clone 1 gave rise to a PCR product. DNA was prepared from Clone 1 grown in CSM-ura-leu-trp liquid medium and screened with additional primers (SEQ ID NOS. 40 and 41; 42 and 43; 44 and 45; 46 and 47; 48 and 49; 50 and 51; 52 and 53; 54 and 55; 56 and 57; 58 and 59; 60 and 61; and 62 and 63) spread throughout the *G. max* chloroplast genome. Desired clones are those that give rise to PCR products of expected size for all reaction. FIG. 3B shows that clone 1 gave rise to a PCR product of expected size in all reactions, indicating that clone 1 harbors clone 04E08 with integrated yeast replication and selection elements, hereafter called Gm-001.

[0152] Gm-001 was transferred to bacterial strain DH10B by electroporation, followed by selection on LB agar medium containing chloramphenicol (34 µg/mL). Isolated transformants were screened by colony PCR using primers (SEQ ID NOS. 40 and 41; 42 and 43; 44 and 45; 46 and 47; 48 and 49; 50 and 51; 52 and 53; 54 and 55; 56 and 57; 58 and 59; 60 and 61; and 62 and 63) spread throughout the *G. max* chloroplast genome and primers (SEQ ID NOS. 64 and 65 and SEQ ID NOS. 66 and 67) specific for integration of the URA3-LEU2 cassette. Desired clones are those that give rise to PCR products of expected size for all reaction. FIG. 3C shows a bacterial clone gave rise to a PCR product of expected size in all reactions, demonstrating that Gm-001 was indeed transferred from yeast into bacteria.

[0153] To further confirm that Gm-001 is stable in the hybrid system, the DNA molecule was isolated and analyzed by restriction mapping. Briefly, bacterial clones were growing to saturation in LB media containing chloramphenicol and collected by centrifugation. The Plasmid Maxi kit (QIAGEN) is used to prepare plasmid DNA from the isolated clones. Each molecule, as well as clone 04E08, was restriction mapped with HindIII. FIG. 3D shows that the isolated clone appear similar to clone 04E08, indicating that Gm-001 is stable in the hybrid system.

[0154] One of skill will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

Example 6

Cloning the *A. thaliana* Chloroplast Genome Using a Hybrid System (Prophetic)

[0155] In this example, a system is established using a hybrid vector to support replication of chloroplast DNA from *A. thaliana* in yeast and bacteria. The *A. thaliana* chloroplast genome DNA is derived from three PAC clones, MAB17, MC13, and MAH2, which are obtained from the Mitsui P1 library (described in Liu et al., *Plant J.*, 7: 351-358 (1995) and Sato et al., *DNA Res.*, 6: 283-290 (1999)). These three PAC clones replicates stably in bacteria and contain partially overlapping sequences that comprise the entire *A. thaliana* chloroplast genome. MAB17, MC13 and MAH2 are digested with restriction endonucleases to generate linear DNA fragments with overlapping 5' and 3' termini. These restriction fragments are transformed into yeast with a hybrid vector backbone containing DNA sequences homologous to adjacent regions in the *A. thaliana* chloroplast genome. A single DNA molecule comprising the entire *A. thaliana* chloroplast genome is created by homologous recombination.

[0156] Standard methods are used to identify desired clones.

[0157] One of skill will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

Example 7

Cloning the *Z. mays* Chloroplast Genome Using a Hybrid System

[0158] In this example, a system is established using a hybrid vector to support replication of chloroplast DNA from *Z. mays* in yeast and bacteria. A hybrid gap-filling vector is created with DNA sequences that have high homology to adjacent regions in the *Z. mays* chloroplast genome. Chloroplast genome DNA is obtained from genomic DNA preparations from *Z. mays* cells. A single DNA molecule comprising the entire *Z. mays* chloroplast genome is created by trans-

forming linearized gap filling vector DNA and chloroplast genome DNA into *S. cerevisiae* using the lithium acetate or spheroplast methods as described in EXAMPLE 2. Homologous recombination takes place in vivo in the transformed yeast cells. Once the target DNA is captured by the vector via homologous recombination, the DNA can be stably replicated in both yeast and bacterial systems.

[0159] Standard methods are used to identify desired clones.

[0160] One of skill will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

Example 8

Cloning the *O. sativa* Chloroplast Genome Using a Hybrid System

[0161] In this example, a system is established using a hybrid vector to support replication of chloroplast DNA from

O. sativa in yeast and bacteria. A hybrid gap-filling vector is created with DNA sequences that have high homology to adjacent regions in the *O. sativa* chloroplast genome. Chloroplast genome DNA is obtained from genomic DNA preparations from *O. sativa* cells. A single DNA molecule comprising the entire *O. sativa* chloroplast genome is created by transforming linearized gap filling vector DNA and chloroplast genome DNA into *S. cerevisiae* using the lithium acetate or spheroplast methods as described in EXAMPLE 2. Homologous recombination takes place in vivo in the transformed yeast cells. Once the target DNA is captured by the vector via homologous recombination, the DNA can be stably replicated in both yeast and bacterial systems.

[0162] Standard methods are used to identify desired clones.

[0163] One of skill will appreciate that many other methods known in the art may be substituted in lieu of the ones specifically described or referenced.

SEQUENCE LISTING

```

<160> NUMBER OF SEQ ID NOS: 67

<210> SEQ ID NO 1
<211> LENGTH: 22908
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic polynucleotide

<400> SEQUENCE: 1

tgaatttatac agatctaact gaggagtaag aaaccccat gtcaaagaaa aacagaccaa      60
caattggcg aacccttaat ccttcaatat taagcgatt tgatagttct tcagcctctg     120
gcgatcgagt cgagcaggta ttcaagttat caactggtcg ccaggccaca tttattgaag    180
aggtataacc tccgaaccag gtgaaagcg atacctttgt tgatcagcat aacaacgggc    240
gtgaccaggc atctttaacg cccaaatcat taaaaagtat ccgaagcact attaagcatc   300
agcaatttta ccctgcaata ggtgttagac gggctacagg gaaaattgaa atttggatg     360
gttcccggcg tcgagcttct gccatcttag agaacgttagg gttgggggtt ttatcacgg  420
accaggagat cagegttcag gaagcgcaaa atttagcgaa agacgttcag acagcattgc  480
agcacagcat tcgagaaata ggtctgcgtt ttagtgeaat gaaaatgtat gggatgatc  540
agaaggatat tgcagccaaa gaagggtgtt ctcaggcgaa ggtcacgcgt gctctccagg 600
cagcagatgc tccggaagaa ttatcgcccc tttccctgt gcagtcggaa ttaacctttt 660
cgactacaa aacgctttgt gctgtggcg acgaaatggg gaacaagaat ttatagttt 720
atcagcttat tcaaaacata tccccggaaa taaacgacat cttatccatt gaagaaatgg 780
ccgaagatga agttaaaaat aaaatcctgc gcttgataac aaaggaagcc tcactactca 840
cgatataagg ttctaaagat aagtccgtag ttactgaatt atggaaattt gaggacaagg 900
atcgctttgc aaggaaagcgc gtgaaaggcc gtgcatttc ttatgagttt aatcgactct 960
caaaagagtt acaggaagaa ctgcacagga tgattggca tatcctttaga aagagcctcg 1020
ataaaaagcc gaagccttaa actttcgcca ttcaaatttc actattaact gactgtttt 1080
aaagtaaatt actctaaaat ttcaaggta aatcgccacg atttcacctt ggattttacc 1140
ttcctccctt cctcccgaaa aaaataaaaa aattgcttgt cacgagaaag tcaacaagtg 1200

```

-continued

actttcaata aaatctcttc cgaaaaggga ttcacacaag tgccttgtt ttaaggaaga	1260
gtaaattgag taacttacgc gaataccaga atcgtattgc agatatcgca aaacgctcta	1320
aagctgtgtc tggctggca agcaactgcgc agttcggtac tgataaccas ttcataaag	1380
atgatgcccgc gcgtgccgca tctatccttg aagctgcacg taaagaccg gttttgcgg	1440
gtatctctga taatgccacc gctcaaatacg ctacagcgtg ggcaagtgccttgcact	1500
acgcccgcacg acataaaatct atgcccgcgc cggaaattctt ggccctctgc caccagacgc	1560
tggaaaactg cctgatagag tccacccgca atagcatggta tgccactaat aaagcgatgc	1620
tggaatctgtt cgcagcagag atgatgagcg tttctgacgg tggttatgcgt ctgcctttat	1680
tcctcgcat gatcctgcctt gttcagttgg gggcagctac cgctgatgcgt tgtaacctca	1740
ttccgggttac gcgtgaccag tccgacatct atgaagtctt taacgtggca ggttcatctt	1800
ttggttctta tgctgctggat gatgttctgg acatgcaatc cgtegggtgt tacagccagt	1860
tacgtcgccg ctatgtgtc gtggcaagctt ccgtatggcac cagcaaaacc gcaaccttca	1920
agatggaaga cttcgaaaggc cagaatgtac caatccgaaa aggtcgact aacatctacg	1980
ttaaccgtat taagtctgtt gttgataacg gttccggcag ctacttccac tcgttacta	2040
atgctgctgg tgagcaaatac actgttacct gctctctgaa ctacaacattt ggtcagattt	2100
ccctgtcggtt ctccaaagcg ccggataaaag gcactgagat cgcaatttgcg acggaaatca	2160
atattgaagc cgctcctgag ctgatcccgc tgatcaacca cgaatgaag aaatacaccc	2220
tgttcccaag tcaagttcggtt atcgcggctg agcacacacggt acaggcggcgt tatgaagcac	2280
agcgtgaatt tggctggac ctgggttccc tacagttccg caccctgaag gaataacctgt	2340
ctcatgaaca ggatatgtcg cgttccgcata tcatgatctg ggcacttgc ggcacccgaca	2400
cctttgacat cgctctgcgc gttaccatgtt cttttgtatgtt atgggcaccc atcattctgt	2460
gcaaaattcca gactgtatata cgccgacatta ttgagcgcgt taaatcttctt ggtgegatgg	2520
ggatgtttgc tggctgtatgtt gcagcatctt tcttcaaaca gttggcgaag gatttcttcc	2580
agccagccga agactatatac cagactccgt atgttcaacta catcggtacc ccatttagga	2640
ccacccacag cacctaacaa aacggcatac gccttctgg aggcttccag cgccatctt	2700
ggaagtgaa cacctgttagc atcgacctgc agggggggggggggcgttgcgt gttcgccctcg	2760
tgaagaaggt gttgctgact cataccagggc ctgaatcgcc ccatcatcca gccagaaagt	2820
gaggggccca cgggttgcgtt gagttttgtt gtaggtggac cagttggta ttttgaactt	2880
tttgcgttgc acggaacggcgtt ctgcgttgc gggaaatgcgtt gttgtatgtt ctttcaactc	2940
agcaaaatgtt cgttatttc aacaaaggccg ccgtcccgatc aagtcaacgtt aatgtctgc	3000
cagtgttaca accaattaac caattctgtat tagaaaaactt catcgacat caaatgaaac	3060
tgcataattt tcatatcagg attatcaata ccataattttt gaaaaaggccg tttctgtat	3120
gaaggagaaa actcaccatgc gcaatgttccat aggatggcaat gatcctggta tcggctgcgt	3180
attccgactc gtccaaacatc aatacaacactt attaatttcc cctcgtaaaa aataagggtt	3240
tcaagtgaga aatcaccatgc agtgcacgtt gaatccgggtt agaatggcaat aagcttgcgt	3300
atttcttcc agacttgcgtt aacaggccatc ccattacgcgt cgtcatcaaa atcactcgca	3360
tcaacccaaac cgttattcatc tcgtgatgtt gcctgagcgtt gacgaaatc gcgatcgctg	3420
ttaaaaaggac aattacaaac aggaatcgaa tgcacccggc gcagggacac tgccagcgca	3480

-continued

tcaacaatat tttcacctga atcaggatata tcttctaata cctggaatgc tgtttcccg	3540
gggatgcag tggtagttaa ccatgcatca tcaggagtag ggataaaatg cttgtatggc	3600
ggaagaggca taaattccgt cagccagttt agtctgacca tctcatctgt aacatcattg	3660
gcaacgctac ctttgccatg tttagaaac aactctggcg catcgggctt cccatacaat	3720
cgtataggatg tcgcacctga ttgcccaca ttatcgcgag cccattata cccatataaa	3780
tcagcatcca tggtaatt taatcgccgc ctgcagcaag acgtttcccg ttgaatatgg	3840
ctcataaacac cccttgtatt actgtttatg taagcagaca gtttattgt tcatgtatgt	3900
atatttttat ctgtgcaat gtaacatcg agattttagg acacaacgtg gctttcccc	3960
ccccccctgc aggtcgatag cagcaccacc attaaatga tttcgaaat cgaaactgac	4020
atttggaaacga acatcagaaaa tagtttaag aaccttaatg gtttggctg tgattctg	4080
accaacgtgg tcacctggca aaacgacat ctttttaggg gcagacatta gaatggata	4140
tccttgaat atatatataat attgtgaaa tgtaaaagg aagaaaagg agaaaatg	4200
acgattgtcta accacctatt gaaaaaaca ataggtcatt aaataatatt gtcaacttca	4260
agtattgtga tgcaaggcatt tagtcatgaa cgcttctcta ttctatatga aaagccggtt	4320
ccggcgctct caccccttctt ttttctccca attttcagtt tgaaaaaggat atatgcgtca	4380
ggcgacctct gaaattaaca aaaaatttcc agtcatcgaa ttgattctg tgcgatagcg	4440
ccccctgtgt ttctcggtat gttgaggaaa aaaataatgg ttgctaagag attcgaactc	4500
ttgcatactta cgatacctga gtattccac agttaactgc ggtcaagata tttcttgaat	4560
caggcgcctt agaccgctcg gccaaacaac caattacttg ttgagaataa gagtataatt	4620
atccatataaa tataacgttt ttgaacacac atgaacaagg aagtacagga caattgattt	4680
tgaagagaat gtggattttg atgtattgt tggattcca ttttaataa ggcaataata	4740
tttaggtatgt agatataacta gaatgtctcc tcgacgctct cccttatgcg actcctgcat	4800
taggaagcag cccagtagta ggttgaggcc gttgagcacc gecgcgcgaa ggaatggtc	4860
atgcaggaggat gggcgccca acgtcccccc ggcacgggg cctgccccca tacccacgcc	4920
gaaacaagcg ctcatgagcc cgaagtggcg ageccgatct tccccatcg tgatgtcgcc	4980
gatataggcg ccagcaaccg cacctgtggc gccgggtatg cccgcacga tgcgccggc	5040
gttagaggatc tacaactcca cttattgtta ggttagattt tccgttagtt gtttattaaat	5100
tgcaataatg gggcgccag ttttggcaac agtgcatttc taccaggaca cctatgagtt	5160
tgcctcatgg ccaaactagag gtgttggaaat tgcattgttataatttgcgatgttgcatt	5220
accctctgaa tcctgcccgtt atccccatt gttcggtatc ttatgttttgcgatgttgcatt	5280
cattaagacg ttgcgttacc gtcatgcaat gcggttagtt atcgaagttt gatatccgc	5340
caatatacgg cgaacgcttg ttcttcaggta aagcatattt cccgcacccgc gcctctactt	5400
tctgcttgcata ctcattgttt tgagtgcgtt tttggataa cccgcacccgc tcagecttgc	5460
ttttgcctt agcgatccat gaagtcaatt ttttggatgtt ggttggccgc gcaacggccgg	5520
aaactgtatct ttttggatgtt ttaacttgcgtt acttcttattt ctttatttgcg acgtcatttgc	5580
gacaggggggat ggggtatca ttttgacatg ggggtgtggta aaaaaatataa aataaagccaa	5640
atgtcttagc gagaacacatc ttaacccctgg ttgcgcgtga agagatctt aatttgcatttgc	5700
ctatcagcgc attttggct tggatgtgcgtt agggccaaaaa ggtatgggtgtaa aaccggatc	5760

-continued

ggtttagcgcg acgttcacgg ttagtgcggta taacaatctc tacagacaga attcccttgt	5820
ttacagcttc acggaatgca cgaacgacgg ttgattggct ataaccagg tctgccgca	5880
tcaggcggtg aggcttgtga atgaagtatt cactgggtt tgccgcgaga tttgcacatt	5940
gcgacaggat atgcccggcg ctacggata gaccggagt tgttacaaag caggccaatt	6000
catagccaga aaaagtaaaa tcgctcatcg ttatacagct caggaaagtg actttagcca	6060
gcattacaat gctggtggtt cttactacgt ctgtagcgc gttgcgcga caggtaccag	6120
cacaccagca tcaagcaatc gcttcatcg ccactgctga ccttgcggg ttatacaggt	6180
cgtgaaagaa atccctgcttc cattgcttgt atcgatcacg gtttcttaa gggtgaataa	6240
cccacgagat atgtattttt gtttggggac gttcctgcgt tcaccgggtt cgatcagaat	6300
tccggttatca cgcaaccagg tgaagagata gtttggccc aggccgagca ctttggcata	6360
gttgcgcatt agaaccggc tggcggtago aacgcgttca gccaattcga cttaggtgc	6420
atccataagc attttttgct ccagccgttgc ctttgcgtt gccaggtcgg cagccaaacg	6480
gagagcttca gggagactct gccaatagc aggtttaat cttccggtgc gatagtgcgt	6540
aaatgtctgg tttacccatca gcccggaaacgc gggagaaatc cagccgttgc actccacagc	6600
gagcaattca tggccaaaag tgccgcgcgc acggccttcg acctgcaggc atgcaagctt	6660
ggcgtaatca tggctatagc tggccgttgc gtggaaattgt tatccgttca caattccaca	6720
caacatacga gcccggaaagca taaagtgtaa agectggggat gctaatgag tgagctaact	6780
cacattaatt gctggcgctt cactggccgc ttccagtcg gggaaacgtt cgtgcagggt	6840
agtccatgtt gtcactctc agtacaatct gctctgtatgc cgcatacgat agccagtata	6900
cactccgtca tgcgttacgtt actgggtcat ggctgcgcgc cgacacccgc caacacccgc	6960
tgcgcgcgc tgcggggctt gtctgttccc ggcatccgtt tacagacaag ctgtgcacgt	7020
ctccgggggc tgcgttgcgtt aggggttttcc accgttcatca ccggaaacgcg cgaggcagct	7080
ggggtaaaggc tcatcagcgtt ggctgttgcgtt cgattcacag atgtctgcct gttcatccgc	7140
gtccagtcg tggatgtttcc ccagaaggctt taatgtctgg cttctgttccaa agccggccat	7200
gttaaggggcg gtttttccct gtttggtcac tggatgcctcc gtttggggg gatttctgtt	7260
catgggggtt atgataccga tggaaacgaga gaggatgttc acgatacggg ttactgttgc	7320
tgaacatgcc cgggttactgg aacgttgcgtt gggtaaacaat cttggcggtat ggatgcggcg	7380
ggaccaggaga aaaatactc agggtaatccat ccagcgcttc gttataacag atgttaggtgt	7440
tccacagggtt agccaggcgc atccgtcgat gcagatccgg aacataatgg tgcaggccgc	7500
tgcgttccgc gtttccagac ttacggaaac acggaaacccg aagaccatcc atgttgcgtc	7560
tcaaggcgtca gacgttttcg agcagcgttgc gtttccatcg tgcgttgcgtt	7620
attctgttccaa ccaatgttgcgtt aaccccgccaa gcttgcgttgc gtttgcgttgc acaggagcac	7680
gatcatgcgc acccggtggcc aggacccaaatc gcttgcgttgc gtttgcgttgc tgcgttgcgtt	7740
ggagatggcg gacgttgcgttgc atatgttgcgtt ccaagggttgc gtttgcgttgc tcaatgttgc	7800
ccgcaagaat tggatggctt cttatcttgg agtgggtaaat cccgttgcgtt ggttgcgttgc	7860
gcttccatcc aggttcgttgc gggccggctt catgcaccgc gacgcaacgc gggggaggcag	7920
acaagggtata gggccggccgc tacaatccat gccaacccgtt tccatgttgcgtt cgcccgagcgt	7980
gcataaaatcg cccgttgcgttgc tggccgttca atgttgcgttgc ttaggttgcgtt aagagccgc	8040

-continued

agcgatcctt	gaagctgtcc	ctgtatggtcg	tcatctaccc	gcctggacag	catggcctgc	8100
aacgcgggca	tcccgtgcc	gccggaagcg	agaagaatca	taatggggaa	ggccatccag	8160
cctcgcgctcg	cgaacgccag	caagacgtag	cccagcgcgt	cggccgcct	gccggcgata	8220
atggcctgt	tctcgccgaa	acgtttggt	gccccaccag	tgacgaaggc	ttgagcgagg	8280
cgctgcaaga	ttccgaatac	cgcaagcgac	aggccgatca	tcgtcgcgct	ccagcgaaag	8340
cggtcctcgc	cgaaaaatgac	ccagagcgct	gccccacct	gtcctacgag	ttgcatgata	8400
aagaagacag	tcataagtgc	ggcgacgata	gtcatgcccc	gccccaccg	gaaggagctg	8460
actgggttga	ggctctcaag	ggcatcggtc	gagcttgaca	ttgttaggacg	tttaaacatt	8520
accctgttat	cccttaggatc	ctacgtatc	gatgaattcg	atcccatttt	tataactgga	8580
tctcaaaata	cctataaaacc	cattgttctt	ctcttttagc	tctaagaaca	atcaatttat	8640
aaatatatattt	attattatgc	tataatataa	atactatata	aatacattta	cctttttata	8700
aatacattta	cctttttttt	aatttgcatg	atttaatgc	ttatgctatc	ttttttat	8760
agtccataaa	acctttaaag	gacctttct	tatggatat	ttatattttc	ctaacaaggc	8820
aatcgccgtc	ataaaacttta	gttgcattacg	acgcctgtgg	acgtcccccc	cttccccctta	8880
cgggcaagta	aacttaggga	tttaatgca	ataaataaaat	ttgtcctctt	cgggcaaatg	8940
aatttttagta	tttaaatatg	acaagggtga	accattactt	ttgttaacaa	gtgatcttac	9000
cactcactat	ttttgttgaa	ttttaaactt	atttaaaatt	ctcgagaaag	atttaaaaaa	9060
taaacttttt	taatctttta	tttatttttt	ctttttcgat	atggattgc	ccaatattat	9120
tcaacaattt	atcgaaaca	gcgttttaga	gccaataaa	attggtcagt	cgccatcgga	9180
tgtttattct	ttaatcgaa	ataatgaaac	ttttttctt	aagcgatcta	gcactttata	9240
tacagagacc	acatacagtg	tctctcgta	agcgaaaatg	ttgagttggc	tctctgagaa	9300
attnaaggtg	cctgaactca	tcatgacttt	tcaggatgag	cagtttgaat	ttatgatcac	9360
taaagcgtac	aatgcaaaac	caatttcago	gctttttta	acagaccaag	aattgctgc	9420
tatctataag	gaggcactca	atctgttaaa	ttcaattgct	attattgatt	gtccatttat	9480
ttcaaaacatt	gatcatcggt	taaaagagtc	aaaatttttt	attgataacc	aactcctgta	9540
cgatatacat	caagatgatt	ttgacactga	attatgggg	gaccataaaa	cttacctaag	9600
tctatggat	gagttAACG	agactcgtgt	tgaagaaaga	ttgggttttt	ctcatggcga	9660
tatcacggat	agtaatattt	ttatagataa	attcaatgaa	atttattttt	tagaccttgg	9720
tcgtgcgtgg	ttagcagatg	attttgtaga	tatatccctt	gttgaacgtt	gcctaagaga	9780
ggatgcacatcg	gaggaaactg	cgaaaatatt	tttaaagcat	ttaaaaaatg	atagacctga	9840
caaaaggaat	tatTTTTAA	aacttgatga	attgaattga	ttccaagcat	tatctaaaat	9900
actctgcagg	cacgctagct	tgtactcaag	ctcgtaacga	aggtcggtac	cttgcgttg	9960
aagggtggcga	cgttaattcgt	tcagcttgc	aatggctcc	agaacttgc	gctgcattgc	10020
aagtttggaa	agaaattaaa	ttcgaatttg	atactattga	caaactttaa	ttttttat	10080
tcatgatgtt	tatgtgaata	gcataaacat	cgttttattt	tttatgggtgt	tttaggttaaa	10140
tacctaaca	tcattttaca	ttttaaaat	taagttctaa	atgtatctt	ttttttaaatt	10200
tgccctgtctt	tataaattac	gatgtgccag	aaaaataaaa	tcttagctt	ttattataga	10260
atttatcttt	atgttattata	ttttataagt	tataataaaa	gaaatagtaa	catactaaag	10320

-continued

cggatgtgc	gcgttatct	taacggaagg	aattcggcgc	ctacgtatac	atactccgaa	10380
ggaggacaaa	tttattttt	gtggtacaat	aaataagtgg	tacaataaat	aaattgtatg	10440
taaacccctt	ccccttcggg	acgtccccctt	acgggaatat	aaatattagt	ggcagttgcc	10500
tgccaacaaa	tttattttt	gtattaacat	aggcagtggc	ggtaccactg	ccactggcgt	10560
cctaataaa	atattgggca	actaaagttt	atcgagtt	taacataggc	agtggcgta	10620
ccactgcccac	tggcgtccctc	cttcggagta	tgtaaacctg	ctaccgcagc	aaataaattt	10680
tattctatTT	taatactaca	atatttagat	tcccgttagg	ggataggcca	ggcaattgtc	10740
actggcgtca	tagtatatac	atattgtaa	agattgacac	cctttaagta	aacattttt	10800
ttaggattca	tatgaaattt	aatggatatt	tggtacattt	aattccacaa	aaatgtccaa	10860
tacttaaaat	acaaaattaa	aagtattagt	tgtaaacttg	actaacatTT	taaattttaa	10920
atttttct	aattatata	tttacttgca	aaatttataa	aaattttatg	cattttata	10980
tcataataat	aaaaccttta	ttcatggttt	ataatataat	aattgtgt	actatgcaca	11040
aagcagttt	agtcccata	atataactat	atataacccg	tttaaagatt	tatTTaaaa	11100
tatgtgtgt	aaaaatgcTT	atTTtaatt	ttatTTata	taagttataa	tatTTatac	11160
acaatgatta	aaattaaata	ataataaatt	taacgtaacg	atgagttgtt	tttttatttt	11220
ggagatacac	gcaatgacaa	ttgegatcg	tacatatacaa	gagaaacgc	catggttcga	11280
tgacgctgat	gactggcttc	gtcaagacgg	tttcgtattt	gtaggTTgg	caggTTattt	11340
actattccct	tgtgcttact	ttgcaactcc	ggtccggcgg	ccgcctcgag	acgacttgc	11400
cgottcatca	gacacggctt	tcctaaccat	caatggtgg	ttttcaggaa	agacgtttaa	11460
agaagtggc	taaagtttat	ttgttgaaga	attggTTTT	tttccattca	aagaattgtt	11520
agggataaaa	ctttgcattt	ttttataatt	tgttataagt	ttttcaact	tatatgttt	11580
taaaaatgc	ttaattgtc	tattaatgc	ttcatTTTGT	aatgtttca	taggtctgc	11640
ttgcgcata	cgcagttt	ccgataactt	gtctgcttgc	tttccggta	ttgagaataa	11700
gtaaagtata	tgatTTaaa	aagtcatgtt	ttgattaaat	ctttttata	tggtttaaaa	11760
cattatggta	tatctaaata	aatttatttt	ttactaaatc	tccaaTTgc	aatttagaga	11820
tataattaaa	actataaaat	tatTTatgtt	aatttgcata	caaATCCAAC	acaaaatgt	11880
ttttatata	ttaacatgtt	aaatttaaca	tatgttaac	aactaaaatt	ctgtAACAGA	11940
gaacaataaa	ataaatgc	gattttgtgt	aatGCCGAAG	tatattata	tactccctt	12000
tcaaaaaat	aaatactctt	gccactaaaa	ttcatttgcc	taggacgtcc	ccttccctt	12060
acgggatgtt	tatatactag	gacgtcccc	tccccttacg	ggatattttat	atactccgaa	12120
ggacgtcccc	ttcgggcaaa	taaattttag	tggcagttgc	ctgccaactg	cctaggcaag	12180
taaaacttag	gattttatg	caataataa	atttgcTTCC	ttacgggacg	tcagtggcag	12240
ttgcctgc	actgccta	ataaatatta	gtggatattt	atatactccg	aaggaggcag	12300
ttacctgc	actGCCGAGG	caaataaatt	ttagtggcag	tggTaccGCC	actgcctgc	12360
ccctccTTcc	cTTcgggca	agtaaaactt	gcatgttgc	gacattaccc	tgttatccct	12420
aggccggcct	aagaaaccat	tattatcatg	acattaac	ataaaaatag	gcgtatcag	12480
aggcccttcc	gtcttcaaga	aattcggtcg	aaaaaagaaa	aggagaggc	caagaggag	12540
ggcattgg	actattgagc	acgtgagtt	acgtgattaa	gcacacaaag	gcagcttgg	12600

-continued

```

gtatgtctgt tattaatttc acaggttagtt ctggtccatt ggtaaaagtt tgcccgtgc 12660
agagcacaga ggccgcagaa tgtgctctag attccgatgc tgacttgctg ggtattata 12720
gtgtgcccaa tagaaagaga acaattgacc cggttattgc aaggaaaatt tcaagtcttg 12780
taaaagcata taaaatagt tcaggactc cgaatactt ggttggcgtg tttcgtaatc 12840
aacctaagga ggatgtttt gctctggta atgattacgg cattgatatac gtccaaactgc 12900
atggagatga gtcgtggcaa gaataccaag agttcctcg tttgccagtt attaaaagac 12960
tcgtatttcc aaaagactgc aacataactac tcagtgcagc ttcacagaaa cctcattcgt 13020
ttattccctt gtttGattca gaagcagggtg ggacagggtg actttttggat tggaaactcga 13080
tttctgactg ggttggaaagg caagagagcc ccgaaagctt acattttatg ttagctggtg 13140
gactgacgcc agaaaatgtt ggtgtatgcgt ttagattaaa tggcggttatt ggtgttgatg 13200
taagcggagg tggagaca aatggtgtaa aagactctaa caaaaatagca aatttcgtca 13260
aaaatgtcaa gaaatagggtt attactgagt agtatttttta taagtattgt ttgtgcactt 13320
gcctgcaggc ctttgaaaaa gcaagcataa aagatctaaa cataaaatct gtaaaataac 13380
aagatgtaaa gataatgcta aatcatttgg ctttttgatt gattgtacag gaaaatatac 13440
atcgcagggg gttgactttt accatttcac cgcaatggaa tcaaacttgt tgaagagaat 13500
gttcacaggc gcatacgccta caatgaccgc attcttgcta gcctttctc ggtcttgc当地 13560
acaaccgcgc gcagcttagt atataaatac acatgtacat acctctctcc gtatcctcgt 13620
aatcattttc ttgtattttat cgtttttcg ctgtaaaaac tttatcacac ttatctcaaa 13680
tacacttatt aaccgtttt actattatct tctacgctga cagtaatatac aaacagtgc当地 13740
acatattaaa cacagtggtt tctttgcata aacaccatca gcctcaagtc gtcaagtaaa 13800
gatttcgtgt tcatgcagat agataacaat ctatatgtt ataatttagcg ttgcctc当地 13860
aatgcgagat ccgttaacc ggaccctagt gcacttaccc cacgttccgtt ccactgtgt 13920
ccgaacatgc tccttcacta ttttaacatg tggattaaat tctaaatcct ctttatatga 13980
tctggccgata gatgttcta agtcattgag gttcatcaac aattggattt tctgtttact 14040
cgacttcagg taatgaaatg agatgatact tgcttatctc atagttact ggcataaatt 14100
tttagtatagg ttaactctaa gaggtgatac ttattttactg taaaactgtg acgataaaac 14160
cggaaggaag aataagaaaa ctgcgaactga tctataatgc ctatccctg taaagagttt 14220
aagctatgaa agcctcgca ttttggccgc tccttaggtg tgctttttt ccaaggacaa 14280
aacagttct ttttcttgag caggtttat gtttcggtaa tcataaaca taaataaatt 14340
atttcattta tggttaaaaa taaaaataa aaaagtattt taaatttttta aaaaagtgt 14400
ttataagcat gtgacccccc gcaagcaatt aaattttgca atttgcattt taggcaaaag 14460
ttactatttc tggctcgtgt aatataatgtt tgctaatgtt aacttttaca aagtgc当地 14520
ggacttagtc aaaagaaatt ttctttaaaaa tatatacgac tagccaaattt agcacttctt 14580
tatgagatatttact ttatataagcc agattttgtt attatatgtt tttacccggc 14640
gaatcatgga catacattct gaaataggta atattctcta tggtgagaca gcatagataa 14700
cctaggatac aagttaaaag ctagtactgt tttgcagtaa ttttttctt ttttataaga 14760
atgttaccac ctaaataagt tataaagtca atagttaaat ttgatatttgc attgtaaaat 14820
accgtaatat atttgcatttca taaaaggct caatgttgcac tagccagcat gtcaaccact 14880

```

-continued

atattgtatca ccgatattag gacttccaca ccaacttagta atatgacaat aaattcaaga	14940
tattcttcat gagaatggcc cagctcatgt ttgacagctt atcatcgata agctttaatg	15000
cggtagttt taacagttaa attgctaacg cagtcaggca ccgtgtatga aatctaacaa	15060
tgcgctcatc gtcatcctcg gcaccgtac cctggatgt gttaggcata gcttggttat	15120
ccgggtactg ccgggcctct tgccggatata cgccattcc gacagcatcg ccagtacta	15180
tggcgtctg cttagcgtat atgcgttata gcaatttctta tgccgcacccg ttctcggagc	15240
actgtccgac cgcttggcc gccgcccagt cctgctcgat tcgctacttg gagccactat	15300
cgactacgcg atcatggcga ccacacccgt cctgtggatc aattcttag tataaatttc	15360
actctgaacc atcttggaaag gaccggataa ttatggaaa ttctttttc aattgtatat	15420
gtgttatgtatca gtaatactttt tcttcaacaa ttaaataactc tcggtagcca agttggttta	15480
aggcgcaaga cttaattta tcactacgga attggcctat taggcctacc cactagtcaa	15540
ttcggggagga tcgaaacggc agatcgcaaa aaacagtaca tacagaagga gacatgaaca	15600
tgaacatcaa aaaaattgtatca aaaaattgtatca aaaaattgtatca aaaaattgtatca	15660
caggaggagc gactcaagcc ttccggaaag aaaataacca aaaaggcatac aaagaaacgt	15720
acggcgttcc tcataattaca cgccatgata tgctgcagat ccctaaacag cagaaaaacg	15780
aaaaataccaa agtgcctcaa ttccgtatcaat caacgattaa aaatattgag tctgcaaaag	15840
gacttgatgt gtgggacagc tggccgctgc aaaacgctga cggAACAGTA gctgaataca	15900
acggctatca cgttgtgttt gctttgcgg gaagccccgaa agacgcgtat gacacatcaa	15960
tctacatgtt ttatcaaaag gtcggcgaca actcaatcga cagctggaaa aacggggcc	16020
gtgtctttaa agacagcgat aagttcgacg ccaacgatcc gatcctgaaa gatcggacgc	16080
aagaatggtc cgggttgc acctttacat ctgacggaaa aatccgttta ttctacactg	16140
actattccgg taaacattac ggcaaacaaa gcctgacaac agcgcaggta aatgtgtcaa	16200
aatctgtatca cacactcaaa atcaacggag tggaaatgtatca caaaacgatt tttgacggag	16260
acggaaaaac atatcagaac gttcggacgt ttatcgatc aggcaattat acatccggcg	16320
acaaccatac gctgagagac cctcactacg ttgaagacaa aggccataaa taccttgtat	16380
tcgaagccaa cacgggaaca gaaaacggat accaaggcga agaatctta tttaacaaag	16440
cgtactacgg cggccggcacg aacttcttc gtaaagaaag ccagaagctt cagcagacgc	16500
ctaaaaaaacg cgtatgtatca gtagtgcacg gcccctcgat tatcatagat ttaaataatg	16560
attacacatt gaaaaaaatgtatca gtagtgcacg gcccctcgat tatcatagat ttaaataatg	16620
tcgagcgcgc gatgttttc aaaatgtatca gcaaatggta ctgttcaact gattcagcgc	16680
gttcaaaaat gacgtatgtatca ggttataact caaacgtat ttagtgcctt gtttatgtat	16740
caaaactttt aaccggccct tacaagccgc tgaacaaaac agggcttgcg ctgcaaatgg	16800
gtcttgcattcc aaacgtatgtatca gtagtgcacg gcccctcgat tatcatagat ttaaataatg	16860
gcaacaatgt ggttataact caaacgtat ttagtgcctt gtttatgtat	16920
caacatttcg gccaagcttc ttaatgtatca tcaaaaggaa taaaacatcc gttgtcaaaa	16980
acagcatcctt ggaccaagga cagctgacag tcaactaata acagcaaaaa gaaaatggcg	17040
atacttcattt ggcattttctt tttatgtatca aacaatgtatca gtagtgcacg gttgtcaaaa	17100
tccacaggac ggggtgtggc gccatgatcg cgtatgtatca gtagtgcacg gttgtcaaaa	17160

-continued

cgagcaggac tggccggcg ccaaaggcgt cgacatgc tccgagaacg ggtgcgcata 17220
gaaatttgcat caacgcatat agcgctagca gcacgcata gtgactggcg atgctgtcg 17280
aatggacgat atcccgcaag aggccccgca gtaccggcat aaccaacgcct atgcctacag 17340
catccagggt gacggtgccg aggatgacga tgagcgcatt gtttagatttc atacacggtg 17400
cctgactgcg ttagcaattt aactgtgata aactaccgca tttaaagctta tcgatgataa 17460
gtctgtcaac atgagaatgt atccggaacc ctaatataa ctgcgtataa tggatgttat 17520
acgaaggat taggtccctc gactacgtcg ttaaggccgt ttctgacaga gtaaaattct 17580
tgagggaaact ttcaaccat tggaaatgg ttcaagaagg tattgactta aactccatca 17640
aatggtcagg tcattgagtg ttttttattt gtgttattt ttttttttag agaaaatct 17700
ccaatataata aatttaggaat catagttca tgatTTTCTG ttacaccta ctttttgt 17760
ggtgccctcc tccttgtcaa tattaatgtt aaagtgaat tcttttcct tatcacgtt 17820
agccattagt atcaatttgc ttacctgtat tcctttacat ctcctttt ctccttctt 17880
ataaatgtat gtagattgcg tatatagttt cgtctaccct atgaacatata tccatTTGT 17940
aatttcgtgt cgtttctatt atgaatttca ttataaagt ttatgtacaa atatcataaa 18000
aaaagagaat ctttttaaagc aaggatttcc ttaacttctt cggcgacagc atcaccgact 18060
tcgggtgtac tggggaaacc acctaatac ccagttctga tacctgcata caaaacctt 18120
ttaactgcattt cttcaatggc cttaccttct tcaggcaagt tcaatgacaa tttcaacatc 18180
attgcagcag acaagatagt ggcgataggg ttgaccttacat tctttggcaa atctggagca 18240
gaaccgtggc atgggtcgta caaaccaat ggggtttct tggctggcaa agaggccaag 18300
gacgcagatg gcaacaacc caagaaacctt gggataacgg aggtttcatc ggagatgata 18360
tcaccaaaaca tgggtgtgtt gattataata ccatttaggt gggttgggtt cttactagg 18420
atcatggcg cagaatcaat caattgtatgt tgaaccttca atgttagggaa ttcgttctt 18480
atggtttccctt ccacagtttt tctccataat cttgaagagg ccaaaacatt agctttatcc 18540
aaggaccaaa taggcaatgg tggctcatgt tggtagggcca tggaaagcgcc cattttgt 18600
attctttgca cttctggAAC ggtgtattgt tcactatccc aagcgacacc atcaccatcg 18660
tcttccttcc tcttaccaaa gtaataaccc cccactaattt ctctgacaac aacgaagtca 18720
gtaccttttag caaattgtgg cttgatttga gataagtcta aaagagagtc ggatgcaaag 18780
ttacatgggt ttaagtggc gtacaatttga agttttttac ggatTTTtag taaaaccttgc 18840
tcaggtctaa cactaccggc acccgccgtt cggaaacgcac agcaaaataag gccagcacag 18900
ccagcgcaagt tgcccggtt gttctgtattt ctgttcttgc tgaattaaac aaaggccgcac 18960
agtaacaaag gacttcattt ataatttttcc ttcaaggagga agacatgtca ttctttctta 19020
cgtaaaaaac agctttgtct tggtagggaga aacttgcgtc tactgggtt cttgttctga 19080
tttgcgeact tgggtgtgtt ggggttgcattt gggaaacgtca tcagctaaag caagccatag 19140
agaaaatttgg cagtcatttgcattt cggctgttta aggaacgtga taagtcata atggatctta 19200
accagaccat tgagacgtatg aacaaaggcag agcaacattt tcacagccag gaagtgaaaa 19260
atgaatcaga acaaggcaag tatgtgtaca ggcaatggaa acgaaaagct gaagttcaga 19320
aacaactggt tgggggggtt aatgttcgcc agcgttattcc tgcgtgacact cagcggttgc 19380
tccggggactt gatcagcgaa tttaacggcg acggccgacaa aggttaacca ccctgtcccc 19440

-continued

```

aaaagtgcattatgttag gatgccagag ttttagcagtg aatatttga tgatctgcc 19500
gcgtatcc tcgatacaga aacgatgctg atggggatta acaggaagaa tcgcaacgtt 19560
aatgattaca accgagctat tagcggtaac taaaagggt ttttatgtct gataaagtaa 19620
cagtaaagca aactatcaac aaagcgactt caatctacaa aattgagcaa atcactgtt 19680
gcaagccagg atctgaacaa taccgtcgtg ctttcgagct tgccgatcag ctgggtttaa 19740
aacacccgga ttgcatttgag catgtatttc cgacctatgc tgatgagcaa tgtactcatg 19800
ttcttaccga agaggatttt ttcagcactg aagaacgaga aggcgtttagt cgctgcattt 19860
gtgtgatttg ttcttcggta agtgtatgatg tattccctaa tggccctgaa tatgggttta 19920
ttggatacca attcctgtac gagggcgatg agcttaatg ctatgaacat ggtttctca 19980
tcgaaagcgt agaataatac gactcccttc caaccggcta cgttggccgg tttttactt 20040
atccacatata tccactggat agatccaata atcagggtcca tacagatccc aatttagatcc 20100
atatacatcc ctgatcgttgc caggccgcgc cacgtctggc ttagaagtgt atcgcgtatgt 20160
gtgtgtggagg gaaaacgatg tgggtgtggag ggataaaaaat gtgtgtgtgac ggggtgttaa 20220
tgtgtgtggc cgggatatacgat gatgtgtgtt gacgggaaag cttgggttagt tatcaccact 20280
tataaaaaact atccacacaa ttccggaaaaa gtaatatgaa tcaatcatat atctccgata 20340
ttctttacgc agacattgaa agtaaggcaaa aagaactaac agttaattca aacaacactg 20400
tgcaagcctgt agcgtttagt cgcttggggg tattcgtgccc gaagccatca aagagcaaaag 20460
gagaaaagtaa agagattgtatg gccacccaaag cgttttccca gttggagata gctaaagccg 20520
agggttacga tgatattaaa atcacccggtc ctgcactcgat tatggataact gatttcaaaa 20580
cgtggatccgg tggcatctac gcgttcagca aatacggctt gtcctcaaac accatccagt 20640
tacgtttca ggaattcgct aaagoctgtg gttcccttc aaaacgtctg gatgcgaaac 20700
tgcggttaac cattcatgaa tcacttggac gcttgcgtaa caagggtatc gcttttaagc 20760
gcggaaaaga tgctaaaggc ggctatcaga ctgggtctgct gaagggtggg cgttttgtatg 20820
ctgaccccttgc tctgtatagag ctggggctgtt attcgttgcattt gttccagctt 20880
attatcgctgt tctgttgcaaa caccacgcct tgcgtgcctt tccgaagaaa gaagtcac 20940
aaggccatata cactttcatc gaaaggccctc cgcagaaccc gttgcgcgtt tcgttgcgc 21000
gaatccgtga ggcgttgcgtt ttgcgttgcgtt ctgttgcgtt gcaaaaacgtt atcattaaga 21060
aagcgtatgtt acaatcggtt atctcgactt ttcttatttgcgtt aagaaaggcc 21120
ggaaaagttt tggatcgatc catttcgtca atccaaagctt gaaactcccc gaaataagtgt 21180
gtgtgtggagg gaaacccgtt taaaaggatg tgggtgtggc gggaggctt tccaaattcc 21240
tggggatgtt gtcgttgcgtt gggggacgcgc cctcgtttt ccctccagca 21300
cacatctgtc catccgtttt tccctccagt gcacatgtttt ttctctgcctt tccctccag 21360
cacacatatt tgataccagc gateccctca cagcacataa ttcaatgcgtt cttccctcta 21420
tcgcacatct tagacttttta ttccctccatc agcacacatc gaagctgcgg ggcaaggccgt 21480
tctcaccaggatgtt tgatagagag tggatcgatc ctggccattt aagcaggaaa tcaccaaaaat 21540
gattcaggctt acaacccgtt cgttgcgtt gggggacgcgc cctcgtttt ccctccagca 21600
aagcgtatgtt tgacacacactt ggcaaggag taagcatgcgtt gagaatgcgtt tggatgcgtt 21660
tctacgcata cattattttt ttatgcgtt aattcaaaaaa tacagcataa 21720

```

-continued

```

aggatgactt tcgatgagtg attccagcca gcttcacaag gttgctcaaa gagcaaacag 21780
aatgctcaat gttctgactg aacaagtaca gttgaaaag gatgagctac acgcgaacga 21840
gttttaccag gtctatgcga aagcggact ggcaaaattt cctctactga ctcgagcgaa 21900
cggttactat gccgtaagtg aaatggaga aaagggttat gtttcgata aacgcctgc 21960
tggctttca atgaaatatg cgatgtcaat tcagaacatc attgacatcat atgaacatcg 22020
cggagtgcca aaataccggg atcgcgtacag cgaagcgtat gtgatttca tctccatct 22080
taaaggcggt gtgtcaaaaa ctgtatcgac ggtttctctg ggcgtatgca tgctgtctca 22140
ccctcatctt cttatggagg atttaaggat tctggttattt gaccttgcgtc cgcaatctc 22200
agcaacgatg ttttaagcc ataaacactc tattggtac gtaaacgcaa catctgcaca 22260
ggctatgtt cagaatgtaa gccgtgaaga gctgttagag gagtttattt ttccctctgt 22320
tgtacctggg gttgacgtta tgcctgcgtc gattgacgt gcctttattt catccgattt 22380
gagagagctg tgcaatgagc atctaccggg tcagaacatc catgcgttcc tgaaagaaaa 22440
tgtgattgtt aagctgaaga gcgattatga ctttacccctt gttgatagtgt gtccctcacct 22500
tgacgccttc ctgaaaaatg ctttggccctt ggccaaatata ctgtttacac ctctggcc 22560
agcaactgtc gatttccact catcgcttaa atacgttgcc cgccttcctg agttggtaa 22620
actcatttcg gatgaggct gcgagtgcca gcttgcgtactt aacattggtt ttatgtccaa 22680
gtttagttaa aaggcagacc ataagtattt ccatagcctt gctaaagaag tgttcggtgg 22740
ggatatgtttt gatgtttcc tccctcgccct tgacggttttt gaacgctgctt gcgagtcttt 22800
tgacactgtt atttcagcta accccggcaac gtatgttggt agtgctgtatg cattgaagaa 22860
cgcgcaattt ggcggaaag attttgcataa agcagttttt gaccgtat 22908

```

```

<210> SEQ ID NO 2
<211> LENGTH: 70
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

```

```
<400> SEQUENCE: 2
```

```

tgacagctta tcatacgattt tctgcattt atccgcttat tatcactaag aaaccattt 60
tatcatgaca 70

```

```

<210> SEQ ID NO 3
<211> LENGTH: 70
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

```

```
<400> SEQUENCE: 3
```

```

ggcagttattt ggtgccctta aacgcgttggt tgctacgcctt gaataaaattt ccgtatgtat 60
aaattaaagt 70

```

```

<210> SEQ ID NO 4
<211> LENGTH: 7507
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic polynucleotide

```

-continued

<400> SEQUENCE: 4

gcggccgcaa	ggggttcgcg	tcagcgggtg	ttggcgggtg	tcggggctgg	cttaactatg	60
cggcatcaga	gcagatttgc	ctgagagtgc	accatatgcg	gtgtgaaata	ccgcacagat	120
gcgttaaggag	aaaataccgc	atcaggcgcc	attcgccatt	caggctgcgc	aactgttggg	180
aaggggcata	ggtgcgggca	tcttcgtat	tacgccagct	ggcgaaaggg	ggatgtgctg	240
caaggcgcatt	aagtgggta	acgcagggt	tttcccagtc	acgacgttgt	aaaacgcacgg	300
ccagtgaatt	gtaatacgac	tcactatagg	gcgaattcga	gctcggtacc	cggggatcct	360
ctagagtcga	cctgcaggca	tgcaagcttg	agtattctat	agtgtcacct	aaatagcttg	420
gcgttaatcat	ggtcatagct	gtttcctgtg	tgaaattgtt	atccgctcac	aattccacac	480
aacatacgcg	ccggaaagcat	aaagtgtaaa	gcctggggtg	cctaattgagt	gagctaactc	540
acattaatttgc	cggtgcgc	actgcggcgt	ttccagtcgg	gaaacctgtc	gtgccagctg	600
cattaaatgaa	tcggccaacg	cgaacccctt	gccccggcc	ggccgtcga	ccaattctca	660
tgtttgacag	tttatcatcg	aatttctgc	attcatccgc	ttattatcac	ttattcaggc	720
gttagcaacca	ggcggttaag	ggcaccaata	actgccttaa	aaaaatttacg	ccccggcctg	780
ccactcatcg	cagttactgtt	gtaattcatt	aagcattctg	ccgacatgg	agccatcaca	840
aacggccatga	tgaacctgaa	tcgcgcagcg	catcagcacc	ttgtcgcc	gcgtataata	900
tttgcggat	gtgaaaacgg	gggcgaagaa	gttgtccata	ttggccacgt	ttaaatcaaa	960
actgggtaaa	ctcacccagg	gattggctga	gacgaaaaac	atattctcaa	taaacccttt	1020
agggaaatag	gcgcagggttt	caccgtaaaca	cgccacatct	tgcgaatata	tgtgtagaaa	1080
ctgcggaaaa	tgcgtcggt	attcaactca	gagcgtgaa	aacgtttcag	tttgcgtatg	1140
aaaaacgggt	taacaagggt	gaacactata	ccatatacacc	agctcacccgt	ctttcatgc	1200
catacggaaat	tccggatgag	cattcatcg	gcgggcaaga	atgtgaaataa	aggccggata	1260
aaacattgtgc	ttattttct	ttacggtctt	taaaaaggcc	gtaatatcca	gctgaacgg	1320
ctgggttag	gtacatttgc	caactgactg	aaatgcctca	aatgttctt	tacgtgcca	1380
ttgggatata	tcaacgggt	tatattccgt	gattttttc	tccattttag	cttccttagc	1440
tcctgaaaat	ctcgataact	aaaaaaatac	gccccgttagt	gatcttattt	cattatgg	1500
aaagttggaa	cctcttacgt	gcgcgtcaac	gtctcattt	cgccaaaagt	tggcccgagg	1560
ctccccggta	tcaacaggga	caccaggatt	tatttattct	gcgaagtgt	cttcgtcac	1620
aggtatttat	tcgcgtataag	ctcatggago	gggtgtaccc	tcgcacagga	aggacagaga	1680
aagcgcggat	ctggaaagt	acggacagaa	cggtcaggac	ctggattggg	gaggegggt	1740
ccggccgtgc	tgctgacgg	gtgacgttct	ctgttccgg	cacaccat	acgttccg	1800
attcctatgc	gatgcacatg	ctgtatgcgc	gtataccgc	gaaagtctg	caaagcctga	1860
tgggacataa	gtccatcgt	tcaacggaa	tctacacgaa	ggttttgcg	ctggatgtgg	1920
ctggccggca	ccgggtgcag	tttgcgtgc	cgaggctgc	tgccgtgcg	atgctgaaac	1980
aattatcctg	agaataaaatg	ccttggcctt	tatatggaaa	tgtgaaactg	agtggatatg	2040
ctgtttttgt	ctgttaaaca	gagaagctgg	ctgttatcca	ctgagaagcg	aacgaaacag	2100
tcgggaaaat	ctcccattat	cgttagagatc	cgcatattat	atctcaggag	cctgtgtatc	2160
gtttatagga	agtagtgttc	tgtcatgtat	cctgcaagcg	gtaacgaaaa	cgatttgaat	2220

-continued

atgccttcag gaacaataga aatcttcgtg cggtgttacg ttgaagtgga gcggattatg	2280
tcagcaatgg acagaacaac ctaatgaaca cagaaccatg atgtggctcg tcctttaca	2340
gccagtagtg ctcgcccag tcgagcaca gggcgaagcc ctcgagttag cgaggaagca	2400
ccagggaaaca gcacttatat attctgctta cacacgtgc ctgaaaaaac ttcccttggg	2460
tttatccact tatccacggg gatattttta taattatttt ttttatagtt ttttagatctt	2520
cttttttaga ggcgcctgtta ggcctttatc catgctgggtt ctagagaagg tggtgtgaca	2580
aattgcctt tcagtgacaa aaatcacccct caaatgacag tcctgtctgt gacaaattgc	2640
ccttaaccct gtgacaaattt gccctcagaa gaagctgttt tttcacaaag ttatccctgc	2700
ttattgactc tttttatattt agtgtgacaa tctaaaaact tgtcacactt cacatggatc	2760
tgtcatggcg gaaacagcgg ttatcaatca caagaaacgt aaaaatagcc cgcgaatcgt	2820
ccagtcaaac gacctcaactg aggccgcata tagtctctcc cgggatcaaa aacgtatgt	2880
gtatctgttc gttgaccaga tcagaaaatc tcatggacc ctacaggaac atgacggat	2940
ctgcgagatc catgttgcta aatatgctga aatattcggg ttgacctctg cggaaaggcag	3000
taaggatata cggcaggcat tgaagagttt cggggggaaag gaagttgggtt tttatcgccc	3060
tgaagaggat gcccggatg aaaaaggcta tgaatctttt ctttgggtta tcaaaccgtgc	3120
gcacagtcca tccaggggc ttacagtgtt acatatacAAC ccataatctca ttcccttctt	3180
tatcggttta cagaaccgggt ttacgcgtt tcggcttagt gaaacaaaag aaatcaccaa	3240
tccgtatgcc atgcgtttat acgatccct gtgtcgtat cgtaaagccgg atggctcagg	3300
catcgctct ctgaaaatcg actggatcat agagcggtac cagctgcctc aaagttacca	3360
gcgtatgcct ctgcgttccgc gcccgttccgc gcaggctgtgt gttaatgaga tcaacagcag	3420
aactccaaatcg cgcctctcat acattgagaa aaagaaaggc cggccagacga ctcataatcg	3480
atttcccttc cgcgatatac cttccatgac gacaggatag tctgagggtt atctgtcaca	3540
gatttgaggg tgggtcgta catttggttt gacctactga gggtaatttg tcacagttt	3600
gtgtttctt tcagctgca tggattttct catacttttta gactgtaat tttaaggaa	3660
gccaaatttg agggcagttt gtcacagttt atttcccttc ctttcccttc gtcatgtgac	3720
ctgatatacg gggtagtttc gtcatcattt atgagggtttt attatcacag ttattactc	3780
tgaattggct atcccggtgt gtacctctac ctggagttt tcccaeggtt gatatttctt	3840
cttgcgtga gcgtaaagac tatctgacag aacagttttt ctttgcgttcc tcggcgttcc	3900
gctcgctatg ctccgttaca cggctgcggc gagcgcttagt gataataagt gactgaggta	3960
tgtgcttc ttatctcctt ttgttagtggt gctcttattt taaacaactt tgccgttttt	4020
tgtatgactt tgcattttgt tgggtctttt cagtaaatttca agatattaa taaaaaaacg	4080
caaagcaatg attaaaggat gttcagaatg aaactcatgg aaacacttaa ccagtgcata	4140
aacgctggtc atgaaatgac gaaggctatc gccattgcac agtttaatga tgacagcccg	4200
gaagcgagga aaataaccgg cgcgtggaga ataggtgaag cagcggattt agttgggtt	4260
tcttctcagg ctatcagaga tgccgagaaa gcaggccgac taccgcaccc ggatatgaa	4320
attcgaggac gggttgagca acgtgttgggt tatacaattt aacaaattaa tcatatgcgt	4380
gatgtgtttt gtagcgcattt ggcacgttgc gaaagacgtat ttccaccgtt gatcgggtt	4440
gctgcccata aagggtggcgt ttacaaaacc tcagttctg ttcatcttgc tcaggatctg	4500

-continued

gctctgaagg ggctacgtgt tttgctcgta gaaggtaacg acccccaggg aacagcctca	4560
atgttatcacg gatgggtacc agatttcatt attcatgcac aagacactct cctgccttc	4620
tatcttgggg aaaaggacga tgtcaacttat gcaataaaac ccacttgctg gccggggctt	4680
gacattatttc ctccctgtct ggctctgcac cgtattgaaa ctgagttaat gggcaaattt	4740
gtgaaggta aactgcccac cgatccacac ctgatgctcc gactggccat tgaaactgtt	4800
gctcatgact atgatgtcat agtattgtac agcgcgctt acctgggtat cggcacgatt	4860
aatgtcgtat gtgctgctga tgtgctgatt gttcccacgc ctgctgagtt gtttgactac	4920
acctccgcac tgcagttttt cgatatgctt cgatgtcgc tcaagaacgt tgatctaaa	4980
gggttcgagc ctgatgtacg tattttgctt accaaataca gcaatagtaa tggctctcag	5040
tccccgtgga tggaggagca aattcgggat gcctggggaa gcatggttct aaaaaatgtt	5100
gtacgtgaaa cggatgaagt tggtaaaggc cagatccgga tgagaactgt ttttgaacag	5160
gccattgtac aacgctcttc aactgggtgcc tggagaaatg ctctttctat ttggaaacct	5220
gtctgcaatg aaattttcga tcgtctgatt aaaccacgct gggagattag ataatgaagc	5280
gtgcgcctgt tattccaaaa catacgctca atactcaacc ggttgaagat acttcgttat	5340
cgacaccagc tgccccgatg gtggattcgt taattgcgcg cgttaggagta atggctcgcg	5400
gtaatgcccata tactttgcct gtatgtggc gggatgtgaa gtttactctt gaagtgcctc	5460
gggggtgatag tggtaaaggc acctctcggt tatggtcagg taatgaacgt gaccaggagc	5520
tgcttactga ggacgcactg gatgtctca tcccttctt totactgact ggtcaacacaga	5580
cacccggcgtt cggcgtcaaga gtagtggcgt tcataaaaa tggcgatggg agtcggcgtc	5640
gtaaagctgc tgcacttacc gaaagtgatt atcgtgttctt ggttggcgag ctggatgtat	5700
agcagatggc tgcattatcc agattgggtt acgattatcg cccaaacaagt gcttataaac	5760
gtgggtcagcg ttatgcacgc cgattgcaga atgaatttgc tggaaatatt tctgegctgg	5820
ctgatgcggaa aaatatttca cgtaagatta ttacccgcgtt tatcaacacc gccaaattgc	5880
ctaaatcgt tggatgtctt ttttctcacc cccggtaact atctgcggcg tcaggatgt	5940
cacttcaaaa agcctttaca gataaagagg aattacttaa gcagcaggca tctaacccttc	6000
atgagcagaa aaaagctggg gtgtatattt aagctgaaga agttatcaact cttttaactt	6060
ctgtgcttaa aacgtcatct gcatcaagaa ctgtttaag ctcacgcacat cagttgttc	6120
ctggagcgc acgtattgtat aaggcgata aaatgggtcgtt taacctggac aggtctcggt	6180
ttcccaactga gtgtatagag aaaattgggg ccattcttaa ggaacttgaa aagccagcac	6240
cctgatgcga ccacgttttta gtctacgttt atctgtcttt acttaatgtc ctggatgttaca	6300
ggccagaaaag cataactggc ctgaatatttcc tctctggggcc cactgttcca ctgtatcg	6360
cggtctgata atcagactgg gaccacggtc ccactcgatcg egtcggctcg attattatcg	6420
tgggaccacg gtcccaactcg tategtcggt ctgattatttta gtctggggacc acggccac	6480
tcgtatcgatc ggtctgataa tcagactggg accacgggtcc cactcgatcg gtggatcg	6540
ttattatgtct gggaccatgg tcccactcgat atcgtcggtc tgattattttag tctggggacca	6600
cggtcccact cgtatcgatcg gtctgattat tagtctggaa ccacgggtccc actcgatcg	6660
tcggatcgat tattatgtctg ggaccacgggtt cccactcgatcg tcgtcggtctt gattatttagt	6720
ctgggaccac gatcccactc gtgttgcgg tctgattatc ggtctggggac cacggccca	6780

-continued

cttgcattgt cgatcagact atcagcgtga gactacgatt ccatcaatgc ctgtcaaggg	6840
caaggattga catgtcgtag taacctgttag aacggagtaa cctcggtgtg cggttatgttg	6900
cctgctgtgg attgctgctg tgcgtctt atccacaaca ttttgcgcac gggttatgtgg	6960
acaaaatacc tggttaccca ggccgtgccc gcacgtaaac cgggctgcat ccgatgcaag	7020
tgtgtcgctg tcgacgagct cgccgagctcg gacatgagggt tgccccgtat tcagtgtcgc	7080
tgatTTTGTAT tgtctgaagt tgTTTTACG ttaagttgtat gcagatcaat taatacgata	7140
cctgcgtcat aattgattat ttgacgtggt ttgatggcct ccacgcacgt tgtgatatgt	7200
agatgataat cattatcaact ttacgggtcc ttccgggtga tccgacaggt tacggggcgg	7260
cgacctcgcg ggTTTCGCT atttatgaaa attttccggt ttaaggcggt tccgttcttc	7320
ttcgtcataa cttaatgttt ttatTTAAA taccctctga aaagaaagga aacgacaggt	7380
gctgaaagcg agcttttgg cctctgtcgt ttccTTTCTC tgTTTTGTC cgtggaatga	7440
acaatggaaag tccgagctca tcgctaataa cttcgatag catacattat acgaagttat	7500
attcgat	7507

<210> SEQ ID NO 5

<211> LENGTH: 44

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 5

ttggTTggta cgtactcgag tgcaagcttg agtattctat agtg	44
--	----

<210> SEQ ID NO 6

<211> LENGTH: 44

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 6

ttggTTggta cgtactcgag tccattgctg acataatccg ctcc	44
--	----

<210> SEQ ID NO 7

<211> LENGTH: 2474

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic polynucleotide

<400> SEQUENCE: 7

tcgcgcgttt cgggtatgac ggtaaaacc tctgacacat gcagctcccg gagacgggtca	60
cacgttgtct gtaageggat gcccggagca gacaagcccg tcagggcgcg tcagegggtg	120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgtt ctgagagtgc	180
accatagggc ggccgcacgc tggaattcta cgtactcgag agtactcgcc ccgcgagctt	240
ggcgtaatca tggcatagc tgTTTCTGT gtggaaattgt tatccgtca caattccaca	300
caacatacga gccggaagca taaagtgtaa agcctgggtt gcctaattgag tgagctaact	360
cacattaatt gcgttgcgt cactgccccgc ttccagtcg ggaaacctgt cgtgccagct	420
gcattaaatga atcggccaac gcgcggggag aggccgtttt cgtattggc gctttccgc	480

-continued

ttcttcgctc actgactcgcc tgcgtcggt cgttcgctg cggcgacgg tatcagctca
ctcaaaggcg gtaatacggt tatccacaga atcagggat aacgcaggaa agaacatgtg
agcaaaaaggc cagaaaaagg ccaggaaccg taaaaaggcc gcgttctgg cgttttcca
taggctccgc cccctgacg agcatcacaa aaatcgacgc tcaagtctaga ggtggcgaaa
cccgacagga ctataaagat accaggcggtt tccccctgga agctccctcg tgcgctctcc
tgttccgacc ctgcggctta ccggataacct gtccgcctt ctccctcgg gaagcgtggc
gttctcat agctcacgct gtaggtatct cagttcggtg taggtcggtc gctccaagct
gggctgtgtg cacgaacccc ccggttcagcc cgaccgcgtgc gccttatccg gtaactatcg
tctttagtcc aaccggtaa gacacgactt atcgccactg gcagcagccca ctggtaacag
gattagcaga gcgaggatgt taggcgggtc tacagagttc ttgaagtgtt ggcctaacta
cggtcacact agaagaacag tattttgtat ctgcgtctg ctgaagccag ttacccctgg
aaaaagagtt ggttagcttt gatccggaa acaaaccacc gctggtagcg gtgggttttt
tgtttgcaag cagcagatta cgccgcagaaa aaaaggatct caagaagatc ctttgatctt
ttctacgggg tctgacgctc agtggAACGA aaactcacgt taagggattt tggtcatgag
attatcaaaa aggatcttca cctagatctt tttaattaa aatgaagtt ttaaatcaat
ctaaagtata tatgagtaaa cttggcttga cagttacca tgcttaatca gtgaggcacc
tatctcagcg atctgtctat ttgcgttcatc catagttgcc tgactccccg tcgttagat
aactacgata cggggagggtt taccatctgg cccagtgct gcaatgatac cgccgacacc
acgctcaccc gctccagatt tatacgcaat aaaccagccca gccggaaaggcc cgagcgcag
aagtggctt gcaactttat ccgcctccat ccagtctatt aattttgttgc gggaaagcttag
agtaagttagt tgcggcgttta atagtttgcg caacgttgcgtt gccattgtca caggcatcg
ggtgtcacgc tcgtcggtt gtaggtcttcc attcagtgcc gggtccaaac gatcaaggcg
agttacatga tccccatgt tggcaaaaaa agcggttagc tccctcggtc ctccgatcg
tgcagaagt aagtggccg cagttgttac actcatgggtt atggcagcac tgcataattc
tcttactgtc atgccccatccg taagatgtt ttctgtact ggtgagttact caaccaagtc
attcttgagaa tagtgtatgc ggcgaccggat ttgtcttgc cccggctcaaa tacgggataa
taccggccca catagcagaa cttaaaagt gctcatcatt ggaaaaacgtt ctccggggcg
aaaactctca aggatcttac cgctgttgcg atccagttcg atgtaaacccca ctgcgtgcacc
caactgtatct tcagcatctt ttactttcac cagcggttctt gggtgagccaa aaacaggaag
gcaaaatgc gcaaaaaagg gaataaggcc gacacggaaa tggtgaaatac tcatacttt
ccttttcaatttatttgcgaa gcatttatca gggttattgt ctcgttgcg gatacatatt
tgaatgtatt tagaaaaata aacaaatagg ggttccgcg acatccccca gaaaagtgcc
acgttgcgttc taagaaacca ttattatcat gacattaacc tataaaaata ggcgtatcac
qadggccctt cgtc

```
<210> SEQ ID NO 8
<211> LENGTH: 7453
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic polynucleotide
```

-continued

<400> SEQUENCE: 8

tcgcgcgttt	cggtgatgac	ggtaaaaacc	tctgacacat	gcagctcccg	gagacggtca	60	
cagcttgcgt	gtaagcggat	gccgggagca	gacaagcccg	tcagggcgcg	tcagcgggtg	120	
ttggcggtgt	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180	
accatagggc	ggccgtactc	gagtgcaga	tttagtattc	tatagtgtca	cctaaatagc	240	
ttggcgtaat	catggtcata	gctgtttcct	gtgtgaaatt	gttatccgct	cacaattcca	300	
cacaacatac	gagccggaag	cataaagtgt	aaagcctggg	gtgcctaattg	agtgagctaa	360	
ctcacattaa	ttgcgttgcg	ctcaactgccc	gctttccagt	cggaaacact	gtcgtccag	420	
ctgcattaaat	gaatcggcca	acgcgaaccc	cttgcggccg	cccggggcgt	cgaccaattc	480	
tcatgttta	cagcttatca	tcgaatttct	gccccatc	cgcttattat	cactaagaaa	540	
ccattattat	catgacattta	acctataaaa	ataggcgtat	cacgaggccc	tttcgttcc	600	
aagaatttcg	gtcgaaaaaa	gaaaaggaga	gggccaagag	ggagggcatt	ggtgactatt	660	
gagcacgtga	gtatacgtga	ttaagcacac	aaaggcagct	tggagtatgt	ctgttattaa	720	
tttcacaggt	agttctggtc	cattggtaa	agtttgcggc	ttgcagagca	cagaggccgc	780	
agaatgtgct	ctagattccg	atgcgtactt	gctgggtatt	atatgtgtgc	ccaatagaaa	840	
gagaacaatt	gacccggta	ttgcaaggaa	aatttcaagt	cttgtaaaag	catataaaaa	900	
tagttcaggc	actccgaaat	acttggttgg	cgtgtttcgt	aatcaaccta	aggaggatgt	960	
tttggctctg	gtcaatgatt	acggcattga	tatcgccaa	ctgcattggag	atgagtctg	1020	
gcaagaatac	caagagttcc	tcggtttgc	agtttattaa	agactcgtat	ttccaaaga	1080	
ctgcaacata	ctactcagt	cagttcaca	gaaacctcat	tcgtttattc	ccttgcgttga	1140	
ttcagaagca	ggtgggacag	gtgaactttt	ggattggAAC	tcgatttctg	actgggttgg	1200	
aaggcaagag	agccccggaa	gcttacattt	tatgttagct	ggtggactga	cgccagaaaa	1260	
tgttgtgtat	gctgttagat	taatggcgt	tattgggtt	gatgttaagcg	gaggtgtgga	1320	
gacaaatggt	gtaaaagact	ctaacaaaat	agcaaatttc	gtcaaaaatg	ctaagaaata	1380	
ggtttattact	gagtagtatt	tatthaagta	ttgtttgtgc	acttgcctgc	aggccttttgc	1440	
aaaaggcaagc	ataaaagatc	taaacataaa	atctgtaaaa	taacaagatg	taaagataat	1500	
gctaaatcat	ttggcttttt	gattgattgt	acaggaaaat	atacatcgca	gggggttgac	1560	
ttttaccatt	tcaccgcaat	ggaatcaa	ttgttgaaga	aatgttcac	aggcgcatac	1620	
gctacaatga	cccgattctt	gctagccctt	tctcggtct	gcaaaacaacc	gccggcagct	1680	
tagtatata	atacacatgt	acatacctt	ctccgtatcc	tcgttaatcat	tttcttgc	1740	
ttatcgctt	ttcgctgtaa	aaacttttac	acacttatct	caaatacact	tattaaccgc	1800	
ttttactatt	atcttctacg	ctgacagtt	tatcaaacag	tgacacat	taaacacagt	1860	
ggtttcttg	cataaacacc	atcagcctca	agtcgtcaag	taaagatttc	gtgttcatgc	1920	
agatagataa	caatctata	gttgtataatt	agcggtgcct	catcaatgcg	agatccgtt	1980	
aacccggaccc	tagtgcactt	accccacgtt	cggtccactg	tgtgccgaac	atgctccctc	2040	
actattttaa	catgtggat	taattctaaa	tcctctttat	atgatctgcc	gatagatagt	2100	
tctaagtcat	tgagggtcat	caacaattgg	atttctgtt	tactcgactt	caggtaatga	2160	
aatgagatga	tacttgc	tta	tctcatagtt	aactggcata	aatttttagta	taggttaact	2220

-continued

ctaagagggtg atacttattt actgtaaaac tgtgacgata aaaccggaaag gaagaataag	2280
aaaactcgaa ctgatctata atgccttattt tctgtaaaaga gtttaagctc tgaaagcctc	2340
ggcattttgg ccgctccatg gtagtgcttt tttccaagg acaaaaacagt ttcttttct	2400
tgagcaggtt ttatgtttcg gtaatcataa acaataaata aattatttca tttatgttta	2460
aaaaataaaaaa ataaaaaaagt atttaaatt tttaaaaaaat ttgattataa gcatgtgacc	2520
tttgcaagc aattaaattt tgcaatttgc gatttaggca aaagttacta tttctggctc	2580
gtgtaatata tigtatgctaa tgtgaacttt tacaaagtgc atatggactt agtcaaaga	2640
aattttctta aaaatataa gcaactagcca atttagca tctttatgag atatattata	2700
gacttttatta agccagattt gtgtattata tgtatttacc cggcgaatca tggacataca	2760
ttctgaaata ggtaatattc tctatggta gacagcatag ataaccatgg atacaaggta	2820
aaagctagta ctgtttgca gtaattttt tctttttat aagaatgtt ccacctaata	2880
aagttataaa gtcaatagtt aagtttgata tttgattgtt aaataccgtt atatatttgc	2940
atgatcaaaa ggctcaatgt tgactagcca gcatgtcaac cactatattt atcaccgata	3000
ttaggacttc cacaccaact agtaatatga caataaattt aagatatttct tcatgagaat	3060
ggcccgctc atgtttgaca gcttatcattc gataagctt aatgcggtag tttatcacag	3120
ttaaatttgtt aacgcgttca ggcaccgtgt atgaaatctt acaatgogct categtcatc	3180
ctcggcaccc tcaccctgga tgctgttaggc ataggcttgg ttatgcgggt actgcgggc	3240
ctcttgcggg atatcgcca ttccgacago atcgccagtc actatggcgt gctgttagcg	3300
ctatatgcgt tgatgcaattt tctatgcgca cccgttctcg gagcactgtc cgaccgctt	3360
ggccgcgcgc cagtctctgtt gatcaattttt ttagtataaa tttcaacttgc aaccatcttgc	3420
gggaccacac cctgttgcgtt gatcaattttt ttagtataaa tttcaacttgc aaccatcttgc	3480
gaaggaccggg ataattttt gaaatctttt tttcaattttt atatgtgtt tttcaacttgc aaccatcttgc	3540
tctttcttca acaattttt aactctcggtt gccaagttgg tttcaacttgc aaccatcttgc	3600
tttatcacta cggaaattttt ttcaggcgta gcaaccaggc gtttaaggc accaataact	3660
gccttaaaaa aattacgccc cgccctgcca ctcatcgat tactgttgc attcatatgg	3720
cattctgcgg acatggaaac catcacaaac ggcgtatgcgat acctgtatgc ccageggcat	3780
cagcaccttgcg tataatattt gcccattttttt gaaacggggg cgaagaagttt	3840
gtccatatttgc gccacgtttt aatcaaaactt ggtgaaactt acccaggatggctgagac	3900
aaaaaacata ttctcaataa acccttttagt gaaataggcc aggttttccat cgtaacacgc	3960
cacatcttgc gaatataatgtt gtagaaacttgc cccgttgcgtt cactccagag	4020
cgatgaaaac gtttcagttt gctcatggaa aacgggttcaaa cactatccca	4080
tatcaccaggc tcaccgttccat tcattggccat acggaaattcc ggatgagcat tcatcaggcg	4140
ggcaagaatgt tgaataaaagg ccggataaaa cttgtgttcaaa ttttttttgc ggttccat	4200
aaaggccgtt atatccagctt gaaacgggttcaaa cttgtgttcaaa ctgactgaaa	4260
tgcctcaaaa ttttttttgc gatgttgcattt ggtatataatgc acgggttcaaa atccagtgat	4320
ttttttttttcc attttagctt ccttagcttcc tgaaaatcttcaaa gataactcaa aaaatacgccc	4380
cggttagtgc gtttttttgc ttttttttgc gatgttgcattt ggtatataatgc acgggttcaaa	4440
tcatttttcgc caaaagggttgg cccagggttcc cccgggtatca acaggacacac caggatttat	4500

-continued

tatttctgcg aagtgtatctt ccgtcacagg tattttatcg cgataagctc atggagccgc	4560
gttaaccgtcg cacaggaagg acagagaaaag cgcggatctg ggaagtgcacg gacagaacgg	4620
tcaggacactg gattggggag gcggttgccg ccgctgctgc tgacgggtgt acgttctctg	4680
ttccggtcac accacatacg ttccgccatt cctatgcgt gcacatgcgt tatgccgtta	4740
taccgctgaa agttctgcaa agcctgtatgg gacataagtc catcagtta acggaagtct	4800
acacgaagg ttttgcgtg gatgtggctg cccggcaccg ggtgcagttt gcgatgccgg	4860
agtctgtatgc gggtgcgtatg ctgaaacaat tattctgaga ataaatgcct tggcctttat	4920
atggaaatgt ggaactgagt ggatatgtctg tttttgtctg taaaacagag aagctggctg	4980
ttatccactg agaagcgaac gaaacagtctg ggaaaatctc ccattatcgt agagatccgc	5040
attattaatc tcaggagcct gtgttagcgtt tataggaagt agtgttctgt catgtgcct	5100
gcaagcggta acgaaaacga tttgaatatg ctttcaggaa caatagaaat cttcgtccgg	5160
tgttacgttg aagtggagcg gattatgtca gcaatggact cgagtaacggc cgccgagcttg	5220
gcgtaatcat ggtcatagct gtttctgtg taaaattgtt atccgtcac aatccacac	5280
aacatacggc cccggaaagcat aaagtgtaaa gcttgggggtt cctaattgtt gagctaactc	5340
acattaatttgcgtt actgtccgtt ttccagtcgg gaaacctgtc gtgcagctg	5400
cattaaatgaa tcggccaaacg cggggggaga gggggttgc gtattggggcg ctcttcgtt	5460
tcctcgtca ctgactcgct ggcgtcggtc gttcggctgc ggcgagcggt atcagctcac	5520
tcaaaggcgg taatacgggtt atccacagaa tcaggggata acgcaggaaa gaacatgtga	5580
gcaaaaaggcc acgaaaaaggc caggaaccgt aaaaaggcccg cggtgtggc gttttccat	5640
aggctccggcc cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac	5700
ccgacaggac tataaagata ccaggcggtt cccctggaa gtcctctgt ggcgtctct	5760
gttcccgaccg tgcccggttac cggataacctg tccggcttcc tcccttgggg aagcgtggcg	5820
ctttctcata gtcacgctg taggtatctc agttcgggtgt aggtcggtcg ctccaagctg	5880
ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ctttatccgg taactatcgt	5940
cttgagtcaccccgtaag acacgactta tcgcccactgg cagcagccac tggtaacagg	6000
attagcagag cgaggatgtt aggcgggtgtt acagagtctc tgaagtgggtt gcctaactac	6060
ggctacacta gaagaacagt atttgggttac tgegctctgc tgaagccagt tacttcgga	6120
aaaaaggttg ttagcttgc atccggcaaa caaaccaccc ctggtagccg tggtttttt	6180
gtttgcaagc acgagattac ggcgcgaaaaaa aaaggatctc aagaagatcc tttgatcttt	6240
tctacggggctt ctgacgctca gtggaaacgaa aactcacgtt aagggtttt ggtcatgaga	6300
ttatcaaaaaa ggatcttcac cttagatcctt ttaaattaaa aatgaagttt taaatcaatc	6360
taaaatgtat atgagtaaac ttggctgtac agttaccaat gcttaatcag tgaggcacct	6420
atctcagcgtc tctgtctatt tcgttcatec atagttgcct gactccccgt cgtgtatata	6480
actacgatac gggagggtt accatctggc cccagtgctg caatgtatcc ggcgacccca	6540
cgctcacccg ctccagattt atcagcaata aaccagccag ccggaaaggccg cgagcgcaga	6600
agtggctctg caactttatc cgcctccatc cagtcttattt attgttgccg ggaagctaga	6660
gtaaatgtttt cggccagttaa tagttgcgc aacgttggtt ccattgtac aggcatcgat	6720
gtgtcacgct cgtcggttgg tatggcttca ttccagtcgg gttcccaacg atcaaggcga	6780

-continued

gttacatgtatccccatgtt	gtgcaaaaaaaaaa	gcccggtagct	ccttcggtcc	tccgatcgtt	6840	
gtcagaagta	agtggccgc	agtgttatca	ctcatggta	tggcagcaact	gcataattct	6900
cttactgtca	tgccatccgt	aagatgcttt	tctgtgactg	gtgagttactc	aaccaagtca	6960
ttctgagaat	agtgtatgct	gcccggagt	tgctcttgcc	cggcgtcaat	acgggataat	7020
accgcgccac	atagcagaac	tttaaaagtg	ctcatcattt	gaaaacgttc	ttcggggcga	7080
aaactctcaa	ggatcttacc	gctgttgaga	tccagttcga	tgttaaccac	tcgtgcaccc	7140
aactgatctt	cagcatctt	tactttcacc	agcggttctg	ggtgagcaaa	aacaggaagg	7200
caaaatgcgc	caaaaaaggaa	aataaggcg	acacggaaat	gttgaataact	catactttc	7260
cttttcaat	attattgaag	catttatcg	ggttattgtc	tcatgagcgg	atacatattt	7320
gaatgtattt	agaaaaataa	acaataggg	gttccgcgc	catttcccg	aaaagtgc	7380
cctgacgtct	aagaaaccat	tattatcatg	acattaacct	ataaaaatag	gcgtatcacg	7440
aggcccttc	gtc					7453

<210> SEQ ID NO 9

<211> LENGTH: 1353

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 9

tgcgcgcgtt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccg	gagacggtca	60
cagcttgtct	gtaagcggat	gccggggagca	gacaaggcccc	tcagggagcg	cgtcagcggg	120
tgttggcggt	tgccgggct	ggcttaacta	tgcggcatca	gagcagattg	tactgagagt	180
gcaccatacc	acagctttc	aattcaattc	atcatttttt	ttttattctt	ttttttgatt	240
tcggttttt	tgaaattttt	ttgattcggt	aatctccgaa	cagaaggaag	aacgaaggaa	300
ggagcacaga	cttagattgg	tatataacg	catatgttagt	gttgaagaaa	catgaaattt	360
cccagtattt	ttaacccaac	tgcacagaac	aaaaacctgc	agggaaacgaa	gataaatcat	420
gtcgaaagct	acatataagg	aacgtgctgc	tactcattct	agtccctgtg	ctgccaagct	480
attnaatatc	atgcacgaaa	agcaaacaaa	cttgcgtgc	tcatggatg	ttcgtaccac	540
caaggaattt	ctggagtttag	tttggcatt	aggtcccaaa	atttgcgttac	taaaaacaca	600
tgtggatatc	ttgactgatt	tttccatgga	gggcacagtt	aagccgctaa	aggcattatc	660
cgccaagttac	aatttttac	tcttgcaga	cagaaaattt	gctgacatg	gtaatacagt	720
caaatttgcg	tactctgcgg	gtgtatcacag	aatagcagaa	tggcagaca	ttacgaatgc	780
acacgggtgt	gtggggccag	gtattgttag	cggtttgaag	caggcggcag	aagaagtaac	840
aaagggaaacct	agggccctt	tgtatgttgc	agaattgtca	tgcaagggt	ccctatctac	900
tggagaatat	actaagggtt	ctgttgcacat	tgcgttgcac	gacaaagatt	ttgttgcgtt	960
ctttattgtct	caaagagaca	tgggtggaa	agatgtttgtt	tacgttgc	tgattatgac	1020
acccgggtgt	ggtttagatg	acaagggaga	cgcattgggt	caacagtata	gaaccgttga	1080
tgtatgtggtc	tctacaggat	ctgacattat	tattgttgc	agaggactat	ttgcaaggaa	1140
aaggatgtgt	aaggtagagg	gtgaacgtt	cagaaaagca	ggctggaaag	catatttgag	1200
aagatgcggc	cagcaaaact	aaaaaaactgt	attataagta	aatgcgtt	tactaaactc	1260
acaattttaga	gttcaattt	aatttatcatc	gttattaccc	tatgcgggt	gaaataccgc	1320

-continued

acagatgcgt aaggagaaaa taccgcatca gga	1353
<210> SEQ_ID NO 10	
<211> LENGTH: 2474	
<212> TYPE: DNA	
<213> ORGANISM: <i>Saccharomyces cerevisiae</i>	
<400> SEQUENCE: 10	
tcctgatgcgtatttctc cttacgcatac tgcggtatatcacccgc atatcgacgg	60
tcgaggagaa cttctagtat atccacatac ctaatattat tgccttatta aaaatggaaat	120
cccaacaatt acatcaaaaat ccacattctc ttcaaaaatca attgtcctgt acttccttgt	180
tcatgtgtgt tcaaaaacgt tatatttata ggataattat actctatttc tcaacaagta	240
attgggttggt tggccgagcg gtctaaaggcg cctgattcaa gaaatatctt gaccgcagtt	300
aactgtggga atactcaggt atcgtaaagat gcaagagttc gaatcttta gcaaccattaa	360
ttttttctt caacataacg agaacacaca ggggcgtat cgcacagaat caaattcgat	420
gactggaaat tttttgttaa tttcagaggt cgcttgacgc atataccctt ttcaactgaa	480
aaatttggggag aaaaaggaaaa ggtgagaggc cggaaccggc ttttcatata gaatagagaa	540
gcgttcatga cttaatgctt gcatacacaat acttgaagtt gacaatattta tttaggacc	600
tattgttttt tccaaataggt ggtagcaat cgtcttactt tctaaactttt cttaccttt	660
acatttcagc aatatataata tataatttcaa ggatatacca ttctaatgtc tgcccctatg	720
tctgccccta agaagatcgt cgttttgcca ggtgaccacg ttggtaaaga aatcacagcc	780
gaagccattaa aggttcttaa agctatttct gatgttcgtt ccaatgtcaa gttcgatttc	840
aaaaatcatt taatttgggg tgctgtatc gatgtacag gtgtcccaact tccagatgag	900
gcgttggaaag cttccaaagaa ggtagatgcg gttttgttag gtgtgtggg tggctctaaa	960
tggggtagccg gtatgtttag acctgaacaa ggtttactaa aaatccgtaa agaacttcaa	1020
ttgtacgcca acttaagacc atgttaacttt gcataccgact ctctttttaga cttatctcca	1080
atcaageccac aatttgctaa aggtactgac ttcttttttt tcagagaatt agtggaggt	1140
atttactttt gtaagagaaaa ggaagacgt ggtgatgggt tgccttgggaa tagtgaacaa	1200
tacaccgttc cagaagtgc aagaatcaca agaatggcccttcatggc cttacaacat	1260
gagccaccat tgccttatttgc tgccttggat aaagctaatg ttttggctc ttcaagatta	1320
tggagaaaaaa ctgtggagga aaccatcaag aacgaattcc ctacattgaa ggttcaacat	1380
caattgtatttgcg catgtccata gtttggatcgtt caacccacccat aatgggtatt	1440
ataataccacca gcaacatgtt tgggtatato atctccgtatc aagcctccgt tatcccgat	1500
tccttgggtt tgggtccatc tgcgtcccttgc gcctcttgc cagacaagaa caccgcattt	1560
ggtttgcgtt aaccatgcca cgggttctgtt ccagatttgc caaagaataa ggttgcacct	1620
atcgccacta tcttgcgttgc tgcaatgtatc ttgaaattgtt cattgtactt gcctgttggaa	1680
ggtaaggccat ttgttggatcgtt aacttttttttgc tggatggatcgtt caggtatcgtt aacttgggtatc	1740
ttaggtgggtt ccaacagttac caccgttgc ggttgcgttgc tgcgttggatcgtt aacttgggtatc	1800
atcccttgcgtt aaaaaggatc tcttttttttgc tggatggatcgtt acataaaactt tataaaatgaa	1860
attcataata gaaacgacac gaaatttacaa aatggatat gttcatagggt tagacgaaac	1920
tataatacgca atctacatac atttatcaag aaggagaaaa aggaggatag taaaggataa	1980

-continued

caggtaagca aattgatact aatggctcaa cgtgataagg aaaaagaatt gcactttaac	2040
attaatattg acaaggagga gggcaccaca caaaaagtta ggtgtaacag aaaatcatga	2100
aactacgatt cctaatttga tattggagga ttttctctaa aaaaaaaaaa atacaacaaa	2160
taaaaaaac ac tcaatgacct gaccatgttga tggagttaa gtcaataacct tccttgaacca	2220
tttccctataa tggtgaaagt tccctcaaga attttactct gtcagaaacg gccttacgac	2280
gtagtcgata tggtgcactc tcagtacaat ctgctctgtat gcccgcatagt taagccagcc	2340
ccgacaccccg ccaacaccccg ctgacgcgc ctgacgggct tgtctgtcc cgccatccgc	2400
ttacagacaa gctgtgaccg tctccggag ctgcatgtgt cagaggtttt caccgtcatc	2460
accgaaacgc gcga	2474

<210> SEQ ID NO 11

<211> LENGTH: 1177

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 11

aattcccggtt ttaagagctt ggtgagcgctt aggagtcaact gccaggatatc gtttgaacac	60
ggcatttagtc agggaaagtca taacacagtc ctttcccgca attttctttt tctattactc	120
ttggccctctt ctagtacact ctatattttt ttatgcctcg gtaatgattt tcatttttt	180
ttttcccccta gcggatgact cttttttttt cttagcgattt ggcatttatca cataatgaat	240
tatacattat ataaagtaat gtgatttctt cgaagaatat actaaaaaat gagcaggcaa	300
gataaaacgaa ggcaaaagatg acagagcaga aagccctagt aaagcgtatt acaaatgaaa	360
ccaagattca gattgcgatc tctttaaagg gtggcccccgtt agcgatagag cactcgatct	420
tcccgaaaaa agaggcagaa gcagtagcag aacaggccac acaatcgcaa gtgatatacg	480
tcccacacagg tatagggtttt ctggaccata tgatacatgc tctggccaaag cattccggct	540
ggtcgctaat cgttgagtgc attggtgact tacacataga cgaccatcac accactgaag	600
actgcgggat tgctctcggt caagtttta aagaggccctt actggcgctt ggagtaaaaa	660
ggtttggatc aggatttgcg cctttggatc aggcactttc cagagegggtt gtatcttt	720
cgaacaggcc gtaacgagttt gtcgaacttg gtttgc当地 ggagaaagta ggagatctct	780
cttgcgagat gatccgcattttt gtttgc当地 ggcttagcaga attaccctcc	840
acgttgttgc tctgcgaggc aagaatgtatc atcaccgtatc tgagatgtcg ttcaaggctc	900
tttgc当地 catabagagaa gccacctcgcc ccaatggatc caacgtt ccctccacca	960
aagggtgttct tatgtgtatc caccgattat ttaaagctgc agcatacgat atatacat	1020
gtgttatatat gtataccatc gaatgtatc aagttatgtatc acgaacagta tgataactgaa	1080
gtatgacaagg taatgtatc ttctatacgat gtcatttgc当地 acgaggcccg ctttctttt	1140
ttcttttgc tttttctttt ttttctttaa gaaactcg	1177

<210> SEQ ID NO 12

<211> LENGTH: 3008

<212> TYPE: DNA

<213> ORGANISM: *Saccharomyces cerevisiae*

<400> SEQUENCE: 12

tggcaattt catgtttctt caacactaca tatgcgtata tataccaatc taagtctgt	60
---	----

-continued

ctccttcctt	cgttcttcct	tctgttcgga	gattaccgaa	tcaaaaaaat	ttcaaggaaa	120
ccgaaatcaa	aaaaaagaat	aaaaaaaaaa	tgtatgttttgc	aaacccccc	cccccccccc	180
gatgcgcgc	gtgcggctgc	tggagatggc	ggacgcgtg	gataatgttct	gccagggtt	240
gggtcgacgc	tacctaaga	gagacgcgtg	cggccgcaag	cttgcgtgcc	tgcagggtcga	300
tcgactctag	aaatcgatag	atctgaatta	attcttgaat	aatacataac	ttttcttaaa	360
agaatcaaag	acagataaaa	tttaagat	attaaatattt	agtgagaagc	cgagaatttt	420
gtAACACCAA	cataacactg	acatctttaa	caacttttaa	ttatgataca	tttcttacgt	480
catgattgt	tattacagct	atgctgacaa	atgactcttg	ttgcgtggct	acgaaccggg	540
taataactaag	tgattgactc	ttgctgacct	tttattaaga	actaaatgg	caatattatg	600
gagcatttca	tgtataaattt	ggtgcgtaaa	atcggtggat	ctctcttctt	agtacatcct	660
actataacaa	tcaagaaaaa	caagaaaatc	ggacaaaaca	atcaagtatg	gattctagaa	720
cagttggat	attaggaggg	ggacaattgg	gacgtatgt	tgttggggca	gcaaacaggc	780
tcaacattaa	gacggtaata	ctagatgtc	aaaattctcc	tgcacaaacaa	ataagcaact	840
ccaaatgacca	cgttaatggc	tcctttcca	atcctcttgc	tatcgaaaaa	ctgatgaaa	900
aatgtgtatgt	gtcaacgatt	gagattggc	atgttgcgt	tccttacacta	aagaatctc	960
aagtaaaaca	tcccaaattt	aaaatttacc	cttctccaga	aacaatcaga	ttgatacaag	1020
acaaatataat	tcaaaaagag	catttaatca	aaaatggtat	agcagttacc	caaagtgttc	1080
ctgttggaa	agccaggatgg	acgttccat	tgaatgttgg	aagagatgg	ggttttccat	1140
tcgtcttgc	gtcgaggact	ttggcatacg	atggaaagagg	taacttgcgtt	gtaaagaata	1200
aggaaatgt	tccgaaagct	ttggaaagtac	tgaaggatcg	tcctttgtac	gccgaaaaat	1260
gggcaccatt	tactaaagaa	tttagcgtca	tgattgtgg	gtctgttaac	ggtttagtgt	1320
tttcttacc	aattgttagag	actatccaca	aggacaataat	ttgtgcactt	tgttatgcgc	1380
ctgcttagat	tccggactcc	gttcaactt	aggcgaagtt	gttggcagaa	aatgcaatca	1440
aatctttcc	cggttgggt	atattgggt	tggaaatgtt	ctattttagaa	acagggaaat	1500
tgcttattaa	cgaaattgcc	ccaaggcctc	acaactctgg	acattatacc	attgtatgtt	1560
gcgtcaactt	tcaatttgc	gctcatttgc	gatcaatattt	ggatttgcca	atgcacaaaga	1620
atttcacatc	tttctccacc	attacaacga	acgccattat	gttcaatgtt	cttggagaca	1680
aacatataaa	agataaaagag	ctagaaactt	gcgaaagagc	attggcact	ccaggttcc	1740
cagtgtactt	atatggaaaa	gagtctagac	ctaacagaaa	agtaggttac	ataaaatatta	1800
ttgcctccag	tatggggaa	tgtgaacaaa	ggctgaacta	cattacaggt	agaactgata	1860
ttccaatcaa	aatctctgtc	gctcaaaatgt	tggacttgg	agcaatggc	aaaccatgg	1920
ttggaaatcat	catggatca	gactctgact	tgcggtaat	gtctgecgca	tgtgegggtt	1980
taaaagattt	tggcggttca	tttgaagtga	caatagtctc	tgctcataga	actccacata	2040
ggatgtcagc	atatgttatt	tccgcaagca	agcgtggat	taaaacaattt	atcgctggag	2100
ctgggtgggc	tgctcaacttgc	ccaggttatgg	tggctgcaat	gacaccaccc	cctgtcatcg	2160
gtgtgcccgt	aaaagggttct	tgtcttagatg	gagtagattc	tttacattca	attgtgcata	2220
tgcccttagagg	tgttccagta	gctaccgtcg	ctattaataa	tagtacgaac	gctgcgtgt	2280
tggctgtcag	actgcttggc	gctttagatt	caagttatac	aacgaaaatg	gaacagttt	2340

-continued

tataaaagca agaagaagaa gttcttgta aagcacaaaa gtttagaaact gtccgttacg	2400
aagcttatct agaaaacaag taatatataa gtttattgtat atacttgtac agcaaataat	2460
tataaaatgtatataacctt ttttaggctt tgttatgatt acatcaaatg tggacttcatt	2520
acatagaaat caacgcttac aggtgtcctt ttttaagaat ttcatacata agatctctcg	2580
aggatccccgggttaccgagtcgaattcgcggccgccccgggttaacccctaggcatgc	2640
actagtggcc taattggcccgacgtcagggtgcactttcgggaaatgtgcgcggaaaccc	2700
ctatTTgtttt atttttctaa atacattcaa atatgtatcc gtcatgaga caataaccct	2760
gataaatgtctcaataatataaaaaagga agagttatgat tattcaacat ttccgtgtcg	2820
cccttattccctttttgcgcattttgcccttctgtttt tgctcacccgaaacgctgg	2880
tgaaaagtaaa agatgtgaa gatcgttgg gtgcacgagt gggttacatc gaactggatc	2940
tcaacagcggtaagatcctt gagagtttc gccccgaaga acgtttcca atgtatgagca	3000
cTTTTAAAG	3008

<210> SEQ ID NO 13
<211> LENGTH: 4879
<212> TYPE: DNA
<213> ORGANISM: *Saccharomyces cerevisiae*

>400 > SEQUENCE: 13

agcagttgc ttctccatg ggaagaggct tctaagtctg aagaagtaaa cagttcttg 60
ctatttcaca cttcctgggt gatggtcaact tgctgcctg aatatata tatgtatgac 120
atatgtactt gtttcttt ttgtgcctt gttacgtcta tattcattga aactgattat 180
tcgatttct tcttgcgtac cgcttctaga ggcacatcgac agtttagcg aggaaaactc 240
ttcaatagtt ttgccagcgg aattccactt gcaattacat aaaaaattcc ggccggtttt 300
ccgcgtgtgac tcaatgtcga aatacctgc taatgaacat gaacatcgcc caaatgtatt 360
tgaagaccccg ctgggagaag ttcaagatata aataagtaaca agcagccaat agtataaaaa 420
aaaatcttag tttattacat ttccctggaaat ttcaatgtcaaa aactgtaat tatagagaga 480
tatcacagag ttactcacta atgactaacg aaaaggctg gatagagaag ttggataatc 540
caactcttc agtgttacca catgacttt tacgcccaca acaagaacct tatacgaaac 600
aagctacata ttcggttacag ctacccatc tcgatgtgcc tcatgatagt ttttctaaca 660
aatacgtgt cgcttgagt gtatggctg cattgatata tagagtaacc ggtgacgtatg 720
atattgttct ttatattgcg aataacaaaa tcttaagatt caatattcaa ccaacgtgg 780
catttatga gctgtattct acaattaaca atgagttgaa caagctcaat tctattgagg 840
ccaatttttc ctttgacgag cttagctgaaa aaattcaaag ttgccaagat ctggaaagga 900
ccccctcgatg gttccgttgc gccttttgg aaaaccaaga tttcaattaa gacgagttca 960
agcatcatt agtggacttt gcttgaatt tggataccag taataatgcg catgtttga 1020
acttaatatta taacagctt ctgttattcga atgaaagagt aaccattgtt gcggaccaat 1080
ttactcaata ttgtactgct ggcgtaagcg atccatccaa ttgcataact aaaatcttc 1140
tgatcaccgc atcatccaaag gatagttac ctgatccaa taaaacttg ggctggcg 1200
atttcggtgg gtgtattcac gacatttcc aggacaatgc tgaagccctc ccagagagaa 1260
ccctgtgtgt ggagactcca acactaaatt ccgacaatgc ccgttcttc acttatcgcc 1320

-continued

acatcaaccg cacttctaac atagttgcc	1380
attatggat taaaacagg	
atccaaagag	
gtgatgtagt gatgatctat tcttctaggg	1440
gtgtggattt gatggatgt gtgatgggt	
tcttcaaagc cggcgcaacc	1500
ttttcagttt tcgaccctgc	
atatccccca gccagaca	
ccatTTactt aggtgttgct aaaccacgtg	1560
ggttgattgt tatttagagct gctggaca	
aat	
tggatcaact agtagaagat tacatcaatg	1620
atgaattgga gattgttca	
agaatcaatt	
ccatcgctat tcaagaaaaat	1680
ggttaccattt aagggtggcaa	
attggacaat ggcgaggat	
ttttggctcc atatgatcac	1740
tacaaagaca ccagaacagg	
tgttgttagtt ggaccagatt	
ccaaacccaac cctatcttc acatctggtt	1800
ccgaaaggat tccttaagggt	
gttcttgta	
gacatTTTC cttggcttat tattcaatt	1860
ggatgtccaa aagggttcaac	
ttaacagaaa	
atgataaaatt cacaatgctg	1920
agcggttattt cacatgttcc	
aattcaaaaga gatatgttta	
caccattatt tttaggtgcc caattgtatg	1980
tcctactca agatgtatatt	
ggtagacccgg	
gccgttttagc ggaatggatg	2040
agtaagtatg	
gttgcacagt taccatttta acacctgcca	
tgggtcaatt acttaactgcc	2100
caagctacta caccatccc	
taagttacat catcggtt	
tttgtgggtga catttaaca aaacgtgatt	2160
gtctgaggtt acaaaccctt	
gcagaaaaatt	
gccgttattgt taatatgtac	2220
ggttaccactg	
aaacacagcg tgca	
ttttttttct tatttcgaag	
ttaaatcaaa aatgacgt	2280
ccaaactttt	
tgaaaaaatt gaaagatgtc atgcctgt	
actggatgtt	
gttgaacgtt cagctactag	2340
ttgttaacag	
gaacgtatgt	
actcaaataat	
gtgttattgg cgaaataggt	2400
gagattttatg	
ttcgtgcagg	
tggtttggcc	
gaaggttata	
gaggattacc agaattgaat	2460
aaagaaaaat	
ttgtgaacaa ctggtttgtt	
gaaaaagatc	
actggatta tttggataag	2520
gataatggtg	
aaccttggag	
acaattctgg	
tttaggtccaa	
gagatagatt gtacagaacg	2580
ggttatttag	
gtcggttatct	
accaaaccgt	
gactgtgaat	
gttgcggtag ggctgtatgt	2640
caagttaaaa	
ttcgtgggtt	
cagaatcgaa	
tttaggagaaa	
tagatacgca	2700
catttccaa	
catccattgg	
taagagaaaa	
cattactta	
gttcgcaaaa	
atgccgacaa	2760
tgagccaaca	
ttgatcacat	
ttatggccc	
aagatttgac	
aagccagatg	
acttgc	2820
ttccaaagt	
gtgttccaa	
aggaggttga	
aactgaccct	
atagttttaagg	
gttcaatcg	2880
ttaccatctt	
ttatccaagg	
acatcaggac	
tttcttaaag	
aaaagatgg	
ctagctatgc	2940
tatgccttcc	
ttgattgtgg	
ttatggataa	
actaccattt	
aatccaaatg	
gtaaagttaa	3000
taagcctaaa	
cttcaatcc	
caactccaa	
gcaattaaat	
ttggtagct	
aaaatacagt	3060
ttctgaaact	
gacgactctc	
agtttaccaa	
tgttgacgc	
gaggtagag	
acttatggtt	3120
aagtatatta	
cctaccaago	
cagcatctg	
atcaccagat	
gattcg	
ttt	
tcgattttagg	3180
ttgtcattct	
atcttggcta	
ccaaaatgt	
ttttacctta	
aagaaaaagc	
tgcaagttga	3240
tttaccattt	
ggcacaattt	
tcaagttatcc	
aacgataaag	
gcctttgcgg	
cgaaaaatttga	3300
cagaattaaa	
tcatcggtt	
gatcatctca	
aggtgagg	
gtcgaaaatg	
tcactgcaaa	3360
ttatcggtt	
gacgccaaga	
aattgggttga	
gacgctacca	
agttcg	
cctctcgaga	3420
atattttgtt	
gaacctaata	
gtgccgaagg	
aaaaacaaca	
attaatgtt	
ttgttacccg	3480
tgtcacagga	
tttctgggtt	
cctacatctt	
tgcagattt	
tttaggacgtt	
ctccaaagaa	3540
ctacagtttcc	
aaagtgtttt	
cccacgtcag	
ggccaaggat	
gaagaagctg	
catttgcaag	3600
attacaaaag	
gcaggtatca	
cctatggtac	
ttggAACGAA	
aaatttgcct	

-continued

caaataattaa agttgtatta ggcgatttat ctaaaagcca atttggctt tcagatgaga	3660
agtggatgga tttggcaaac acagttata taattatcca taatggtgcg ttagttcact	3720
gggttatcc atatgc当地 ttgagggatc caaatgttat ttcaactatc aatgttatga	3780
gcttagccgc cgtagccaaag ccaaagtctt ttgactttgt ttccctccact tctactctt	3840
acactgaata ctacttaat ttgtcagata aacttggtag cgaaggaaag ccaggcattt	3900
tagaatcaga cgatataatg aactctgaa gcgggctcac ttggatgat ggtcagtcca	3960
aatgggctgc tgagtagatc attagacgtg caggtgaaag gggctacgt gggtgtattt	4020
tcaagaccagg ttacgtaaaca ggtgcctctg ccaatggttc ttcaaaacaca gatgattct	4080
tatttagatt tttgaaaggt tcagtccat tagttaagat tccagatatc gaaaattccg	4140
tgaatatggt tccagtagat catgtgtctc gtgtgttgt tgctacgtct ttgaatcctc	4200
ccaaagaaaa tgaattggcc gttgtcaag taacgggtca cccaaagaata ttattcaaag	4260
actacttgcatacttacac gattatggtt acgtgtcgaa aatcgaaagc tattctaaat	4320
ggaagaaatc attggaggcg tctgttattt acaggaatga agaaaatgcg ttgtatcctt	4380
tgctacatcat ggtcttagac aacttacatc aaagtaccaa agctccggaa cttagcgata	4440
ggaacgcgcgt ggcacatctta aagaaagaca ccgcacatggac aggtgttgat tggctaaatg	4500
gaatagggtgt tactccagaa gaggtggta tatatatggc atttttaac aagggtggat	4560
tttacactcc accaactcat aatgacaaac ttccactgcc aagtatagaa ctaactcaag	4620
cgcacataaag tctatgtct tcagggtctc gtgtctgtgg aagctccgcgca gcaatgttgc	4680
gtttagcatt acgtatgata tgtccatgtt caataattaa atatgaatgggaaagac	4740
tttagcttc ttccgggtgt gtcactttaaa aactccgaga ataataatata ataagagaat	4800
aaaaatattttt tagttaataa agaactgtaa atcagctggc gtttagtgc taatggcagc	4860
ttcatcttgg tttattgtt	4879

<210> SEQ ID NO 14
<211> LENGTH: 1475
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic polynucleotide
<400> SEQUENCE: 14

ttaggtctag agatctgttt agcttgcctc gtccccgcgg ggtcacccgg ccagegacat	60
ggaggcccag aataccctcc ttgacagtct tgacgtgcgc agctcagggg catgtatgtt	120
ctgtcgecccg tacatttagc ccatacatcc ccatgtataa tcatttgcattt ccatacattt	180
tgtatggccgc acggcgcgaa gcaaaaattt cggctctcg ctgcagaccc gcgagcagg	240
aaacgcctcc ctcacagacg cgttgaattt tccccacgcg ggcgcctgtt agagaatata	300
aaaagggttag gatggccac tgaggttctt cttcatata ctccctttta aatcttgc	360
aggatacgt tctcacatca catccgaaca taaacaacca tgggtaaaggaa aagactcac	420
gtttcgagggc cgcgattaaa ttccaacatg gatgtgtattt tataatgggtt taaatgggt	480
cgcgataatg tcggcaatc aggtgcgaca atctatcgat tttatggaa gcccgtgcg	540
ccagagttgt ttctgaaaca tggcaaggtt agcggttcca atgtatgttac agatgagatg	600
gtcagactaa actggctgac ggaattttagt cctcttccgc ccatcaagca ttttatccgt	660

-continued

actcctgatg atgcattgtt actcaccact gcatcccc gaaaaacagc attccaggta	720
ttagaagaat atcctgattc aggtgaaaat attgttcatg cgctggcagt gttcctgcgc	780
cggttgcatt cgattcctgt ttgttaatgt cctttaaca gcatcgctt atttcgctc	840
gctcaggcgc aatcacgaat gaataacggt ttgggttcatg cgagtatgtt tgatgacgag	900
cgtaatggct ggcctgttga acaagtctgg aaagaaatgc ataagcttt gccattctca	960
ccggatttag tcgtcaactca tggtgatttc tcacttgata accttatttt tgacgagggg	1020
aaatataatag gttgtattga tggggacga gtcggaatcg cagaccgata ccaggatctt	1080
gccatccatat ggaactgcct cggtgagttt tctccttcat tacagaaacg gcttttcaa	1140
aaatatggta ttgataatcc tgatatgaat aaattgcagt ttcatttgat gctcgatgag	1200
tttttctaatac cagttactgac aataaaaaaga ttcttggttt caagaacttg tcatttgat	1260
agttttttta tattgttagtt gttctatattt aatcaaattgt tagcgtgatt tatattttt	1320
ttcgcctcga catcatctgc ccagatgcga agttaagtgc gcagaaagta atatcatgcg	1380
tcaatcgat gtgaatgctg gtcgtataac tgctgtcgat tcgataactaa cggccatc	1440
cagtgtcgaa aacgagctc cgagaaccct taata	1475

<210> SEQ ID NO 15	
<211> LENGTH: 36	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer	
<400> SEQUENCE: 15	
ttggttcccg ggtcgccgtt ttccggatgt acgggt	36

<210> SEQ ID NO 16	
<211> LENGTH: 42	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer	
<400> SEQUENCE: 16	
ttgggttgcg acccgccgtg atgcggattt ttctccttac gc	42

<210> SEQ ID NO 17	
<211> LENGTH: 36	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer	
<400> SEQUENCE: 17	
ttggttcccg ggtcctgatg cggttatttc tcctta	36

<210> SEQ ID NO 18	
<211> LENGTH: 36	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer	
<400> SEQUENCE: 18	
ttggttcccg ggtcgccgtt ttccggatgt acgggt	36

-continued

```
<210> SEQ ID NO 19
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 19
ttgggttcccg ggaagcttgc atgcctgcag gtcgat          36

<210> SEQ ID NO 20
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 20
ttgggttcccg ggagcagttg ctttctccta tggaa           36

<210> SEQ ID NO 21
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 21
ttgggttcccg ggttaggtct agagatctgt ttagct          36

<210> SEQ ID NO 22
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 22
ttgggttgtcg acggccggcc actagttcgc gcgttcggg gatgacggtg      50

<210> SEQ ID NO 23
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 23
ttgggttgtcg acggccggcc actagttgat gcggatattt ctccttacgc      50

<210> SEQ ID NO 24
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 24
ttgggttgtcg acggccggcc actagtgtatc ctgcagagat cttatgtatg      50

<210> SEQ ID NO 25
<211> LENGTH: 50
```

-continued

```

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 25
ttgggttgcg acggccggcc actagttaca ataaaccaag atgaagctgc      50

<210> SEQ ID NO 26
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 26
ttgggttgcg acggccggcc actagttatt aagggttctc gagagtcgt      50

<210> SEQ ID NO 27
<211> LENGTH: 4024
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic polynucleotide

<400> SEQUENCE: 27
tatggatgaa cgaaaatagac agatcgctga gataggtgcc tcactgatta agcattggta      60
actgtcagac caagtttact catataact ttagattgat ttaaaaacttc atttttaatt      120
taaaaggatc taggtgaaga tccttttga taatctcatg accaaaaatcc cttaacgtga      180
gttttcgttc cactgagcgt cagaccccgt agaaaagatc aaaggatctt cttgagatcc      240
ttttttctg cgcgtaatct gctgottgca aaaaaaaaaa ccaccgatc cagcggttgt      300
ttgtttgcgc gatcaagagc taccactt tttccgaaag gtaactggct tcagcagagc      360
gcagatacca aatactgttc ttcttagtga gccttagtta ggccaccact tcaagaactc      420
tgttagcaccc cctacataacc tcgctctgtt aatcctgtta ccagtggctg ctgccagtg      480
cgataagtgc tgtcttaccg gggtggactc aagacgatag ttaccggata aggcgacgc      540
gtcgggctga acgggggggtt cgtgcacaca gcccagctt gagcgaacga cctacacgca      600
actgagatac ctacagcgtg agctatgaga aagcgccacg cttcccgaaag ggagaaaggc      660
ggacaggtat ccggtaagcg gcagggtcgg aacaggagag cgcacgaggg agcttccagg      720
ggaaaacgcc tggtatctt atagtcctgt cgggttgcg cacctctgac ttgagcgtcg      780
atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgcccacg acgcggcctt      840
tttacggttc ctggccttt gctggcctt tgcacatg ttcttcctg cgttatcccc      900
tgattctgtg gataaccgta ttaccgcctt tgagtgagct gataccgctc gccgcagccg      960
aacgacccgag cgcagegagt cagtgagcga ggaagcggaa gagcgeccaa tacgcaaacc      1020
gcctctcccc ggcgttggc cgattcatta atgcagctgg cacgacaggt ttcccgactg      1080
gaaagcgggc agtgcgcga acgcaattaa tgtgagttt ctcactcatt aggccaccca      1140
ggctttacac tttatgcttc cggctcgat gttgtgtgaa attgtgagcg gataacaatt      1200
tcacacagga aacagctatg accatgatta cgccaagctc gggccgcct cgagacccg      1260
taaacagata ggaatggacg ttgcctagaa agaatcaatt ttagtgagcg tccccttacg      1320
ggacaataaa ttaattgtt gcctcgccata tcggctaaca agtttggcag tggcaggcca      1380

```

-continued

ctgccactga cgcccttccc cttcccccttc gggacaataa ataaatttgt tgcctgc当地	1440
caaatttatt tattgttatta acattgttatt aacatagaat gtttacatac tccgaaggag	1500
gacaatattt tttattgtgg taccgccact gcctgcttcc tccttccct tccccttc当地	1560
ggatataaat ataggcaag taaacttago ataaacttta gttacccaat atttatatac	1620
tccgaaggac gtcagtggca gttgcctgccc aactgcctat gcaactaaag tttatcg当地	1680
tatataaaata tagaataaaa tttatttgct gcgcttagca gtttacatac tcccaagttt	1740
acttgcccga aggggaagga ggacgtcccc ttacgggaca ataaataaat ttgtgc当地	1800
ccaaacaaattt tattttattgtt attaacatcc taatataaat attagtggac gtccc当地	1860
ggcaaatgaa ttttagtggta tattttatata ctccgaagga ggcagttgcc tgccaactgc	1920
ctaggcaagt aaacttagga gtattaaaat aggacgcccgg tggcagttgg accgccc当地	1980
cctatgttaa tactgcata aacttttagtt gcccgaaggg gtttacatac gtcgacatta	2040
ccctgttatac octtaggatcc tacgtacagt ggcagttggta ccggccactgc ctgtatgta	2100
aacattctat atttatatac tcctaagttt acttgcccga aggggaagga ggcaactgccc	2160
actaaaaatttcc atttgcccga aggggacgtc cacttactga attactattt attttctcta	2220
gagcaagaaa atagctgttg agcgggttag acctttctg gtctgc当地 tacatgttaac	2280
aagctaaccctt aaagaataac attcaactat aaaactaaat gattttattt tatgatgtc	2340
atgtttttcc ctaaaatata tttatttgac ataaatataat ttatgtgata taatataattt	2400
aaatgttattt aaaattttcc aacaatttttt aaatttataatttccggacaga ttattttagg	2460
atcgtcaaaa gaagttacat ttattnagaa ctatgataca aaataccatc atgctaccaa	2520
tggtagtaac tgggtttgtt ttagttggattt attatttata ttataaggca gtgtgc当地	2580
tacaatacaccc tccaaaccagt tatgttaatttgc attcaactgc attagatgtt tatgtatgtc	2640
aatctacgta tactttaaag aaaattttac tagctcatgc ctgtacgtt aataggttat	2700
gcaccttggaa ctttcaagcc agtgccacgg cttattnaa tcagggtggaa gcttccgggt	2760
ttttacaagg ctttggctt ttaggaaact tattgtgtttt attagatggg gccc当地	2820
ccgggtccggc ggccgc当地 ctttgcactc tcagttacaat ctgtctgtat gccc当地	2880
taagccagcc cccacaccccg ccaacaccccg ctgacgc当地 ctgacgggct tgc当地	2940
cggtccgc ttacagacaa gctgtgacccg tctccgggag ctgc当地 ctttgc当地	3000
caccgtcatac accgaaacgc gcgagacgaa agggcctgt gatacgc当地 tttttagg	3060
ttaatgtcat gataataatg gtttcttgc当地 cgtcagggtgg cactttccgg ggaaatgtgc	3120
gcccccccccc tattttttttaa tacattcaaa tatgtatccg ctc当地 gagac	3180
aataaccctg ataaatgctt caataatatt gaaaaaggaa gagttatgagt attcaacatt	3240
tccgtgtcgc ctttatttccc ttttttgc当地 cttttgc当地 tcttgc当地 gtc当地	3300
aaacgc当地 gaaagtaaaa gatgtgaaatcagttggg tgc当地 gagtg gtttacatgc	3360
aactggatctt caacagccgtt aagatcccttgc aggtttccgg cccgaaagaa ctttccaa	3420
tgtatgacac ttttaaagttt ctgtatgtc ggc当地 ggttacatgc gacgc当地	3480
aagagcaactt cggccgc当地 atacactattt ctc当地 gagtaatgc cttgggttag tactcaccag	3540
tcacagaaaa gcatcttacg gatggcatga cagtaagaga attatgc当地 gtc当地	3600
ccatgatgtca taacactgc当地 gccaacttac ttctgacaac gatcgagga cc当地	3660

-continued

taaccgcctt	tttgacacaac	atgggggatc	atgtaaactcg	ccttgatcgt	tgggaaccgg	3720
agctgaatga	agccatacca	aacgacgago	gtgacaccac	gatgcctgta	gcaatggcaa	3780
caacgttgcg	caaactatta	actggcgaac	tacttactct	agcttcccg	caacaattaa	3840
tagactggat	ggaggcggat	aaagtgcag	gaccacttct	gcgctcgcc	cttccggctg	3900
gtctggttat	tgctgataaa	tctggagccg	gtgagcgtgg	gtctcgcgtt	atcattgcag	3960
cactggggcc	agatggtaag	ccctccgt	tcgtagttat	ctacacgacg	gggagtcagg	4020
caac						4024

<210> SEQ ID NO 28

<211> LENGTH: 1382

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic polynucleotide

<400> SEQUENCE: 28

tttcatgtt	tgacagctta	tcatcgataa	gtttaatgc	ggtagtttat	cacagttaaa	60
ttgctaaccgc	agtcaaggcac	cgtgtatgaa	atctaacaat	gcccgtatcg	tcatcctcg	120
caccgtcacc	ctggatgctg	taggcatagg	cttggttatg	ccggtaactgc	cgggcctctt	180
gccccatatac	gtccattccg	acagacatcg	cagtcactat	ggcggtgtgc	tagegctata	240
tgcgttgcgt	caatttctat	gcccacccgt	tctggagca	ctgtccgacc	gtttggccg	300
ccggccatgc	ctgctcgctt	cgctacttgg	agccactatc	gactacgoga	tcatggcgcac	360
cacacccgtc	ctgtggatcc	tctacgcccgg	acgcacatcg	gcccgcatac	ccggggccac	420
agggtgcgtt	gctggcgctt	atatcgccga	catcaccat	ggggaaagatc	gggctcgcca	480
cttcggggctc	atgagcgctt	gtttagggct	gggtatgggt	ggaggccccc	tggccgggg	540
actgttggcc	gcatcttcct	tgcacatgcacc	attccctgc	ggggcggtgc	tcaacggcct	600
caacctacta	ctgggctgt	tcctaattgc	ggagtcgcac	aaggagagac	gtcgaccat	660
gcccttggaa	gccttcaacc	cagtcagctc	cttccgggtt	gcccggggca	tgactatcg	720
cgccgcactt	atgactgtct	tctttatcat	gcaactcgta	ggacaggtgc	cgccagcgct	780
ctgggtcatt	ttcggcgagg	accgtttcg	ctggagcg	acgtatgc	gcctgtcgct	840
tgcggatttc	ggaatcttc	acgcacatcg	tcaagccttc	gtcaactggc	ccggccacaa	900
acgttccggc	gagaagcagg	ccattatcg	cgcatggcg	gcccacgcgc	tgggtacgt	960
cttgctggcg	ttcgcgacgc	gaggctggat	ggccctcccc	attatgttc	ttctcgcttc	1020
cgccggccatc	gggatgcccc	cgtgcggcc	catgctgtcc	aggcaggtag	atgacgacca	1080
tcaaggacac	cttcaaggat	cgctcgccgc	tcttaccac	ctaacttcga	tcattggacc	1140
gtgtatcgatc	acggcgat	ttatcgatcg	ggggatcgatc	ggggatcgatc	tttttttttt	1200
tgtatggcc	gcctataacc	ttgtctgcct	cccccggtt	cgtcgcgggt	catggagccg	1260
ggccacatcg	acctgaatgg	aaggccggcg	cacccgtatcg	acggattcac	cactccaaga	1320
attggagcca	atcaattctt	gcggagaact	gtgaatgcgc	aaaccaaccc	ttggcagaac	1380
at						1382

<210> SEQ ID NO 29

<211> LENGTH: 679

-continued

```

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic polynucleotide

<400> SEQUENCE: 29

cccaatggca tcgtaaagaa catttgagg catttcagtc agttgctcaa tgtacctata      60
accagaccgt tcagctggat attacggcct tttaaagac cgtaaagaaa aataagcaca      120
agttttatcc ggccttatt cacattcttgc cccgcctgat gaatgctcat ccggaattcc      180
gtatggcaat gaaagacggt gagctggtga tatggatag tgttcacccct tgttacacccg      240
tttccatga gcaaactgaa acgtttcat cgctctggag tgaataccac gacgattcc      300
ggcagtttctt acacatatacgcaagatg tggcgtgtta cggtgaaaac ctggcctatt      360
tccctaaagg gtttattgag aatatgttt tcgtctcagc caatccctgg gtgagttca      420
ccagtttga tttaaacgtg gccaatatgg acaacttctt cgccccgtt ttcaccatgg      480
gcaaataatttacgcaaggc gacaagggtgc tcatgccgtt ggcatcatttgc gttcatcatg      540
ccgtttgtga tggcttccat gtcggcagaa tgcttaatga attacaacag tactgcgtg      600
agtggcaggg cggggcgtaa ttttttaag gcagttatgg tgcccttaa acgcctggtt      660
gctacgcctg aataagtga      679

<210> SEQ ID NO 30
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 30

ttggttcccg gggatataaa tacattcaaa tatgtatcc      39

<210> SEQ ID NO 31
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 31

ttgggtcccg gggatatacat ccttttaat taaaaatgaa g      41

<210> SEQ ID NO 32
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 32

ttgggtcccg gggatatactt ctcatgtttg acagcttatac at      42

<210> SEQ ID NO 33
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 33

```

-continued

aaccaacccg gggatatcat gttctgccaa gggttggttt gc	42
<210> SEQ ID NO 34	
<211> LENGTH: 42	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer	
<400> SEQUENCE: 34	
ttggttcccg gggatatccc caatccaggt cctgaccgtt ct	42
<210> SEQ ID NO 35	
<211> LENGTH: 42	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer	
<400> SEQUENCE: 35	
aaccaacccg gggatatctc acttatttagt gcttagcaac ca	42
<210> SEQ ID NO 36	
<211> LENGTH: 42	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer	
<400> SEQUENCE: 36	
ttggttcccg gggatatctt gccgggtgac gcacaccgtg ga	42
<210> SEQ ID NO 37	
<211> LENGTH: 42	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer	
<400> SEQUENCE: 37	
aaccaacccg gggatatctg ccgatggccg cggcggttg ac	42
<210> SEQ ID NO 38	
<211> LENGTH: 60	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer	
<400> SEQUENCE: 38	
aaaaaatgaat ttgaattcta attgaatatg tttgttgtt tggggtaata actgatataa	60
<210> SEQ ID NO 39	
<211> LENGTH: 60	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Primer	
<400> SEQUENCE: 39	
accccttgaa tcactactac agtacttgat tcaagggtt ctggaatcat agttcatga	60

-continued

<210> SEQ ID NO 40
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 40

acaatccact gccttgcattcc

20

<210> SEQ ID NO 41
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 41

gctatgcattg gttccttggt

20

<210> SEQ ID NO 42
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 42

gcttccattt gatctctgca c

21

<210> SEQ ID NO 43
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 43

gccatcgat tcgttttagt tc

22

<210> SEQ ID NO 44
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 44

aaccatattt taccggaaagc

20

<210> SEQ ID NO 45
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 45

acggctcaggc agtcaattct

20

<210> SEQ ID NO 46
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 46

aagatcgac tacaatcacc gta

23

<210> SEQ ID NO 47
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 47

gggatcaaag tacccatttt gt

22

<210> SEQ ID NO 48
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 48

caaaagacgg ggagactcaa

20

<210> SEQ ID NO 49
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 49

tggcatgtt cttggaaact c

21

<210> SEQ ID NO 50
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 50

ctaaccggaa gtgccttccttg

20

<210> SEQ ID NO 51
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 51

tgtcgagaag tgcaaaatcg

20

<210> SEQ ID NO 52
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 52

-continued

tttccggaat atgagtggtgc	20
 <pre><210> SEQ ID NO 53 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer</pre>	
<400> SEQUENCE: 53	
tatcaagtgc gtccttcgt	20
 <pre><210> SEQ ID NO 54 <211> LENGTH: 25 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer</pre>	
<400> SEQUENCE: 54	
cgacgagata tattatcatt tttag	25
 <pre><210> SEQ ID NO 55 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer</pre>	
<400> SEQUENCE: 55	
aaaactgctc ggcaacagtc	20
 <pre><210> SEQ ID NO 56 <211> LENGTH: 24 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer</pre>	
<400> SEQUENCE: 56	
tgcaatagaa aaagaatcca gaaa	24
 <pre><210> SEQ ID NO 57 <211> LENGTH: 21 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer</pre>	
<400> SEQUENCE: 57	
agtttgctgt atgcggacta a	21
 <pre><210> SEQ ID NO 58 <211> LENGTH: 20 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Primer</pre>	
<400> SEQUENCE: 58	
gggaaaaacgca caaagattgc	20

-continued

<210> SEQ ID NO 59
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 59

tagaagaagg gggtttgtgc

20

<210> SEQ ID NO 60
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 60

tttctgcccc aagaacagag

20

<210> SEQ ID NO 61
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 61

tcgaattccc acatgtgaa

20

<210> SEQ ID NO 62
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 62

ccggaagata tatccacatc g

21

<210> SEQ ID NO 63
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 63

cccgccctcct acatatttga

20

<210> SEQ ID NO 64
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 64

ggggcccacgga aatttattta

20

<210> SEQ ID NO 65
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
 <223> OTHER INFORMATION: Primer

 <400> SEQUENCE: 65

 ctgttgacat tgcgaagagc

20

<210> SEQ ID NO 66
 <211> LENGTH: 21
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Primer

<400> SEQUENCE: 66

 tgtcgccaa gaagttttaa a

21

<210> SEQ ID NO 67
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Primer

<400> SEQUENCE: 67

 tgggatccaa aggaaacttg

20

What is claimed is:

1. A vector comprising all essential genes of a plastid genome from a vascular plant, wherein said vector further comprises at least one yeast selection marker sequence, a yeast centromere sequence, a yeast autonomously replicating nucleotide sequence, a bacterial origin of replication and at least one bacterial selection marker, wherein said vector provides for stable replication of said plastid genome in a yeast and a bacterial cell.
2. The vector of claim 1, wherein said plastid genome is at least 135 kb in size.
3. The vector of claim 1, wherein said plastid genome is at least 150 kb in size.
4. The vector of claim 1, wherein said plastid is a chloroplast.
5. The vector of claim 1 wherein said genome comprises at least about 90% of a chloroplast genome.
6. The vector of claim 1, wherein said yeast selection marker is at least one of the group consisting of HIS3, TRP1, URA3, ADE2 and LEU2.
7. The vector of claim 1, wherein said bacterial selection marker is at least one of the group consisting of tetracycline resistance, ampicillin resistance, chloramphenicol resistance, kanamycin resistance and neomycin resistance.
8. The vector of claim 1, wherein said bacterial origin of replication is a P1 or F' origin of replication.
9. The vector of claim 1, wherein said plastid genome is from a vascular plant selected from the group consisting of soybeans, tomatoes, potatoes, wheat, rice, corn, barley, oats, rye, cotton, *Arabidopsis*, tobacco, and legumes
10. The vector of claim 1, wherein said plastid genome is a chloroplast genome.
11. The vector of 10, wherein said chloroplast genome is from a soybean.

12. The vector of claim 1, wherein said vector comprises at least one pair of yeast selection markers.

13. The vector of claim 12, wherein said at least one pair of yeast selection markers allow for positive and negative selection.

14. The vector of claim 13, wherein each yeast selection marker pair comprises a URA3 gene.

15. The vector of claim 14, wherein said yeast selection marker pair further comprises a yeast selection marker selected from the group consisting of a LEU2 gene, a HIS3 gene, an ADE2 gene, a LYS2 gene and a kanMS6 gene.

16. The vector of claim 12, wherein each of said at least one pair of yeast selection markers further comprise at least one bacterial selection marker.

17. The vector of claim 17, wherein said bacterial selection marker is selected from at least one member of the group consisting of an ampicillin resistance gene, a tetracycline resistance gene, a chloramphenicol resistance gene and a gentamycin resistance gene.

18. A method for making a recombinant plastid genome comprising, constructing a vector of claim 1, introducing said vector into yeast, inserting, deleting or both inserting and deleting at least one nucleotide sequence into or from said vector by homologous recombination to produce a modified vector containing said recombinant plastid genome, isolating said modified vector from said yeast, introducing said isolated modified vector into bacteria, and amplifying said modified vector in said bacteria.

19. The method of claim 18, wherein said plastid genome is a chloroplast genome.

20. The method of claim 19, further comprising introducing said modified genome into a chloroplast of a vascular plant.

21. The method of claim **20**, wherein said chloroplast is incapable of photosynthesis or made incapable of photosynthesis prior to introduction of said modified genome.

22. The method of claim **21**, wherein said chloroplast is capable of photosynthesis after introduction of said modified genome.

23. The method of claim **18**, wherein said at least one nucleotide sequence codes at least one exogenous protein.

24. A host cell comprising a vector of claim 1.

25. The host cell of claim **1**, wherein said host cell is a bacterial cell or a yeast cell.

26. The host cell of claim **1**, wherein said host cell is a cell of a vascular plant.

27. A modified plastid genome produced by the method of claim **18**.

28. The plant comprising the modified plastid genome of claim **27**.

* * * * *